Proceedings of the GIS Research UK 17th Annual Conference GISRUK 2009

University of Durham
1st – 3rd April 2009

Editor: David Fairbairn

Proceedings of the GIS Research UK
17th Annual Conference
GISRUK 2009

University of Durham
1st – 3rd April 2009

Published April 2009, University of Durham

ISBN: 978-0-900974-58-8

Welcome

The vitality of Geographic Information Science research in the UK shows no sign of abating, and the GISRUK series of conferences, which has been running annually since 1993, continues to meet the needs of researchers in both theoretical and applied areas of the discipline. The 2009 GISRUK conference demonstrates the breadth of our subject: within these Proceedings you can find contributions which address the widest range of geographical enquiry, and we feel sure that there is much here to inform, intrigue and inspire those with interests in geospatial data handling.

The Proceedings form the record of this 17th annual GISRUK conference, held in Durham. As in every previous year, GISRUK can boast of visiting a new venue. Durham has a long and continuous history of excellence in geographical study, and along with sister departments in both Newcastle and Northumbria universities, it has contributed much to advances in geographical knowledge, both locally and globally. The environment of North East England is a stimulating and exciting one in which to ply one's trade as a geographer, and the adoption of GI methods, technologies and concepts has led to high-quality research (as recognised by our most recent RAE results) in North East universities, in all areas concerned with geospatial data. We are grateful to Durham University for acting as host for this meeting. The local organising committee – Dr Chris Dunn (Durham), Dr Bruce Carlisle (Northumbria), Dr Seraphim Alvanides (Newcastle) and myself – have also benefitted from the professional assistance of Event Durham (notably Stina Maynard and Judith Aird) and David Hume who has produced these Proceedings.

As usual, we also thank the generous sponsors of GISRUK. At a time of economic belt-tightening, we are indebted to those sponsoring ogranisations featured on the back cover of this volume for their continued support of GISRUK.

In addition, I am grateful to the reviewers of the papers contained in this volume. The excellence of the papers presented in these pages is testament both to the authors and to the reviewers who have given their time and expertise in suggesting improvements. Over 100 abstracts were submitted and we feel that a high-quality conference has resulted from the reviewing process. The standard format for GISRUK publications places high demands on authors – the limit of 1500 words can feel constricting – but I hope that readers of this volume will find value in the concise descriptions and explanations provided. All this work re-iterates my first sentence - the vitality of Geographic Information Science research in the UK shows no sign of abating.

Welcome to Durham!

David Fairbairn
Chair, Local Organising Committee
School of Civil Engineering & Geosciences, Newcastle University

Personnel

Seraphim Alvanides [1, 3]
Katherine Arrell [2, 3]
Angela Baker [2]
Bruce Carlisle [1, 3]
Jane Drummond [2]
Chris Dunn [1, 2, 3]
David Fairbairn [1, 2, 3]
Bruce Gittings [2, 3]
Peter Halls [2, 3]
Glen Hart [2, 3]
David Hume [1]
Phil James [3]
David Lambrick [2, 3]
Jeremy Morley [2]
Nick Mount [2, 3]
Gary Priestnall [2, 3]
Nick Tate [2, 3]
Adam Winstanley [2, 3]
Stephen Wise [2, 3]
Jo Wood [2, 3]

[1] Local Organising Committee
[2] GISRUK National Committee
[3] Reviewers

Contents

Session 2A ***Transport network analysis***

Session 2B ***Spatial clustering***

Session 3A ***GIS for health data***

Session 3B ***Spatial statistics for modelling***

Session 4A ***Public participation and GIS***

Session 4B ***Tools for spatial data handling I***

Session 10B *Social applications of GIS*

Poster session

The effect of travel distance on patient non-attendance at hospital outpatient appointment: a comparison of straight line and road distance measures

Mark Dusheiko[1], Peter J Halls[2] , William Richards[3]
[1] National Primary Care Research and Development Centre, Centre for Health Economics, University of York, Heslington York YO10 5DD.
Tel: +44 1904 321 40
Email: mad105@york.ac.uk
[2] University of York, Heslington York YO10 5DD
[3] Hull/York Medical School, University of York, Heslington York YO10 5DD.

KEYWORDS: Non-attendance, outpatients, travel distance, access, healthcare.

1. Introduction

Patient non-attendance at hospital outpatient appointment is a major concern for the health care system. It results in an inefficient use of health care resources due to underutilised capacity within hospitals. For patients, non-attendance can result in worse health outcomes, and is a costly externality to other patients who experience unnecessarily long waiting times (Sharp *et al*, 2001; Hamilton *et al* 2002). Non-attendance may also be an indication of access inequity and a contributing factor to the observed inequality in access and utilisation of specialist care, and health inequality (Morris et al, 2005; van Doorslaer, Koolman and Jones, 2002). In the UK NHS there are no patient charges for hospital care. But access to outpatient care imposes other costs on patients: they face a waiting time after being referred and incur travel costs and time costs on the day of the appointment. There is considerable variation in the travel distances faced by patients accessing specialist secondary care (Damiana et al., 2005). However, little is known about the effects of distance on patient use of outpatient services and in particular the extent to which distance affects non-attendance.

The accessibility of secondary care services is an important policy issue. Since January 2006 the UK NHS has introduced a policy of patient choice of secondary care provider with patients offered a choice of at least 3 hospitals and time for an outpatient appointment. The policy, however, assumes that patients are able to exercise choice, given the constraints on the geographic location of hospitals and their propensity to travel (Greener, 2007). Choice and access has been enhanced by increasing use of private sector hospitals and new diagnostic and treatment centres as well as a recent attempt to move some types of care into a community or primary care setting. These policies, however, are offset by the trend towards greater centralisation of secondary care provision (Mungall et al 2005.) and the creation of fewer and larger specialist hospitals (Maybin, 2007). This research will inform current policy about the travel distances faced by patients, and the extent to which distance affects their ability to access outpatient specialist care.

2. Objective

To estimate the association between travel distance and patient non-attendance at first outpatient appointment using straight line and road travel distances. We investigate whether distance has a differential effect between urban and rural residents and across patient subgroups, such as the elderly or socio-economically deprived. We examine the extent to which straight line distances and road distances differ in their explanatory and predictive power.

3. Data

A nationally representative sample of over 550,000 patients resident in an English Lower Super Output Area, registered with an English General Practitioner, and booked to attend a first outpatient appointment at an English NHS hospital trust was obtained from Hospital Episode Statistics (HES) data for the financial year 2005/06.. Post code grid references for the main hospital site of treatment and the population weighted centroid grid reference of the Lower Super Output Area in which the patient was resident were used to calculate straight line distances. Individual patient characteristics and appointment details (such as specialty of treatment) were obtained from HES. Non attendance was recorded if the patient cancelled or did not show up for the appointment Census data, Indices of Multiple Deprivation and benefit claims data from ONS was attributed to the patient at LSOA level. GP practice characteristics and quality of care obtained from General Medical Statistics and Quality and Outcomes Framework data.

3. Methods

We estimate a multilevel logistic regression model for the probability of patient non-attendance. The model is used to estimate the effect of distance on the probability of non-attendance conditional on observable individual patient level characteristics, GP practice characteristics and hospital effects. Fixed effects models used to control for unobserved hospital characteristics associated with non-attendance. We use a number of goodness of fit measures (pseudo R^2, information criteria, deviance residuals, positive predictive power in out of sample testing) to assess the goodness of fit of straight line and road travel distances. We also investigate differences in the effect of our distance variables between urban and rural areas.

Straight line, Pythagoran, distances can be effective measures in an immediate locality but cannot take into account geographical barriers. A measurement between, for example, Grimsby and Hull will take a path directly across the Humber estuary and make the distance appear much shorter than the reality experienced by the patients. Hitherto, measurements of actual travel distance, together with estimates of the journey time, have been limited in their application due to the slowness of the prevailing methodology. Shortest-path distances are almost always computed using Dijkstra's Problem 2 algorithm (1959), which computes the distance from a start node to the end node for every possible path and selects the shortest for the result. For networks as complex as the British road system, this can be very slow. A York study of some 800 cases, where both distance and travel time were estimated using ArcGIS took some 55 seconds per path to compute, 12 hours in total (Halls, 2008).

Car (1997) introduced the concept of Hierarchical Spatial Reasoning (HSR) to subdivide a complex road network into a set of hierarchical networks. Still employing Dijkstra's algorithm, she seeks to ascend the network hierarchy as quickly as possible and only descend on approach to the destination. This was shown by Car, Taylor and Brunsdon (2001) to deliver a significant improvement in performance over the traditional Dijkstra-based methodology. Halls (2008) reports the implementation of a methodology based on Car (1997), Dechter and Pearl (1985) and O'Connor (2002). This replaces the Dijkstra algorithm in Car's HSR approach with the *best-first* A* algorithm and calculates travel time alongside distance. He reports computing the 800 case problem, which took 12 hours in ArcGIS, in just 3 minutes. This kind of performance makes the computation of 'actual' road distance and times a practical proposition for studies involving hundreds of thousands of cases, such as this.

The calculation of road distances and times from LSOAs to outpatient departments is performed by taking the five Postcodes within the LSOA and nearest the centre and north, south, east and west extremities. This takes account of the larger, rural LSOAs where entry points into the major road network differ across the area. The calculation for the central Postcode will be for a similar location to that used for the straight line Pythagoran distances for urban environments, although it may differ for sparsely populated rural LSOAs.

4. Results

Preliminary analysis on an English national sample of 586040 outpatient appointments estimated average straight line travel distance to be 9.5km, road distance 14.8km and travel time 11.6 minutes. Figure 1 plots road distances against straight line distances. There is a high degree of correlation (0.95) but road distances tend to be higher than expected given the straight-line distance. 10% of individuals did not attend their first outpatient appointment. The probability of non-attendance was highest in adult mental health and lowest in radiology. Initial estimates on the full sample found that all distance measures had a significant positive effect on non-attendance. A 100% increase in straight-line distance was associated with a 7.3% (95% CI: 5% - 9.6%) increase in the odds of non-attending. A 100% increase in road distance was associated with a 6.4% increase in the odds of non-attendance (95% CI: 4.4% - 8.4%), and a 100% increase in travel time was associated with a 6.3% (95% CI: 4.3% - 8.3%) increase in the odds. The model using straight line distance had the best model fit.

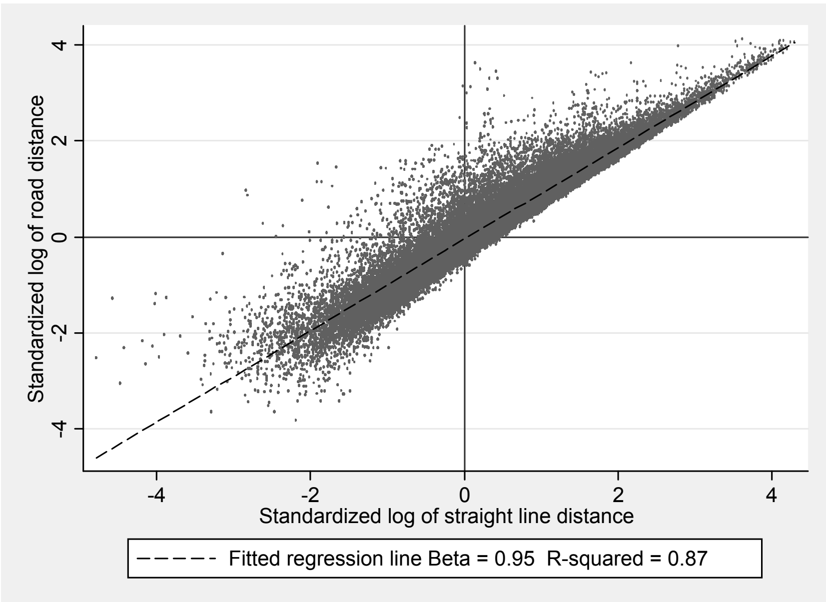


Figure 1. Scatter plot of standardized logarithm of road distance against the standardized logarithm of straight line distance

5. Discussion

Our initial analysis suggests that travel distance is a significant predictor of non-attendance. The magnitude of the effect, however, is small and very similar across different distance measures. We intend to extend the analysis by comparing the effects of distance between urban and rural communities as well as for economically deprived communities, and those with low car ownership who rely on public transport.

Our work with road distances was performed initially using the Bartholomew digital road map data, supplied by HarperCollins Publishers. Work is underway to replace this with Ordnance Survey Integrated Transport Network data, which will enable more realistic route determination through the provision of more detailed junction information. In future we intend to incorporate public transport networks and to incorporate travel cost explicitly in the analysis.

5. Acknowledgements

NPCRDC and CHE receive funding from the Department of Health (DoH) for policy research. The views expressed are those of the authors and not necessarily those of the DoH. We thank the HES team and the Information Centre for Health and Social Care for providing the hospital records data.

References

Car A (1997), *Hierarchical spatial reasoning: theoretical consideration and its application to modelling wayfinding*. Published PhD Thesis: Department of Geoinformation, Technical University of Vienna.

Car A, Taylor G. and Brunsdon C (2001) An analysis of the performance of a hierarchical wayfinding computational model using synthetic graphs *CEUS* **25(1)** pp69-88.

Damiani M, Propper C and Dixon J (2005) Mapping choice in the NHS: cross sectional study of routinely collected data. *BMJ* **(330)** pp284

Dechter R and Pearl J (1985) Generalized Best-First Search Strategies and the Optimality of A* *Journal of the Association for Computing Machinery* **32(5)** pp505-536

Dijkstra EW (1959) A note on two problems in connection with graphs. *Numerische Mathematik* **(1)** pp269-271.

Greener I (2007) Are the assumptions underlying patients choice realistic?: a review of the evidence *British Medical Bulletin* **83(1)** pp249-258

Halls PJ (2008) Assessing Variations In Access To Services Through Travel Time *Annals of Geomatics* **VI(1)** pp97-106

Mungall IJ (2005) Trend towards centralisation of hospital services, and its effect on access to

Hamilton W, Round A and Sharp D (2002) Patient, hospital, and general practitioner characteristics associated with non-attendance: A cohort study. The British journal of general practice **52(477)** pp317-319.

Maybin Jo (2007). *The reconfiguration of hospital services in England*. London: Kings Fund Available online at http://www.kingsfund.org.uk/publications/briefings/the_1.html (last accessed on 29th February 2008).

Morris S, Sutton M and Gravelle H.(2005), Inequity and inequality in the use of health care in England: an empirical investigation. *Social Science & Medicine* **(60)** pp1251–1266

O'Connor C (2002). *Wayfinding for Footbal Fans: Spatial Algorithms and Data*. Unpublished MSc Dissertation, Department of Computer Science, University of York.

Sharp D and Hamilton W (2001) Non-attendance at general practices and outpatient clinics *BMJ* **(323)** pp1081 -1082.

Van Doorslaer E, Koolman X and Puffer F (2002) Equity in the use of physician visits in OECD countries: has equal treatment for equal need been achieved? in *Measuring Up: Improving Health Systems Performance in OECD Countries*. OECD, Paris.

Biography

Mark Dusheiko obtained a B.Sc in Economics from the University of Warwick in 1997 and a M.Sc. in Health Economics from the University of York in 1998. Since then he has worked as a research fellow at the National Primary Care Research and Development Centre, Centre for Health Economics, University of York.

Peter Halls is the University of York's GIS Advisor and has many years experience in the GIS arena.

William Richards obtained a B.Sc. in Economics from the University of Bristol and is currently a 3rd year medical student at the Hull York Medical School.

Accessibility Analysis by Generalised Cost in a GIS Framework

Alistair Ford, Stuart Barr

School of Civil Engineering and Geosciences, Cassie Building, Newcastle University,
Newcastle upon Tyne. NE1 7RU
Tel. (+44) 0191 222 7121
a.c.ford@ncl.ac.uk, www.ceg.ncl.ac.uk

KEYWORDS: accessibility, networks, transport, generalised cost

1. Introduction

The Tyndall Cities Programme is a major UK research project examining the impacts of climate change on cities of the future and undertaking an integrated assessment of the adaptation and mitigation strategies available to planners and managers. As an integral part of this project, a land-use transport interaction model is used to allow the forecasting of future population and employment patterns within the city of London, and thus predicting the impact of such change on land-use, flood risk, water scarcity and urban heat. This paper describes the work undertaken as part of this programme to develop a set of spatial accessibility measures for transport throughout London, by a number of different modes. The spatial accessibility measures developed and their corresponding analytical routines are used within the Tyndall Cities programme to characterise a suite of potential transport futures for the city of London. In this paper we present results that show how several different infrastructure investment futures modify the spatial accessibility cost across the city of London at an inter-ward level.

2. Accessibility Measures

The measurement and prediction of transport accessibility is key to many applications in urban planning, including the assessment of future transport infrastructure and land-use modelling (Liu and Zhu, 2003). The measurement of accessibility can be seen as the attempt to characterise the cost of travel between an origin and a destination in space, often seen as being inversely-proportional to the distance, time or cost of travel between two locations (Miller, 1999). The origins and destinations in this study were selected to be the 2001 Census Area Statistics (CAS) wards for the Greater London Authority (GLA) area, giving 633 zones to and from which to measure accessibility. These zones were chosen due to the readily-available data relating to population (from the Census) and employment (NOMIS) data which could be used in the land-use transport interaction model.

Since a number of different modes of transport, both public and private, were to be considered in this study, it was determined that using a simple time or distance measure would not fully characterise the cost of travel between two locations; often there is a choice between a slower but cheaper mode and a faster but more expensive one. Therefore, the cost of travel was expressed as Generalised Cost, taking into account both time and monetary components of any journey in one unified value. The UK Department for Transport's 'Transport Analysis Guidance' website WEBTAG (WEBTAG, 2008) defines Generalised Cost as "the sum of both the time and money cost" for a journey. The notion of 'Value of Time', assigning a monetary cost for a unit of time and vice versa, allows the direct comparison and combination of both components of journey cost.

3. Computing Generalised Cost

The computation of Generalised Cost differs between public and private transport. When travelling in private transport, all of the costs of the journey are incurred directly by the individual, including the cost of fuel, maintenance of the vehicle, parking costs and the time of travel. Public transport charges the user a ticket cost which covers all monetary components of the journey, with only the time components requiring calculation. The cost of private travel is often very closely correlated to the distance of the journey (Ortuzar and Willumsen, 2001), since most of the monetary components of the Generalised Cost are incurred per unit distance (see below). The cost of public transport is often more closely linked to the time of the journey, since any monetary costs tend to be paid once at the start of the journey. Fares tend to be paid for a journey dependent on geographical zones, time limits or flat fares for any length of journey. Thus the monetary cost is more dependent on the location of the origins and destinations and the routes chosen between (*ibid*).

Equation 1 gives the WEBTAG recommended formulation for the Generalised Cost of a private journey, G_{car}, typically by car, expressed in units of time.

$$G_{car} = (V_{wk} * A) + T + D*VOC/(occ*VOT) + PC/(occ*VOT) \tag{1}$$

Where:

A = Access time to network (walk time to car);

V_{wk} = walking weight;

T = journey time in car;

VOC = vehicle operating cost per km;

VOT = Value of Time;

D = distance in km;

PC = Parking Costs; and,

Occ = number of vehicle occupants.

Equation 2 gives the WEBTAG recommended formulation of the Generalised Cost of a journey by public transport, G_{PT}, expressed in units of time.

$$G_{PT} = (V_{wk} * A) + V_{wt} * W + T + F/VOT + I \tag{2}$$

Where:

A = access time (walk time) to network;

V_{wk} = walking weight;

V_{wt} = waiting weight;

W = total waiting time for the journey;

I = interchange penalty (not modelled);

F = fare paid for journey;

VOT = Value of Time; and,

T = journey time of transport.

It can be seen that the Generalised Costs of both private and public journeys are reliant upon both the access time to the network and the journey time incurred whilst travelling in the vehicle. The access time is weighted by a value to represent the disincentive of a person to walk a great distance to access a transport service. This weight is typically one to two times higher than the perceived cost of in-vehicle time (WEBTAG, 2008). Both the access time and journey time components of the Generalised Cost for all modes were computed within a GIS framework. There are a number of possible additional components to generalised cost, such as capacity of vehicles, which are not included in this study but which are seen as very important next steps for the development of such work.

4. GIS Network Models of Transport Modes

ArcGIS Network Analyst was employed to construct network representations of the four modes of transport used in this study; heavy rail, light rail (Tube, Docklands Light Railway, Tramlink), bus and private car (road). Each of these modes was modelled separately, with interchange between modes not represented. Allowing interchanges would have provided a more accurate model of the transport system as a whole, but this work required the flexibility to implement differing future scenarios for each mode and thus the modes were kept as separate networks. These networks were utilised to analyse shortest paths between each pair of origin and destination zones in terms of time, and then VBA scripts developed within ArcGIS to incorporate these into the full generalised cost calculations outlined above. Evaluating the shortest paths in time terms ensured that the speed of the journey was taken into account even if the distance was not the shortest possible. The distance of each least time path was recorded alongside its total travel time.

The generalised costs for heavy and light rail were both computed using the same method, since their network architecture and patterns of behaviour are very similar. Each mode was modelled by a network of lines and stations, both being derived from OS Strategi 1:250,000 scale data in the case of heavy rail, and from TfL's own data for light rail. Least cost paths were computed between zones via stations on the network using Network Analyst (which employs the Dijkstra Algorithm).

Journeys from an origin ward (described by its centroid) must first reach a railway station in order to join a rail service. This is assumed to be the nearest station in order to simplify computation, although it is recognised that the closest station may not always offer the least cost result for travel (i.e. the walking distance to the station may be smaller but the on-train time may be greater). In this study the distance is assumed to be a straight-line journey from centroid to station at a given walking speed. Once at the origin station, the journey is then traced along the rail network to the destination station, adhering to any speeds, and therefore times, thereon. The journey from the rail network to the destination zone centroid follows the same rules.

Whilst this simple case makes computation straightforward, there are a number of special cases which are not catered for in this representation. 10% of zones in London have more than one heavy station within it and over half have no station at all. For light rail, one in three zones has more than one station and one in five has no station. Therefore, provision must be made for cases other than the simplest one outlined above. If both the origin and destination zones have less than two stations then the computation is as with the simple case, however if either the origin or destination zone, or both, has two or more stations within it then all possible route combinations between the zones are computed and the average distance or cost computed.

Generalised Costs for the road mode were calculated assuming all journeys are made by car (since walking and cycling are not included in this work). As with the heavy and light rail modes above, a network dataset was created within ArcGIS from Ordnance Survey ITN data (RoadLink and RoadNode feature types). For this work only links with a value of Motorway, A Road, B Road and Minor Road in their *Descriptive Term* attribute were included in order to reduce complexity and ensure shortest paths were not directed through local streets. Since these Local Streets were removed from the modelling, access to the network was from a zone's centroid to the nearest road link in a straight line.

The shortest path through the network was computed from the nearest access point to the origin zone centroid, to the corresponding nearest point in the destination zone. Journey times for each link in the network were calculated based on London Travel Report average speeds, and the shortest route based on time, rather than distance, was computed. The distance of this journey was also recorded, however, allowing input into the fuel and non-fuel components of the Vehicle Operating Costs in computation of Generalised Cost of private travel.

Generalised Costs for the bus mode were computed based on a network created from data supplied by Jacobs Babtie Consultants. This network represented the journey times and distances between major stops along all bus routes in London. Once created, the network was traversed in much the same way as the road mode described above. The Generalised Costs for the bus mode did not include Vehicle Operating Costs, however, the only monetary cost being a £1 fare regardless of distance travelled.

5. Results of Generalised Cost Computation

The resultant Generalised Costs for all modes were output as Origin-Destination matrices with 633*633 entries, each entry being the Generalised Cost between the origin on the row and the destination in the column. Intra-zonal costs were output on the diagonals where appropriate; for the road and bus modes they were computed as ¾ of the minimum cost to another zone, whilst for rail and light rail modes they were only computed if at least two stations are present within a single zone. The outputs can also be written as Origin-Destination pairs, allowing their display within the GIS by joining to the original zonal geography.

A number of differing future transport scenarios were utilised to test the cost computation framework, implementing a number of the Transport for London T2025 investment plans (TfL, 2006). These allowed various levels of expenditure, with corresponding infrastructure improvements to the current transport networks, to be defined and the effects on the overall Generalised Cost of accessibility to be investigated. A comparison of the results of one such investment scenario is given in Figure 1. It can be seen clearly that areas of increased investment (the new Light Rail lines in red and yellow) reduce the Generalised Cost of travel to neighbouring zones.

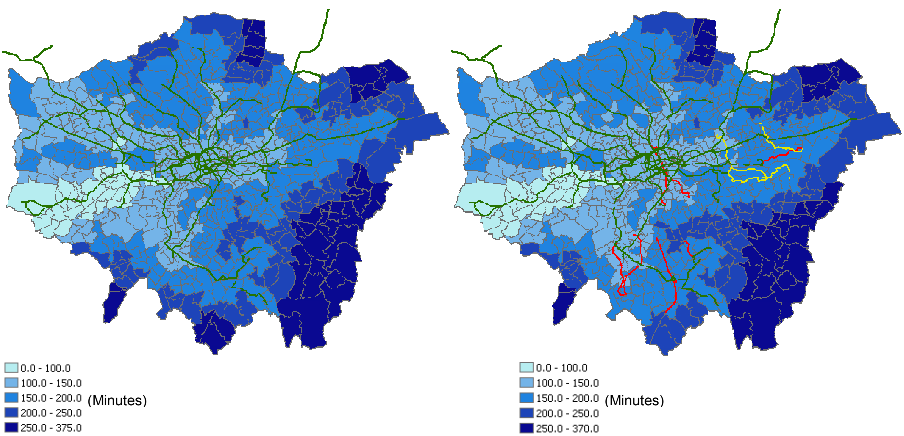


Figure 1. Comparison of baseline and high investment scenarios in the Light Rail mode showing the effects on patterns of Generalised Cost in minutes.

6. Conclusions

This paper presented a GIS framework for the computation of Generalised Costs of travel from origin and destination zones along transport networks. The computation of such costs was outlined, and the implementation of a framework utilising publically-available data to facilitate this computation presented. A number of special cases were highlighted, along with the means of catering for them within the computation framework. Some examples of the resultant calculated costs were given in the context of future infrastructure investment and changes in the underlying transport networks.

7. Acknowledgements

This work was completed as part of the Tyndall Centre for Climate Change Research's Cities Programme, alongside Professor Michael Batty and Stephen Evans of UCL's Centre for Advanced Spatial Analysis. Professor Jim Hall and Dr Richard Dawson of the School of Civil Engineering and Geosciences also provided vital guidance and assistance.

References

Liu S, Zhu X (2004), "Accessibility Analyst: an integrated GIS tool for accessibility analysis in urban transportation planning" *Environment and Planning B: Planning and Design* **31**(1) 105 – 124

Miller, H J (1999), Measuring space-time accessibility benefits within transportation networks: Basic theory and computational procedures. *Geographical Analysis* 31: 1–26

Ortuzar and Willumsen (2001), Modelling Transport, John Wiley and Sons, pp178-198

Transport for London (2006), T2025: Transport vision for a growing world city, http://www.tfl.gov.uk/assets/downloads/corporate/T2025-new.pdf

WEBTAG, (2008), Transport Analysis Guidance, http://www.webtag.org.uk

Biography

Alistair Ford is a research assistant in the School of Civil Engineering and Geosciences at Newcastle University (UK) investigating climate change in cities through GIS. He is currently part of the Tyndall Centre for Climate Change Research investigating land-use change in London to 2100 and the effects of this on water resources, flooding and emissions.

Dr Stuart Barr is a lecturer in GIS in the School of Civil Engineering and Geosciences at Newcastle University specialising in computational urban systems and urban Earth observation. His research interests include spatial, structural and statistical pattern recognition for integrated GIS/EO analysis and urban land-use and transportation modelling.

A Method For Visualisating Journey to Work Patterns Based on Choropleth Mapping and a Physical Analogy

Chris Brunsdon[1], Jon Corcoran[2], Jing Li[1]

[1]Department of Geography,
University of Leicester,
University Road,
Leicester LE1 7RH
cb179@le.ac.uk

[2]University of Queensland Social Research Centre
Level 4, General Purpose North 3 (Bld 39A),
St Lucia Campus,
University Of Queensland,
QLD 4072

KEYWORDS: Travel to work, Spatial Interaction, Visualisation, Flow Mapping,Queensland

1 Introduction

The visualisation of journey-to-work patterns has long proved difficult, largely because such patterns are formed from attributed four-dimensional data. That is, any origin-destination pair consists of a pair of geographical coordinates $\{(x_1,y_1),(x_2,y_2)\}$ with the attribute of the number of trips made from the origin to the destination. On a map, each trip can be represented as a line[1] , whose thickness can indicate the number of trips, but graphics of this kind can become easily cluttered. Suppose there are n spatial units used to count both the origins and destinations. Then there are n^2 origin-destination pairs and potentially this many lines on a graphic. If one is analysing, say, ward based data in Greater London, UK or SLA-based data in Queensland, Australia the number of lines will run to tens of thousands. Although there are some ways in which this problem may be overcome - for example only drawing lines where the trip count exceeds some threshold, or overlaying drawing lines with partial transparency, it is felt there these approaches could be usefully augmented by some alternative approach. A suggestion for one such approach is outlined here.

2 A Physical Analogy

Travel to work data is essentially information about flows - in this case flows of people along particular ‘channels’ that connect the origin to the destination. A number of physical situations also involve flows - for example these occur in fluid mechanics, and also in the study of electricity. Here an analogy is made to the latter example. If the area under study is thought of as an electric circuit, then one could imagine wires joining each origin to each destination, and the travel-to-work flows as being electric currents flowing in these wires. Using this analogy, a system of travel-to-work patterns might be represented by the diagram

[1]...or possibly an arrow to indicate direction, since swapping origin and destination results in a different trip pattern

in Figure 1.

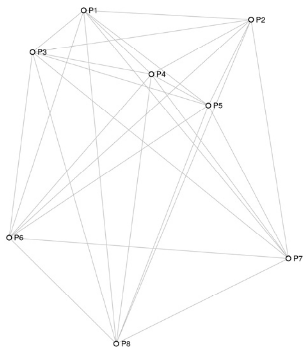


Figure 1: Journey-to-Work System as a Wiring Diagram

The points $\{P_1...P_8\}$ are situated at the centroid of each geographical zone, and each wire (shown as a grey line) connects one each pair of points. The wires are assumed to be insulated, except at the end points, so that no electrical contact is made between a pair of crossing lines. The idea of this analogy is to find a set of voltage potentials $\{V_1....V_8\}$ at each of the points that will cause currents to flow in the 'wires' in proportion to the net flows of journeys to work between each pair of locations. More generally, it would be desirable to do this for $\{V_1...V_n\}$ in an *n*-zone system. For a given wire connecting zone *i* to zone *j*, Ohms law gives the current in the wire as

$$I_{ij}=(V_i-V_j)/R \tag{1}$$

where I_{ij} is the current (in Amperes) and R is the resistance of the wire (in Ohms). Furthermore, the resistance R is proportional to the length of wire:

$$R\mu d_{ij} \tag{2}$$

and so

$$I_{ij}\mu(V_i-V_j)/d_{ij} \tag{3}$$

where d_{ij} is the distance between P_i and P_j. Recall that it is only required that the current is only *proportional* to journeys to work, so that

$$F_{ij} \mu (V_i - V_j)/d_{ij} \quad (4)$$

where F_{ij} is the flow of commuters between zones i and j. Since it is only required that the modeled currents to be *proportional* to the flows, any convenient constant of proportionality may be chosen. Here, for convenience, the value 1 will be used and thus the equation

$$F_{ij} = (V_i - V_j)/d_{ij} \quad (5)$$

will apply for the remainder of this article.

An equation of this form exists for each wire, and although it is not generally possible to solve all of the equations perfectly[2] , this method attempts to find $\{V_1,...,V_n\}$ to satisfy these equations as closely as is possible.

Essentially, greater potential difference across a wire leads to greater amounts of flow. Considering all of the points where voltage potential is given, in general flow is from higher potential places (sources) to lower ones (sinks). Since there is just one voltage potential attached to each zone, choropleth maps of voltage potentials should give an overview of which places are sources of trips, and which are sinks. Although this may not reflect the full richness of journey to work information in a region, experiment suggests that if this method is used in conjunction with existing approaches, it provides a useful way of uncovering trends, and of comparing changes in pattern over time.

3 Example

Here, we demonstrate the approach using journey to work data for the Queensland region, based on journey-to-work counts between 300 Statistical Local Areas (SLAs). Counts are available for 1996, 2001, and 2006. For each year, there are 90,000 journey-to-work counts, as each SLA can be an origin or a destination. Potential maps for the three time periods are shown in Figure 2. These highlight a number of ‘sink’ potential regions, which attract commuters (shown in red) and a number of blue ‘source’ regions, generally hinterlands, that provide commuters. Intensity of colour represents strength of potential of either sign, with the same colour scale applying for all three maps. One notable, small subsystem appears in the north, with Gympie as a source and the surrounding region as a sink. Although present in all three time periods, this is notably more marked in 2001 and 2006.

[2]essentially because there are $n(n-1)/2$ equations and only n voltage potentials.

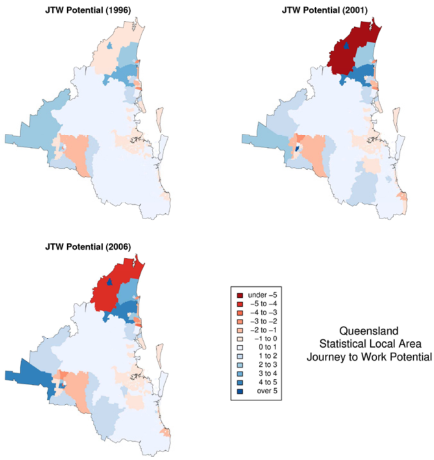


Figure 2: Potential Maps for Queensland Journey to Work Areas

The relationship between Gympie and its surrounding region is shown in more detail in Table **Error! Reference source not found.**. From this, it can be seen that there is a jump in net flow from Gympie to the surrounding region between 1996 and 2001, with flow to the surrounding region increasing by around 50% in the 1996 to 2001 period, and flow in the opposite direction decreasing dramatically from 1,285 in 1996 to 248 in 2001. This trend reverses to some extent in 2006, to give 404 trips, but even this is below one third of the 1996 figure.

Table 1. Journeys to work for Gympie (SLA code 315102535) and the surrounding region (SLA code 315102532)

Year	Origin	Destination	
		315102532	315102535
1996	315102532	2,362	1,285
	315102535	1,977	3,591
2001	315102532	2,012	248
	315102535	2,903	4,809
2006	315102532	2,724	404
	315102535	2,919	5,035

The potentials for all three sets of flows are also illustrated in a scatterplot matrix (Figure 3). The diagonal line indicates the relationship $x=y$, that is, observations lying on this line show no change between the two time periods. The outlying points on the 1996 vs. 2006 and 1996 vs. 2001 scatterplots correspond to the figures for Gympie discussed above. Also of note is the fact that the approximate slope of the points plotted on the 1996 vs. 2006 matrix, which appears steeper than the $x=y$ line. This suggests that typically larger potentials in 1996 have got even larger in 2006, and smaller potentials have got even smaller - in short the ‘sourceness’ or ‘sinkness’ of areas has generally increased over the decade from 1996 to 2006.

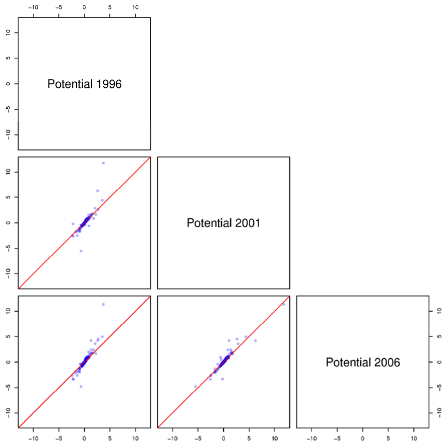


Figure 3: Potential Scatterplot Matrix for Queensland Journey to Work Areas

4 Conclusions

The method of potentials outlined here may be thought of as an exercise in data reduction, or projection. There are n^2 flows, but this technique attempts to summarize this information in n data items. This is helpful, as interactions between pairs of data points are difficult to visualise, but this method allows the information to be depicted a a choropleth map. In the example, it was shown how the approach can identify individual patterns (the changes in the journey to work patterns around Gympie) as well as long term trends (the illustration of the intensification of effects). These observations would be harder to make purely on inspection of the raw journey to work tables. However, as the method is in essence a projection from n^2-dimensional space to n-dimensional space, some information will inevitably be lost. In particular, the method only considers *net* travel to work flows, rather than those is specific directions. Also, information relating to journeys to work where the origin and destination are in the same SLA is not considered in this analysis. Initially it is proposed to consider the latter simply by mapping the degree of ‘containment’ that each SLA has - that is, the percentage of journeys to work originating in a given SLA who destination is also in that SLA. In addition, to consider in-flows and out-flows from SLAs separately, it is proposed to apply the potential-based approach outlined here to the in-flow and out-flow matrices separately. Finally, it may also be of value to consider population (or population in employment) based cartograms to visualise the potential maps, so that greater detail may be given to more populated, urban areas, where the physical extents of SLAs are notably smaller.

Biography

Chris Brunsdon is currently Professor of Geographical information at the University of Leicester. His interests include the development of geocomputational algorithms, and spatial analysis approaches and statistical inferential techniques to a number of 'real world' problems in human and physical geography - such as the analysis of house prices, crime patterns, and the monitoring of water quality.

Flow-based geographies in North East England

John Mooney

NERIP
Stella House, Goldcrest Way
Newburn Riverside
Newcastle upon Tyne NE15 8NY
Tel: (+44) 191 229 6810
Fax: (+44) 191 229 6201
Email: john.mooney@nerip.com
www.nerip.com

KEYWORDS: Commuting Flows, Urban / Rural Definitions, City Regions

1.1 Introduction

In a recent OECD territorial review, concerns were expressed about the level of spatial awareness amongst policy-makers in North East England. The region's major governmental institutions asked their regional observatory, NERIP, (North East Regional Information Partnership) to come to an understanding of how the different urban centres, and the more rural parts of the region, interact with each other. A set of linked reports (NERIP 2006a; 2006b; 2007a; 2007b) was subsequently produced, supported by novel geographies of the region.

The geography of a country or region can be conceived of not just as a set of places but also as a collection of flows between these places. This conception underpinned the work outlined in this paper, which was largely based upon the 2001 Census commuting matrices. The following is, of necessity, a brief overview - the report *Spatial Analysis of Economic Flows in North East England* (NERIP 2006b) contains more detailed information on the production of these geographies.

1.2 Urban / Rural Split

Production of the first report (NERIP 2006a) – which compares the rural and urban areas of the North East – required a simple, 3-way split of ward and higher-level geographies. As the focus of the larger set of reports was on the economic aspects of the region, it was felt that the geographical definitions used should differentiate between urban areas, "core" rural areas and those rural areas having a strong connection to / dependence upon urban areas. The identification of the third category presented a particular problem as the official urban / rural geographies used in England - the Rural Definition (Bibby & Shepherd, 2004) and Classification (Shepherd, 2005) - are based on patterns of residential settlement, rather than the strength of economic ties.

The solution adopted used the Rural Definition as a starting point. The Definition distinguishes between one urban and a number of rural morphologies. The approach adopted by NERIP took this as a starting point, and began with a simple urban / rural split by collapsing all of the rural morphologies into a single category. This category was then subdivided according to the percentage of each individual area's out-commuters travelling to an urban area..

The three resulting classes of (in this case) ward are:

Urban	Urban in the existing definition.
Fringe	Not urban in the existing definition. At least 50% of workers in the ward commute to an Urban ward.
Rural	Not urban in the existing definition. Fewer than 50% of workers in the ward commute to an Urban ward.

The definition is easily understood and, as Figure 1 shows, gives an intuitively correct "picture" – the major urban areas being surrounded by a "halo" of fringe wards.

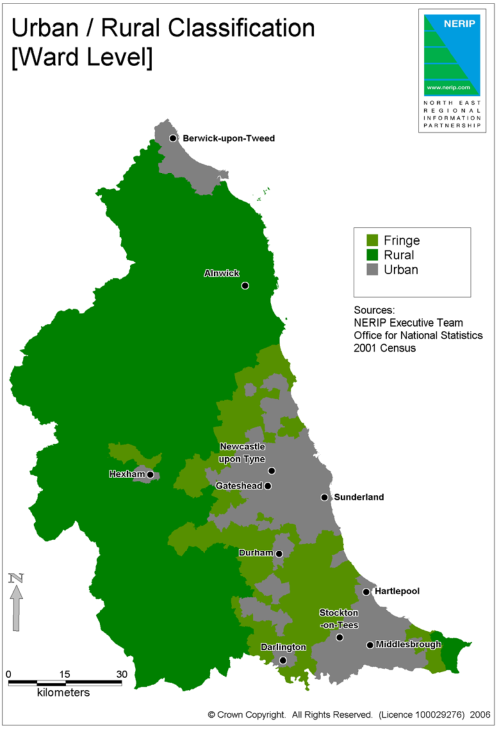


Figure 1.

1.3 Commuting Zones

A second report (NERIP 2006b) was largely concerned with economic flows around the region. Mapping commuter flows at a district level is unsatisfactory as it obscures localised patterns, but smaller geographies lead to unreadable, over-complex, flow maps. An innovative method of grouping small geographical units according to their commuting patterns was developed to establish a new large area geography that reflects the complexity of the region.

If flows between a large set of small areas are known then it is possible to apply a set of rules that group these small areas into larger ones according to the strength of flow between them. The Centre for Urban and Regional Development Studies (CURDS) at Newcastle University pioneered such work in the UK with Travel To Work Areas (Bond & Coombes, 2007). This high level geography is created by grouping small areas according to their commuting links, with the resulting groups satisfying criteria of minimum population size and self-containment (the proportion of commuters who work in the same area that they live / the proportion of jobs filled by people living in the same area).

The Commuting Zones (or CZ) have been produced by a similar method that does not impose criteria on the larger geographies produced. Instead, it requires that each small geographical unit (2001 Census wards, in this case) is placed within the group with which it has the strongest mutual commuting links.

Figure 2 shows the Commuting Zones of the North East. It is important to recognise that what shapes each CZ is not just the power (in terms of attracting commuters) of each individual town or city but the relative power and location of all such centres. Durham's CZ is constrained by nearby minor centres such as Peterlee and Spennymoor as well as by the major urban centres of Sunderland and Tyneside. Hexham, by contrast, has no local rivals and dominates a large area – despite being a much less significant commuter destination.

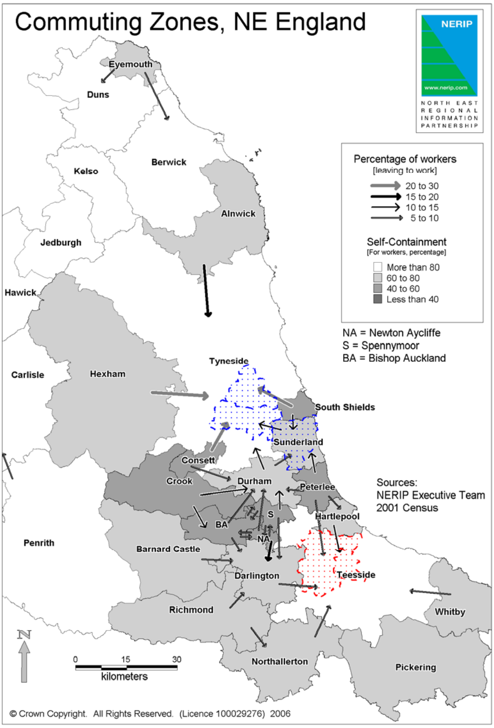


Figure 2.

1.4 Influence mapping

The final pair of reports (NERIP 2007a; 2007b) looked at the city-regions of the North East. City regions are not an established geography and are usually referred to in vague terms, but in this case the practicalities of data gathering meant that their boundaries had to be firmly defined. As part of this process, a series of maps based on rankings of destination choice were produced. These illustrate the patterns of influence of major centres in a polycentric environment.

A common method of defining city regions or metropolitan areas is to nominate certain areas as cores and then identify peripheries by examining the commuter flows into these cores from outside. In this type of "top-down" analysis the initial nomination of cores is a key step, but one that is not examined in this paper. For the analysis described here the cores suggested in a recent UK Working Paper (ODPM, 2006) - were used. These cores, made up of Local Authority Districts are shown as dotted areas on Figure 2. Where a core LAD fell within a or overlapped with the geographic units being used for analysis (in this case CZ), that entire zone was considered to be a core.

Having defined urban cores, the Working Paper then uses a threshold of commuter flow to establish which outlying areas can be thought of as the periphery. Instead of setting a threshold, NERIP have examined the ranking of flow destinations. This approach tries to account for the presence of flows to centres within the region other than those nominated as cores. The result, Figure 3, demonstrates the primacy of Tyneside and Teesside in the region, but also shows that they do not entirely dominate it. The most marked example of this is in the South West, where the influence of Teesside is limited by the presence of Darlington and Bishop Auckland. The case is less stark in the centre of the region, but Figure 2 demonstrates that there is a difference between, for example, Alnwick and Durham in terms of their relationship to Tyneside. Both have commuting flows into the larger zone, but – as Figure 2 shows - Durham is a destination in its own right. Considering Durham (or, to give another example, Sunderland) only as a "satellite" of Tyneside (which would be the case if using a threshold of commuter flow) neglects the city's relationships with other areas.

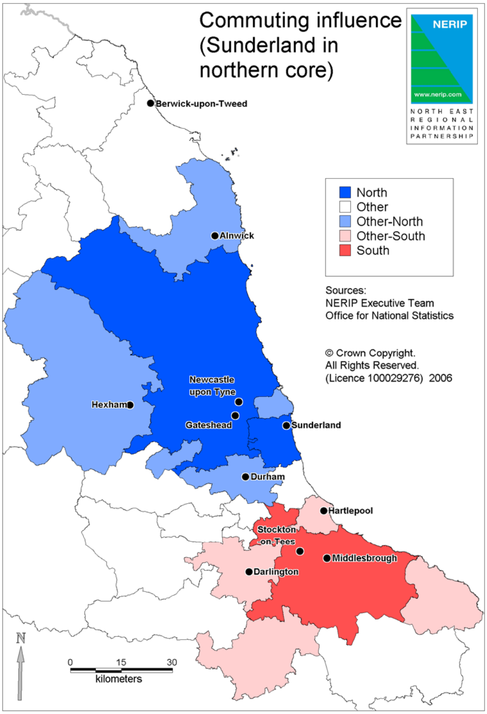


Figure 3.

The Influence mapping technique was also applied to commercially-supplied datasets illustrating travel to retail and leisure developments. The resulting maps, combined with the urban rural split shown in Figure 1, were used to derive a city-region geography for use in the reports.

1.5 Conclusions

The work presented here has direct application in the development of regional policy in the North East. UK regional policy focuses strongly on the economy and there is a consequent need for visualisations of economic activity. From a methodological perspective, the following points can be made:

- Attempts to draw a boundary of "urban" and "rural" are inherently subjective and sensitive to matters of scale. The three-way definition described here is intended as a complement to the official 2004 definition, to be used where the relationship to urban areas is of particular interest.

- The commuting zone methodology would benefit from further development, with a possible second stage providing internal detail for the very large zones.

- The influence mapping technique provides an easy to implement alternative to simple threshold mapping when examining the hinterland of major urban centres. The outputs reflect the influence of minor centres not nominated as destinations in their own right.

2. Acknowledgements

This work was undertaken using data supplied by the following organisations:
- Experian plc
- Office for National Statistics
- Ordnance Survey

3. References

Bibby, P. & Shepherd, J. (2004) *Developing a New Classification of Urban and Rural Areas for Policy Purposes - the Methodology* Office for National Statistics web publication, available from:
http://www.defra.gov.uk/rural/ruralstats/rural-defn/Rural_Urban_Methodology_Report.pdf
Bond,S. & Coombes, M. (2007) *2001-based Travel-To-Work Areas Methodology* Office for National Statistics web publication, available from:
http://www.statistics.gov.uk/geography/downloads/2001_TTWA_Methodology.pdf
ODPM (2006) A Framework for City-Regions - Working Paper 1: Mapping City-Regions CLG web publication, available from:
http://www.communities.gov.uk/publications/citiesandregions/framework2
NERIP (2006a) *Rural & Urban Comparisons In North East England* NERIP web publication, available from:
http://www.nerip.com/reports_briefing.aspx?id=223
NERIP (2006b) *Spatial Analysis of Economic Flows in North East England* NERIP web publication, available from:
http://www.nerip.com/reports_briefing.aspx?id=259
NERIP (2007a) *The Southern City Region - City Region Comparisons in North East England* NERIP web publication, available from:
http://www.nerip.com/Reports_Briefing.aspx?id=351
NERIP (2007b) *The Northern City Region - City Region Comparisons in North East England* NERIP web publication, available from:
http://www.nerip.com/Reports_Briefing.aspx?id=351
Shepherd, J. (2005) *Defra Classification of Local Authority Districts and Unitary Authorities - A Technical Guide* Defra web publication, available from:
http://www.defra.gov.uk/rural/ruralstats/rural-defn/LAClassifications_technicalguide.pdf

4. Biography

John Mooney formerly worked on the Norwich Union national flood map project. His doctoral research was on the generation of spatially and temporally detailed population data for use in the risk assessment of major industrial hazards, and was sponsored by the Health and Safety Executive.

Surnames as indicators of cultural regions in the UK

James A. Cheshire, Pablo Mateos and Paul A. Longley
Department of Geography, University College London, Gower Street, London, WC1E 6BT
james.cheshire@ucl.ac.uk, p.mateos@ucl.ac.uk, plongley@geog.ucl.ac.uk

KEYWORDS: Family Names, Regionalization, Spatial Clustering, Isonymy

1. Introduction

In recent years the geography of surnames has become increasingly researched in genetics, epidemiology, linguistics and geography (Colantonio et al., 2003, Mateos and Tucker, 2008). Despite this interest, there has never been a comprehensive study of the UK's regional surname patterns. This paper presents a tentative look at the geographic distribution of surnames in the UK, and undertakes a preliminary assessment of the regional geography that this reveals.

1.1 The regional geography of surnames

The need to define uniform regions is well established within geography (Vidal de la Blache, 1926, Hartshorne, 1939). Interest in automated regionalization methods can be traced back to the 1960s and 1970s (see Monmonier 1973 and Bunge 1966). In important respects, the development and application of these ideas was limited by the computing power available; a problem now overcome. This, along with the availability of data at finely disaggregated spatial units, has facilitated a renaissance of interest in regionalization techniques.

Most Anglo Saxon surnames remain spatially concentrated in the areas where they first came into popular usage in the 12th and 13th centuries. Zelinsky (1997) argues there are rich research opportunities through investigation of the geography of personal names and questions why so little has been done in this area. Studies of UK surnames began in the late 1970s (see Lasker, 1977), and continued during the following decades (see Mascie-Taylor and Lasker, 1984, Sokal et al., 1992, and Lasker, 1998). However, these studies were not national and only analysed samples of surnames. To date, there has been no national study of the regional geography of UK surnames. However, several methods for dialectometrics, genetics and linguistics have been developed and applied using recent national level data from a several countries; for example in the Netherlands (Manni et al., 2008) and Switzerland (Rodriguez-Larralde et al., 1998). This work applies, and briefly evaluates, a number of these regionalization methods for establishing UK surname regions.

1.2 Measuring name distance between areas

Lasker's coefficient of isonomy is widely used for surname studies within genetics to determine the degree of proximity between two geographical areas in 'surname space' (Colantonio, 2003). It is selected here as the base measure of surname similarity. Isonomy refers to possession of the same surname (Lasker 1977). The Lasker coefficient of isonomy compares the probability of this occurring within a given location with the probability of this occurring within another (Fox and Lasker, 1983). The method is conventionally used in population genetics to investigate the degree of in-breeding within populations. The coefficient of isonomy (R) is calculated as:

$$R_{xy} = 0.5 \sum_i x_i \, y_i \qquad (1)$$

where x and y are a pair of locations, i a particular surname, and x_i and y_i the frequency of each surname as a proportion of the total population in locations x and y.

In order to compare the surname characteristics of each location relative to all the others in a country, the Lasker coefficient (R) values are converted to a distance measure between each pair of locations.

Higher values represent locations further apart in 'surname space'. The *Lasker Distance* is calculated as the negative value of the logarithm of the Lasker coefficient of isonomy between localities (Rodriguez-Larralde et al. 1998):

$$L_{x,y} = -\log_e(2R_{x,y}) \qquad (2)$$

This metric produces a distance matrix between all pairs of locations in the study area, in which distances are symmetric (i.e. $L_{x,y} = L_{y,x}$) with a zero diagonal ($L_{x,x} = L_{y,y} = 0$).

2. Data and Methods

The surname frequencies used in this project come from the enhanced 2001 UK Electoral Register. In addition to the names and addresses for UK nationals aged 17 or over who are (or are about to become) eligible to vote in UK or European elections it includes data collected about other individuals not registered to vote, or sourced from commercial surveys and credit scoring databases. The data represent 45.6 million people resident in the UK in October 2001 comprising a total of XX million unique surnames. Further information and analysis on this dataset are provided by Longley et al. (2006) and McElduff et al. (2008).

The basic areal unit of analysis in this study is the District, an administrative area corresponding to the Local Authority level in the hierarchy of the UK local government. There are 436 Districts in the UK, with an average population size of approximately 105,000 inhabitants. Districts are considered to represent sufficient internal homogeneity in terms of cultural ties to be used as the basis of this regionalization exercise.

The Lasker Distances for each pair of districts were calculated across the XX million surnames following Equations (1) and (2) to produce a distance matrix of 436 by 436 values.

3. Analysis: Automatic detection of cultural regions

Initially, the *Monmonier* algorithm (Monmonier, 1973, Jombart, 2008) was used to establish the existence of what Manni et al. (2004) term 'barriers' to surname propagation within the population. The algorithm was implemented by creating a network of Delaunay triangles, used to calculate a Lasker distance between district centroids. The Monmonier algorithm then works from areas of 'steepest gradient', that is greatest distance, through a series of 'left-right' decisions until it reaches another boundary or the external boundary of the map (Manni et al., 2004). This process is iterated a number of times in order to produce a series of lines or barriers between areas of greatest Lasker Distance.

Second, Ward's (1963) hierarchical clustering method was applied to detect regional clusters of surnames. Ward's clustering is formulated in order to find compact, spherical clusters within data of certain dimensions (Gordon, 1987); in this case the two dimensional Lasker Distance matrix. The hierarchical clustering calculates a partitioning of the districts into clusters of similar districts (Kleiweg, 2006) - as shown in Figure 1. In Figure 2 the thematic map on the left represents Ward's (1963) clustering of all 436 Districts, and the one on the right treats the 32 London districts as a single entity (see below).

Third, undertaking multidimensional scaling (MDS) in two and three dimensions provided a comparison to hierarchical clustering. MDS widely used within genetic, surname and linguistic regionalization research (see for example Manni et al. 2008 and Smith et al., 1984), reduces data dimensionality. The technique, based on Tobler's First Law of Geography (Tobler, 1970), provides a statistical comparison of data by locating similar items in a multidimensional space (Wikelmaier, 2003). In addition to 2D (see Figure 3) and 3D plots, the results of the MDS were interpreted as a set of red, green, and blue values and used to plot a thematic map of isonomy (Figure 4). These procedures were completed in the R package iL04 (Kleiweg, 2004, R Development Core Team, 2008).

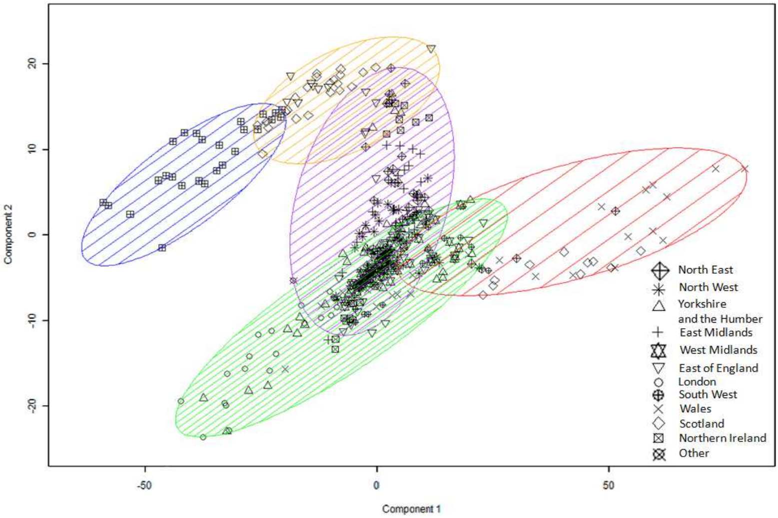


Figure 1: Ward's Hierarchical Clustering plot using data from all UK Districts.
The Ward's cluster tree was partitioned into 5 clusters represented by the shaded ellipses. Higher shading density represents more tightly clustered districts. Each district is given a symbol representing the Government Office Region it falls under. The 5 clusters correspond well to North/South England, Wales, Scotland and Northern Ireland.

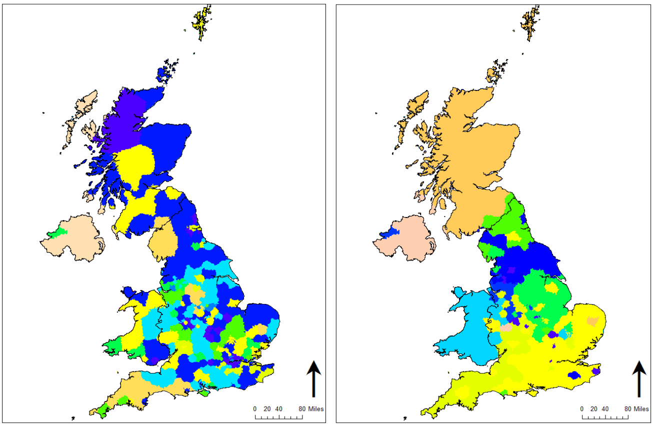


Figure 2: Maps of Ward's hierarchical clustering with 10 partitions with (left) and without (right) London.
The effect of the high migrant population in the London Districts is clearly illustrated by the amount of 'noise' in the left hand map. Removing London (right) presents a much clear picture of Anglo Saxon surnames The uniqueness of London Districts distorts the Lasker Distances of the rest of the UK (left), while amalgamating them to a single district them smoothes the patterns and reveals the regional surname structure of the UK (right).

4. Discussion

Despite its apparent effectiveness in a number of regionalization studies (Manni and Barrai, 2001, Manni et al, 2004, Manni et al., 2008), the *Monmonier* algorithm produced disappointing results in this context. London's diverse surname structure, consequent upon a large migrant population, created too much noise for a study of Anglo Saxon surnames. All the barriers were plotted around each London District and their contiguous suburbs. Aggregating the surname frequencies and populations of each of the London districts into a single, homogeneous, district reduced the noise, but the districts adjacent to London remained problematic. The longest barriers were detected between Northern Ireland and Great Britain, Scotland and England, Wales and England and South West England and the rest of the UK.

More effective methods in this context were multidimensional scaling (MDS) and hierarchical clustering. The latter method provided a general insight into the data and further highlighted the distorting effect of London on the Lasker distances of the UK. This can be seen in Figures 1 and 2. The Ward clustering results plotted in Figure 1 clearly distinguishes Northern Ireland, Scotland, Wales, North England and South England.

The application of a range of data reduction methods to the surname distances of the entire population across the UK unearths striking commonalities in surname structure within certain regions. The clearest examples in Figures 1 to 4 are Northern Ireland, Wales, Scotland, Northern England and London. The populations living in islands appear isolated from mainland Great Britain, as can be seen from the unique colour values ascribed to these areas in Figure 4. The principle of isolation by distance (Barrai et al., 2002) is also evident in Figure 4 as a colour gradient between South Western and North Eastern areas. A number of anomalous districts appear in the regionalization methods used here, such as the Wyre Forest District to the west of England.

5. Conclusion

This work is the first attempt to establish surname regions throughout the UK based on a near complete population analysis. This work provides much potential in studying population structure at different geographical scales in addition to establishing historical cultural regions and the extent of domestic migration; something the authors intend to research using 1881 British Census data.

In conclusion, by calculating the Lasker Distance for each surname occurrence in the 2001 enhanced electoral roll, and tentatively applying a number of regionalization techniques, this study has begun to unearth the underlying surname regions in the UK. These warrant further investigation as they may represent genetic and cultural variations within the UK population.

Acknowledgments

This work was undertaken as part of an ESRC CASE award in partnership with ESRI (UK) awarded to James Cheshire.

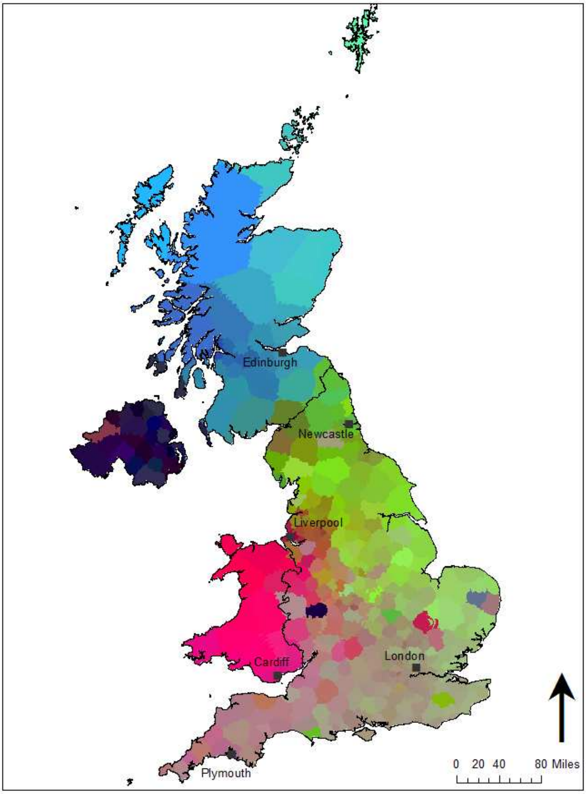


Figure 4: Map of three dimensional MDS values per district.
Colours are assigned by transferring the District's 3D coordinates in MDS space to a RGB colour cube. X, Y, Z on the MDS axes become R, G, B in the colour cube allowing a colour value to be assigned. The final colours in the map therefore represent the similarity of Districts in surname space (based on Lasker Distance) after MDS. The closer the colours in the RGB cube, the closer the Lasker Distances between districts.

References

Barrai, I. Rodriguez-Larralde, A. Manni, F. Scapoli, C. 2002. "Isonymy and Isolation by Distance in the Netherlands". Human Biology 74, 2: 263-283.

Bunge, W. 1966. "Gerrymandering, geography and grouping." *Geographical Review* 56, 2: 256-263.

Colantonio, S., Lasker, G., Kaplan, B., Fuster, V. 2003. "Use of surname models in human population biology: a review of recent developments." *Human Biology* 75, 6: 785-807.

Gordon, AD. 1987. "A Review of Hierarchical Classification" *Journal of the Royal Statistical Society. Series A (General).* 150, 2: 119-137.

Hartshorne, R. 1939. *The Nature of Geography*. Lancaster, Association of American Geographers.

Jobling MA. 2001. In the name of the father: surnames and genetics. Trends in Genetics 17(6): 353-357.

Jombart, T. 2008 adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403-1405. Fox, W., Lasker, G. 1983. "The distribution of surname frequencies". *International Statistical Review.* 51: 81-87.

Kleiweg, P. 2006. iL04: An interface to RuG/L04, software for dialectometrics and cartography. R package version 1.13. http://www.let.rug.nl/~kleiweg/L04/

Lasker, G. W. 1977. "A coefficient of relationship by isonomy: a method for estimating the genetic relationship between populations." *Human Biology*, 49: 489-493.

Lasker, G. W. 1998. "Calculating components of the Coefficient of Relationship." *Annals of Human*

Biology, 25, 5: 501-502.
Longley, P., Webber, R., Lloyd, D. 2006. "The Quantitative Analysis of Family Names: Historic Migration and the Present Day Neighbourhood Structure of Middlesbrough, United Kingdom." *Annals of the Association of American Geographers*. 97, 1: 31-48.
Manni, F. Barrai, I. 2001. Genetic Structures and Linguistic Boundaries in Italy: A Microregional Approach. *Human Biology*. 73, 3: 335-347.
Manni, F. Heeringa, W. Toupance, B. Nerbonne, J. 2008. "Do Surname Differences Mirror Dialect Variation?" *Human Biology* 80, 1: 41-64.
Manni, F., Guérard, E., Heyer, E. 2004. "Geographic patterns of (Genetic, Morphologic, Linguistic) variation: How barriers can be detected by using Monmonier's Algorithm." *Human Biology* 76, 2: 173-190.
Mascie-Taylor, C. G. N. and Lasker, G. W., 1984. "Geographical distribution of common surnames in England and Wales." *Annals of Human Biology,* 12: 397-401.
Mateos, P. and Tucker D. 2008. "Forenames and Surnames in Spain in 2004." *Names: A Journal of Onomastics* 56, 3: 165-184
McElduff, F., Mateos, P., Wade, A., Cortina Borja, M. 2008 What's in a name? The frequency and geographic distributions of UK surnames. *Significance*, 5(4) 189-192
Monmonier, M. 1973. "Maximum-Difference Barriers: An Alternative Numerical Regionalization Method." *Geographical Analysis* 5, 3: 245-261.
R Development Core Team.2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
Rodriguez-Larralde, A., Scapoli, C., Beretta, M., Nesti, C., Mamolini, E., Barrai, I. 1998. "Isonomy and the genetic structure of Switzerland. II. Isolation by distance." *Annals of Human Biology* 6: 533-540.
Smith, M.T. , Smith, B. L., Williams, W. R. 1984. "Changing isonymic relationships in Fylingdales parish, North Yorkshire, 1841-1881." *Annals of Human Biology* 11, 5: 449-457.
Sokal, R. R., Harding, R. M., Lasker, G. W. and Mascie-Taylor, C. G. N. 1992. "A spatial analysis of 100 surnames in England and Wales." *Annals of Human Biology* 19, 5: 445-476.
Tobler, W. 1970. "A computer movie simulating urban growth in the Detroit region." *Economic geography* 46, 2: 234-241.
Vidal de la Blache, P. 1926. *Principles of Humam Geography.* New York. Henry Holt and Company.
Ward, J. 1963. "Hierachical grouping to optimize an objective function". *Journal of the American Statistical Association* 58, 301:236-244
Wikelmaier, F. 2003. "An Introduction to MDS." http://perception.inrialpes.fr/~Arnaud/indexation/mds03.pdf. Accessed 23/02/09
Zelinksy, W. 1970. "Cultural variation in personal name patterns in the Eastern United States". *Annals of the Association of American Geographers* 60, 4: 743-769.

Biography

James is undertaking a PhD funded by a CASE award in partnership with ESRI (UK). He is based in UCL's department of Geography. James graduated with First Class Honours in Geography from the University of Southampton and his research interests include regionalization methods, GeoWeb2.0 and contemporary surname distributions.

Pablo Mateos is Lecturer in Human Geography in the Department of Geography at University College London (UCL). He obtained his PhD in Geography from UCL (2007) and MSc in GIS and Human Geography from the University of Leicester (2004).

Paul Longley's research interests concern the use of GIS and quantitative methods in urban analysis. He is Editor-in-Chief of Computers, Environment and Urban Systems and reviews editor of Environment and Planning B as well as Deputy Director of Centre for Advanced Spatial Analysis at UCL.

Using the Analytic Hierarchy Process to prioritise candidate improvements to a geovisualization application

David Lloyd, Jason Dykes, Robert Radburn
giCentre, City University London, EC1V 0HB
Telephone: +44 (0)20 7040 0212, Fax: +44 (0)20 7040 8584
Email: {d.lloyd | jad7 | sbbd476}@soi.city.ac.uk, Web: www.soi.city.ac.uk/~at775

KEYWORDS: geovisualization, analytic hierarchy process, crime and disorder reduction

1. Introduction

Crime and disorder reduction (CDR) research analysts ('analysts') in a UK local authority have generated candidate improvements for enhancing geovisualization prototypes designed using human-centred methods (Lloyd et al., 2007; Lloyd et al., 2008). Prioritising these is an important process and may require modification to established decision support techniques due to the nature of geovisualization. We explore this issue through the Analytic Hierarchy Process (AHP) (Saaty 1977) examining both the resulting priorities and their consistency.

2. Approach

Three crime analysts suggested ~350 explicit and implicit improvements to prototypes in seven experiments that enable analysts to explore crime attributes (absolute and relative numbers) spatially (using choropleth shading) and temporally (using glyphs). When coded and grouped the ~120k transcribed words yield 35 possible improvements. A clear task is to prioritise these possible improvements, initially unconstrained, and then in the context of limited development resource in order to direct development. Approaches to the first of these include multi-criteria decision analysis (MCDA) (Dodgson et al., 2000), GIS-based MCDA (Malczewski 2006), and the well established (Wasil and Golden 2003) Analytic Hierarchy Process which has been widely used in prioritising software development (Karlsson and Ryan 1997). AHP participants prioritise from a list by relating every possible pair of combinations. An overall score and ranking are produced for each item, along with a consistency ratio for each user.

Our 35 possible improvements would need too many pairwise comparisons for completion in a reasonable time. We reduce this number by grouping (Karlsson et al.,1997) and use pairwise group comparison to subsequently relate the group results. Analysts consider improvement groups in turn: 'data-related' (dealing with aggregation, filtering, context); 'interface-related' (system behaviour, complexity, speed); 'interaction-related' (readability, orientation, scale, legend) and 'new' (novel visualization tools and displays). Two analysts score preferences on each pairwise comparison within each of the four groups and then for the four groups themselves using an integer divergent scale (Karlsson and Ryan 1997). Comments made during the test are noted, and analysts asked about the process retrospectively. The perspective of 'geovisualization expert' ('expert') was provided by Dykes who had participated in the human-centered development process and undertook the AHP.

3. Findings

3.1 Quantitative findings

Marked similarities are noted in the rankings of the desirability of the 35 possible improvements prioritised by the two CDR analysts (Pearson coefficient 0.50, significant at 0.01 level; 2 tailed, n=35). This is not the case with expert's rankings, which are significantly different from both analysts'. Figure 1 shows analysts' and expert's rankings as parallel plot small multiples, conditioned by improvement group. Analysts' priorities are skewed towards 'data-related' improvements and against 'new' items. The expert's priorities are more evenly distributed, and incline towards 'interaction related' and against 'interface related' choices.

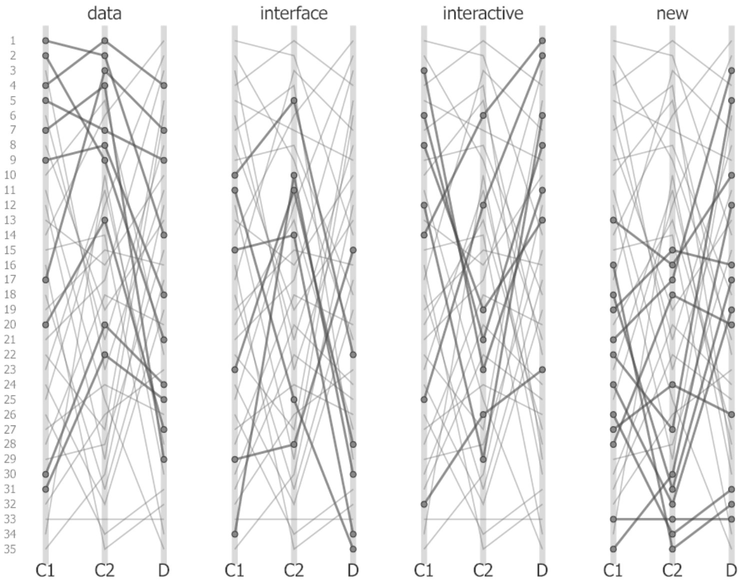


Figure 1. Parallel plots showing the rank of candidate improvements from 1 (top) to 35 (bottom), for each group, The same plot is shown four times with each group highlighted in turn: 'data' (10 lines highlighted corresponding to 10 'data' improvements), 'interface' (6), 'interactive' (7) and 'new' (12). Rankings for CDR analysts (C1 and C2) and expert (D) are shown left to right within each plot.

Saaty (1980) considers an AHP 'consistency ratio' of < 0.1 acceptable; "*in practice, however, consistency ratios exceeding 0.10 occur frequently*" (Karlsson and Ryan, 1997). Those achieved here range from 0.03 to 0.21 for data-, interface- and interaction-related possible improvements, but the results from the 'new-related' group are noticeably less consistent, ranging from 0.43 to 0.69. Analyst C1 is more consistent than the others throughout. C1's relative preferences across the 35 possible improvements are not as strong as those of C2 and D, as measured by the Gini coefficient (C1: 0.27; C2: 0.48; D: 0.42), calculated from Lorenz curves (Lorenz, 1905) of the same data.

Table 1. AHP consistency ratios for the four different groups and overall group comparison of the 35 possible improvements (low is more consistent).

User \ Group	'data'	'interface'	'interaction'	'new'	group comparison
CDR Analyst 'C1'	0.03	0.09	0.04	0.49	0.07
CDR Analyst 'C2'	0.06	0.20	0.21	0.69	0.06
Geovis expert 'D'	0.16	0.10	0.21	0.43	0.04

3.2 Qualitative findings

CDR analysts spent considerable time before the AHP exercise clarifying details of the 'new' candidate improvements. This resonates with our problems mediating geovisualization possibilities to these analysts (Lloyd et al., 2007) and parallels the difficulties experienced in identifying 'undreamed of' requirements (Robertson 2001). Comments made during the task include concerns about individual's consistency; concerns at the descriptions provided not differentiating sufficiently for some comparisons; and unprompted explanations being given for scores.

The CDR analysts found the AHP to be efficient and meaningful - preferred candidate improvements were successfully identified. The 'expert' experience was less positive - a tendency to focus on the process and one's own consistency rather than the detail of the improvements was noted through participation, as were difficulties in interpreting improvement descriptions consistently. Achieving consistency was an important aim for all users, and two of the three participants were concerned after awarding scores of '1' frequently in succession. One of the analysts summed up their understanding of the proposed tools as "*a guess on the back of what you are telling me*", indicating that earlier difficulties reported in mediating geovisualization to analysts continue and may be exacerbated by the coding and grouping required for AHP.

4. Conclusions

The two analysts have very different dispersions and different consistency ratios, but their rankings are indistinguishable, supporting the notion that the AHP is robust. The priorities of the expert are markedly different despite the high levels of engagement between analyst and expert throughout the human-centred development process. Geovisualization applications are predominantly 'expert' driven (Fuhrmann et al., 2005) and so the discrepancies in terms of priorities are an important finding that should be explored further with other analysts and 'expert' developers.

Given the poor consistency in ranking 'new-related' improvements, such rankings clearly cannot be relied upon to indicate priorities within this group. But the fact that 'new' candidate improvements are ranked inconsistently by all subjects suggests particular uncertainty about their nature and/or possible benefits. The issue may be one of communication and interpretation - unfamiliar improvements are more difficult to describe, communicate and interpret consistently with the coding, grouping and succinct descriptions required for pairwise comparison in the context of ~350 possibilities. Including the kinds of complex novel visual features typical of geovisualization as possible improvements may thus affect the working of the AHP. This is despite our efforts to expose the CDR analysts to geovisualization techniques and prototypes over an extended period and providing detailed descriptions prior to and during the AHP process. The time spent by the analysts at the outset and the qualitative data lend weight to this conclusion, confirming our earlier findings on difficulties in mediating geovisualization to these users (Lloyd et al., 2008). We also note the understandable focus of the analysts on prototype improvements that have the most bearing on their current activities rather than on innovation. This may be another limitation of the AHP, as we have previously observed these analysts being more open to innovation when not asked to prioritise - indeed all 350 candidate improvements were suggested by these users working with geovisualization prototypes in our human-centred design process (Lloyd et al., 2008).

Consequently, future application of AHP in geovisualization might variously:

- involve all parties in the AHP concurrently so that concepts can be discussed and interpretations clarified - AHP as a collaborative process to mediate shared understanding of priorities
- provide visual descriptions/stimuli with demos, videos or presentations prior to and during the process so that the candidate improvements are agreed
- use fewer, more specific, candidate improvements - sampling rather than aggregation
- run the AHP against different scenarios to establish (for example) current and future priorities
- weight the results by analyst based on criteria such as consistency (from the consistency ratio) or dispersion (from the Gini coefficient)

A variant of the classic knapsack problem allowed us to determine how the AHP output can help prioritise possible improvements under the constraint of different value solutions and developer costs. Results reveal that the analysts focus just as strongly on known functionality when development resources are limited, even when current tasks provide opportunity for beneficial geovisualization (Lloyd et al., 2007).

Whilst showing how a decision support technique can be successfully employed, we suggest that the nature of geovisualization may cause difficulties for those seeking to differentiate between candidate improvements, and may not provide an unambiguous development roadmap. Approaches to developing prototypes rapidly in collaboration with prospective users through 'patchwork prototyping'

(Jones et al., 2007), or establishing requirements in ways that involve creativity (Maiden et al., 2004) may be beneficial in resolving the different perspectives identified here.

5. Acknowledgements

We gratefully acknowledge EPRSC and Leicestershire County Council support via CASE award EP/C547438/1, the help of the LCC CDR team, and the anonymous reviewers for their comments.

Biography

***David Lloyd** is a PhD candidate researching the use of human-centred techniques in geovisualization at the giCentre, City University London.*

***Jason Dykes** is a Senior Lecturer in Geographic Information at the giCentre, City University London with interests in geovisualization techniques, tools, design, application and evaluation.*

***Robert Radburn** is a Senior Research Officer at LCC and holds an ESRC UPTAP research fellowship at the giCentre, City University London to develop capacity for visual analysis in local government.*

References

Dodgson, J., Spackman, M., Pearman, A. & Phillips, L.(2000), pp38-61 Multi-criteria analysis manual

Fuhrmann, S., Ahonen-Rainio, P., Edsall, R., Fabrikant, S., Koua, E., Tobon, C., Ware, C. & Wilson, S. (2005), Making Useful and Useable Geovisualization: Design and Evaluation Issues, in J. Dykes (ed.), *Exploring Geovisualization*, Chapter 28, p556, Elsevier, Oxford.

Jones, M.C., Floyd, I.R. & Twidale, M.B. (2007), Patchwork Prototyping with Open-Source Software, in K. St.Amant & B. Still (eds), *The Handbook of Research on Open Source Software: Technological, Economic, and Social Perspectives*, pp133-8, IGI Global, Hershey.

Karlsson, J., Olsson, S. & Ryan, K. (1997). Improved practical support for large-scale requirements prioritising. *Requirements Engineering*, **2(1)**, 51-60.

Karlsson, J. & Ryan, K. (1997). A cost-value approach for prioritizing requirements. *Software, IEEE*, **14(5)**, 67-74.

Lloyd, D., Dykes, J. & Radburn, R. (2007), Understanding geovisualization users and their requirements – a user-centred approach *GIS Research UK 15th Annual Conference*, ed. A.C. Winstanley, Maynooth, Ireland.

Lloyd, D., Dykes, J. & Radburn, R. (2008), Mediating geovisualization to potential users and prototyping a geovisualization application, *GIS Research UK 16th Annual Conference*, ed. D. Lambrick, Manchester, England.

Lorenz, M.O. (1905). Methods of Measuring the Concentration of Wealth. *Publications of the American Statistical Association*, **9(70)**, 209-219.

Maiden, N., Gizikis, A. & Robertson, S. (2004). Provoking creativity: imagine what your requirements could be like. *IEEE Software*, **21(5)**, 68-75.

Malczewski, J. (2006). GIS based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science*, **20(7)**, 703-726.

Robertson, S. (2001). Requirements trawling: techniques for discovering requirements. *International Journal of Human-Computer Studies*, **55(4)**, 405-422.

Saaty, T.L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, **15(3)**, 234-281.

Saaty, T.L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill New York.

Wasil, E. & Golden, B. (2003). Celebrating 25 years of AHP-based decision making. *Computers and Operations Research*, **30(10)**, 1419-1420.

Moving to Real-Time Segmentation: Efficient Computation of Geodemographic Classification

Adnan, M., Singleton, A.D., Brunsdon, C., Longley, P.A.

University College London, Department of Geography, Gower Street, London, WC1E 6BT.
Tel: +44 (0)20 7679 0510 Fax: +44 (0)20 7679 0565
Email m.adnan@ucl.ac.uk, a.singleton@ucl.ac.uk, plongley@geog.ucl.ac.uk

KEYWORDS: Geodemographics, GIS, Clustering, Infrastructure, Geoweb 2.0

1. Introduction

Geodemographic classifications are created using computationally intensive clustering algorithms which partition the records of large multidimensional datasets into groupings which optimise homogeneity across attribute space. In their current form, commercial geodemographic classifications are created in silos by expert producers, most prevalently with closed methods and little documentation of the data inputs; the weighting and normalisation procedures; or, the exact methods of clustering[i]. In this paper we present preliminary work that examines the creation of more responsive and open geodemographic information systems. There are a number of motivations for this work. Current classifications are created using static data sources which do not necessarily reflect the dynamics of population change in modern cities. Data is increasingly available in near real-time and could be integrated to create more temporally responsive systems. Application specific classifications have been successfully demonstrated across a variety of areas, and there are many more sectors which could potentially benefit if the methods of construction and interpretation were more accessible. This paper represents initial work on clustering efficiency which will be essential if classifications are to be created within a time period acceptable to end users. Ultimately it is hoped that these users will have access to an online tool which manages both the specification, creation and testing of segmentations tailored for specific purposes. This begins to represent a new method of creating geodemographic classifications which fundamentally challenges the prevailing view of segmentation produced by a limited number of expert producers.

2. The Real Time Challenge

The world is becoming increasingly urbanized, complex and connected. These changes are driving a demand for better information that can be used to make effective decisions about the organisation, flows and connectivity of people, processes and place. While decennial census have in the past been appropriate to monitor these changes, the rate and scale of current population change is making these large surveys increasingly redundant. Through better integration of a range of data sources it is proposed that new horizons can be opened up for analyzing the different characteristics of populations and their behaviours. Key to making effective choices across a range of spatial problems is the ability of decision support tools to present areal data from a range of attributes in an understandable format. For example, one may be interested in a local measure which represents school attainment, deprivation and GP referrals for obesity. The dimensions in this example could all be measured independently from national coverage datasets, however, there are a series of challenges related to their amalgamation into a single measure including considerations of scale differences, data normalisation, weighting, methods to reduce dimensions and presentation/ visualisation.

2. Computing Bespoke Indicators

The creation of these real time generalisations raises a range of computational challenges relating to the integration of large and possibly disparate spatial databases, data normalisation and optimisation for fast transactions. This preliminary work assesses the suitability of three different clustering algorithms for integration in an online environment where processing efficiency must be maximised. The finest level in most geodemographic classifications are created using the k-means algorithm which attempts to find a set of cluster centroids minimising expression (1) below.

$$V = \sum \sum_{i=1}^{k} (x_j - \mu_i)^2 \qquad (1)$$

Where there are k clusters, and μ_i is the mean centroid of all the points x_i in cluster *i*.

The k-means algorithm initially randomly locates a set of *k* seeds within a data matrix and then allocates all data points to their nearest seed. A new mean cluster centroid is calculated for each cluster, and a new partitioning of the data points made based on the new nearest centroid. The centroids are then recalculated for the new clusters, and the algorithm repeats these steps until convergence (when data point cluster switching ceases). However, k-means produces classifications which are unstable in that the initial random assignment of seeds creates final models with variable performance (Singleton and Longley, 2008). As such, in order to optimise a classification, a model is required to run multiple times in order to select the “best” result from the sample. This process is computationally inefficient, firstly because each k-means takes time to compute, and secondly because the results from only one cluster analysis is saved. This inefficiency is unacceptable when users may be waiting for results through an online tool, thus, processing time must be optimised. For example, if the input data for the Office for National Statistics Output Area Classification (n=41 variables) are clustered on a high specification computer where k =52, then the processing time for this to converge is around 1.4 seconds. For this classification to be run 10,000 times, a user may well expect to wait for 14,000 seconds (3.88 hours) to obtain results. Considerable work has been done to improve the efficiency of K-means. For example, Reynolds *et al,* (2004) describe a new way of choosing the initial seeds, describing the algorithm as “k-means++”. This method selects initial centres based on the density of data points which improved the overall processing time because data points converge more quickly. A more radical method of improving classification efficiency is to supplement k-means with alternate classification procedures. Three such techniques are Partitioning Around Medoids (PAM) (Kaufman and Rousseeuw, 1990), Clustering Large Applications using Clara (PAM) (Kaufman and Rousseeuw, 1990), and clustering using a genetic algorithm(GA) (Painho and Bação, 2000; Maulika and Bandyopadhyay, 2000).

PAM is represented in Equations 2. This algorithm attempts to assign points from within a multidimensional data matrix into clusters based in their “nearness” to a series of randomly selected representor points. Unlike k means, representor points are actual data points from within the data matrix, rather than any point within the Euclidean space. In PAM “nearness” is calculated using a pre-computed dissimilarity matrix across all variables and data points within a data set. This has improved efficiency over k-means because it reduces the on-the-fly distance calculations, and additionally is less sensitive to outlier values as the averages essentially utilise a median rather mean in the optimisation procedures. Effectively, this minimises

$$V = \sum \sum_{i=1}^{k} | x_j - \mu_i | \qquad (2)$$

The pre-computation of a distance matrix is memory intensive and PAM struggles when applied on large data sets. Thus, Kaufman and Rousseeuw (1990) developed a sampling algorythm called Clustering Large Applications (Clara). Clara draws multiple samples of the dataset, applies PAM on each sample, and returns its best clustering as output. In this paper we will call this algorithm as Clara (PAM). As demonstrated by Brunsdon (2006), GA can also be combined with PAM to offer further improvements in classification efficiency by supplementing the initial random representor point selection with a genetic algorithm which generates multiple possible sets of representor points. Through a *breeding* procedure these are gradually refined to create a set of representor points which are optimised given a specific data set.

3. Classification Comparison

This paper provides a comparison between K-means, Clara (PAM), and Genetic Algorithm (GA). Two metrics are compared: computation efficiency (time) and classification optimisation efficiency using average silhouette width. The aim of this analysis is to examine which type of classification procedure would be most appropriate for computing real time geodemographic segmentations online. To compare K-means, Clara (PAM), and GA we used the input data for the National Statistics Output Area Classification (Vickers and Rees, 2007) aggregated at three geographical levels (Output Area (OA), Lower Super Output Area (LSOA), and Ward for London.

3.1 Measuring "Time efficiency"

In order to measure "time efficiency" of the algorithms we ran K-means, Clara (PAM), and GA for 1-100 cluster solutions on the three different geographic levels for London, and then compared the time taken for each algorithm to converge on a specified frequency of clusters. Figure 1-Figure 3 shows the relationship between computation time and the number of clusters (1-100) by running K-means, Clara (PAM), and GA on the three geographic aggregations of a dataset covering London.

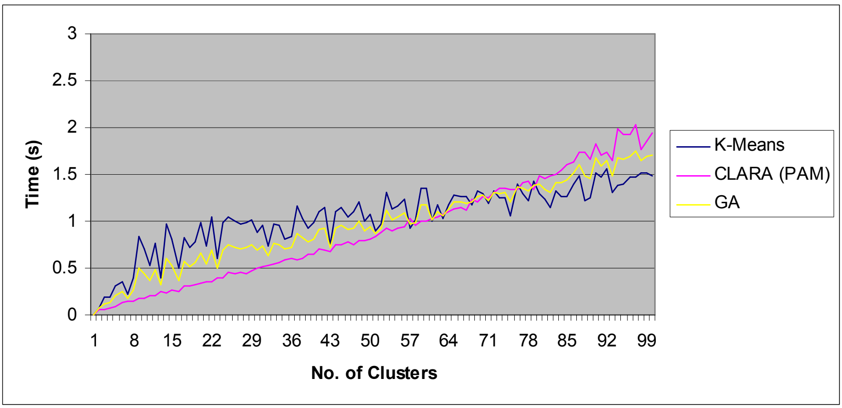


Figure 1: OA level results for the three clustering algorithms

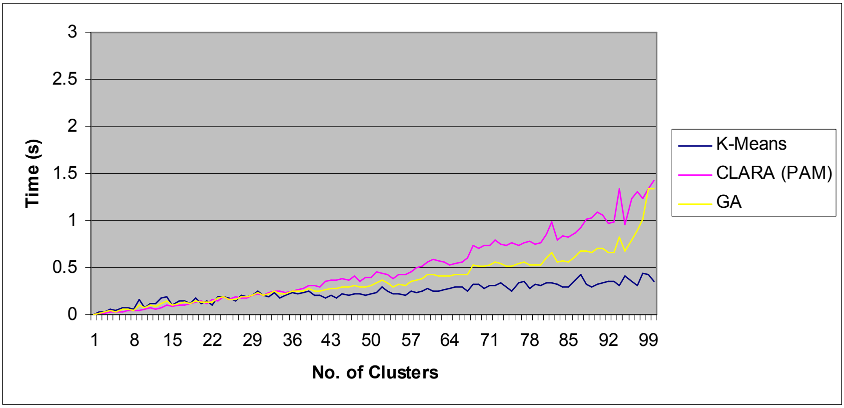


Figure 2: LSOA level results for the three clustering algorithms

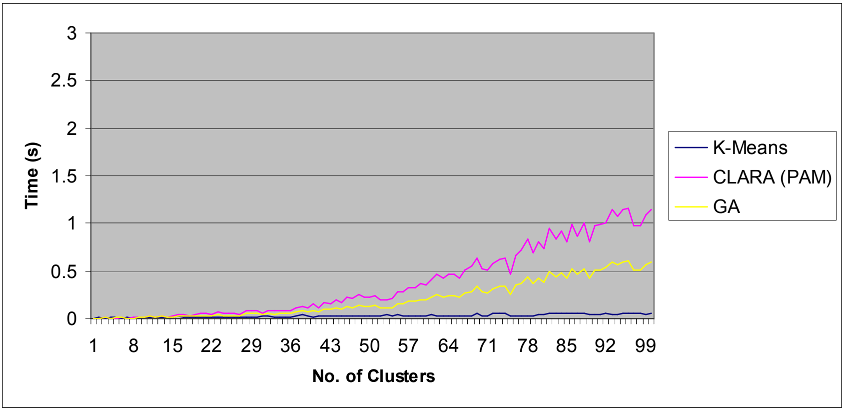


Figure 3: Ward level results for the three clustering algorithms

These charts show that for a small number of data points (LSOA and Ward) K-means runs faster than Clara (PAM) and GA. However, for OA (which has approx 22,000 records) Clara (PAM) and GA algorithms both run faster than K-means.

3.2 Measuring efficiency based on "Average Silhouette Width"

Reynolds and Richards et al (2006) demonstrated how average silhouette width (Kaufman and Rouseeuw, 1990) could be implemented as a method of comparing clustering efficiency. They present the following equation for silhouette width:

$$S(k) = \frac{x(k) - y(k)}{\max(x(k), y(k))} \tag{3}$$

Where $y(k)$ is the average distance of k from all other objects in the cluster C_k. For each other $C \neq C_k$ average distance of k from the object C is given by $d(k,C)$. $x(k)$ is the smallest result after computing $d(k,C)$ for all clusters $C \neq C_k$. The mean of $S(k)$ for all objects k, is said to be the "Average Silhouette Width" of that cluster solution. $S(k)$ ranges between 1 for a good clustering solution and -1 which would be a bad clustering solution (Reynolds and Richards et al, 2006). Thus, to measure relative efficiency of the algorithm optimisation procedures, the average silhouette widths were calculated for K-means, Clara (PAM), and GA for 1-100 cluster solutions on the three different levels of geographies for London. Figure 4-Figure 6 shows the relationship between average silhouette width and cluster frequency by for K-means, Clara (PAM), and GA on the three levels of geographic data for London.

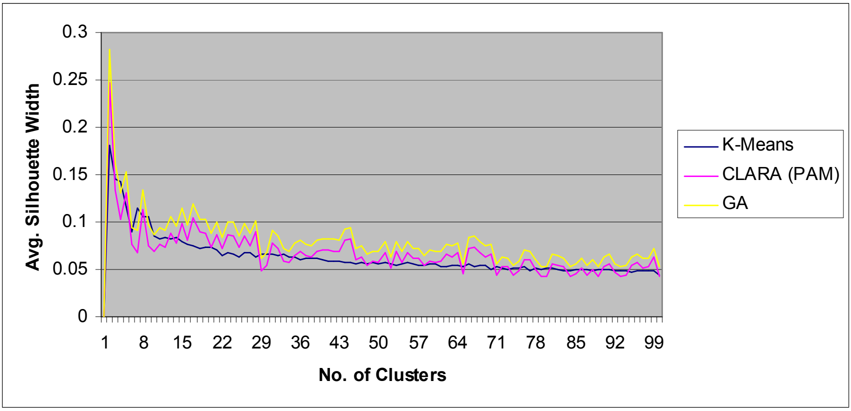


Figure 4: OA level results for the three clustering algorithms

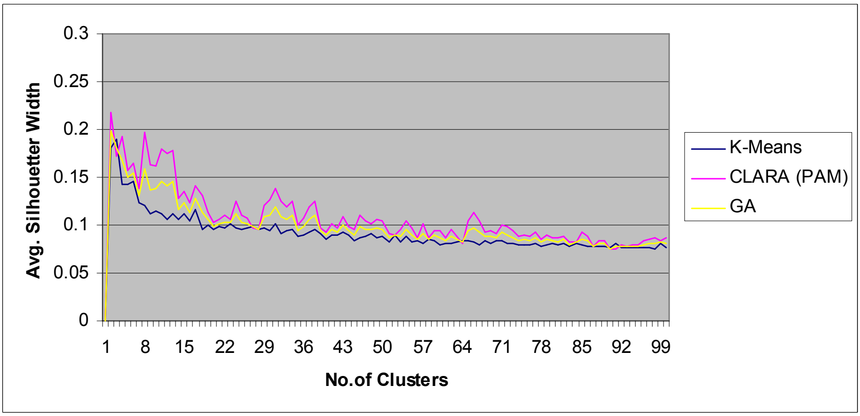


Figure 5: LSOA level results for the three clustering algorithms

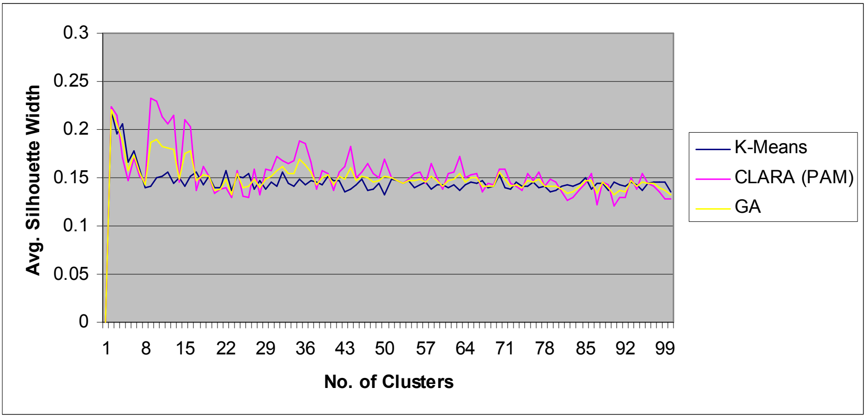


Figure 6: Ward level results for the three clustering algorithms

These results indicate that for larger data sets, GA works better than K-means and Clara (PAM). However, for smaller data sets Clara (PAM) still gives better results than K-means and GA.

4.0 Conclusion and Future Research

This paper does not purport to provide conclusions on the overall superiority of any particular clustering algorithm, but aims to evaluate a range of techniques for their suitability in a specific application of creating online geodemographic classifications. Where the best possible partitioning of the data is required, GA works better for larger data sets, but Clara (PAM) is good for smaller data sets. For speed of computation, Clara (PAM) works better than GA and K-means for larger datasets. However, K-means could be a better choice for smaller datasets. For an online geodemographic classification system, these algorithms could be chosen on the fly, based upon dataset size and user inputs. Before any substantive conclusions can be drawn from this research further testing is required to compare other clustering techniques, and in particular those methods which could be used to make K-means run faster on larger datasets – such as those methods which optimise initialisation procedures.

5.0 References

Brunsdon, C. (2006) A Cluster based approach to the zoning problem using and extended genetic algorithm. Proceedings of the GIS Research UK 14th Annual Conference, The University of Nottingham, United Kingdom, 5-7th April 2006.

Kaufman , L., Rousseeuw, P.J. (1990) Finding Groups in Data. New York: John Wiley & Sons, Inc.

Maulika, U., Bandyopadhyay, S. (2000) Genetic algorithm-based clustering technique. Pattern Recognition, 33(9), 1455-1465.

Office of National Statistics (2008). Ness data Exchange. From http://www.neighbourhood.statistics.gov.uk/HTMLDocs/downloads/NeSS-Data-Exchange-Technical-Implementation-Guide-v1.0.pdf.

Painho, M., Bação, F. (2000) Using Genetic Algorithms in Clustering Problems. Proceedings of Geocomputation 2000, University of Greenwich, United Kingdom, 23 - 25 August 2000.

Reynolds, A.P., Richards, G., Rayward-Smith, V.J. (2004) The Application of K-Medoids and PAM to the Clustering of Rules. Lecture Notes in Computer Science. 3177/2004, 173-178.

Reynolds, A.P., Richards, G., et al. (2006). Clustering Rules. A Comparison of Partitioning and Hierarchical Clustering Algorithms.

Singleton, A.D., Longley, P.A. Creating open source geodemographic classifications for Higher Education applications, 2008.

Vickers, D.W. and Rees, P.H. (2007). Creating the National Statistics 2001 Output Area Classification. Journal of the Royal Statistical Society, Series A. 170(2), 379-403.

Biography

Muhammad Adnan is working as a UCL Spatial Literacy Research and Computing Officer, and is also a first year PhD student. His research interests concern how data from a variety of public domain sources can be integrated to develop real time geodemographic classifications for public service delivery and decision making.

Dr Alex D Singleton completed his PhD in November 2007 and is now a Research Fellow in the Department of Geography and Centre for Advanced Spatial Analysis at University College London.

Professor Chris Brunsdon, Department of Geography, University of Leicester. Chris Brunsdon's research interests include the development geocomputation and spatial statistical techniques, particularly applied to social and economic data.

Professor Paul Longley, Department of Geography, UCL. Paul Longley's research interests concerns the use of GIS and quantitative methods in urban analysis. He is the outgoing Editor-in-Chief of Computers, Environment and Urban Systems and reviews editor of Environment and Planning B as well as Deputy Director of Centre for Advanced Spatial Analyses at UCL.

---

[i] One exception is the National Statistics Output Area Classification

Estimating Patients' Exposure to Traffic in General Practice service areas

Eleni Sofianopoulou[1], Tanja Pless-Mulloli[1], Stephen Rushton[2]

1 Institute of Health and Society, Newcastle University, William Leech Building, UK
Tel. +44 191 246 4840
e.sofianopoulou@ncl.ac.uk, http://www.ncl.ac.uk/ihs/people/profile/e.sofianopoulou
[2]Institution for Research on Environment and Sustainability, Newcastle University, William Leech Building, UK
KEYWORDS: traffic exposure, GP service area, kernel, splines interpolation, GIS and health

1. Introduction

Numerous spatial health data are collected by General Practices (GPs) however their utility in environmental epidemiology is limited because they are not linked to environmental datasets. One of the main reasons is the dissimilar spatial units that data are provided in. As a consequence a source of environmental health information is under-exploited. We aim to define a spatial unit that depicts a GP service area and estimate a measure of traffic exposure within it, as a first step to understanding the likely impact of pollution on patients' health. Our study covers the GPs in Newcastle and North Tyneside Primary Care Trusts (PCT), from January 2002 until July 2006. The aim of this study is quite timely as it covers the objectives of current policies relevant to re-use and exploitation of spatial public sector data (Commission of the European Communities 2004).

1.2 Data

We accessed the postcodes of the patients registered per GP practice via the North East Public Health Observatory (NEPHO), after getting ethical approval from the Newcastle and North Tyneside PCTs. We were provided with patients' data for the 1st of April for each year of interest and assumed that it remained the same for the rest of the year. The annual numbers of patients per PCT are presented on Table 1.

Table 1 Number of patients per PCT per year

Year	*Newcastle PCT (35 GPs)*	*North Tyneside PCT (29 GPs)*
2002	230,464	173,123
2003	236,181	179,391
2004	244,023	185,314
2005	252,845	191,711
2006	264,614	199,545

The road network was accessed via the Digimap website of EDINA database. We downloaded the A roads, B roads and Minor roads by Ordnance Survey Meridian 2 map, at 1:50,000 scale. We also accessed traffic flows data from 120 monitors that count the numbers of vehicles 24hours a day. The data were provided by the Tyne and Wear Traffic and Accident Data Unit Transport Centre, based on Gateshead City Council. Data analysis was conducted using R 2.7.1 statistical software (CRAN) with the nlme pakage (Pinheiro J., Bates D. et al. 2008) and ArcGIS (ESRI) with the Hawth's Tool extension (Beyer 2004) and Spatial Analyst extension.

2. Methods

2.1 General Practice service areas

Kernel estimation was developed to obtain a smooth estimate of a probability density from an observed sample of observations. We used the kernel analysis to estimate the area where the majority of patients are expected to be present per GP. Prior to using the kernel analysis, we estimated the service area using the minimum convex polygons, created by the postcodes of all patients registered per GP. Figure 1 shows an example of the difference between the GP service areas based on minimum convex polygon and kernel analysis.

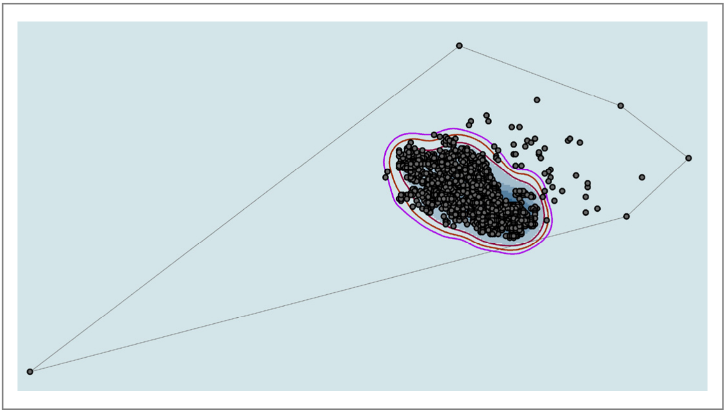


Figure 1. GP service area based on minimum convex polygon and kernel analysis (curved borders)

The minimum convex polygon includes areas that patients do not live, whilst the area depicted by kernel analysis represents the core of the area that the patients are living in. We found the latter to be more representative of GP service areas. In

Figure ***1***, are presented three kernel contours, which stand for the 95%, 98% and 99% of probability density function. We estimated them in order to check the sensitivity of kernel on different x% probability distributions.

Kernels have been used a lot in ecology for summarising spatial data and have recently been employed for health research (Petersen J., Atkinson P. et al. 2008). There are various types of kernels, but the quartic kernel is the most widely used, as discussed by (Bailey and Gatrell 1995). The formula for this is given in Equation1

$$\lambda_\tau(s) = \sum_{h_i \le \tau} \frac{3}{\pi}(1 - h_i^{\ 2} / \tau^2) \qquad (1)$$

where,
$\lambda\tau(s)$ is the intensity at the point of estimate s
h_i is the distance between the point s and the observed event location s_i
τ is the bandwidth which is sampled around point s
The region of influence within which observed events contribute to $\lambda\tau(s)$ is therefore a circle of radius τ centred on s.

2.2 Traffic exposure

We accessed monthly traffic flow data for approximately 120 monitoring points. After editing the dataset, the monitoring sites with complete data for our study period were 50 out of 120. Those sites were located on A-roads and on two major B-roads. The variability of traffic flows, amongst A-roads was significant (Figure 2) and was the incentive for using the traffic flow data rather than the type of road or length of type of road, as indicator of exposure to traffic pollution. In Figure 2 is presented the variability of traffic flows per type.

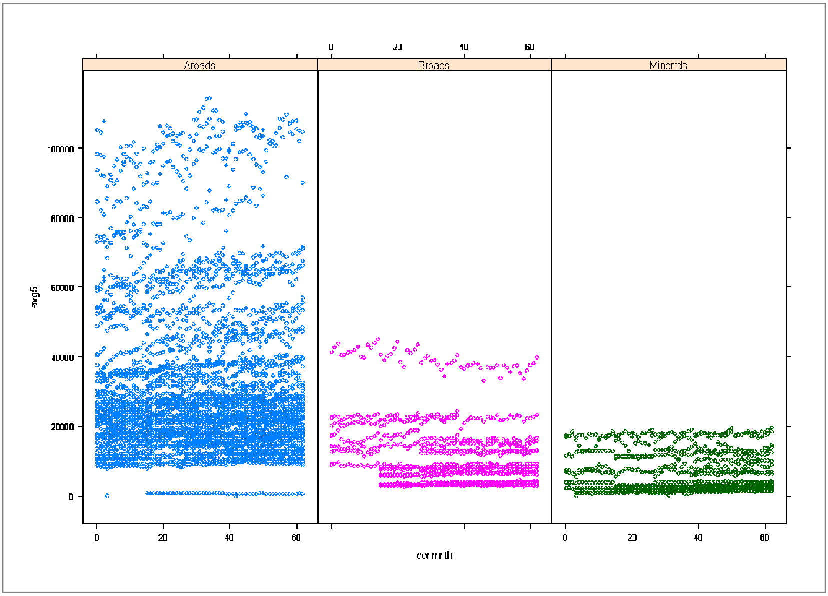


Figure 2. Variability of traffic flows per Road type (A, B, Minor Roads)

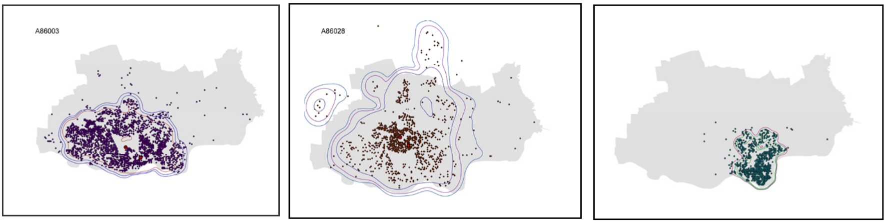
We analysed the traffic flows and estimated monthly traffic flows in time and space using splines interpolation technique. The interpolation of traffic flows was carried out within the main road network on the study area. Therefore we employed a type of splines that produce a minimum curvature surface at the desired row and column space (Terzopoulos 1988). This technique has been used in the past in geology and geosciences (Smith W.H.F. and Wessel P. 1990; Zoraster 2003).

3. Results

3.1 Kernel analysis

Service areas were produced for five years for each one of the 64 GPs, giving a total of 325 service areas. Little variation on borders of service areas per GP was observed over the five years period. We, also, selected the contour representing the 98% probability density distribution, as the boundary of this area is more representative of patients' spatial distribution. Figure 3 shows a few examples of the GP services areas from GPs that belong to Newcastle and North Tyneside PCT.

a.



b.

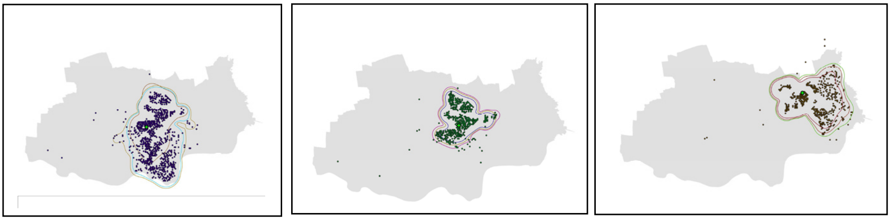


Figure 3. a) GP service areas in Newcastle PCT, b) GP service areas in North Tyneside PCT

3.2 Splines interpolation

We interpolated the traffic flows, on the main road network, for each one of the 55 months of the study period. In Figure 4 is presented a map with the road network, the monitoring sites and the interpolated traffic flows. The darker colour in the road network symbolizes areas with more traffic flows, while the lower traffic density is represented by lighter shades of blue.

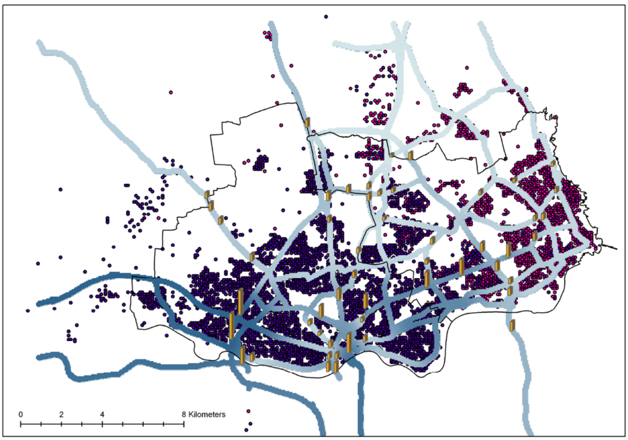


Figure 4. Interpolation of traffic flows within the study area

The output of the interpolation was validated by running the spline model after removing one of the 50 monitoring sites and then comparing the predicted value to the original data on that site. This process was conducted for 10 randomly selected monitoring sites, which is 20% of the total number of traffic monitors. The spline output was also assessed in time by repeating the analysis for those 10 points for 6 randomly selected months.

The degree of correlation between the interpolated traffic flows and observed flows ranges from 88% to 92%. In Table 2 are presented the results of correlation of the 10 randomly selected monitoring sites in the six different months. In order to validate further the splines model, we used regression to assess the relationship of the interpolated values against the observed values, using mixed-effects model. Mixed effects models are appropriate when we have repeated measurements per subject. In this case the traffic flows that were assessed for six months were grouped per monitoring site. The results show a significant relationship between the interpolated and observed traffic flows (p<0.000).

Month	*Correlation Coefficient*
Mar-02	0.88
Sep-02	0.92
Oct-03	0.91
Dec-04	0.89
Jul-05	0.91
May-06	0.90

Table 2. Correlation coefficients of observed traffic flows against estimated values for 10 monitoring sites over months

After the completion of the validation process, we estimated the traffic flows within each one GP service area separately. In total, 55 monthly traffic flow indices were estimated per GP, by extracting and summing traffic flows per GP service area, giving an overall of 3520 patients' traffic exposures. Figure 5 presents a boxplot of the estimated traffic exposures per GP service area, which shows considerable variation among them.

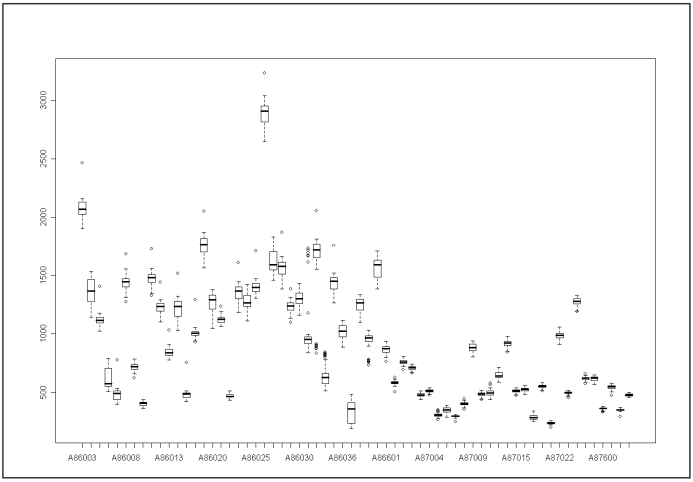


***Figure 5**. Boxplot of estimated traffic exposure per GP service area*

4. Conclusions

- Kernel analysis can create representative GP service areas that can be used as the spatial unit of analysis for data collected per GP.
- Traffic flows when interpolated produce a continuous map, which allows the analysis of traffic flows per GP service area.
- Overall, the estimation of traffic exposure of patients registered per GP can be used to investigate the relationship between traffic exposure and various health outcomes.

5. Acknowledgements

We would like to thank the Colt Foundation for funding this research. We would also like to thank the North East Public Health Observatory as well the Newcastle and North Tyneside Primary Care Trusts for providing us with the patients' data. Finally we are thankful to Tyne and Wear Traffic and Accident Data Unit Transport Centre for providing us with the traffic flow data.

References

Bailey, T. C. and A. C. Gatrell (1995). Interactive spatial data analysis, Harlow Essex, England: Longman Scientific & Technical, 1995.

Beyer, H. L. (2004) Hawth's Analysis Tools for ArcGIS available at: http://www.spatialecology.com/htools

Commission of the European Communities (2004). Proposal of the European Parliament and of the Council - establishing an infrastructure for Spatial Information in the Community (INSPIRE). **COM(2004) 516 final**.

Office of Public Sector Information (2005) Public Sector Information 2005 No.1515 available at: http://www.opsi.gov.uk/si/si2005/20051515.htm

Petersen J., Atkinson P., et al. (2008). "Teenage pregnancy-new tools to support local health campaigns." Health & Place **15**(1): 300-7.

Pinheiro J., Bates D., et al. (2008). The nlme Package. R-project. R-core.

Smith W.H.F. and Wessel P. (1990). "Gridding with continuous curvature splines in tension." Geophysics **55**(3): 293-305.

Terzopoulos, D. (1988). "The computation of visible-surface representations." IEEE Transactions on Pattern Analysis and Machine Intelligence **10**(4 (July)): 417-438.

Zoraster, S. (2003). "A surface modeling algorithm designed for speed and ease of use with all petroleum industry data." Computers & Geosciences, **29**(9): 175-182.

Calculating a Walkability Index for the UK: trials and tribulations

Thomas Burgoine[1], Seraphim Alvanides[1], Amelia A. Lake[2]

[1]Geography, Newcastle University; [2]HNRC, Newcastle University
t.h.s.Burgoine@ncl.ac.uk

KEYWORDS: Environmental justice, obesogenic environments, walkability, GIS

1. Introduction

Levels of obesity have increased three-fold in the last twenty years and predictions for the future of the global obesity problem are becoming ever more pessimistic (Foresight 2008). This is a trend that, if left unaddressed, is only set to worsen by 2010 – see figure 1.

Obesity and overweight are the result of energy imbalance between consumption (by food intake) and expenditure of calories (through being physically active). It has been suggested that when it comes to interventions for tackling obesity, a "neighbourhood based approach could add to traditional individual level obesity interventions, which often ignore the environmental context that shapes our behaviours" (Black and Macinko 2008, 2). This neighbourhood effect is often referred to as one aspect of the 'obesogenic environment', a concept led by the notion that our surroundings can drive an "automatic, unconscious influence…[upon] behaviour" (Brug *et al* 2006, 528).

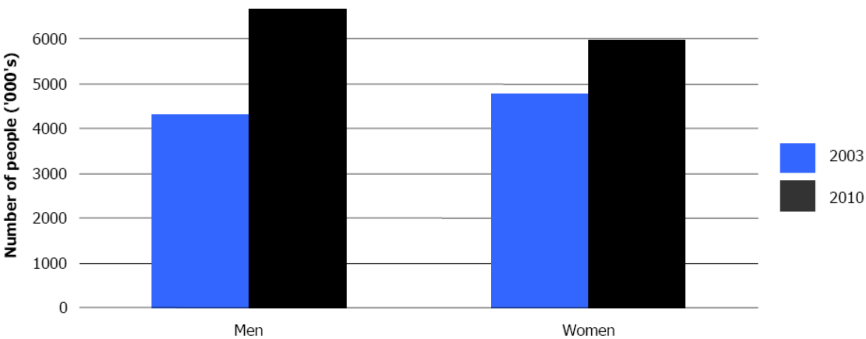


Figure 1. Estimated number of obese people (UK), 2003 and 2010, by sex

Source: Zaninotto *et al* (2006)

Changing geographies across space mean that the obesogenicity of the environment is unlikely to remain uniform through space. Consequently, questions of 'environmental justice' abound: could it be possible that some populations are actually predisposed to obesity simply because of the 'obesogenic' environment in which they happen to live? (Sexton and Adgate 1999). This work sought to examine the (changing) geography of obesogenicity throughout North East England through creating a theoretical map of the obesogenic environment – an obesogenic index. Giving due recognition to the disparities in levels of obesity found throughout the UK, how would the geography of North East England's (obesogenic) environment differ across space, if at all?

2. Components of a walkability index

Walking is one way of increasing calorific expenditure with minimum effort, in tandem with other interventions to tackle obesity (such as food intake and other forms of physical activity). For the purposes of the study, the built environment was assessed vis-à-vis the 'walkability' of the neighbourhood (comprised of a measure of residential density, street connectivity and land use mix). Individually, it is hypothesised that these determinants act tangibly upon health by making walking "feasible and appealing" (Lovasi *et al* 2008, 2), however walkability measures have thus far been predominantly operationalised in the US and Australia.

The elements of the index, along with the index itself was calculated at the Lower Super Output Area (LSOA), areal level, using boundary data available from EDINA Digimap. LSOAs were deemed appropriately sized for this study as they allowed a sufficient level of detail (useful in a relatively constrained study area such as the North East alone) to be obtained whilst still allowing for scrutinisation of larger patterns and trends.

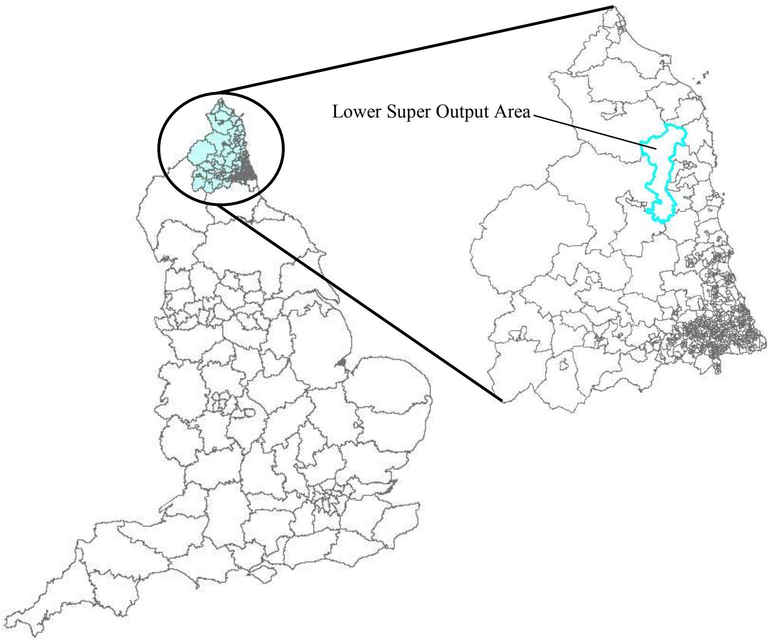


Figure 2. The study area of North East England, complete with an illustration of a LSOA

Highly populated neighbourhoods are usually thought to include mixed-use development and an increased variety and choice of retail activities. This is often hypothesised to result in "shorter, [more] walkable distances between complementary shops and restaurants", which, confounded by the amplified difficulty in car parking is thought to increase walkability (Leslie *et al* 2007, 117). Residential density was calculated based on the work of Leslie *et al* (2007) – equation 1.

$$Number\ of\ Household\ Spaces \Big/ Total\ Area\ of\ Domestic\ Buildings\ (m^2) \qquad (1)$$

The total number of household spaces (the total number of houses) within any given LSOA was sourced from the 2001 Census. Information on the total area of domestic buildings was sourced from the Generalised Land Use Database (GLUD), available through Neighbourhood Statistics.

Areas that are highly connected possess a myriad of street intersections thus increasing the probability that a direct route will be available between two points and thus increasing convenience and the propensity to walk (Frank and Engelke 2005). Connectivity was calculated here using an OS Meridian 1:50000 map downloaded from EDINA Digimap. The number of street intersections was summed within each LSOA (see figure 3) and equation 2 applied to adjust for the varying size of LSOAs within the study area.

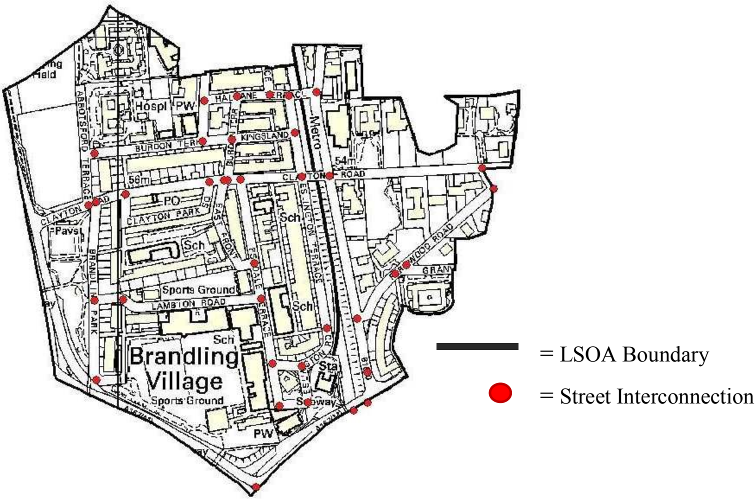


Figure 3. An illustration of how OS Meridian highlights road interconnections within one LSOA

$$\textit{Number of Street Interconnections} \big/ \textit{Total Area of LSOA } (m^2) \qquad (2)$$

An increasingly varied mixture of land uses within an area is usually related to "shorter distances between residences and destinations such as stores and workplaces" (Owen *et al* 2007, 388). This improved proximity between origin and destination is likely to augment the propensity for walking. Land use mix was determined here using the GLUD (which divides LSOAs into land parcels, areas of land that share a common use), and the following entropy calculation (equation 3) based on the work of Leslie *et al* (2007),

$$-\frac{\sum_k (p_k \ln p_k)}{\ln N} \qquad (3)$$

whereby, k is the type of land use, p is the proportion of the LSOA devoted to that specific land use, and N is the total number of land use categories. For the purposes of this study (and due to data constraints), three types of land use were incorporated into the equation: 'residential', 'non-residential' and 'green spaces'. The resultant entropy scores range between 0, which denotes homogeneity in land uses, and 1, where land uses are distributed equally between the 3 land use categories.

3. Results and evaluation of the Index

Once each element of walkability had been calculated, the results were ranked and divided into quintiles. The three individual indicators were then combined to create a single obesogenic index, with regards to walkability, operationalised by summing the quintile scores for each of the components and then once again breaking down these total scores into quintiles (Wheeler 2004). The obesogenic index, showing combined quintile scores for our LSOAs, was mapped using ArcGIS 9.1 (figure 4).

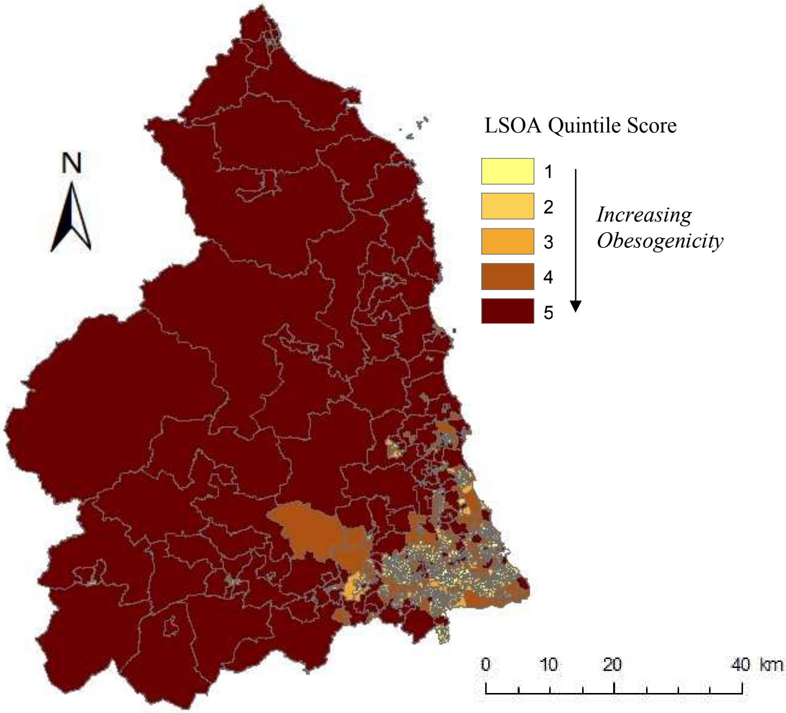


Figure 4. The walkability index for the North East study area

The most salient observation is that the majority of the LSOAs to the West of the study area fall within the top 20% in terms of their theoretical encouragement of obesity; it should be noted that the majority of these LSOAs are strictly defined as rural. There is also significant clustering evident intimating that LSOAs of similar obesogenicity are likely to be found in close proximity, and this is unlikely to be due solely to the methodology utilised in the creation of the index. Pockets of relatively less obesogenicity can be witnessed throughout the region, most notably in the South-East (Newcastle city centre); figure 5 explores the 'urban' obesogenic environment of this Newcastle extent, which as we can see, is a lot more varied in terms of its theoretical obesogenicity. There are many cases of 'very' obesogenic locales bordering those in the least obesogenic 20% of LSOAs. Whilst these stark divisions may exist in reality, they could be due to the use of quantiles that can tend to exaggerate differences between areas and thus these sharp distinctions should perhaps be seen as indicative of a more gradual change through space. Overall, this more populous sector of the study area is regarded as substantially less obesogenic, relative to the less populated 'rural' areas noted above.

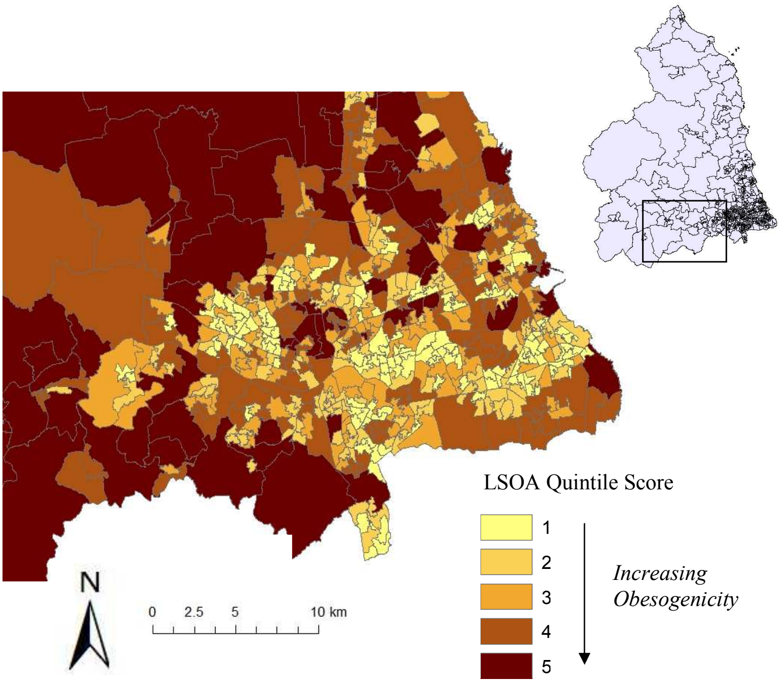


Figure 5. The walkability index for the Newcastle extent of the study area

Further analysis was made after the constituent parts of the index were mapped individually, however this is not detailed here due to space constraints (for maps, see figure 6). Levels of correlation were very high between elements of the index; the street connectivity indicator was found to covary most with the overall index (r_p=0.663, $p<0.001$), suggesting that individuals in poorly connected communities are highly likely to be exposed to the most all-round obesogenic environments.

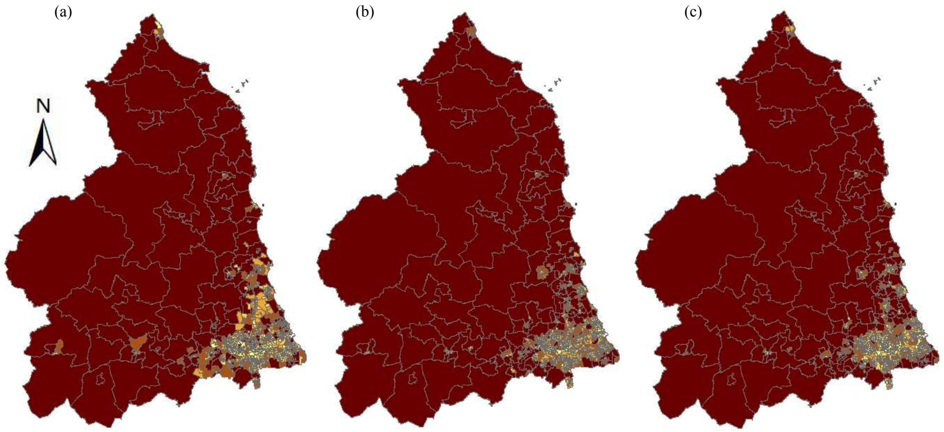
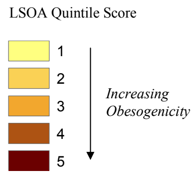
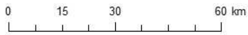


Figure 6. Obesogenicity through (a) residential density, (b) street connectivity and (c) land use mix.





ArcGIS 9.1 was then used in conjunction will Google Earth in an attempt to determine how walkability manifests itself in reality. Figure 7 shows an area of theoretically low walkability, whereas figure 8 shows a LSOA of theoretically high walkability within the study area.

= LSOA Boundary

Figure 7. Illustration of low walkability, Berwick Upon Tweed 002E LSOA

Image courtesy of Google Earth © 2008 Infoterra Ltd & Bluesky

= LSOA Boundary

Figure 8. Illustration of high walkability, Newcastle Upon Tyne 016C LSOA

Image courtesy of Google Earth © 2008 The GeoInformation Group

Whilst the assessment of high walkability does appear correct upon visual inspection (with an inferred, dense population and well connected street network), it is clear that the difference between high and low walkability areas could be fundamentally due to rurality/urbanity, a flaw that is perhaps inherent in the metric utilised here and a methodological concern to be addressed in future work. On the other hand, these areas may not in fact be very walkable, with farmland often notoriously inaccessible to the public despite residential proximity.

Despite some methodological concerns, this research still succeeded in portraying a varying (obesogenic) environment, through walkability, for the North East of England. This research is novel in the UK context where there is a lack of studies that objectively define the walkability of the environment in this way. Such an index will be used in future work, in conjunction with ecological and individual level data on obesity itself, in order to scrutinise the link between individual and environment and to ascertain the extent to which these areas are in fact ‘obesogenic’.

4. Acknowledgements

We would like to thank the Economic and Social Research Council (ESRC) for supporting this research.

References

Black, J. L. and Macinko, J. (2008) “Neighbourhoods and obesity” *Nutrition Reviews*, 66(1), pp. 2-20
Brug, J., Van Lenthe, F. J. and Kremers, S. P. J. (2006) “Revisiting Kurt Lewin: how to gain insight into environmental correlates of obesogenic behaviours” *American Journal of Preventive Medicine*, 31(6), pp. 525-529
Foresight (2008) “Tackling Obesities: Future Choices – Obesogenic Environments – Evidence Review” London: Government Office for Science
Frank, L. D. and Engelke, P. (2005) “Multiple impacts of the built environment on public health: walkable places and the exposure to air pollution” *International Regional Science Review*, 28(2), pp. 193-216
Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A. and Hugo, G. (2007) “Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes” *Health and Place*, 13, pp. 111-122
Lovasi, G. S., Moudon, A. V., Pearson, A. L., Hurvitz, P. M., Larson, E. B., Siscovick, D. S., Berke, E. M., Lumley, T. and Psaty, B. M. (2008) “Using built environment characteristics to predict walking for exercise” *International Journal of Health Geographics*, 7(10), pp. 1-13
Owen, N., Cerin, E., Leslie, E., duToit, L., Coffee, N., Frank, L. D., Bauman, A. E., Hugo, G., Saelens, B. E. and Sallis, J. F. (2007) “Neighbourhood walkability and the walking behaviour of Australian adults” *American Journal of Preventive Medicine*, 33(5), pp. 387-395
Sexton, K. and Adgate, J. L. (1999) “Looking at environmental justice from an environmental health perspective” *Journal of Exposure Analysis and Environmental Epidemiology*, 9, pp. 3-8
Wheeler, B. W. (2004) “Health-related environmental indices and environmental equity in England and Wales” *Environment and Planning A*, 36, pp. 803-822
Zaninotto, P., Wardle, H., Stamatakis, E., Mindell, J. and Head, J. (2006) “Forecasting obesity by 2010”, *Department of Health*, (http://www.dh.gov.uk/assetRoot/04/13/86/29/ 04138629.pdf, 5/2/07)

Biography of principal author

Thomas Burgoine is a first year Geography PhD student at Newcastle University, with this research being completed as part of his MA dissertation. Thomas was awarded a 1+3 studentship by the ESRC in 2007 and is the student representative on the Association for the Study of Obesity (ASO) committee.

A spatial accuracy assessment of an alternative circular scan method for Kulldorff's spatial scan statistic

Simon Read[1], Peter Bath[1], Peter Willett[1], Ravi Maheswaran[2]

[1]Department of Information Studies, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, UK
Tel. 0044(0)114 2222630 Fax 0044(0)114 2780300
simon.read@sheffield.ac.uk
[2]School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

KEYWORDS: epidemiology, 'spatial scan statistic', 'point process', 'case-control study', SaTScan

1. Introduction

This paper concerns the Bernoulli version of Kulldorff's spatial scan statistic, and how accurately it identifies the exact centre of approximately circular regions of increased spatial density in point data. We present an alternative method of selecting circular regions that appears to give greater accuracy. Performance is tested in an epidemiological context using manifold synthetic case-control datasets. A small, but statistically significant, improvement is reported. The power of the alternative method is yet to be assessed.

2. Research Background

The spatial scan statistic (Kulldorff 1997), implemented via SaTScan (www.satscan.org), has been used extensively by epidemiologists as a tool for cluster detection. The Bernoulli version of the spatial scan statistic aims to detect localised clusters in binary labelled spatial point data, ideal for use with case-control studies. Each point is a 'case' (incident of disease) or 'control' (random sample of the non-diseased population) (Rothman 2008).

Kulldorff (1997) proved that if one wishes to locate a specific region where each data-point has a higher (or lower) probability of a being a case than a data-point outside, then the spatial scan statistic (Section 2.1) is the most powerful test. However, no universally accepted optimum method for finding candidate regions exists. Some propose scanning for arbitrarily shaped regions, which is flexible but computationally expensive: see Tango & Takahashi (2005) for an example, and citations of others. SaTScan offers a computationally efficient circular scan (Section 2.2), similar in ways to Openshaw's GAM (1987), and we examine an alternative to this. (Section 3).

2.1. Kulldorff's Spatial Scan Statistic (Bernoulli Version)

Consider a study of N points across a region A. C are cases, the remainder controls. Let Z be any sub region of A, containing n points, c of which are cases. Let A_c be all A outside Z. Let p be the probability (*risk*) that any point in Z is a case. Let q be the probability that any point in A_c is a case. Null hypothesis H_0 is $p=q$; alternative hypothesis H_A is $p>q$. Given N, the likelihood of any particular Z occurring under H_A is:

$$L_A(Z) = (\frac{c}{n})^c (1-\frac{c}{n})^{n-c} (\frac{C-c}{N-n})^{C-c} (1-\frac{C-c}{N-n})^{(N-n)-(C-c)} I(\frac{c}{n} > \frac{C-c}{N-n}) \qquad (1)$$

Where I is the indicator function.

Under H_0 the likelihood is uniform for all *Z*:

$$L_o = (\frac{C}{N})^C (\frac{N-C}{N})^{N-C} \quad (2)$$

For any *Z*, the *spatial scan statistic* (a.k.a. *likelihood ratio*) is defined as:

$$LR(Z) = \frac{L_A(Z)}{L_o} \quad (3)$$

Let *Z'* be the *Z* with highest *LR(Z)*. *Z'* is the most likely disease cluster. Statistical inference for *Z'* (rejection of H_0) is obtained by Monte Carlo testing, but isn't necessary in this investigation: *Z'* is always the most significant location of increased risk, and we place only one location of increased risk in our model (Section 4).

2.2. Original (SaTScan) Method of Finding Circular Candidates for *Z'*

SaTScan identifies circular or elliptical candidates for *Z'* (most likely cluster). We considering circular. For each point in *A*, concentric circles are drawn with radii just large enough to reach a case, as Figure 1. Circle in *A* with the highest likelihood ratio is declared *Z'*.

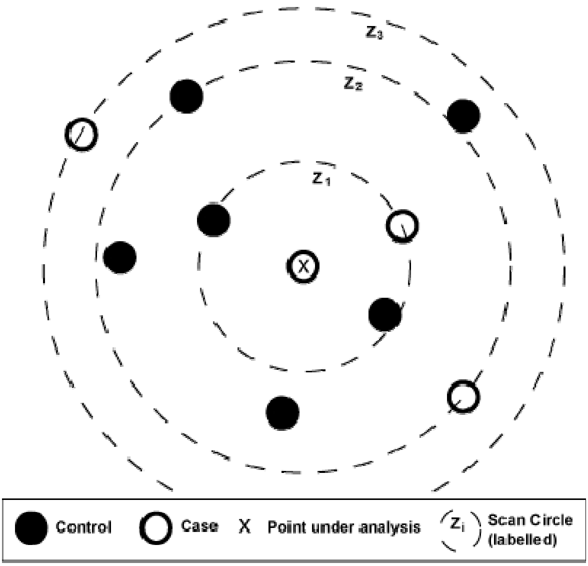


Figure 1. Example of Original Circular Scan Method

3. Our Alternative Method of Finding Circular Candidates for *Z'*

For each point in *A*, circles are drawn whose rim intersects the point and a case, diameter being exactly equal to distance between them, as Figure 2. Circle in *A* with the highest likelihood ratio is declared *Z'* (most likely cluster).

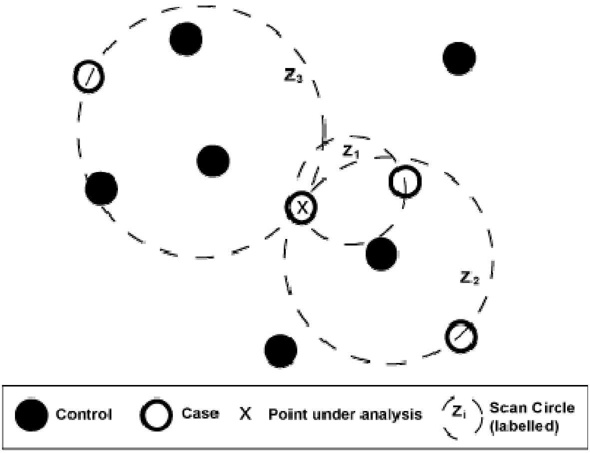


Figure 2. Example of Alternative Circular Scan Method

Both methods generate *C(N-1)* circles. The alternative method screens out *C(C-1)/2* duplicates. Circles of the alternative method are smaller than the original, so usually contain fewer points (smaller *n*). This is important: circles where *n* is a large proportion of *N* may be too big to be useful clusters. SaTScan, by default, screens out circles where *n* > *N/2*. So, although the original method generates more circles, more are also screened out. A larger number of smaller circles clearly facilitates a more detailed analysis, possible leading to more accurate locating of disease sources.

4. Accuracy Testing Methodology

Original and alternative methods are competitively tested on synthetic case-control datasets. The stochastic nature of disease distribution means one must compare methods on manifold datasets to draw a meaningful conclusion. Synthetic data provides this multiplicity. To add realism, the underlying intensity of the control distribution is matched to the 2001 population density of Trent, a UK region containing urban and rural areas, and features like coastal towns and ribbon developments. The same intensity is used to generate cases, with the injection of a single randomly located (and oriented) hypothetical source of multiplicative increased risk. Figure 3 shows an example randomization, based on a 500x500 grid for computational convenience. The risk source takes one of two distributions, both Gaussian, see Figure 4. The non-symmetric type could represent a wind-blown pathogen. The peak of each distribution is aligned with the location of the source. The height of the peak symbolises the source's *maximum relative risk* (MRR). For instance, MRR=5 means a data-point at the centre of the source is 5 times more likely to be a case than a data-point far away from the source. Relative risk decreases smoothly as distance from the source increases. To remove edge-effects, only source locations well inside the region boundary are permitted. Note clusters are more obvious when the source is in a densely populated area.

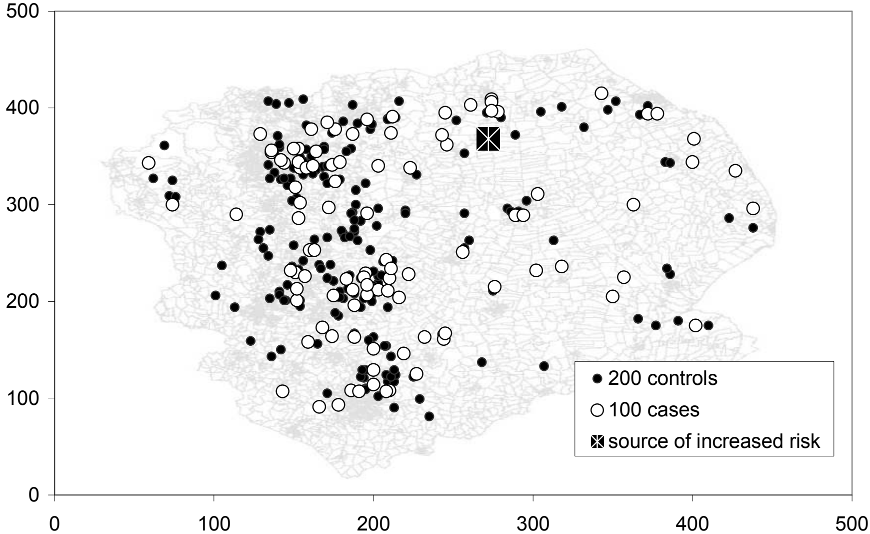


Figure 3. Example of Randomised Synthetic Case-Control Distribution

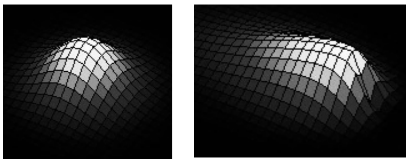


Figure 4. Examples of Symmetric (left) and Non-Symmetric (right) Sources of Increased Risk

The alternative method is coded in C++ as a self-contained object, passed only the case-control locations. The original method is the batch version of SaTScan, downloaded from www.satscan.org. Relative performance is discussed below.

5. Test Results.

For each dataset, the centre of the most likely circle produced by each method is recorded as its 'guess' at the position of the source. The calling program records the distance (in grid units) between guess and actual source location, as the *Estimation Error* (*EE*). For *m* datasets, based on a grid of *width* units across, *Mean Estimation Error Difference* is defined:

$$MEED = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{EE_{i-alternative} - EE_{i-original}}{width}\right) \tag{4}$$

where $EE_{i\text{-}alternative}$ is the *EE* of the alternative method when applied to the i^{th} dataset, and $EE_{i\text{-}original}$ is correspondingly that of the original.

Eight runs of m=5000 datasets ($width$=500, N=300, C=100) were used to test four MRR values for symmetric and non-symmetric sources, results in Table 1. Note *MEED* is expressed as a percentage, negativity indicating the alternative method is more successful. p-values are based on rejection of the null hypothesis *MEED*=0.

Table 1. Comparative Mean Estimation Error Difference (MEED).

Disease source shape	Indicator	*MRR=1.5*	*MRR=3*	*MRR=5*	*MRR=10*
Symmetric	*MEED*	-0.89%	-1.02%	-1.05%	-0.55%
	p-value	*<0.00001*	*<0.00001*	*<0.00001*	*0.0001*
Non-symmetric	*MEED*	-0.76%	-0.72%	-0.76%	-0.31%
	p-value	*<0.00001*	*<0.00001*	*<0.00001*	*0.0162*

6. Discussion and Future Research

Table 1 shows that, in every run, the alternative method's 'guesses' tend to be slightly, but significantly, closer to the actual source than those of the original. Although small in terms of the grid units, this could equate to several kilometres in real studies. This supports the argument in Section 3, that a greater number of smaller circles promotes greater spatial accuracy. Interestingly, the alternative method appears to perform better (relative to the original) at lower MRR levels. In our model at least, lower MRR leads to a smaller cluster area, which is naturally detected more accurately by a smaller circle. This property may be useful, as the relative risk of real disease causal factors is typically small (Rothman 2008).

One important omission here is power analysis. Increased computing resources are required, as the Monte Carlo testing required to obtain cluster significance causes a x1000 runtime increase. We will address this in the near future.

We propose investigating a similar alternative to the elliptical version of SaTScan, and also applying the concept to temporal dimensions.

7. Acknowledgements

We thank the Medical Research Council for funding Simon Read.

8. References

Kulldorff, M. (1997). "A spatial scan statistic". *Communications in Statistics - Theory and Methods*, 26(6), 1481-1496.

Openshaw, S., et al. (1987). "A Mark 1 Geographical Analysis Machine for the automated analysis of point data sets." *International Journal of Geographical Information Systems*, 1(4), 335-358.

Rothman, K.J., et al. (2008). *Modern Epidemiology*. Philadelphia: Lippincott, Williams and Wilkins.

Tango, T. & Takahashi, K. (2005). "A flexibly shaped spatial scan statistic for detecting clusters". *International Journal of Health Geographics*, 4(1), 11.

9. Biography

Simon Read (principal author/programmer) is a second year research student, interested in the spatial processing of epidemiological information, particularly methods for identifying local disease clusters. Simon has an MSc in Information Systems, and previously published a paper on the design issues around creating an image browser for Google Earth.

Using GIS derived variables to identify factors affecting physical activity levels in children

Iain Lake[13], Andrew Jones[1], Natalia Jones and Jenna Panter[1], Florence Harrison[1], Esther van Sluijs[2] and Simon Griffin[2]

[1]School of Environmental Sciences, University of East Anglia, Norwich, NR4 7JT
[2]MRC Epidemiology Unit, Institute of Metabolic Science, Box 285
Addenbrooke's Hospital, Cambridge, CB2 0QQ
[3]Author for Correspondence, Tel. +44 1603 593744 Fax +44 1603 591327 I.lake@uea.ac.uk,
http://www.uea.ac.uk/env/people/facstaff/lakei

KEYWORDS: obesity, physical activity, data redundancy research priorities

1. Introduction

Worldwide over a billion people are overweight. Most research on obesity has been focussed upon individual interventions in spite of evidence showing that obesity is largely due to environmental factors (Jain, 2005). These can be collectively termed “obesogenic environments” which are acknowledged to have played a part in the reduction of physical activity levels in the population (Government Office for Science, 2007). Tacking these environments is important as this intervention is likely to affect multiple pathways within the obesity system (Government Office for Science, 2007).

To provide evidence for such interventions, studies are required on the association between physical activity and where individuals live. However, only a minority of such studies have used objective environmental measures (Humpel et al., 2002). GIS provides one method of creating objective measures, but few studies have harnessed this ability. There are many benefits to using GIS (Leslie et al., 2007).

When using GIS there are several issues that need careful consideration. GIS provides the ability to produce many variables to describe the environment where someone lives (Lake et al., 2000). However, many of these will be highly correlated and so the effort required calculating all of these may be questioned. Also the resources required to create different GIS variables varies greatly between variables. Are the more resource intensive variables required? The aims of this paper were:

- To examine collinearity between GIS based environmental variables and determine whether a subset might provide comparable variability
- To identify the most important GIS variables for differentiating between children’s physical activity levels
- To provide recommendations to future studies on where resources savings could occur
- To highlight improvements in GIS software and data sources that would be most useful for future studies

2. Study System

The SPEEDY project aims to identify patterns and determinants of physical activity and diet in 9-10 year olds. During the 2007 Summer term over 2000 9-10 year old children were recruited from 90 schools in Norfolk. Each child was required to wear an accelerometer for a week and the children and their parents were required to complete a detailed questionnaire on health related behaviour. The children’s height and weight was also measured.

3. Methods

GIS variables were created to describe the environment of each child's home. These were then used as explanatory variables of the amount of moderately vigorous physical activity (MVPA) the child undertook at weekends. Using GIS, 114 environmental variables were created. Some were based upon the point at which the child lived, others upon the neighbourhood around their home. Neighbourhoods were defined based upon the roads within a 10 minute walk of the property. All variables were grouped into seven categories.

- Facilities and services – Walking distance from each property to its nearest of a number of different facilities. Additionally facility density within the neighbourhood was calculated.
- Ease of walking – Effective walkable area, road junction density and connected node ratio (Forsyth, 2006) were calculated.
- Infrastructure for walking / cycling – Density of street lights, pavements and roadside verges were calculated within each neighbourhood.
- Neighbourhood surroundings –The percentage of different land uses within each neighbourhood was calculated alongside a measure of diversity. Each property was also classified into urban-rural categories (Bibby and Shepherd, 2001).
- Traffic Safety – For each neighbourhood the road density and the proportion of A roads were calculated. The density of road accidents, the class of road at the front of each property and whether this was a cul-de-sac or a through road was elicited. Finally, for non-through roads a proxy measures for traffic volume was estimated by modelling the flow of vehicles from addresses to through roads.
- Home Characteristics – The size of each property's garden was calculated by manually identifying the garden boundary. Each property was also queried to investigate whether other addresses were located with the same property.
- Neighbourhood quality – Information about the quality of each neighbourhood was obtained using information from the 2001 census and crime data from the police.

The analysis proceeded in two stages. Firstly all variables were correlated with all others within their category and any correlated by more than 0.7 identified. This provides an indication of data redundancy within variable groupings. A full multivariate model was then produced to examine the impact of the explanatory variables upon physical activity. A forward entry regression technique was performed by adding the most significant variable in turn.

4. Results

The data redundancy analysis is presented in table 1 with the multi-variable analysis in Table 2.

Table 1. Data Redundancy with the explanatory variables

Category	Total number of explanatory variables	Number remaining within category once correlations > 0.7 removed
Facilities and services	38	28 (74%)
Ease of walking	3	2 (75%)
Infrastructure for walking / cycling	4	4 (100%)
Neighbourhood surroundings	25	13 (52%)
Traffic Safety	14	11 (79%)
Home Characteristics	2	2 (100%)
Neighbourhood quality	27	16 (59%)
Totals	114	76 (66%)

Table 2. Environmental factors associated with physical activity in children during the weekend[1]

Explanatory Variables	Coefficient	95% CI	P
Traffic Level – relative to non-through road with quantiles 1 to 2 of simulated traffic. All private roads			(0.0011)
Non-through road, quantiles 3 to 5 of simulated traffic	-0.042	-0.085 - 0.00065	0.054
Non-through road, quantiles 6 to 9 of simulated traffic	-0.053	-0.093 - -0.012	0.011
Non-through road, quantile 10 of simulated traffic	-0.119	-0.178 - -0.0603	0.000
Through road classified as local street	-0.023	-0.065 - 0.018	0.273
Trough road classified as minor roads or B-road	-0.057	-0.094 - -0.019	0.003
Through road classified as A-Roads	-0.061	-0.131 - 0.0095	0.090
Location – relative to urban area			(0.0010)
Town and Fringe	0.054	0.026 - 0.081	0.000
Village	0.042	0.012 - 0.073	0.007
Hamlets & isolated dwellings	0.049	0.0036 - 0.095	0.034
Deprivation (IMD) score for super output area household is located in	0.0013	0.00028 - 0.0023	0.012

[1]The model also included variables to control for the bmi of the child, their sex, the average temperature that weekend and the amount of rainfall and the socioeconomic status of their parent

Adjusted r-squared	9.37%
Change in adjusted r-squared over adjusted model	1.81%
Numbers of observations	1552

5. Discussion and conclusions

114 environmental variables were created and these varied in the effort required to acquire the input data and calculate the variables within GIS. The most difficult data to obtain was the land use map which required combining information from Ordnance Survey Master Map with data from the CEH Land Cover Map. This process was time consuming and computationally intensive. Research into more efficient ways of joining such data or convincing data providers to provide such comprehensive land use maps would be helpful.

Focussing upon the difficulty of extracting the environmental variables using GIS, the overlay of the polygon neighbourhoods with other polygon layers (e.g. land use map, pavement polygons) was time consuming. Such operations are possible using the GUI of GIS programmes, but the data volumes involved overwhelmed these capabilities. Therefore, these could only be achieved through GIS programming techniques. Improvements to the operation of polygon in polygon overlays could significantly speed up and simplify such research.

An analysis of the correlations between the explanatory variables indicated that 1/3 were correlated ($r > 0.7$) with other variables indicating that many were measuring similar features. This was most apparent with the facilities and services variables, although because these were straightforward to create this duplication does not have large resource implications. It is a greater issue with the neighbourhood variables describing the land uses surrounding each property. These were difficult to create, and the measures describing urban land use were highly correlated. Omitting some of these variables could lead to time savings without losing important information.

In the multi-variable model 3 environmental variables were significant but these added less than 2% to the explanatory power of the model. These suggest that either our variables are poor measures of the environment, or that environmental variables do not play a major role in accounting for differences between MVPA in children. However, the magnitudes of effects are noteworthy. They indicate that living on moderately busy roads reduces MVPA by over 5%, residing outside an urban area is associated with a 5% increase in MVPA and living in a deprived area is associated with an increased MVPA of around 8%.

One feature of the multi-variable model is that none of the variables describing the child's neighbourhood were significant. A future research priority should therefore, be to examine how GIS based approaches can be used to calculate further environmental measures at local scales (e.g. property type, size of enclosed garden). The insignificance of these variables may also indicate that our neighbourhoods were incorrectly sized. Future studies should examine objectively (e.g. GPS tracking, questionnaires) the appropriate neighbourhood size for 9-10 year old children. An alternative would be to calculate environmental variables at a variety of spatial scales but the resources implications of this should not be underestimated.

In the multivariable model three variables were significant. The urban-rural measure and the IMD score were straightforward to obtain and calculate. However, the traffic level outside the child's home was one of the most difficult variables to create due to the lack of actual or modelled traffic volumes on small roads. Developing techniques to model robustly traffic volumes on all road types should be a priority.

6. Acknowledgements

The research is supported by a grant from the National Prevention Research Initiative. We also acknowledge the help of a large number of organisation who provided data for the project.
We are also extremely grateful to all the study participants and their parents for filling out the questionnaires, wearing an actigraph and providing us with a sample of their saliva.

7. References

BIBBY, P. & SHEPHERD, J. (2001) Developing a New Classification of Urban and Rural Areas for Policy Purposes. London, Office for National Statistics.

FORSYTH, A. (Ed.) (2006) *Environment and Physical Activity; GIS Protocols Version 4*, University of Minnesota, Metropolital Design Centre.

GOVERNMENT OFFICE FOR SCIENCE (2007) Tackling Obesities: Future Choices. London, Foresight.

HUMPEL, N., OWEN, N. & LESLIE, E. (2002) Environmental factors associated with adults' participation in physical activity. A review. *American Journal of Preventive Medicine,* 22**,** 188-199.

JAIN, A. (2005) Treating obesity in individuals and populations. *British Medical Journal,* 331**,** 1387-1390.

LAKE, I. R., LOVETT, A. A., BATEMAN, I. J. & DAY, B. (2000) Using GIS and large-scale digital data to implement hedonic pricing studies. *International Journal of Geographical Information Science,* 14**,** 521-541.

LESLIE, E., COFFEE, N., FRANK, L., OWEN, N., BAUMAN, A. & HUGO, G. (2007) Walkability of local communities: Using geographic information systems to objectively assess relevant environmental attributes. *Health and Place,* 13**,** 111-122.

8. Biography

Dr Lake is a lecturer in environmental sciences at the University of East Anglia. His research interests include environment and human health, climate change, and health and geographic information systems.

GP catchments and the characteristics of patients: a 'postcode lottery'?

Daniel James Lewis[1], Pablo Mateos[1], Maurizio Gibin[1,2] and Paul Longley[1]

[1] Department of Geography, University College London, Gower Street, London, WC1E 6BT
daniel.lewis@ucl.ac.uk, p.mateos@ucl.ac.uk, plongley@geog.ucl.ac.uk
[2] Department of Geography, Birkbeck, University of London, Malet Street, Bloomsbury, London, WC1E 7HX
m.gibin@bbk.ac.uk

KEYWORDS: Health Geography, General Practice, Catchment Areas, Accessibility, Patient Choice

1. Introduction

The aim of this paper is to explore some of the socio-spatial dimensions of family doctor 'choice' as recorded by general practice (GP) lists in Southwark, London. This is particularly relevant in the context of current National Health Service (NHS) initiatives to widen patient choice in secondary and tertiary care (NHS 2008a). The NHS exists to provide universal and equal access to health care to everyone, and although a misnomer, the term 'postcode lottery' does establish a correct connotation that the nature and level of care that patients receive does vary spatially (Campion-Smith 2007). In the context of patient choice, NHS guidelines refer to 'local' GP providers, while also noting that "[t]here are a number of reasons why you may not be able to register with your chosen GP. For example, the practice may be full or you may live too far away" (NHS 2008b). Jenkins and Campbell (1996) set out the relationship between catchment size and location of GP surgeries as encapsulated in NHS guidelines. These have the effect of limiting the areas within which patients can register. Robson (1995: 73) notes that "[f]or most people the convenience of the nearest surgery is the main determinant of where they register".

This paper develops a case study of the London Borough of Southwark to identify population groups and neighbourhoods that appear to deviate from the assumption of choice of a GP purely on the grounds of geographical proximity. Then follows an investigation of the revealed preferences of different groups with respect to GP practices. The analysis seeks to ascertain whether there are any demographic, socioeconomic or ethnic biases in the catchment area characteristics of Southwark GPs and whether it is possible to build up likelihood profiles for GPs with particular characteristics in similarly realised urban environments.

2. GP choice and accessibility

In cities, where geographic accessibility to health care is less important than in rural settings, focus switches to describing and analysing 'localities' (Shortt and Moore 2006). Whilst this promotes the health of 'communities', the London Borough of Southwark has 48 General Practices serving the needs of a diverse population of almost 270,000 (ONS, 2007). Therefore, before the possibility of certain groups, or 'communities', seeking healthcare from particular GPs can be entertained, the issues surrounding choice of doctor must be drawn out. By this it is understood that the characteristics of the populations that use one doctor rather than another, or which groups are more likely to travel further to visit a doctor, must be delineated. Shortt et al (2005) note that there is an increasing need for "GPs to possess more and more information about their registered population" (p 2715) and this study will help in this respect.

The core, pragmatic consideration for choosing a GP is geographic accessibility in terms of closest travel time from one's residence. This is of course not invariably the case in reality; there are many other factors behind why a person might choose a GP other than their closest one. If geographic accessibility, expressed as travel time distance, can be controlled for, then the remaining population traits governing choice of GP should become more evident.

3. Building a Model for GP Accessibility

The outcomes of assignments of patients to GPs can be investigated by comparing the characteristics of patients of any given practice with those that might be expected based solely on the criterion of geographical accessibility. The latter is termed here 'theoretical catchment'. In order to calculate such theoretical catchments, a model is constructed using a travel time distance matrix and Southwark's patient register.

Individual level data relating to GP registrations are collected and stored in the NHS 'Exeter' system which monitors activity nationally across primary care trusts (PCTs). Transport for London's (TfL) base public transport accessibility index (PTAI) data are used to estimate travel times using public transport (i.e. bus, tube or train). This dataset is created by dividing London into 100m grid cells, each with a measure of accessibility to transport service access points (i.e. bus stops). The measure is calculated assuming a walking speed of 4.8kph and a likely average waiting time of half of the interval between services. Then, the average travel time using the public transport network is calculated to all service access points. The resultant values for Southwark Output Areas (OAs) are based on an aggregated spatial join, where the points in the 100m grid fall within particular OA boundaries.

In addition to modelling GP catchments based on accessibility by public transport, a constraint is imposed whereby each GP can only be assigned a theoretical population of the same size as its current registered population.

The model works with General Practices that have a unique spatial location only. This means that General Practices that house multiple General Practioners are treated as one; similarly, for General Practices that have more than one spatial location (i.e. multiple surgeries operated by the same General Practioner) each location is treated as a unique General Practice. This leads to 44 unique GP locations. The total population of each OA used in the model was derived from the total population in Southwark's patient register.

Each GP is assigned a theoretical catchment via a location-allocation style algorithm (see Densham and Rushton 1992). The model runs thus:

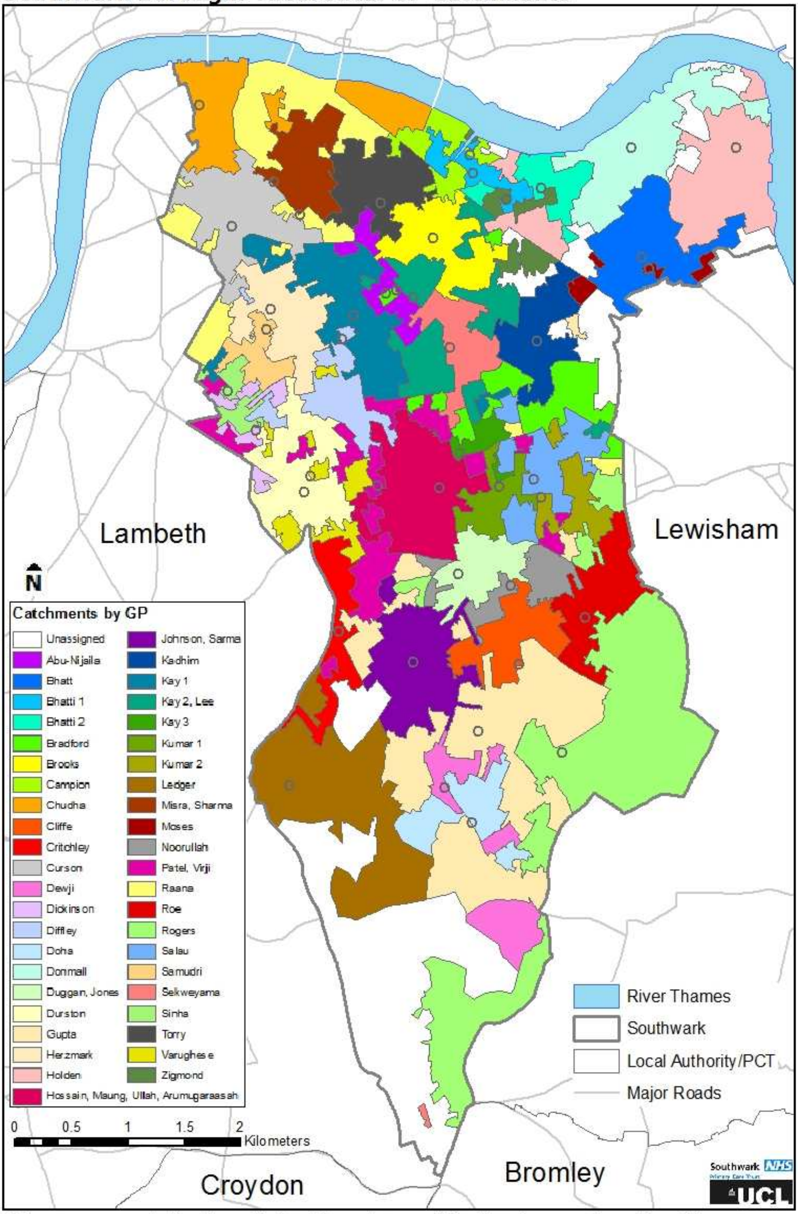
a) The travel time from each OA to all 44 GPs in Southwark is stored in a table where each record contains any possible pairing of OA and GP (i.e. 846 OAs times 44 GPs= 37224 records).
b) This table is then sorted by travel time (accessibility) in ascending order.
c) Starting from the first record, the first OA is assigned to its closest GP.
d) The next OA that has not being already assigned, is then allocated to its closest GP, repeating this step until the total population assigned to a GP reaches its GP size constraint (actual total population)
e) When this constraint is reached, the OA is then assigned to the next closest GP with available capacity.

Figure 1 shows the geographical visualisation of the resulting theoretical catchment areas, coloured by GP. White areas have remained unassigned due to GP populations having been reached; this seems to be an effect caused by people living in Southwark using GPs in other PCTs.

4. Using the Model to Analyse Patient and GP Characteristics

The analytical aspect of this paper falls into two part, using the GP catchment accessibility model defined above. First, the differences between the theoretical and actual registered populations are considered using the London Output Area Classification (LOAC), a derivative of the ONS Output Area Classification (OAC) devised by Petersen (2007) based upon clustering of London data only. Second, these same catchments are compared using a classification of ethnicity based on personal names (Mateos et al, 2007).

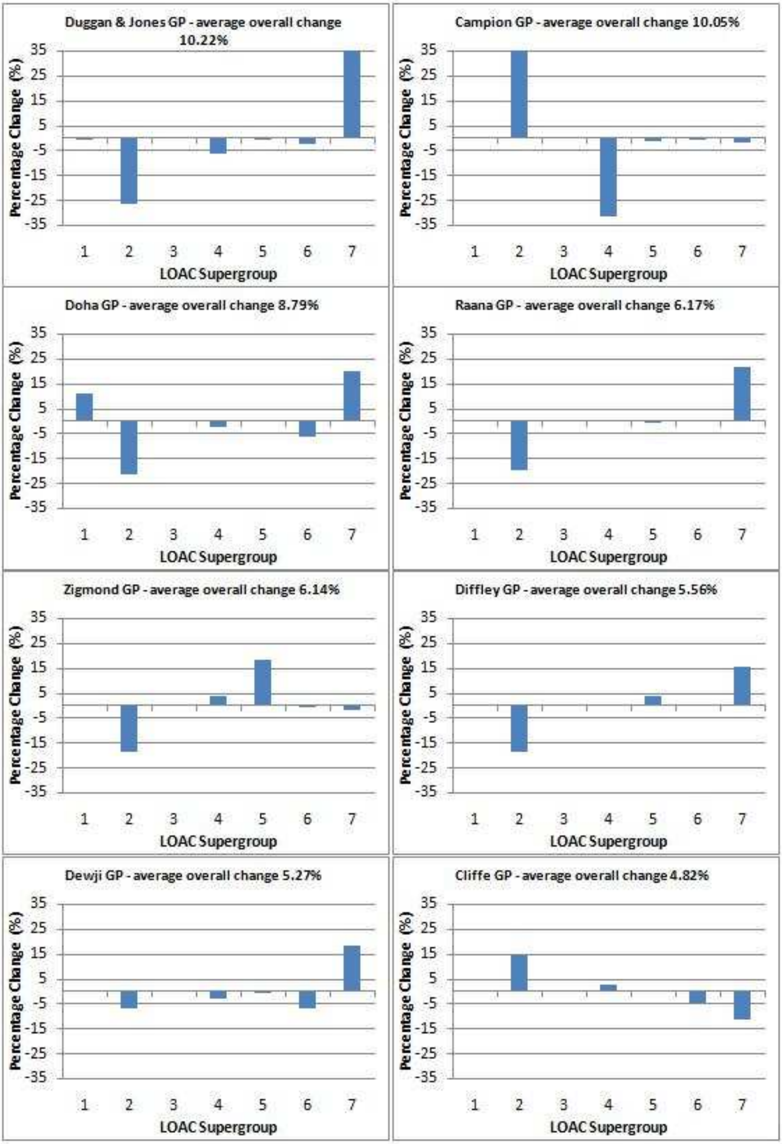
Southwark Borough: Theoretical GP Catchments



This map represents the theoretical catchment areas of Southwark general practices. These were created by a measure of public transport accessibility in minutes at the OA level.

The location of practices themselves are represented by a circle. All practices were preallocated to the OA that they fell within.

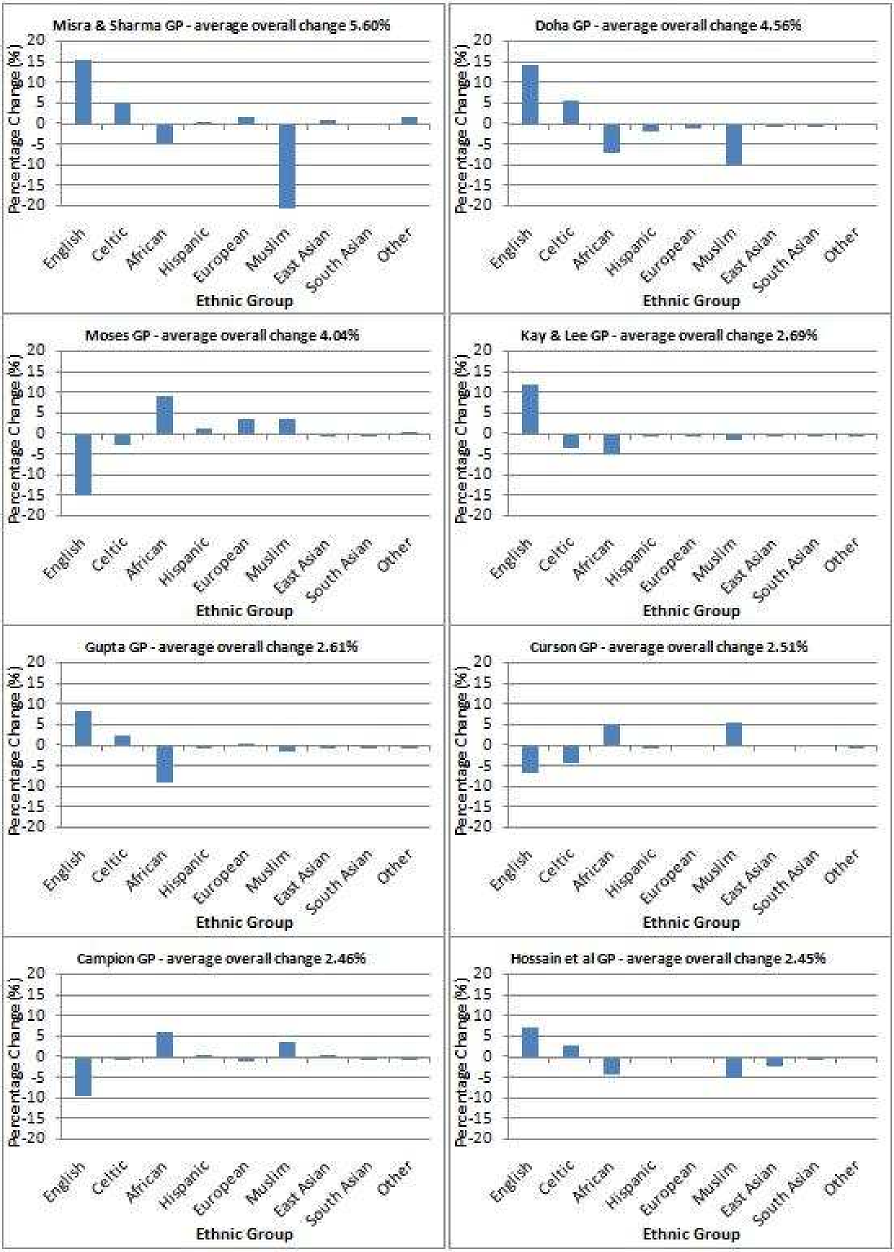
Figure 1: Theoretical GP Catchments for Southwark, London



<u>**Supergroup Meanings (from Vickers et al 2005)**</u>
1 – Blue Collar Communities **2** –City Living **3** – Countryside **4** – Prospering Suburbs
5 – Constrained by Circumstance **6** – Typical Traits **7** - Multicultural

A positive percentage for a LOAC supergroup class indicates a larger population belonging to this supergroup in the theoretical than in the actual catchments, and vice versa for negative values.

Figure 2: LOAC Variations between baseline and observed catchment estimates (Top 8 by overall difference)



A positive percentage represents a larger population size belonging to a particular ethnic group in the baseline than in the observed catchment, and vice versa for negative values.

Figure 3: Ethnic variations between baseline and observed catchment estimates (Top 8 by overall difference)

4.1 Variation between baseline and actual catchment characteristics according to LOAC Supergroup

The LOAC classification was devised by Petersen et al (2007) as a way of analysing and assessing ‘neighbourhood types’ in London’s population. LOAC is more sensitive to the distinctive characteristics of the capital than the widely used national OAC classification created by Vickers et al (2005). The 7 top level groups, the ‘super groups’, of the LOAC were assigned to both the actual registered populations and the theoretical catchment populations and compared for evidence of any major differences.

Figure 2 shows the top 8 GPs according to the difference in all LOAC supergroups from the theoretical to the actual catchments. GP *Campion* has over a third (35.2%, or 1,400 people) more patients in the ‘City Living’ neighbourhood supergroup than might be expected from the theoretical model and around 1300 people (31.6%) fewer in the ‘Prospering Suburbs’ supergroup. Similarly, GP *Raana* attracts 2100 fewer patients from the ‘multicultural’ neighbourhood supergroup than the model suggests it should.

The indication here is that residents of certain neighbourhood types are more or less likely to choose particular GPs. In the case of ‘the postcode lottery’ it may suggest that when a GP’s list is ‘full’, then the GP is more able to consciously select patients from a particular, preferable, neighbourhood type.

4.2 Variation between baseline and observed catchment characteristics according to ethnicity

Mateos et al (2007) describe a way of classifying individuals based upon the combination of their forenames and surnames. Using this methodology, the baseline and observed GP populations can be segmented according to their most probable ethnicity, and variations between the two can be analysed in a similar fashion as in the previous section, results of which are shown in Figure 3.

Figure 3 shows considerable differences in the ethnic composition of the theoretical and actual GP populations. For example GPs *Misra* and *Sharma* have a far greater registered Muslim population (about 1100 people) than expected; likewise *Doha* was expected to accounts for over 700 more ethnically English people than is observed in reality.

These differences may represent a choice of GP based upon ethnic lines, or word of mouth networks within ethnic communities. Conversely, it could also reflect a preference of GPs for particular ethnic groups, which if proved correct would carry significant consequences in terms of discrimination.

5. Conclusions and Future Work

This paper has developed a case study of the London Borough of Southwark that identifies population groups and local areas that deviate from the assumption that GP choice is purely based on geographical proximity to a patient’s residence. The GP catchment model developed here presents innovative features in that the analysis is performed at the individual level, using the actual patient register, and that real travel time through the public transportation network is used as an accessibility measure. This has allowed initial exploration into the potential reasons behind GP selection whether by choice or constraint, looking at GP selection by patients (demand effects) as well as patient selection by GPs, or “cherry-picking” (supply effects). Both processes are probably intertwined in creating the “postcode lottery” geography of primary healthcare mentioned.

The model presented needs further improvement and future work will concern two areas. Firstly, refinement of the catchment accessibility model, and secondly, the analysis of variations between the baseline and observed catchment populations of each GP.

Improving the model will mean using smaller areal units than OA, which seem to affect the quality of the output. Similarly the modelling process will be redefined to account for issues of contiguity and compactness. The hope is that the concurrent analysis will allow for a movement from analysing GP practices, to the catchments that specific General Practioners themselves have.

The ultimate goal is also to develop this methodology beyond the boundaries of Southwark PCT to the whole of London, or even to several urban areas/conurbations.

References

Campion-Smith, C (2007) "Ethics and Primary Care" in Leathard, A and McLaren, S eds. *Ethics: Contemporary Challenges in Health and Social Care*, The Policy Press, pp. 69-81
Densham, P and Rushton, G (1992) "Strategies for solving large location-allocation problems by heuristic methods", *Environment and Planning A*, **24**(2) 289 – 304
Jenkins, C and Campbell, J (1996) "Catchment areas in general practice and their relation to size and quality of practice and deprivation: a descriptive study in one London borough" *BMJ* 313:1189-1192
Mateos, P Webber, R and Longley, P (2007) The cultural, ethnic and linguistic classification of populations and neighbourhoods using personal names. *Working paper 116. Centre for Advanced Spatial Analysis* (UCL), London. Available at: http://www.casa.ucl.ac.uk/working_papers/paper116.pdf
NHS (2008a) "About Patient Choice" available at http://www.nhs.uk/choices/Pages/Aboutpatientchoice.aspx last accessed 8/12/2008
NHS (2008b) "How do I register with a GP?" available at http://www.nhs.uk/chq/Pages/1095.aspx?CategoryID=68&SubCategoryID=158 last accessed 8/12/2008
ONS (2007) "Key Population and Vital Statistics", Brown, S & Redgwell, G Eds. **Series VS No 33, PPI No 29** 2006 Data, Palgrave Macmillan
Petersen, J Longley, P and Atkinson, P (2007) "Health Applications for Open Geodemographics", *GISRUK Proceedings 2007*, available at: http://ncg.nuim.ie/gisruk/materials/proceedings/PDF/3A2.pdf last accessed 8/12/2008
Robson, J (1995) "Editorial: GPs, Patients and the distance between them" *BMJ* **310**:73-74

Shortt N and Moore A (2006) Functional localities: an integrated spatial approach towards health care locality definition. *GeoJournal* **67**: 27-40

Shortt N K, Moore A, Coombes M and Wymer C (2005) Defining regions for locality health care planning: a multidimensional approach. *Social Science & Medicine* **60**: 2715-27
Vickers, D Rees, P and Birkin, M (2005) "Creating the National Classification of Census Output Areas: Data, Methods and Results", *Working Paper 05/2 July 2005, School of Geography, University of Leeds*.

Biography

Daniel James Lewis is a PhD student in the Department of Geography at University College London (UCL). He has an MSc in Geographic Information Science from UCL, and a BA in Geography from the London School of Economics.

Pablo Mateos is Lecturer in Human Geography in the Department of Geography at University College London (UCL). He obtained his PhD in Geography from UCL (2007) and MSc in GIS and Human Geography from the University of Leicester (2004).

Paul Longley is Professor of Geographic Information Science in the Department of Geography, UCL. His research interests concerns the use of GIS and quantitative methods in urban analysis. He is Deputy Director of Centre for Advanced Spatial Analyses at UCL and UCL Site Manager of the 'Spatial Literacy in Teaching' (Splint) Centre of Excellence in Teaching and Learning.

Implementing Grid-enabled GWR for teaching

Rich Harris[1], Daniel Grosse[2], Chris Brunsdon[3]

[1]School of Geographical Sciences, University of Bristol, University Road, Bristol. BS8 1SS.
Tel. +44 117 954 6973
rich.harris@bris.ac.uk
[2] Centre for e-Science, Lancaster University, C Floor, Bowland Tower South, Lancaster, LA1 4YT.
[3] Department of Geography, University of Leicester, Leicester, LE1 7RH

KEYWORDS: Geographically Weighted Regression (GWR), distributed computing, e-social science, National Grid Service, spatial statistics

1. Introduction

Geographically Weighted Regression (GWR) is a method of statistical regression that allows modelled relationships to vary across geographical space (Fotheringham et al., 2002). The technique successfully has been demonstrated in a variety of applications, However, a constraint on the use of GWR is that its computational requirements grow exponentially with the length of the data set (the number of observations, n).

E-social science offers a way to overcome that constraint. In the UK, multi processor computer resources are available to academic communities for computationally-intensive analyses of large datasets. These are promoted by the National Centre for e-social science (NCeSS) and by the National Grid Service (NGS), the latter of which aims “to provide coherent electronic access for UK researchers to all computational and data based resources and facilities required to carry out their research, independent of resource or researcher location” (www.grid-support.ac.uk). In other words, access to “grid” or distributed computing.

In this paper we report briefly on a project that has “grid-enabled” GWR, using conventional desktop R software to connect and to run models on the NGS. Our focus, however, is less on that project, and more on a successor project, funded by SPLINT (see acknowledgments), that allows the potential for grid-enabled GWR to be demonstrated and taught to students without the security and registration issues that access to the NGS infrastructure otherwise entails.

2. Implementing grid-enabled GWR in R

An objective of the NCeSS-funded research upon which this paper draws was to create a way of grid-enabling GWR that is as interoperable as possible with existing software. To achieve this, we built on a library for running GWR in the computing and statistical package, R (http://cran.r-project.org/). This is the *spgwr* library, developed by Bivand and Yu, that provides functions for calibrating the bandwidth and for calculating the geographically weighted regression parameters.

A first reason for choosing R is that it is open source and freely available. A second was that a previous NCeSS project called SABRE in R had involved the Lancaster University Centre for e-Science developing a parallel implementation of SABRE (a program for the statistical analysis of binary, ordinal and count recurrent events) as R Objects. That project had used middleware called GROWL “to provide user friendly access to GRID resources for applications accessible from a desktop computer” (www.ncess.ac.uk/research/quantitative/cqess/growl/).

Using GROWL technology, a package was developed that allows GWR to be run on a desktop PC using the existing *spgwr* library but for which the actual data processing occurs remotely on the National Grid infrastructure. The package was entitled *multiR* and is actually a client/server system providing a means to submit a group of tasks for processing on multiple and remote systems. These systems could be processors on a local high performance cluster, a Condor pool or, in the case of the research project, the NGS. The *multiR* client interface is distributed as a package for R and its usage is similar that of the standard R function *lapply*. The idea of *multiR* is to provide a means of invoking an R function multiple times and with varying arguments, where the result of the function is evaluated on multiple processors. By doing so, R becomes a programming environment for coarse grained parallel processing.

Figure 1 outlines the principle of *multiR* and its three tier architecture. Clients use R to define the function and use *multiR* to submit a job to the *multiR* server. The *multiR* server then delegates these tasks to whatever resources it can employ, managing the interface between the user's computer and the NGS.

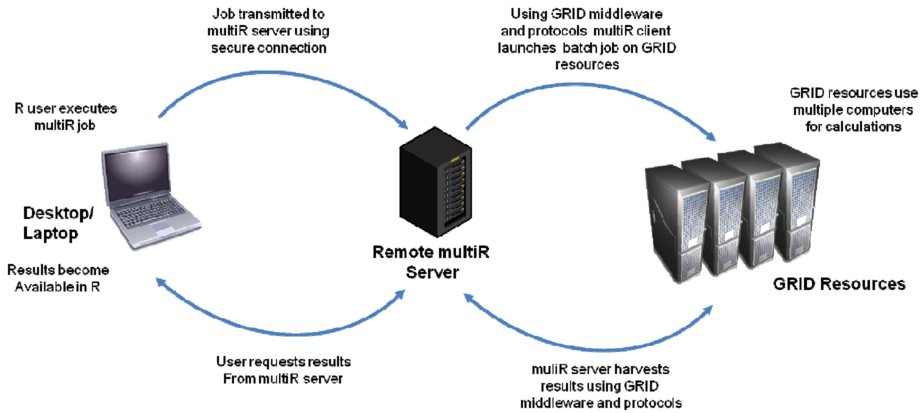


Figure 1. The three tier client/server architecture employed by *multiR*.

In addition to *multiR*, the accompanying *spgwr.dist* library contains the functions required for grid-enabled GWR (dist is an abbreviation of distributed). A typical session in R begins as:

```
> library(spgwr.dist)
# loads the spgwr.dist and multiR packages
> session <- multiR.session("stats-grid.hpc.lancs.ac.uk", "50000",
+ "~/mycertificate.p12", "~/multiR.CA.pem ")
# identifies the user's security credentials to use the NGS and also
the multiR server
```

The analysis then continues in a way similar to using the *spgwr* packgage. Where, in *spgwr*, the bandwidth for GWR is calculated on the user's desktop using a function of the form

```
> bandwidth = gwr.sel(y~x, atribute_data, georeferences)
```

for the grid-enabled version we use

```
>    bandwidth    =    gwr.sel.dist(session,    y~x,    atribute_data,
```

```
georeferences, max.processors)
```

Similarly, where the model is fitted in *spgwr* using

```
> gwr.model = gwr(y~x, atribute_data, georeferences, bandwidth)
```

it is fitted in *spgwr.dist* using

```
> gwr.model = gwr.dist(session, y~x, atribute_data, georeferences,
bandwidth, max.processors)
```

The only difference from the user's perspective is that the additional parameter "session" contains the information required to connect to the *multiR* server – including the user's security credentials – and the parameter "max.processors" specifies a maximum number of processors the GWR analysis should use.

3. Implementing Grid-enabled GWR for teaching

The three tier client/server architecture shown in Figure 1 is effective for a researcher who wants to use the NGS. However, for a student, it is less useful. The principal reason for this is that access to the NGS requires both that an application be made and reviewed, detailing the nature of the project to undertaken using the NGS, and that identity/security checks be undertaken before the applicant is given access – involving, for example, showing a passport to a local administrator.

The application process is documented and initiated at http://www.grid-support.ac.uk/ and can take up to two weeks. Although there are clear reasons for the process, it must be regarded as a sizeable barrier to anyone who wants to try out or to learn about grid-enabled GWR without making a more formal commitment. For a class session, to register all the students is unlikely to be feasible.

To overcome this we currently are developing a 'slimmed down' multiprocessor system that will give the appearance of using the NGS but actually is self-contained and a step removed from it. In terms of Figure 1, it stops at the *multiR* server but still gives the impression of operating in the same distributed environment of the NGS.

The advantage of a self contained system is that it does not require the user to register for the NGS. Instead, user certificates can be generated on demand. This makes it more practical for teaching and demonstration purposes; we currently are preparing teaching material around its use.

4. Conclusion

The National Grid Service and e-social science in general hold promise for the development and evolution of spatial statistics within a distributed computing environment. To date, however, the focus of e-social science has been on the research side and less on teaching. For computational e-social science to flourish it needs to engage with a new generation of users in a way that is accessible for teaching purposes and does not present off-putting barriers such as registration to the NGS. The project this paper presents is, we believe, unique and the first time anyone has sought to mimic the potential for using the NGS (in this case, for geographically weighted regression) without actually connecting to it.

5. Acknowledgements

The initial research was funded by the National Centre for e-social science / ESRC as research project RES-149-25-1041. The teaching material is being developed under a Fellowship generously funded by SPLINT (Spatial Literacy in Teaching: http://www.le.ac.uk/gg/splint/).

References

Fotheringham, AS, Brunsdon, C, & Charlton, M (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying* Relationships. John Wiley & Sons, Chichester.

Biography

Rich Harris is a Senior Lecturer at the University of Bristol. He is interested in the contribution of faith and religion to political debate and governance, the causes and persistence of socio-spatial inequalities, and how space is conceived and incorporated in statistical models (e.g. agency and structure, social interaction and the generation of 'neighbourhood effects').

Daniel Grosse is a software developer at the Centre for e-Science, Lancaster University.

Chris Brunsdon is currently Professor of Geographical information at the University of Leicester. His interests include the development of geocomputational algorithms, and spatial analysis approaches and statistical inferential techniques to a number of 'real world' problems in human and physical geography – such as the analysis of house prices, crime patterns, and the monitoring of water quality.

Utilising scenarios to facilitate a multi-objective land use model: the Broads, UK, to 2100

Paul Munday[1], Andy Jones[1], Andrew Lovett[1] and Paul Dolman[1]

[1]School of Environmental Sciences, University of East Anglia, Norwich, Norfolk, NR4 7TJ
Tel.: +44 (0)1603 591343; Fax.: +44 (0)1603 591327
paul.munday@uea.ac.uk

KEYWORDS: Multi-objective, multi-criteria, land use, scenarios, resolution

1. Introduction

Substantial and continuous land use change has occurred within the UK over hundreds of years due to societal, economic and environmental pressures and it is sensible to expect that changes are likely to continue in the future, which will have consequences for a wide variety of human- and environmental-based systems (Ratcliffe, 1984). Perhaps the greatest pressures upon land use and hence landscape change in the near future can be attributed to socio-economic and climatic perturbation. Consequently, some research (e.g. Holman and Loveland, 2002) has attempted to quantitatively model the impacts of these pressures at the regional spatial scale. Scenarios are often used to qualitatively describe some of the potential changes which are expected to occur and as a method for dealing with uncertainty in the decision-making process (IPCC, 2000). However, model outputs and spatial scenarios are often of a poor resolution and hence do not lend themselves well for application within local landscapes. This paper focuses upon the utilisation of scenarios and regional-scale land use change data to facilitate the production of a model of land use change, at high spatial resolution, within an environmentally sensitive wetland landscape, in Norfolk, UK.

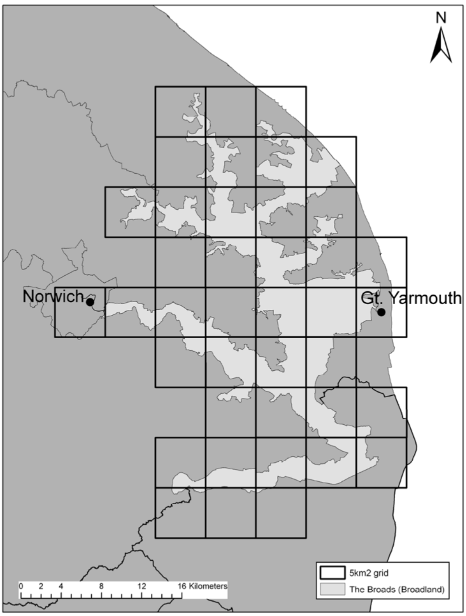


Figure 1. Broadland study area. © Crown Copyright. All rights reserved. Ordnance Survey 2008.

2. Study area

The Broadland (Figure 1) landscape comprises grazing marshes, fen and woodland that support numerous threatened and scarce species of flora and fauna of high conservation concern. It contains 28 Sites of Special Scientific Interest (SSSIs), amounting to 7,000 ha in total, which benefit from protection, either as Special Protection Areas (SPAs) or Special Areas of Conservation (SACs). Pressures for change in Broadland include a multitude of social, economic and environmental factors which may influence future land use, including climate change and changing agricultural policy and associated support measures. Thus, owing to its complex management, environmental sensitivity and competition for land use, the need to identify potential future landscape change within the region is pressing.

3. Application of GIS in Multi-Criteria Decision Analysis (GIS-MCDA)

GIS has been utilised in MCDA for a variety of different purposes including locating facilities (e.g. Kao, 1996), generation of alternative location-allocation patterns (e.g. Cova and Church, 2000) and identification of alternative patterns of land use suitability (Brookes, 1997). Indeed, there has been a substantial increase in GIS-MCDA research over the last 15 years, with an associated increase in popularity of MCDA modules in GIS systems such as IDRISI (Eastman et al., 1993). However, there is a distinct lack of research into the use of both climate and socio-economic scenarios in facilitating multi-objective/multi-criteria decision problems. Such a shortage is likely to be due to emphasis being placed upon the technicalities of integrating GIS and MCDA rather than operational validation of using GIS-MCDA to solve real-world spatial allocation problems (Malczewski, 2006).

3.1. Land use data

A variety of scenario-based projects exist which have produced quantitative land use datasets for application at a range of spatial scales (e.g. PRELUDE – Hoogeveen et al., 2005; ACCELERATES – Abiltrup et al., 2006, and; RegIS – Holman and Loveland, 2002). RegIS provides qualitative scenarios (downscaled to be applicable within East Anglia) and quantitative land use data with areal coverage of arable, permanent grassland, urban and woodland extent within 5 km^2 grids across East Anglia. In total, 40 5 km^2 grids cover the Broadland study area (Figure 1). RegIS utilises 1995 (baseline) and 2050 as time-points and provides land use data for four different scenarios; two socio-economic scenarios (SESs) and two ‘climate-only’ scenarios. The Regional Enterprise (RE) SES represents a future world in which economic development takes precedence over environmental concerns which is realised through weakened environmental policy and values. RE is coupled with a High climate change scenario (see Hulme and Jenkins, 1998) to represent the greatest pressure upon the economy and the environment. The Global Sustainability (GS) SES represents an opposing world in which environmental protection is given higher priority than economic development. GS is coupled with a Low climate scenario which represents the lowest pressure upon the economy and the environment. Both ‘climate only’ scenarios are utilised to distinguish changes attributable only to anthropogenically-induced climate change.

Despite the richness of outputs, the RegIS land use data does not specify the location of each land use type within individual grids. Therefore this research is undertaken to develop a methodology that enables the output of scenarios, such as RegIS, to be downscaled and re-repeated at a local level which is suitable for the generation of detailed maps and complex visualisations.

4. Methodology

The methodology developed has two main stages. In the first stage, the suitability of different parts of the study area was determined for a range of land uses based on a detailed set of criteria. In the second stage, land use trends were identified; future land uses were allocated and maps were produced depicting these predicted landscape trends.

Initially, a vector land use map was created, for the study area, utilising two existing land use datasets (LCM2000 and Ordnance Survey's 1:1250 OS Mastermap product) to provide a baseline (1995) map. Eight land use categories were identified, including the four categories for which data was provided by RegIS, plus recreation, roads, uncultivated land and water. To check for accuracy, 1000 randomised points were generated and the land use type at that point validated with 2004 aerial photography (*c*. 88 % successfully identified). Data was converted to raster format with a cell resolution of 5 m^2.

4.1. Multi-Criteria Evaluation and Multi-Objective Allocation

A range of allocation criteria were identified and each was set to be either a constraint or factor. Constraints are areas not suitable for the objective (land use) in question. Factors are generally continuous in nature (e.g. distance or slope gradient) which indicate the relative suitability of certain areas. The criteria utilised in modelling land use data from the RegIS project were adopted (where applicable) in the first instance. For example, in the context of identifying cells suitable for agriculture, physical constraints (e.g. slopes over 11 %) were adopted whilst factors included agricultural land grade and current land uses and presence outside of the floodplain. Criteria were identified and this method was repeated for all other land uses. Factors were standardised using a linear scaling method and combined using a weighted linear combination approach (see Eastman et al., 1995). Criteria were then combined to form a suitability map representing individual cell suitability for the objective in question on a byte (0-255) data scale. Then, areal totals for each land use category were allocated according to RegIS land use data. This stage represented the development of a land use map for each 5 km^2 grid for the year 2050. A sensitivity analysis of factors was performed with variation being constrained to a handful of fields (representing *c*. 5 % of grid area).

4.2. Identify and project land use trends

The methodology utilised a Cellular Automata/Markov Chain Analysis (CA-MCA) based decision support tool to identify and iteratively project land use trends between two time-points (in this case 1995 and 2050). The coupling of both processes allowed a spatial dimension to be added to the modelling process whereby cells in proximity to existing land uses were weighted more heavily than those farther away. The module was run for 50 iterations with a user-generated 7 x 7 contiguity filter in order to generate a land use map for the year 2100. Figure 2 provides an example of the mapped output generated.

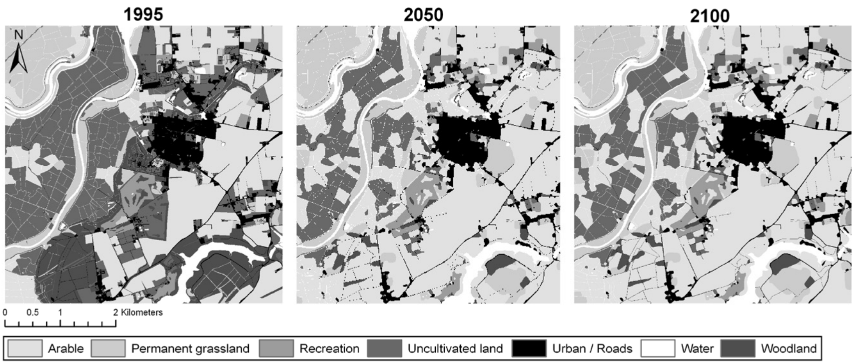


Figure 2. Example of 1995 - 2100 land use change within a 5km^2 grid under the Global Sustainability/Low climate scenario. Note: the large urban area located off-centre is the town of Belton. © Crown Copyright. All rights reserved. Ordnance Survey 2008.

Due to the cellular nature of the modelling approach, areas of the eight land uses were able to be replicated within 0.01 % (or 0.25 ha) of areal totals prescribed by the RegIS land use data for 2050 and these may then be flexibly projected into the future. As a means of validation, stakeholders and experts were invited to comment upon the degree to which the associated scenario drivers could manifest themselves into the balance of land uses depicted by the maps. The conversion of flood-prone land for agriculture was highlighted as a concern; however, we suggest that this is a shortcoming of localising more national/regional-scale scenarios, and land use data, to be applicable within local landscapes as they do not explicitly allow for such local constraints.

A number of key benefits of this approach are evident. The land use maps allow the attribution of scenario storylines to individual land use categories which may aid future landscape planning decisions. Further, this work has allowed the detailed depiction of potential future landscapes that may result from alternative scenarios of contrasting policy and other drivers. The approach also represents the first locally explicit realisation of coarser regional-scale land use change data using an integrated GIS-MCDA methodology. The methodology developed here may be applied within a range of contrasting landscapes and environments. It is envisaged that map outcomes, and the methodology presented here, may be utilised to facilitate studies of habitat fragmentation and connectivity, biodiversity impacts and landscape visualisation.

5. Summary

The coarse modelling resolution of some regional-scale scenario-based projects can render outputs inapplicable within local, often environmentally sensitive, landscapes. Improving data resolution allows us to investigate alternative potential futures at greater spatial detail thereby providing vital input into policy and future decision-making. A GIS-MCDA based approach can provide a means of producing maps of future landscapes which are of a high spatial resolution from coarse input data.

6. Acknowledgements

The first author is in receipt of an ESRC/NERC Ph.D studentship award (PTA-036-2006-00006). The authors would like to acknowledge Dr. Ian Holman for providing datasets on which the analysis is predominantly based.

References

Abildtrup, J., Rosato, P., Gylling, M., Fekete-Farkasne, M., Audsley, E., Giupponi, C. and Rounsevell, M.D.A. (2006) Socio-economic scenarios for agricultural land use studies. *Environmental Science and Policy*, **9**, 101-115.

Cova, T.J. and Church, R.L. (2000) Contiguity constraints for single-region site search problems. *Geographical Analysis*, **32**, 306-329.

Brookes, C.J. (1997) A parameterised region-growing programme for site allocation on raster suitability maps. *International Journal of Geographical Information Systems*, **11**, 375-396.

Eastman, J.R., Kyem, P.A.K., Toledano, J. and Jin, W. (1993b) GIS and Decision Making (Gemeva: UNITAR).

Eastman, J.R., Jin, W., Kyem, P.A.K., Toledano, J. (1995) Raster procedures for multi-criteria/multi-objective decisions. *Photogrammetric Engineering & Remote Sensing*, **61 (5)**, 539-547.

Holman, I., Loveland, P. (Eds.). (2002). Regional Climate Impact Studies in East Anglia and North West England: Technical Report. Final report to MAFF, DETR and UKWIR. Soil Survey and Land Research Centre, Silsoe. 360pp.

Hoogeveen, Y., Volkery, A., Henrichs, T., Ribeiro, T. (2005). Land use scenarios for Europe – Modelling at the European Scale. Background Report for the European Environments Agency (EEA). 75pp.

Hulme, M. and Jenkins, G.J. (1998). *Climate Change Scenarios for the United Kingdom: Scientific Report*. UK Climate Impacts Programme Technical Report No. 1, Climatic Research Unit, Norwich, 80pp.

Intergovernmental Panel on Climate Change (IPCC). (2000). Special Report on Emissions Scenarios (SRES). Working Group III, IPCC. Cambridge University Press: Cambridge. 595pp. Available at: <http://www.grida.no/climate/ipcc/emission/> [accessed 10th May, 2005].

Kao, J.J. (1996) A raster-based C program for siting a landfill with optimal compactness. *Computers and Geosciences,* **22**, 837-847.

Malczewski, J. (2006) GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographic Information Science*, **20 (7)**, 703-726.

Ratcliffe, D.A. (1984) Post-Medieval and recent changes in British vegetation: The culmination of human influence. *New Phytologist*, **98**, 73-100.

Biography

The first author has a BSc in Geography and Physical Geography from Oxford Brookes University and an MSc in Climate Change from the University of East Anglia. The first author is a third-year Ph.D. student researching socio-economic and climate change impacts upon wetland environments. His supervision committee consist of Dr. Andy Jones, Prof. Andrew Lovett and Dr. Paul Dolman.

Visualising spatio-temporal crime clusters in a space-time cube

Tomoki Nakaya[1], Keiji Yano[2]

[1]Department of Geography, Ritsumeikan University
56-1 Tojiin-kita-machi, Kita-ku, Kyoto 603-8577, Japan
Tel: +81-(0)75-467-8801 Fax: +81-(0)75-467-8825
E-mail: nakaya@lt.ritsumei.ac.jp
URL: http://www.ritsumei.ac.jp/~nakaya

KEYWORDS: snatch-and-run offence, space-time scan statistics, kernel density, volume rendering, 3D GIS

1. Introduction

Spatial epidemiological methods are now widely used for analysing crime events, particularly the so-called hotspots, i.e., spatial clusters of crime, with the aid of GIS and spatial statistics. It is also well known that such spatial clusters are recognised as moving or transient objects that reflect diffusion and/or displacement of crime events (Paulson and Robinson, 2004). Thus, mapping the crime events in different time periods often reflects distributional changes of crime clusters over those time periods. However, it is not easy to interpret the spatio-temporal (dis)continuity in crime clusters from a series of conventional crime maps in different time periods.

The aim of this study is to examine the possibilities of three-dimensional mapping of crime event data in a space-time cube in order to visually comprehend the temporal duration as well as the spatial extent of the crime clusters simultaneously. A space-time cube is a three-dimensional space composed of two geographic dimensions and one time dimension (Figure 1). The concept originates from time geography diagrams used for the visual explanation of continuous space-time sequence of geographic objects.

We mainly explore two statistical methodologies as a set of exploratory spatio-temporal data analysis (ESTDA) tools in a 3D GIS environment: (1) three-dimensional kernel-density by volume rendering (ray tracing) and (2) space-time scan statistics for detecting significant space-time cylindrical clusters. Both methodologies are used to visualise spatio-temporal domains with a high density of crime by 3D GIS and related technologies. We evaluate these methodologies by applying them to a dataset of snatch-and-run offences reported in Kyoto City, Japan.

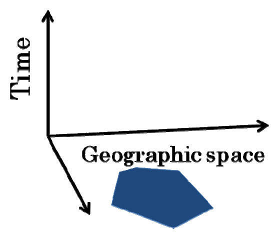


Figure 1. Space-time cube

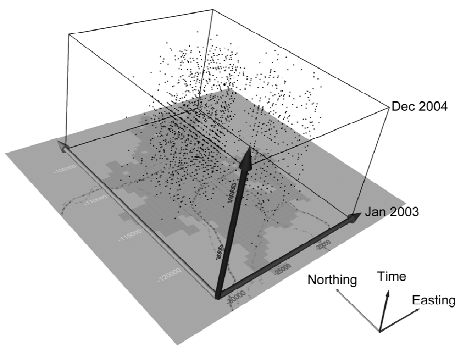


Figure 2. Space-time spot mapping of crime events

2. Data

Snatch-and-run offence is a type of robbery committed on roads: offenders snatch purses or bags and escape by riding a motorcycle in most situations. The dataset used in the study comprises 1855 occurrence points of snatch-and-run offences reported to the Kyoto Prefectural Police during the period 2003–2004. Each occurrence point is shown by a pair of x-y geographic coordinates and the month of occurrence. Figure 2 shows the data points mapped in a space-time cube where *i*th offence is plotted on a three-dimensional coordinate system (x_i, y_i, t_i).

3. Spatio-temporal kernel density

Kernel density estimation is commonly used to transform a set of point distribution into a density surface for the effective interpretation of a two-dimensional point distribution. Assuming an orthogonal relation between space and time dimensions, Brunsdon et al. (2007) have proposed a spatio-temporal three-dimensional kernel density estimate as follows:

$$\hat{f}(x,y,t)=\frac{1}{nh_s^2h_t}\sum_i K_s\left(\frac{x-x_i}{h_s},\frac{y-y_i}{h_s}\right)K_t\left(\frac{t-t_i}{h_t}\right) \tag{1}$$

where $\hat{f}(x,y,t)$ is the density estimate at location (x, y, t), n is the number of events, and h_s and h_t are spatial and temporal bandwidths, respectively. Kernel functions of both K_s and K_t are defined here using the Epanecknikov kernel, which is used in ArcGIS (ESRI Inc.):

$$K_s(u,v)=\begin{cases}\frac{2}{\pi}\left(1-\left(u^2+v^2\right)\right) & \left(u^2+v^2\right)<1\\ 0 & \text{otherwise}\end{cases}$$
$$K_t(w)=\begin{cases}\frac{3}{4}\left(1-w^2\right) & w^2<1\\ 0 & \text{otherwise}\end{cases} \tag{2}$$

Volume rendering (ray tracing) is used to visualise the hyper-surface of density, wherein the two-dimensional projections of a shaded volume are created by using a hypothetical source of light. The degree of transparency is controlled to be high in low density regions so that high density domains appear as a shaded volume with colours graded by density values. Figure 3 shows the space-time density of the crime events in the dataset by using volume rendering.

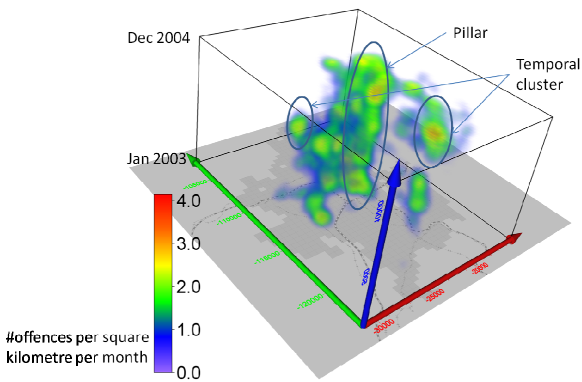


Figure 3. Space-time density of snatch-and-run offences visualised by volume rendering

4. Space-time scan statistics

Spatial scan statistics is devised to statistically detect significant clusters by exhaustively scanning over space using moving circular windows with different radii. Following the same concept of exhaustive search by using moving windows, space-time scan statistics employs cylindrical search windows in a space-time cube in order to seek significant spatio-temporal clusters (Kulldorf et al., 1998). A cylindrical domain is defined as

$$\{(x,y,t) \mid (x-x_c)^2+(y-y_c)^2 \le r^2, t_s \le t \le t_e\} \tag{3}$$

where (x_c, y_c) is the geographical centre of the domain, r is the radius of the domain, and [t_s, t_e] is the domain period. By assuming that the expected number of crimes is spatially constant in the build-up area of the city, we applied Poisson model of space-time scan statistics to a dataset consisting of crime event counts aggregated in each 500m grid cell. The maximum radius of the cylindrical search window is set as 1 km based on a preliminary investigation of cluster sizes. The result as visualised in a space-time cube is shown in Figure 4.

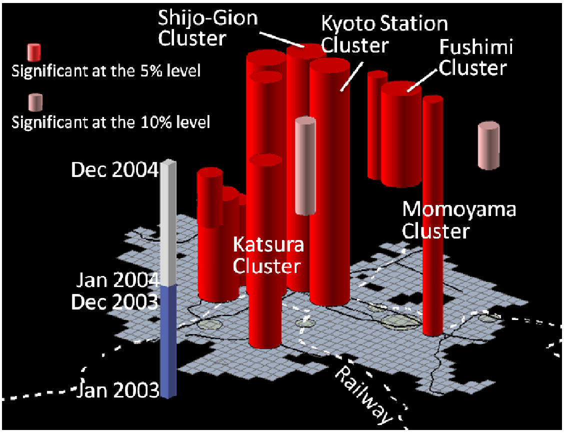


Figure 4. Significant space-time cylindrical clusters of snatch-and-run offences detected by space-time scan statistics

5. Comparative evaluation of the two methodologies

In general, both the three-dimensional mapping methodologies shown in Figures 3 and 4 reveal similar spatio-temporal tendencies of the crime clusters. As shown in Figure 3, high density domains appear as pillars around the Kyoto station and the central part of the city in the space-time cube. This implies that the density of crimes reported in these geographical regions is constantly high. In contrast, transient/temporal spatio-temporal clusters are identified around several suburban railway stations. These two aspects of spatio-temporal clusters can be clearly identified from the space-time scan statistics as well (Figure 4).

An overall trend is that transient clusters moved from the northern to the southern part via the western part of the city. This indicates a displacement phenomenon of crime: local crime preventive actions in the northern part might have pushed out the crime clusters to the western part and then to the southern part of the city.

Another interesting result suggesting a displacement phenomenon is observed in the temporal relationships between the cluster regions of Katsura and Momoyama. We calculate the Spearman rank correlation of monthly-series of crime occurrence between each pair of crime cluster regions detected by space-time scan statistics, and a significant negative correlation between the Katsura and Momoyama cluster-regions is found. As shown in Figure 5, clusters of snatch-and-run offences appear alternatively in the two regions. Offenders change their activity to avoid local police activities that are often intensified after an outbreak of crime begins in the local region. However, acquiring the local knowledge of a different region would be expensive for the offenders. Considering that the level of local police activities decrease after a certain period, it is reasonable for offenders to target the known geographic regions alternately with a sufficient time interval and avoid local police activities in order to commit crimes efficiently.

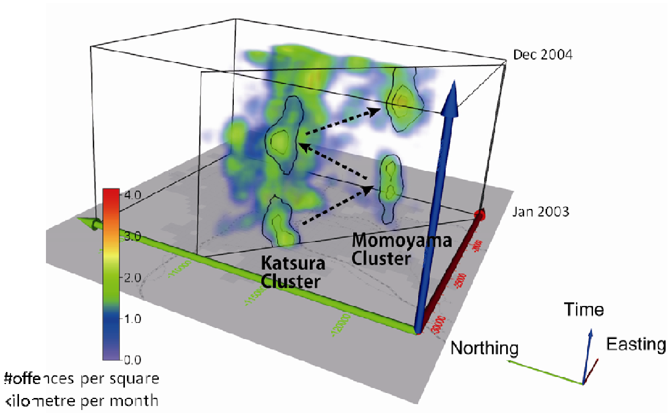


Figure 5. Alternating occurrence of crime clusters in a pair of cluster-regions

6. Conclusions

(1) Three-dimensional crime mapping enables effective simultaneous visualisation of the geographical extent and duration of crime clusters/hotspots. This method is particularly more useful in identifying the geographical diffusion and movement of crime clusters/hotspots as compared to the traditional temporal crime mapping carried out using cross-sectional maps with arbitrary time intervals.

(2) Three-dimensional kernel mapping and space-time scan statistics are complementary to each other. Space-time scan statistics rigidly specify crime clusters domains that can be used for secondary analyses such as evaluation of the socio-environmental and temporal characteristics of the detected domains. It should be, however, noted that the method assumes that space-time domains of crime clusters are cylindrical. Three-dimensional kernel density mapping provides fuzzy domains with a high density of crime, but it can be used as a basis to assess the validity of the assumption of spatial scan statistics and to investigate detailed space-time sequences of crime clusters.

(3) An empirical analysis of the snatch-and-run offences dataset of Kyoto City revealed consecutive clusters during the study period in the central part of the Kyoto City and around Kyoto Station as well as transient clusters around several suburban railway stations. Temporal differences in the transient clusters suggesting a so-called ‘displacement’ phenomenon indicate the necessity to monitor crime events and effects of crime preventive actions in a widespread space-time context.

Acknowledgements

We would like to thank members of *Hanzai Jyosei Bunseki Shitu* (Department of Crime Occurrence Analysis) of the Kyoto Prefectural Police for providing the dataset for this study.

References

Brunsdon C, Corcoran J and Higgs G (2007) Visualising space and time in crime patterns: A comparison of methods *Computers, Environment and Urban Systems* **31** pp52-75

Kulldorff M, Athas W, Feuer E, Miller B and Key C (1998) Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos *American Journal of Public Health* **88** pp1377-1380

Levoy M (1988) Volume rendering: display of surfaces from volume data. *IEEE Computer Graphics & Applications* **8(3)** pp29-37

Paulson DJ and Robinson MB (2004) *Spatial Aspect of Crime: Theory and Practice* Pearson Education.

Biography

Tomoki Nakaya is Associate Professor of Geography at Ritsumeikan University with a special interest in spatial statistics and geovisualisation, particularly used in health & crime geography. Keiji Yano is Professor of Geography at Ritsumeikan University. He is the leader of the Virtual Kyoto Project that has been set up to construct three-dimensional urban landscapes of Kyoto City by using GIS and VR technologies.

‘What’s in your backyard?’ A usability study by persona

Rachel Alsop

Department of Civil, Environmental and Geomatic Engineering, University College London

KEYWORDS: Environmental Information (EI), Usability Engineering (UE), Web Mapping, Personas, Human-Computer Interaction (HCI)

1. Introduction

As society becomes more concerned with the environment and its protection, it is increasingly important to ensure that the public have access to environmental information (EI). Because so much of this information is provided in geographic form (Haklay 2002), the importance of on-line environmental web mapping applications cannot be understated. The Environment Agency’s (EA) ‘What’s in your backyard (WIYBY)?’ website is an example for such a system. This study investigates how different personas (potential users identified by the EA) access EI and interact with the WIYBY site (Human-Computer Interaction). The primary objective was to undertake a usability assessment of WIYBY to investigate how it is meeting the needs of the EA’s personas and identify any improvements which might be made to the site.

2. Background

The WIYBY website provides a good example for Public Environmental Information Systems (PEIS) as Haklay (2002) noted. His research highlighted a wide variety of problems with PEIS, both with WIYBY and other applications that were evaluated. Many of these problems were usability problems that prevent the public from successfully using the site. Since the study, the website has been revamped two or three times, so it is worth revisiting it.

WIYBY is a web-based, mapping facility, whose interface provides the user with access to a GIS database of multi-layered, environmental information. Figure 1 highlights the capabilities of the interface using the WIYBY flood map. The interface enables users to:

1. Zoom in and out
2. Pan, using the tool or by clicking on the map and dragging
3. Double click on the map and information on the flood risk in that area will be displayed

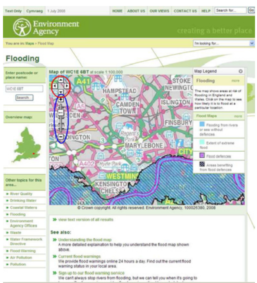


Figure 1. Accessing the WIYBY capabilities (Environment Agency)

3. Method

The EA has carried out an internal requirement analysis, which led to the creation of five personas that need to be used in the design of the system. Of these, three personas were used in the study and are shown in Figure 2. Ajay (At Risk) is concerned that he is at risk of flooding and wants to know more about the quality of his local environment, Roy (Recreation) is concerned with the EA’s involvement with river management and Rebecca (Researcher) is concerned with the EA's data (Gerry McGovern 2007).

Figure 2. The Environment Agency's Personas (Gerry McGovern 2007)

Five people from each persona were recruited to take part in the study following Neilsen's (1994) suggestion that this simplifies 'user testing while gaining almost the same benefits as … from more elaborate tests'. In his later work Nielsen (2003) indicates that representative users are required to undertake representative tasks and be observed. Medyckvj-Scott (1991) has emphasised that it is essential when striving to understand users, that we understand their behaviour and thinking whilst using a system.

The test procedure involved a pre-study questionnaire establishing participants' demographic information and how they matched their persona. This was followed by three typical tasks to be carried out on the internet, tailored to the personas but following the same pattern:

Task 1. Find environmental information on the web, starting from UCL homepage
Task 2. Find a map of the information
Task 3. Find information on the Environment Agency's website

This was followed by a post-study questionnaire and interview/pluralistic walkthrough designed to establish the user's views on usability. Screen recording software was used throughout to record users' responses.

The usability of WIYBY was assessed by measuring effectiveness, efficiency and satisfaction. Effectiveness is measured by whether users are able to complete tasks. For the participants to succeed in the tasks they must have accessed some form of EI, found a map of the information and finally reached the WIYBY site in a set time. The results of the tasks were later translated into usability metrics to establish a persona success rate. Efficiency was measured by the amount of time it took users to complete a task. Satisfaction was measured by a count of positive and negative points made whilst participants carried out the tasks, and through the analysis of the post-study interview and questionnaires. Additionally the different sites visited by users were categorised and coded to discover patterns and develop a general internet usage model by persona.

4. Results and Discussion

The study showed that participants matched the persona profiles in their demographic characteristics, qualifications and internet use. Different personas spent varying amounts of time on the different categories of web pages, for example, Researchers spent more time filtering data from the Google search engine. Figure 3 shows the percentage of time spent by Researchers on different categories of web pages and the average percentage of time spent on different web pages by all personas.

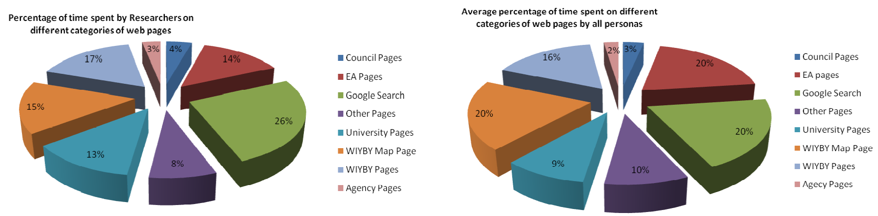


Figure 3. Percentage of time spent by Researchers on different web pages and the average of all personas

When establishing users' effectiveness the study showed that the WIYBY site had a user success rate of 49.83%; this is an average across all personas (Table 1 shows individual success rates) but when this is compared to Skarlatidou (2005) showing that basic use of Google Maps scored 83%, we can conclude that the general experience of users on WIYBY was one of failure.

	Overall success rate	**Success rate (Task 3)**
Researcher	69.4%	54.2%
Recreation	43.3%	40.0%
At Risk	56.7%	55.0%

Table 1. Success rate by persona

Efficiency was also a problem for the At Risk persona taking an average of 3.8 times longer on tasks than the quickest participant. Satisfaction with a system is shown in Table 2 which is based on a count of positive and negative comments. Most negative comments were made about the map page and included 'not great', 'not very clear', 'rubbish' and 'old fashioned'.

Persona	**Positive Points**	**Negative Points**
Researcher	4	16
Recreation	1	7
At Risk	0	12

Table 2. Count of positive and negative points

The three different personas studied and their expectations and assumptions of the WIYBY site revealed how it could be improved in the future. Recreation highlighted the problems in finding the site and knowing the meaning of 'What's in your backyard?' It would therefore be advisable to change the name of the site to something that is more intuitive to the general public. Links to the WIYBY site should also be made more obvious from the EA's homepage. Researchers identified the confusion that arose between the zoom and learn more capabilities. This needs to be demystified and a distinction made of how to use these functions, so that at a first glance users understand what will happen when they click the mouse on the map. Finally, the At Risk category highlighted problems with the scale of the maps. When the highest zoom level is reached a feedback should indicate this and the largest scale should give more spatial detail.

Back in 2002 an assessment of the site concluded that 'in its current form, it holds the potential to alienate users through the maps or text rather than improve awareness' (Haklay 2002:208), and the problems highlighted in this study lead us to believe that this may still be the case. It is hoped that WIYBY will consider these weaknesses and the ways in which to improve the site, in order to ensure that users have a more effective, efficient and satisfying experience. The study has also shown how beneficial it is to use personas in usability engineering and HCI; the use of personas is critical to this understanding and essential for successful and effective web mapping sites.

6. Acknowledgements
I would like to thank Muki Haklay for not only inspiring this study but for his supervision throughout.

Stephen White, for his initial input to the study and the Environment Agency for allowing the study to take place.

My sincere gratitude to all participants in the study and to the businesses which helped me to recruit them.

Minerva Bath Rowing Club,
Bath Cricket Club,
Bath Sports & Leisure Centre,
And The White Hart Inn, Widcombe

References

Gerry McGovern (2007) *Get to know our customers: Make our website work for the people who matter,* prepared for the Environment Agency

Haklay, M., (2002) 'Public Environmental Information Systems: Challenges and Perspectives', Department of Geography, University College London

Medyckyj-Scott, D., (1991) 'GIS and The Concept of Usability' in Mark, D.M., and Frank, A.U., (eds) *User Interfaces from Geographic Information Systems: Report on the Specialist Meeting,* pp.105-110

Nielsen, J., (1994) *Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier,* Retrieved 05/06/08 from http://www.useit.com/papers/guerrilla_hci.html

Nielsen, J., (2003) *Usability 101: Introduction to Usability*, Retrieved 05/06/08 from http://www.useit.com/alertbox/20030825.html

Skarlatidou, A., (2005) "The current state of the online mapping industry: A usability study of Multimap's public service and that of its competitors", Department of Geomatic Engineering, University College London

Biography

Rachel Alsop studied Geography and Development Studies at Sussex University before moving onto an MSc in GIScience at UCL in 2007. Recently completing the MSc she was awarded the RICS prize for best dissertation. Rachel has a keen interest in Environmental GIS.

Appropriation of Public Park Space
A GIS-based case study

Frank O. Ostermann[1]

[1]Department of Geography, University of Zurich, Winterthurerstr. 190, 8057 Zurich, Switzerland
Tel. +41 44 635 5154 Fax +41 44 635 6848
Frank.ostermann@geo.uzh.ch, www.geo.uzh.ch/nfp54

KEYWORDS: public park, space appropriation, case study, quantitative spatio-temporal analysis

1. Introduction

This paper will present selected elements of a recently concluded research project. The key aims were to improve knowledge about individual and aggregated human appropriation of public park space, and subsequently identify key factors to improve design and management of parks. The focus is on the more applied components of the project, i.e. the extensive data collections in a case study, the analysis methods employed, and results that are transferable to other cities and parks.

2. Social Sustainability and Urban public parks – opportunities and challenges

Surveys have shown that citizens consider urban parks to be an important element for their well-being, even if used only occasionally (Tinsley and Croskeys 2002; GrünStadtZürich 2006). They are also places where urban citizens can learn important values such as coexistence, cooperation, and tolerance by experiencing and living cultural diversity (Garcia-Ramon, Ortiz et al. 2004).

However, patterns of design and management (Forsyth and Musacchio 2005; Low, Taplin et al. 2005) and informal processes of displacement and exclusion oppose general access and equal participation (Manning and Valliere 2001; Chiesura 2004), thereby reducing diversity and endangering social sustainability (Owens 1985; Paravicini 2002; Thompson 2002; Brandenburg, Arnberger et al. 2006). To ensure socially sustainable park use, a prerequisite are non-discriminatory access and equal chances for participation (Bundesamt für Statistik, BFS et al. 2003). This research postulates that this is expressed in a heterogeneous, diverse usage and composition of visitors.

This paper looks for processes of exclusion on two scales: The meso-scale of a neighbourhood, where the composition of sampled park visitors is compared to an expected composition based on the neighbourhood population; and on the micro-scale, where the spatio-temporal distribution of park visitors within a single park is subject to quantitative analysis.

In order to identify strategies of design and management that foster socially sustainable appropriation of public parks, we need more knowledge about the actual usage and appropriation. Park usage studies have been mostly in the form of off-site surveys, neglecting direct observations to find out more about how parks are actually used (GrünStadtZürich 2005; Fischer, Stamm et al. 2006).

3. Zurich Case Study

This study employs a pragmatist, mixed methods approach, using both qualitative and quantitative methods sequentially and iteratively where appropriate (Creswell 2003; Morgan 2007). Part of the project was the development of a simple yet efficient quantitative spatial model to capture and represent the complex interpersonal processes of human space use and appropriation on the micro level (Ostermann and Timpf 2007).

An important aspect of this work is its empirical foundation of the modelling and analysis. In order to collect data, three public parks were observed over the span of three years. The case study was undertaken in close collaboration with the administrative department responsible for the design and maintenance of public parks, GrünStadtZürich. The parks to be observed were selected for their function in the city context as neighbourhood parks, and their suitability for observations. This included characteristics such as size and visibility. The observations were realized over a period of three years, including a pilot study. Each of the three parks was observed on 7-14 days for 2-4 hours. As two parks were observed on consecutive years, this amounts to almost 150 hours of observations. In total, over 8000 park visitors were recorded. A new, digital observation method was developed, which allowed the direct encoding of the observational data using TabletPCs and standard GIS software, reducing observer bias. The age and gender of the park visitors and the type, time and location of their activities were recorded into a database. The data is considered representative at a larger scale weeks and seasons. The uncertainty introduced by the observations was acknowledged, and the quality of the data judged sufficiently exact for analysis.

4. Spatio-temporal analysis methods

Subsequently, several established quantitative analysis methods were reviewed for their suitability for the micro-scale. The analysis of the original discrete point data is possible with established spatial analysis methods: Mean centres, SDEs, nearest neighbour index and kernel density estimates are straightforward and provided meaningful results on several scales. The temporal analysis had to remain a primarily qualitative visual one. The complex nature of human spatial usage, appropriation, and interaction makes a data mining approach to detect hidden causes and effects very challenging. A pattern could be returned because of user-introduced bias instead because of an actual case of domination and exclusion. One would have to augment the data with the motivation of the park users, so that one could determine why a certain reaction like relocation has occurred. In addition to the detailed analysis within parks, the composition of the visitor sample was also compared to the neighbourhood population by employing Chi-Square-Tests.

Finally, the multitude of visualization techniques employed (qualitative dot maps, space-time-cubes, animations, density surfaces, small multiples) were evaluated for their suitability in a knowledge generation and knowledge dissemination context.

5. Results

Within the scope of this contribution, only some results can be highlighted. In summary, while some of the activity patterns detected resemble expected ones, others contradict patterns observed and reported in the literature by other research projects. For example, the analysis results on the micro-scale of park usage showed no indication of a suggested domination of open spaces by male visitors (Paravicini 2002). To the contrary, the dynamic activities were located at the periphery of the open spaces in all parks, and there was no male domination in the centre, as the following exemplary figure shows:

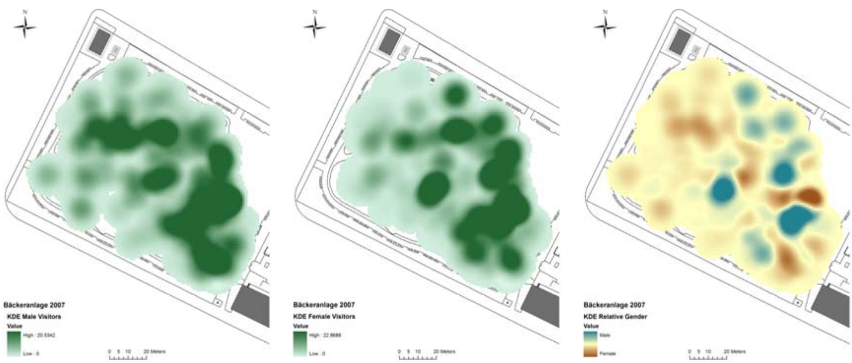


Figure 1: Absolute Density of Males (left), Females (Center) and relative Density (right)

At the meso-scale of neighbourhoods, the analysis showed a statistically highly significant under-representation of elderly visitors. The following table shows this exemplary for one observed park (normalized values for observed sample and expected population):

	Male	Female	p Gender	Children	Adults	Seniors	p Age
Observed	51.4	48.6	0.62	17.9	76.1	5.95	< 0.00
Expected	53.9	46.1		16.1	61.9	22	

Table 1: Savera-Areal 2007 Chi-Square Test

While differences in gender are not statistically significant at an aggregated temporal scale, there is a temporally correlated underrepresentation of females during the evenings (Ostermann 2009). Also, some elements of the park infrastructure proved to be strong attractors for certain visitor groups. Clearly, each user group seems to have certain preferences with regard to the park infrastructure. Therefore, a diverse infrastructure gives the heterogeneous user groups the possibility to participate.

7. Acknowledgements

This research (www.geo.unizh.ch/nfp54) is conducted as part of a National Research Program of the Swiss National Fund (no.54: Sustainable Development of the Built Environment; www.nfp54.ch), as well as in cooperation with and financially supported by Green City Zurich (Grün Stadt Zürich, department responsible for planning and maintaining public parks; http://www.stadt-zuerich.ch/internet/gsz/home.html).

References

Brandenburg, C., A. Arnberger, et al. (2006). Prognose von Nutzungsmustern einzelner Besuchergruppen in urbanen Erholungsgebieten. CORP 2006 & Geomultimedia06.

Bundesamt für Statistik, BFS, et al., Eds. (2003). Monitoring der Nachhaltigen Entwicklung MONET. Schlussbericht Methoden und Resultate. Nachhaltige Entwicklung und regionale Disparitäten. Neuchâtel, Office fédéral de la statistique.

Chiesura, A. (2004). "The Role of Urban Parks for the Sustainable City." Landscape and Urban Planning **68**: 129-138.

Creswell, J. W. (2003). Research Design: qualitative, quantitative, and mixed methods approaches, Sage Publications, Thousand Oaks.

Fischer, A., H. Stamm, et al. (2006). Die Nutzung von Pärken, Grünanlagen und Naherholungsgebieten in Zürich. Zürich.

Forsyth, A. and L. Musacchio (2005). Designing Small Parks: A Manual Addressing Social and Ecological Concerns, Wiley, Hoboken.

Garcia-Ramon, M. D., A. Ortiz, et al. (2004). "Urban planing, gender and the use of public space in a peripherical neighbourhood of Barcelona." Cities **21**(3): 215-223.

GrünStadtZürich (2005). Wirkungsbilanz Parkanlagen. Zürich, GrünStadtZürich.

GrünStadtZürich (2006). Das Grünbuch der Stadt Zürich. Zürich, GrünStadtZürich.

Low, S., D. Taplin, et al. (2005). Rethinking Urban Parks: Public Space and Cultural Diversity, University of Texas Press.

Manning, R. E. and W. A. Valliere (2001). "Coping in outdoor recreation: Causes and consequences of crowding and conflict among community residents." Journal of Leisure Research **33**(4): 410-426.

Morgan, D. L. (2007). "Paradigms Lost and Pragmatism Regained." Journal of Mixed Methods Research **1**(1): 48-76.

Ostermann, F. (2009). Modelling, Analyzing and Visualizing Human Space Appropriation -

A Case Study on Three Urban Public Parks in Zurich, Switzerland. Department of Geography. Zürich, University of Zürich.

Ostermann, F. and S. Timpf (2007). Modelling Space Appropriation in Public Parks. AGILE 2007, Aalborg, Danmark, AGILE.

Owens, P. L. (1985). "Conflict as a social interaction process in environment and behaviour research: the example of leisure and recreation research." Journal of Environmental Psychology **5**(3): 243-259.

Paravicini, U. (2002). Neukonzeption städtischer öffentlicher Räume im europäischen Vergleich, Books on demand; Hannover.

Thompson, C. W. (2002). "Urban open space in the 21st century." Landscape and Urban Planning **60**(2): 59-72.

Tinsley, H. E. A. and C. E. Croskeys (2002). "Park Usage, Social Milieu, and Psychosocial Benefits of Parks Use Reported by Older Urban Park Users from Four Ethnic Groups." Leisure Sciences **24**: 199-218.

Biography

The author has studied Geography at the Universities of Hamburg (Germany) and Geneva (Switzerland), graduating with a Diploma in 2004. Currently, he is completing his PhD at the University in Zürich. His main focuses of interest are urban geography, human mobility and accessibility, and modelling/analysis with GIS.

Analysing perceptions of inequalities in rural areas of England

Steve Cinderby[1], Annemarieke de Bruin[2], Meg Huby[2], Piran White[3]

[1]Stockholm Environment Institute, University of York, Heslington, UK
[2]Department of Social Policy and Social Work, University of York, Heslington, UK
[3]Environment Department, University of York, Heslington, UK

Tel. (+44 1904 432994) Fax (+44 1904 432897)
Email (steve.cinderby@sei.se), web (www.sei.se/relu)

KEYWORDS: Participatory GIS, Vignettes, Inequality, Inequity, Rural England

1. Introduction

This paper describes new research from a Research Council UK (RCUK) Rural Economy and Land Use (RELU) Programme funded project, Social and Environmental Inequalities in Rural Areas (SEIRA), which investigates both data and methods to identify inequalities in rural England.

The paper describes the methodologies used in the third stage of the project to examine how rural residents in England experience and perceive conditions in the 21st century countryside. The project adapts participatory mapping techniques to capture these local understandings spatially by combining them with vignettes that are used in social research. The stakeholder information is then analysed in a GIS using spatial references of participants with the aim to establish whether or how place plays a role in perceptions of unfairness or injustice. The paper describes the novel participatory method and highlights some of the preliminary findings of the GIS analysis of perceptions of inequalities in rural England.

2. The quantitative analysis of inequalities

The first two stages of the SEIRA project identify inequalities in rural England (as defined by Bibby and Shepherd (2004)) using a combination of social and environmental information derived from various national datasets. The rural-urban definition is based on population density linked to settlement morphology. At the Lower Super Output Area (LSOA) English census geography level this enables us to identify rural areas and classifies settlements into a two level hierarchy based on ‘sparsity’ and settlement size. The SEIRA project has compiled forty individual datasets at the LSOA level. Statistical techniques have then been applied to extract four underlying factors representing social, environmental and economic conditions. Gini coefficients have been used to quantify inequalities in these factor scores within counties and to identify those exhibiting the greatest variation.

Having completed this quantitative assessment of inequalities the next question is whether local residents experience and perceive such inequalities. Do they perceive an unfair or unjust distribution of social, economic or environmental goods and services? Four counties, distributed across England, are identified as having high levels of inequality: Northumberland, South Yorkshire, Buckinghamshire and Devon. Within these counties specific locations are identified where the LSOA factor data indicate relatively poor social and economic conditions but high variation in terms of the physical environment. In both areas in each county a cross section of the local population was recruited to attend discussion groups about what it is like to live in rural England in the 21st century.

3. Perceptions of inequality

3.1 Recruiting participants: Avoiding the usual suspects!

In order to encourage participation from a wide cross-section of residents and to avoid recruiting only the so called 'usual-suspects' (who are active in their local areas and typically come forward to represent their communities' viewpoint) the project team used a variety of approaches. One effective recruitment technique used involved on-street – or in this case – on-market rapid consultation using maps to attract interest. This captured spatial snapshots of residents' viewpoints on the positive and negative aspects of their local surroundings but more importantly provided an opportunity for the research team to invite passers-by to more in-depth meetings at later dates. The team also accessed participants through a variety of existing groups including the Women's Institute, local sports teams and clubs. On occasion individual interviews were conducted with participants who could not attend group meetings.

3.2 Vignettes linked to participatory mapping

The vignette technique is well established in social research (Bryman, 2008). Vignettes are designed to elicit participants' reactions to narrative scenarios based upon hypothetical (or anonymised real) situations. Vignettes have been used in a variety of contexts and are commonly employed when the topics for discussion are potentially complex, sensitive or embarrassing. The vignettes used in the SEIRA project were constructed to gain comments on the options available to different rural residents housing needs. One involved a family moving from an urban area that had childcare and transport requirements. The other was constructed around the story of a young woman on a low income hoping to move out from the rural family home but staying in a rural location. These vignettes were employed in order to allow people to discuss conditions in their locality in relation to the needs of hypothetical characters rather than having to describe their own personal situations.

Using these vignettes in our project allowed us to encourage participants to discuss conditions of income, employment, transport and housing without first person reference to their own situations. While participants describe conditions in relation to the needs of vignette characters, their viewpoints and framing of the problems indicated their specific experiences and local knowledge. The vignettes were designed to encourage participants to consider conditions in rural areas from a variety of perspectives that may be different from their own.

The novelty of the application of vignettes in this research was there combination with participatory mapping. Participants identify places that might be suitable locations to meet the requirements of the hypothetical characters. They mark locations on a large scale map of the local region after which they explain why they see conditions in the places marked as either suitable or not for the people in the vignettes. The ensuing discussions allow the capturing of considerable detail on opinions and views about conditions in rural England. This spatial component makes it possible to compare participants' perceptions with the variables from the first part of the project.

4. GIS Analysis of Vignette Participatory Mapping

The quantitative analysis has resulted in four indicators capturing rural conditions. 'Disadvantage' incorporates income deprivation together with disadvantage in education and employment, poor mental well-being, fuel poverty and problems related to access to housing, such as affordability. 'Inaccessibility' is an indicator of areas further away from schools and leisure activities and where farming is often subsidised. 'Eco-desirability' indicates areas that have a high diversity in vegetation and wildlife and where house prices and business activity tend to be high. These four factors have then been mapped for the four fieldwork counties.

The recordings of the community discussions have been transcribed and analysed using a qualitative software package MaxQDA. The transcripts were prepared for analysis through coding using a conceptual framework derived from the four quantitative factors. The place names and spatial comments in the transcripts have been digitised using Google Earth and the spatial interface of MaxQDA.

In the coded transcripts themes such as transport, availability of facilities and health provision have been analysed with corresponding comments about the fairness of their spatial distributions. A distinction was made between those issues related to specific local circumstances as compared to themes related to living in rural England that resonated across the four counties. For this paper we will focus on the generic themes.

Remarks that are indicative of people's perceptions of a specific condition of rural living have been linked with their position on the maps of the four factors. This overlay has then been assessed for areas where the underlying data depicts an unequal distribution that is perceived by residents as unfair. These are then the themes and locations England where the potential for environmental and social injustice suggests a need for rural policy intervention.

5. Acknowledgements

The research forming the basis for this report was funded by the UK Research Councils' Rural Economy and Land Use (RELU) Programme (Project: RES 224-25-0062 and RES 229- 25-0004). RELU is funded jointly by the Economic and Social Research Council, the Biotechnology and Biological Sciences Research Council and the Natural Environment Research Council, with additional funding from the Scottish Executive Environment and Rural Affairs Department and the Department for Environment, Food and Rural Affairs.

References

Bibby, P. & Shepherd, J. (2004). *Developing a New Classification of Urban and Rural Areas for Policy Purposes - the Methodology*. Office for National Statistics.

Bryman, A. (2008). *Social Research Methods*. Third edition, Oxford University Press.

Biography

Steve Cinderby is Deputy Director of the Stockholm Environment Institutes York Centre. Steve was part of the team that developed the SECRA dataset. He is currently collaborating on an expansion of the dataset and utilising it for the identification of inequalities. These will then be ground-truthed with rural communities using participatory GIS techniques.

Annemarieke de Bruin is Research Assistant in the project team of SEIRA at the University of York. With her background in GIS and Tropical Land Use she is responsible for the projects data gathering and spatial analysis. She is involved in the development and execution of the participatory GIS activities.

Meg Huby is a Senior Research Fellow at the University of York. She worked with Steve Cinderby in developing the SECRA dataset and is the principal investigator on the SEIRA project. Her research experience reflects longstanding interests in the linkages between social and environmental problems and policies.

Piran White is a Reader in the Environment Department at the University of York, and is a co-investigator on the SEIRA project. His research interests span wildlife ecology, biodiversity and ecosystem function, ecosystem health and social and environmental inequalities.

Visualising Ecosystem Service Values in maps, films and dance

Kate Moore

Department of Geography, University of Leicester, University Road, Leicester LE1 7RH
Tel. 0116 252 3823 Fax 0116 252 3854
mek@le.ac.uk

KEYWORDS: Ecosystem services, Participatory GIS, Participatory video, Kenya

1. Introduction

Communities living subsistence lifestyles in developing countries are profoundly aware of the values that they attribute to the elements of the environment in which they live. The abundance and quality of food, fuel, and water; spiritual connections to landscapes, the annual patterns of rains and drought, the medicinal properties of native plants and trees are essential to human well-being. But how does this spatial and ecological knowledge of local people relate to scientific understanding of the environment? How can local values be communicated to national and international policy makers? Through a study with a Tugen community in the Rift Valley of Kenya using participatory mapping and video, the values that local people attribute to the environment are recorded and visualised alongside maps and knowledge of ecosystem services from local conservationists. A number of visualisation tools are employed and evaluated in visualising and communicating this knowledge.

2. Ecosystem services

People depend on a healthy environment to support life and livelihoods. Valuing and maintaining ecosystem services is therefore essential for human well-being now and in the future. Yet the rapid growth in human populations and their increasing exploitation and influence on the world's resources has led to escalating change in the health, productivity and function of ecosystems worldwide. Economic and social evaluation of ecosystem services based on the functions that ecosystems provide is being heralded as a way forward to incorporate the true economic value of these services into policy making and planning (Costanza et al 1997; Daily 1997; Daily 2000; MEA 2005). This concept goes further than solely monitoring biological diversity by valuing the functions of an ecosystem as a whole as well as the component parts.

An international working group, the Millennium Ecosystem Assessment (MEA) catalogued the benefits provided by ecosystems for human well-being (MEA 2005). Ecosystem services are broadly categorised into four main groups: provisioning services (food, water, wood, etc); regulating services (those that affect climate, water quality, health, etc); cultural services (recreational, aesthetic, and spiritual); and the underlying supporting services (soil formation, nutrient cycling, photosynthesis, etc) (MEA, 2005). (Figure 1) This approach offers a means to link livelihoods with biodiversity conservation in an approach that gives value to the previously hidden contributions of ecosystems to human life.

Ecosystem services have been mapped in various studies mainly at global, national or regional scales (WRI 2007; Troy and Wilson 2008; Egoh et al 2008). In Kenya, previous work has mapped ecosystem services in connection with human well-being at the national scale (WRI 2007). However all these studies still reduce and generalise the issues and the sense of community values on the ground is to a great extent masked. Ecosystem services and performance will differ with scale (Martin-Lopez et al 2009). For instance biochemical and genetic resources will serve human health and well-being for synthesis of medicines and indirect disease regulation at national scales but locally the medicinal herbs are a directly used product. Therefore mapping of the knowledge of the environment and the values attributed to ecosystem services at the local scale is as important as generalised mappings of wider ecosystems. Traditional knowledge and values of environmental resources and biodiversity is often couched in terms of usage and functionality (Roba and Oba 2009). Mapping with stakeholders at a local level will aid policy makers to understand the complex and often conflicting priorities and values attached to ecosystem services.

Provisioning Services **Products obtained from ecosystems**	**Regulating Services** **Benefits obtained from regulation of ecosystem processes**	**Cultural Services** **Nonmaterial benefits obtained from ecosystems**
• Food • Fresh water • Fuel • Fiber • Genetic resources • Biochemicals	• Air quality regulation • Climate regulation • Water regulation • Erosion regulation • Water purification • Disease regulation • Pest regulation • Pollination • Natural hazard regulation	• Spiritual and religious • Aesthetic • Recreation and ecotourism
Supporting Services **Services necessary for the production of all other ecosystem services** • Soil formation • Nutrient cycling • Primary production		

Figure 1. Categorisation of ecosystem services (adapted from MEA 2005)

3. The Study area

This study concentrates on one community living within the catchment associated with Lake Bogoria National Reserve, Kenya. Sandai sub-location lies approximately 10 kilometres north of Lake Bogoria in the Central Rift Valley. The Tugen people living there are mainly agropastoralist who communally graze cattle and goats and grow subsistence crops on small farms. They rely heavily on the different functions of the environment, particularly water supply and water quality, and production of food for humans and livestock. Around 60% of the population lives below the dollar a day poverty line. There is some diversification of income in small enterprises connected to ecotourism, honey production and cash crops such as maize and water melons.

The MEA framework (2001) designated ten categories of ecosystem of which the Sandai sub-location contains three: Inland Water, Drylands and Cultivated Land. Rivers and wetlands in the form of swamps are particularly important ecosystem components for direct products and cultural services. However, there is severe pressure on the swamps due to encroachment by agriculture, cutting of reeds, diversion of water for irrigation and over-grazing by increasing livestock numbers. The area is semi-arid, with regular periods of drought. The landscape is a mix of acacia shrub and scrub with patches of cultivated land. Many areas show signs of degradation and soil erosion.

The Sandai sub-location is included in the Management Plan for the Lake Bogoria catchment (LBNR 2007) which not only covers the designated reserve but the community areas within the wider catchment. A WWF initiative, the Lake Bogoria Integrated Catchment Programme, (WWF 2008) also seeks to integrate the management within the Lake Bogoria greater catchment area. Understanding the use and human impacts on the environment is therefore a concern for the local residents, the local authority who manage the park and NGOs. Mapping ecosystem services from a variety of perspectives is broadening the knowledge exchange between all stakeholders.

4. Mapping ecosystem services – survey methodology

A comprehensive study has been undertaken of the cultural values and economic importance of different ecosystems in the Sandai sub-location. Focus groups taken were formed from different sections of the community: chiefs, old men, farmers, women, youths and local conservation officers. A mixed methodology using a questionnaire, participatory mapping (Minang and McCall 2006) and participatory video (Lunch and Lunch 2006) for qualitative analysis and visualisation was used with the groups to elicit knowledge and values attributed to the ecosystems used by the community. Walked transects using GPS were used to gather data from different ecosystems. For each sample site ecosystem services were identified and ranked to ascertain the relative value of services to human well-being for the past, present and predicted future. Satellite imagery from Landsat ETM was used as base maps to estimate the aerial extent of each component of the ecosystem. Knowledge of the influence of greater watershed ecosystems was also noted.

Participatory mapping has developed rapidly within development practice and also as an effective research tool. Participatory video adds a complementary component to understanding spatial knowledge. "Participatory Video (PV) is a set of technologies to involve a group or community in shaping and creating their own film" (Lunch and Lunch 2006: 10). Like participatory mapping it initiates a process of analysis of local knowledge and practices but it can provide an aspatial understanding of space and place. The process can be an effective tool for communities to express and communicate their own perspectives on issues and also to engage people in active research.

5. Visualizing ecosystem service values - discussion

In order to facilitate communication and understanding between different stakeholders a number of visualisation strategies were employed.

Firstly the participatory video of the survey transects was constructed consisting of interviews between participants and film of the environmental factors that the groups considered important. This formed a linear account of the mapping process and embedded a deep understanding of the community perceptions of place. Cultural values and representations of place such as in dance were also recorded.

Secondly video clips, photos and annotations are embedded within Google Earth (Google Earth 2008) for display of information and communication to a wide audience. This method offers a form of outreach whereby the community has a global visibility.

Thirdly, an interactive mapscape (MScape 2008) is being constructed to be used on a geo-enabled mobile device. Instead of tagging the transect walk to its true location in Kenya, the transect can be transformed to be used at any place on the earth. The user can then walk within his/her own world but reflect upon the ecosystems and places portrayed from another in images, video and audio. The sights, sounds and challenges faced in an African village can be transposed to a walk in an English park. It is envisaged that inter-cultural understanding may be raised by the sensory experience of walking through the mapscape.

Evaluation of these different visualizations will be undertaken through discussion with conservationists and participants in the UK and Kenya.

This research has attempted to take an interdisciplinary approach to combine knowledge and values of ecosystem services. It is a work in progress so final analysis and discussion of the results are still being formed.

The process of mapping and visualising ecosystem services has been of value to the Sandai community for internal discussion as well as external communication of the environmental challenges they face. However, it is noted that "the mere act of quantifying the value of ecosystem services cannot by itself change the incentives affecting their use or misuse" (MEA 2001). The challenge is to apply the greater understanding of the functions of ecosystems into sustainable practices.

6. References

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon. B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P. and van den Belt, M. (1997) The value of the world's ecosystem services and natural capital. *Nature* **387** 253-260

Daily, G.C. (Ed.), 1997. *Nature's Services: Societal Dependence on Natural Ecosystems*. Island Press, Washington, DC.

Daily, G.C. (2000) Management objectives for the protection of ecosystem services. *Environmental Science & Policy* **3** 333–339

Egoh, B., Reyers, B., Rouget, M., Richardson D. M., Le Maitre D.C., van Jaarsveld, A.S. (2008) Mapping ecosystem services for planning and management. *Agriculture, Ecosystems and Environment.* **127** pp135–140

Google Earth (2008) http://earth.google.com/ Accessed December 2008

Lake Bogoria National Reserve (2007) *Lake Bogoria National Reserve Integrated Management Plan 2007 – 2012*. Kenya: WWF Eastern Africa Regional Programme Office and Lake Bogoria National Reserve.

Lunch, N. and Lunch, C. (2006). Insights *into Participatory Video: A Handbook for the field*. http://www.ids.ac.uk/ids/particip/dbdocs/PVhandbook.pdf Accessed January 2008

Martin-Lopez, B., Gomez-Baggethun, E., Lomas, P.L. and Montes, C. (2009) Effects of spatial and temporal scales on cultural services valuations. *Journal of Environmental Management* **90** 1050-1059

Millennium Ecosystem Assessment (2001) *Ecosystems and Human Well-being: A Framework for*

Assessment. Island Press, Washington, DC.

Millennium Ecosystem Assessment (2005) *Ecosystems and Human Well-being: Synthesis*. Island Press, Washington, DC.

Minang, Peter A.; and M. K. McCall (2006) PGIS and local spatial knowledge utility enhancement for community carbon planning. *Participatory Learning and Action* **54**, 85-91.

MScape (2008) http://www.mscapers.com/ Accessed December 2008

Roba, H.G. and Oba, G. (2009) Community participatory landscape classification and biodiversity assessment and monitoring of grazing lands in northern Kenya. *Journal of Environmental Management* **90** 673-682

Troy, A. and Wilson, M.A. (2008) Mapping ecosystem services: Practical challenges and opportunities in linking GIS and value transfer. *Ecological Economics* **60** 435-449

World Resources Institute; Department of Resource Surveys and Remote Sensing, Ministry of Environment and Natural Resources, Kenya; Central Bureau of Statistics, Ministry of Planning and National Development, Kenya; and International Livestock Research Institute. (2007) *Nature's Benefits in Kenya, An Atlas of Ecosystems and Human Well-Being.* Washington, DC and Nairobi: World Resources Institute.

WWF (2008) *Lake Bogoria Integrated Catchment Programme* http://www.panda.org/about_wwf/where_we_work/project/projects/index.cfm?uProjectID=KE0066&source=ge Accessed December 2008

Acknowledgements

This research is supported by an ESRC/NERC funded PhD studentship. I would like to thank the Sandai community and staff at Lake Bogoria National Reserve for their help and participation in this project. Particular thanks to Sue Page and Clare Madge, my supervisors, and William Kimosop, Chief Warden LBNR.

Biography

Kate Moore is currently a postgraduate researcher at the University of Leicester investigating community conservation initiatives in Kenya using participatory GIS. Previously she was GIS Officer at Leicester University, lecturing in cartography and GIS and was a researcher on the HEFCE funded Virtual Field Course Project.

Circuit theory in naturalistic landscapes: how does resistance distance compare to cost-distance as a measure of landscape connectivity?

Thomas R. Etherington[1], E. Penelope Holland[2]

[1]Central Science Laboratory, Sand Hutton, York, YO41 1LZ, UK
Tel. +44(0)1904462230 Fax +44(0)1904462111
t.etherington@csl.gov.uk
[2]Landcare Research, PO Box 40, Lincoln 7640, New Zealand

KEYWORDS: circuit theory, connectivity, cost-distance, least-cost modelling, resistance distance

1. Introduction

Landscape connectivity is a species-specific measure of how a landscape facilitates or impedes movement between two locations on that landscape (Tischendorf and Fahrig 2000). Being able to measure landscape connectivity is a useful tool for a range of applications in wildlife management from designing protected areas for conservation, to understanding wildlife disease spread.

A connectivity measure that has become popular is the cost-distance value derived from least-cost modelling within a geographic information system (GIS) (Adriaensen *et al.* 2003). The method is based upon a friction surface, which is a raster GIS map in which each cell describes the permeability of different parts of the landscape being studied to the species of interest. This friction surface can be used to derive the most efficient route, balancing distance and friction, between locations. The path of maximum efficiency has a cost-distance value that is a combination of the distance that would be travelled and the cost of the landscape friction traversed.

However there is a fundamental issue with cost-distance as a measure of connectivity, as it does not take into account either alternative routes or the width of routes. With this issue in mind, a new approach has been developed based on circuit theory methods developed in electrical engineering (McRae *et al.* 2008). Circuit theory can be adapted to provide a measure of connectivity by treating a friction surface, like that used in least-cost modelling, as an electrical circuit in which areas of higher friction are analogous to higher electrical resistance. By doing this a connectivity measure called the resistance distance can be calculated between two points.

What is of significance is that unlike cost-distance, resistance distance provides a much more intuitive measure of landscape connectivity, as when the landscape provides either more or wider connections, resistance distance decreases (Figure 1).

The significant advantages of resistance distance over cost-distance have been proved within networks and landscapes with a large number of impermeable barriers (McRae *et al.* 2008). What is unknown is how resistance distance may compare to cost-distance within more naturalistic landscapes, in which bounds between areas of higher and lower landscape permeability is not so clear-cut. To address this question we compared the cost-distance and resistance distance values between random points on virtual landscapes of varying spatial autocorrelation.

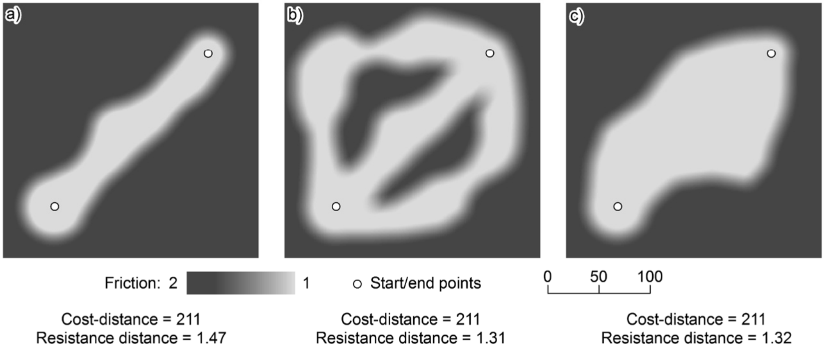


Figure 1. In these three simple landscapes, connectivity would be expected to be worst in landscape a) and better in landscapes b) and c) because of more and wider areas of low friction landscape respectively. The fundamental flaw of cost-distance is highlighted, as it measures connectivity to be the same in all three landscapes. Conversely the resistance distance connectivity measure recognises the variation in landscape structure, with smaller values for landscapes b) and c) indicating greater connectivity as would be expected.

2. Methods

The midpoint displacement method (Fournier *et al.* 1982) was used to create three landscapes, each with dimensions of 100 × 100 cells, with spatial autocorrelation ranging from low to high (h = 0, 0.5 and 1 respectively). Four friction values were used (1, 10, 100 and 1000) to cover a range of permeability levels from no barrier (1) through to virtual barrier (1000). The values of the cells in each landscape were replaced in rank order with friction values to describe the varying permeability of the landscape, resulting in three landscapes with equal friction values and differing only in spatial distribution. Ten iterations of these random landscapes were generated for each level of spatial autocorrelation.

Ten cells were chosen at random within each landscape, and resistance distance and cost-distance were calculated between all pair-wise combinations of these cells. Cost-distances were calculated using the Spatial Analyst extension of ArcGIS (version 9.3), while resistance distances were calculated using the open source Circuitscape software (version 3.0) (http://www.circuitscape.org). Resistance distances were calculated using a Moore neighbourhood to match the diagonal connections applied in least-cost modelling. Linear regressions were fitted to compare resistance distance and cost-distance at each level of spatial autocorrelation. Landscape was included as an interaction term to account for variations between random iterations of the landscape. To assess how cost-distance is related to resistance distance, coefficients of determination (r^2) were calculated. The r^2 values indicate the fraction of the total variation in cost-distance that can be explained by resistance distance. High r^2 values would indicate that there is little to be gained from using resistance distance over cost-distance.

3. Results

To illustrate the process used, we show one of the landscapes created for each level of spatial autocorrelation with the random points overlaid (Figure 2). For these examples, the relationships between pair-wise resistance distance, and pair-wise cost-distance for each landscape have also been plotted (Figure 3) which show a clear positive relationship between cost-distance and resistance distance. This trend continued across all 10 iterations, as regardless of landscape spatial autocorrelation, nearly all of the variation in cost-distance could be explained by a linear regression with resistance distance (h=0 r^2=0.88, h=0.5 r^2=0.94, h=1 r^2=0.90).

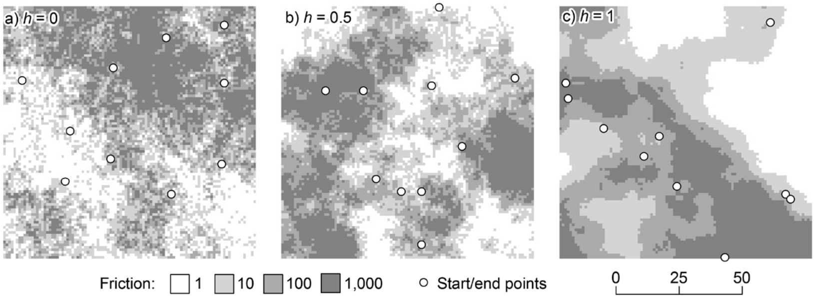


Figure 2. An example of the random landscapes generated that were used as friction surfaces, showing how the three levels of spatial autocorrelation used affected the distribution of friction values. The random points that were generated for each landscape for use in calculating pair-wise cost-distance and resistance distance connectivity values are also shown.

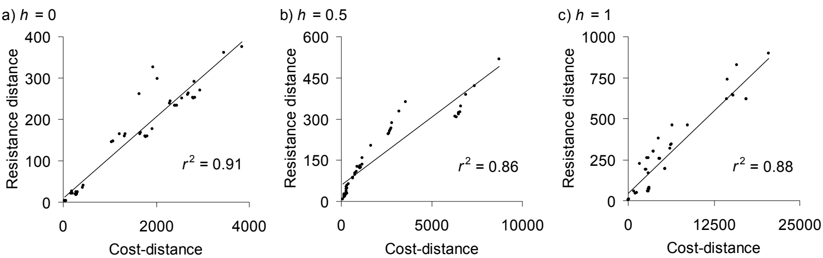


Figure 3. The linear correlations between the pair-wise cost-distance and resistance distance connectivity values for the landscapes in Figure 2 show a strong positive relationship between cost-distance and resistance distance regardless of landscape heterogeneity.

4. Discussion

Previous analyses that have shown a dramatic difference between cost-distance and resistance distance have focused on networks, or on landscapes that have a network character due to impermeable landscape elements forming bottlenecks to connectivity (McRae *et al.* 2008). This brief analysis has demonstrated that the difference between cost-distance and resistance may not be so clear-cut in landscapes with a more naturalistic form.

It would appear that resistance distance does not necessarily provide a dramatic improvement over cost-distance in naturalistic landscapes. This is encouraging for previous works that have used cost-distance as a connectivity measure in relatively open landscapes, as their results would remain comparable regardless of the technique used.

However, given that resistance distance will be no worse than cost-distance, but has the potential to perform better than cost-distance within landscapes more akin to networks, its use where logistically possible is probably preferable when trying to describe landscape connectivity between locations.

Although this paper has focused on applications in landscape ecology, the methods made available by circuit theory could have much wider use in other areas of research that currently make use of least-cost modelling. Given that circuit theory is based on the same inputs as least-cost modelling, and software to calculate resistance distance is freely available, transition, or at least experimentation, should be relatively easy and encouraged where the size and number of connections in a system is of importance.

References

Adriaensen F, Chardon JP, De Blust G, Swinnen E, Villalba S, Gunlinck H and Matthysen E (2003) The application of 'least-cost' modelling as a functional landscape model *Landscape and Urban Planning* **64(4)** pp233-247

Fournier A, Fussel D and Carpenter L (1982) Computer rendering of stochastic models *Communications of the ACM* **25(6)** pp371-384

McRae BH, Dickson BG, Keitt TH and Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation *Ecology* **89(10)** pp2712-2724

Tischendorf L and Fahrig L (2000) On the usage and measurement of landscape connectivity *Oikos* **90(1)** pp7-19

Biography

Tom Etherington is the GI Scientist at the Central Science Laboratory government research agency. With a background in landscape ecology and GIS, which included post-graduate research in Canada, he now conducts spatial analyses revolving around modelling wildlife distributions and movements for use largely in wildlife disease management in the UK.

Pen Holland is a postdoctoral researcher at Landcare Research, New Zealand, primarily working on spatial population modelling techniques to assist invasive mammal management. Her particular interests lie in the representation of space and landscape in population models, and parameterization under uncertainty.

Constructing a standards-based geographical infrastructure for European history

Humphrey Southall[1], Paula Aucott

[1]University of Portsmouth, Department of Geography,
Buckingham Building, Lion Terrace, Portsmouth PO1 3HE, UK
Tel. +44 (23) 9284 2500 Fax +44 (23) 9284 2512
Humphrey.Southall@port.ac.uk, www.gbhgis.org

KEYWORDS: geo-ontologies, gazetteers, boundaries, historical GIS, Europe.

1. Introduction

The Great Britain Historical GIS is the largest historical GIS in Europe. Its construction has gone through three main phases. The first, presented at GISRUK 1997, was a relatively conventional system limited to historical administrative boundaries and statistics from census reports and similar sources, implemented using ArcGIS (Gregory and Southall, 1998). The second, presented at GISRUK 2003, was funded mainly by the UK National Lottery and developed a new architecture, implemented mainly in Oracle, in which both statistics and boundaries became attributes of a gazetteer of historical administrative units. Content was broadened to include both geo-referenced scans of historical maps and much descriptive text about places, and all this was made publicly accessible via the web site *A Vision of Britain through Time.*.

This paper describes further developments since 2004 which have extended the system so that it is no longer limited to Great Britain, and formalised the gazetteer of administrative units as a geo-spatial ontology. The new software infrastructure is entirely open-source: Postgres, PostGIS, Mapserver and OpenLayers, supporting a revised web site to be launched in 2009.
. Despite the addition of historical maps and text, the system is still primarily concerned with defining the changing system of counties, districts and parishes, and assembling and presenting statistical data about these units. At the heart of the system is a single "data table" which currently holds 11,746,887 data values – numbers – all in one column, with other columns identifying the date (**when)**, the geographical area covered (**where**), the statistical report the number was taken from (**source**) and the meaning of the number (**what**):

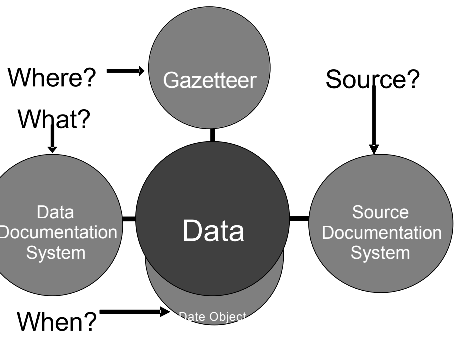


Figure 1. GBH GIS Overview

The values held in these other columns are mainly identifiers linking to a series of major sub-systems.

The Source Documentation System is an inventory of 5,131 table that have appeared in British census reports since 1801, and as the system also stores the column and row within the source table that a number came from it is able to reconstruct selected tables. The Data Documentation System is a relational implementation of the Data Documentation Initiative, and is further described in Southall (1998). The main focus here is the "gazetteer" of administrative units, and why it is not a conventional GIS.

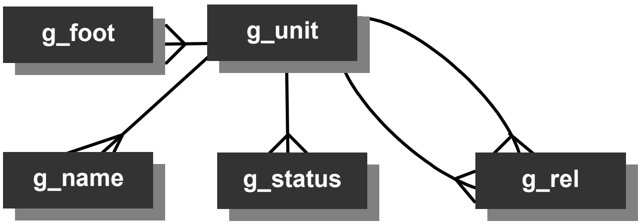
2. Organising geographical knowledge

Traditional vector-based GIS distinguishes spatial data – points, lines and polygons – and attribute data, the former providing a framework for the latter. Where the attribute data are primarily social statistics, the spatial framework usually consists primarily of polygons defining the boundaries of the administrative areas used as reporting units. There are a number of problems with this approach in historical contexts. Firstly, and especially when working with data from the early 19th century and before, locations are the most uncertain part of the available information: the first British census was in 1801, but the Ordnance Survey completed the mapping of administrative boundaries only in the 1880s. In cultural heritage research, it may even be desirable to include "places" whose existence is questionable, such as Camelot and Avalon. Secondly, even when reliable boundary mapping is available, constructing digital boundary data may be prohibitively expensive: while most academic GIS researchers obtain their data ready-digitised, mainly from government agencies, historical GIS researchers must often construct their systems from paper sources. Even with a total budget to date of £2.2m., the GB Historical GIS holds statistical data for several major reporting geographies which have not yet been mapped.

Building parish-level historical systems is expensive and is generally possible only if the results will serve several audiences. One key audience is in the archives sector, but their central requirement is for a *Geographical Name Authority*. For example, the lottery funding for the GBHGIS required the creation of an on-line replacement for the reference books identified by the National Council on Archives (1997), while EU funding involved a collaboration with the national archives of Estonia and Sweden. However, a third problem with conventional GIS systems is that they provide limited facilities for working with geographical names. In historical contexts, a given unit may have a whole series of different names in different sources. For example, one particular parish in Cardiganshire appears in one census or another with all the following names: 'Llanfihangel y Creuddyn Uchaf', 'Eglwysnewydd', 'Eglwys Newydd', 'Llanfihangel y Croyddin Upper', 'Upper Llanfihangel y Creuddyn', 'Upper Llanfihangel y Creuyddyn', 'Upper Llanfihangel y Croyddin', 'Upper Llanfihangel y Croyddyn'. Many parishes in Estonia have names in three different languages, Estonian, Russian and German, and researchers need a systematic record of all the different names, as a new source may use any of them.

All these factors argued for the construction of a different kind of system which drew on research by information scientists into Knowledge Organisation Systems (KOS). The simplest KOS is a *word list*; and that could be a list of place-names. Even this is of some value as a controlled vocabulary in a cataloguing system, but once the set of terms is of any size it needs some notion of hierarchy, and to include both preferred and alternative terms. Such a structure is a *thesaurus*. One example is the Unesco Thesaurus (http://databases.unesco.org/thesaurus), while the *Thesaurus of Geographical Names* (http://www.getty.edu/research/conducting_research/vocabularies/tgn) is a gazetteer developed by the Getty Information Institute organised as a thesaurus. A *polyhierarchic thesaurus* allows a more complex structure. An *ontology* goes further in two ways: it is an enumeration of concepts, not terms, and it can include a variety of relationship types.

The new architecture developed for the GBH GIS is a true ontology because it enumerates, currently, 77,464 administrative units; it then holds 123,395 names for them; and links them via 240,122 relationships of nine different types, the most common being "IsPartOf". The database model is quite simple, as shown below, although many small additional metadata tables define typologies, etc. The "status" table holds more detailed information on the legal status of units, for example distinguishing between modern districts and Unitary Authorities. The "footprints" table is what makes this a geo-spatial ontology, as it holds 75,986 polygons or multi-polygons defining administrative boundaries. However these are held for only 37,774 of the units. The tables are linked together by ID numbers, not names, and these IDs also appear in the data table described above. A broadly comparable system has been built by semantic web researchers for Finland (Kauppinen et al, 2008), but it is somewhat smaller, covering 616 "places".



3. Supporting standards

While GIS technology has developed mainly around *ad hoc* standards defined by software companies, collaboration with the cultural heritage sector and funding through digital library programmes required a much greater adherence to open, published standards. The Dublin Core Metadata Initiative (http://dublincore.org) has defined a *lingua franca* for merging catalogue information from museums, libraries and archives but their "coverage" element, designed to capture geography, usually records overall scope but not granularity. The Encoded Archival Context (EAC) standard has been developed specifically to support archival name authorities, but despite this specifically including place-names the standard makes no provision for including locational data.

To avoid being forced to support these inappropriate standards known to the funders, the project had to identify alternative more appropriate standards. This obviously included the family of standards developed by the Open Geospatial Consortium (http://www.opengeospatial.org), and the historical mapping within the Vision of Britain web site is accessed via an OGC Web Map Server which can also be used by other web sites. However, the OGC standards are too narrowly focused on traditional geographical information and three other standards are also supported to cover geographical names, statistics and place-name-rich text.

The Alexandria Digital Library, based at the University of California Santa Barbara, has developed a suite of standards for digital gazetteers (http://www.alexandria.ucsb.edu/gazetteer). Their Gazetteer Content Standard defines a gazetteer including many variant names, polygonal footprints defined in GML and much date information. However, their Feature Type Thesaurus defines only four kinds of administrative unit, and their requirement that every entity have at least a point coordinate is replaced in the GB system by a requirement for at least one "IsPartOf" relationship.

The Data Documentation Initiative provides a framework for recording the meaning of statistics, linking frequency counts back to the original microdata. There is however a problem with integrating geography into this framework, which is currently being investigated by a joint project with the Census Dissemination Unit (MIMAS).

The system contains c. 10m. words in two kinds of texts. Material from 19th century descriptive gazetteers, including the first edition of Bartholomew's *Gazetteer of the British Isles*, divides easily into discrete entries for named places which are held as rows in a database table. However, the GBHGIS also includes the largest collection anywhere of historical British travel writing and these have been marked-up to identify place-names following the Text Encoding Initiative Guidelines (http://www.tei-c.org). An on-the-fly parser then presents the text as HTML, converting place-names into hyperlinks and also creating a map showing the places mentioned in each text extract.

References

Gregory I and Southall HR (1998) Putting the Past in Its Place: the Great Britain Historical GIS. Carver S (ed) *Innovations in GIS 5.* Taylor and Francis, London.

Kauppinen T, Väätäinen J, Hyvönen E. (2008) Creating and using geospatial ontology time series in a semantic cultural heritage portal. Bechhofer S et al (eds) *Proceedings of the 5th European Semantic Web Conference 2008.* ESWC 2008, LNCS 5021, Tenerife, Spain', pp. 110-123: http://data.semanticweb.org/conference/eswc/2008/paper/105

National Council on Archives (1997). *Rules for the Construction of Personal, Place and Corporate names*. http://www.ncaonline.org.uk/materials/namingrules.pdf

Southall HR (2008). Visualization, data sharing and metadata, pp. 259-75 in Dodge M, McDerby M and Turner M (eds.) *Geographical Visualization: Concepts, Tools and Applications*. Wiley, Chichester.

Biography

Humphrey Southall is a Reader in Geography at the University of Portsmouth, and Director of the Great Britain Historical GIS. His original research was on Britain's evolving labour markets but he is now more involved in trying to make good (geographical) content on the Internet easier to find than junk.

Paula Aucott is a Senior Research Associate at Portsmouth University and Team Leader on the Great Britain Historical GIS. She has an MA in English Local History from Leicester University and an MSc in Geographical Information Systems from Portsmouth University. She is now an expert on Estonia's administrative geographies.

Factors Impacting Fear in an Urban Environment

Ben Calnan[1], Claire Ellul[1]

[1]Cities Institute, London Metropolitan University,
Ladbroke House, 62-66 Highbury Grove, London, N5 2AD
Tel. +44 (0)20 7133 5105 Fax +44 (0)20 7133 5123
b.calnan@londonmet.ac.uk , http://www.citiesinstitute.org

KEYWORDS: transport safety urban environment fear

1. **Introduction**

Much of the research into the urban environment and crime focuses on reducing the opportunity for crime – for example the Safer Places report (ODPM, 2004) lists attributes including *ownership*, *activity, surveillance* and *management and maintenance* as being relevant for crime prevention. A similar list forms the basis of Crime Prevention Through Environmental Design (CPTED) activities (see, for example, Crime Stoppers 2008).

However, a second, perhaps less well known, area of research - the link between features in an urban environment and the fear felt by people walking through that environment – should also be considered. This is of greater relevance when considering social exclusion in the context of access to public transport for research being carried out as part of the AUNT-SUE project[1]. In fact, thirteen percent of people report feeling *unsafe*, and 21% reporting feeling *a bit unsafe*, when walking in their local area after dark. Sixty-eight percent of those aged 60 and over say they never walked in their local area after dark (Town *et al.* 2003). These figures can be contrasted with the fact that, in 2002/03, the average risk of becoming a victim of any crime was only 4.1%, with the highest risk being for young men aged 16 to 24, at 15.1% (Town *et al.* 2003).

The *Design Against Crime* report (Town *et al.* 2003) lists a number of features of an urban environment that are related directly to fear, including *isolation*, *lack of easy surveillance*, *poor lighting*, *lack of orientation* and *lack of opportunities to avoid threats*. Additional factors include litter, graffiti and poor cleanliness. Nasar and Jones (1997) also include factors relating to physical entrapment (barriers to escape) and concealment (obstructions to the view of the pedestrian, behind which an attacker could potentially hide) again mentioning graffiti, vacant buildings, litter and vandalism.

Last year, we presented a mechanism using a Geographical Information System and these lists of urban features to generate an Index of Permeability (IoP) indicating areas in an urban environment where people would, theoretically, feel more or less safe to venture (Ellul and Calnan, 2008). This involved a survey of the neighbourhood to identify the presence or absence of the relevant features, followed by the identification, through means of a series of questionnaires, of how important each feature type was in the context of generating fear. Isovists (which show the area visible from a particular point, see Figure 1 and Davies *et al.* 2006) were used to identify any relevant urban features that could be seen from each one of a regular grid of points within a neighbourhood and an aggregated weighting generated for each point. The IoP surface was generated from the points using an Inverse Distance Weighting (IDW) algorithm.

Here, we present an update on the progress of this research, describing the initial validation of these surfaces – do the surfaces generated by our automated system actually correspond to surfaces of fear

[1] Accessibility and User needs in Transport for Sustainable Urban Environments (AUNT-SUE, 2007). AUNT-SUE is collaboration between London Metropolitan University, University College London and Loughborough University along with partners in the public and private realm with the aim to develop and pilot policies, methods and tools to support accessible transport planning and inclusive transport. The ultimate aim of this collaboration is to develop a tool kit that will help support decision makers, establish benchmarks and incorporate inclusion into policies.

felt by pedestrians?

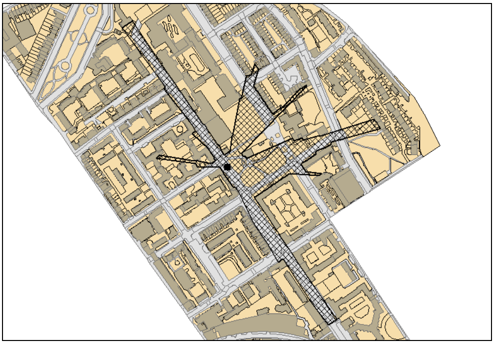


Figure 1 *- A 360 Degree Isovist around a Point*

2. Validating the Index of Permeability

Following IoP generation, focus groups were organised to measure the extent to which the surface model reflects people's fears within an urban area. Four focus groups (a total of 28 people) were held with 'hard to reach' residents in the test area (Somers Town, in London) – i.e. those that, perhaps, are more likely to be socially excluded.

- Single Parents
- Women under 27
- Older People
- Bangladeshi men under 27

These groups were identified through contacts with Camden Council (the location of our test bed area) and the validation process formed part of a wider exercise designed to elicit comments on the public realm and problems in the area. They represented members of the local population that the Council felt were harder-to-reach through conventional methods and are also representative of the target audience for this research, particularly in the context of transport exclusion:

> "*in the UK patterns of car ownership mean that those on low income, elderly, children, certain ethnic minorities groups and to a lesser extent women will tend to experience systematic exclusion from facilities"* (Hine and Grieco 2003, cited in Azmin Fouladi 2007)

Firstly the attendees were asked to fill out a weightings questionnaire rating the contribution of urban features to fear. Secondly, a paper-based participatory mapping exercise was carried out. Individuals were asked to place dots on a map of the area, with a red dot for an area that made them feel unsafe, a yellow dot for OK and a green dot for safe.

The results were interpolated using an Inverse Distance Weighting algorithm, which calculates values for a point on a surface by examining the values of known points around it, assigning more importance to closer points than those further away. This algorithm was selected as the chosen implementation environment (ArcGIS 9.2) allowed the identification of 'boundaries' for the surface, creating an IoP that was confined to road surfaces. This was considered particularly important as the features visible from one point in one street (and the resulting IoP weighting) are not visible around the corner, and thus should not be included in any resulting interpolation for that area. In other words, it cannot be assumed that fear is smooth and unbroken over the entire area.

Both the focus group surfaces (one for each group and one combined surface) and the IoP surface (see Ellul and Calnan 2007) were then reclassified and the focus group surfaces were then subtracted. The map below (**Figure 2**) shows the result when subtracting the combined Focus Group (i.e. results from all 28 people) surface from the IoP.

3. Results and Analysis

There is a match of approximately 47% between the two surfaces, represented by the green areas in **Figure 2**. The areas in red indicate areas where the IoP overestimates the level of fear in an area, and in blue where underestimation occurs.

Figure 2 *- An SDI and Focus Group surface comparison.*

Areas around the edge of the map were more prone to discrepancies, which can be explained by the fact that focus-group data was only collected for the target area whereas the Isovists considered areas outside the target area (as these can still be seen from inside). Discrepancies were also noticed where there were a number of negatively weighted urban features (e.g. windows with closed curtains or blinds), and consequently lower IoP values (i.e. the area is seen as generating more fear). Areas which are busy or well connected appear to have the closest matches, for instance the main road traversing south to north, and the areas in the south.

A number of factors also impacted the results obtained:

- Issues arose with the map-reading (and occasionally English language) skills of focus group members - many did not fully understand the maps or felt they were unclear.
- The dots were placed at the focus group member's discretion, leading to uneven coverage in some areas.
- The meetings were collaborative, individual fears (which are very personal) may have been lost to the group's more dominating members.
- The total number of focus-group respondents was small (although higher than expected, given the hard-to-reach groups that were involved).

However, perhaps the most significant outcome of the work with the focus groups was the fact that, as group members were local residents, their local knowledge dominated their views on the area. Comments captured during the exercise included "gangs hang out around this school so it is not safe", and "I am not comfortable walking past this pub as the drinkers are rowdy". Thus an area identified as neutral by the IoP (in front of a secondary school, to the north of the map) was given a negative score by the focus groups.

From this anecdotal evidence it can be suggested that it is this local knowledge, rather than features in the urban environment such as graffiti, litter or vandalism, that can increase or decrease access to public transport and consequent social inclusion nor exclusion.

4. Current and Future work

To investigate the importance of local knowledge further, a second phase of validation is currently being undertaken with non-residents of Somers Town. To provide a more systematic approach to data capture, subjects (three to date) have been taken around a set route that covered the whole area, marking down at set points how they felt. Initial results (from walk-abouts carried out during the day) suggest that strangers to Somers Town feel that the area was not particularly frightening or intimidating, is predominantly residential and is rather empty of people during the day. The only areas where urban features did have an impact on levels of fear were alleyways. This may be influenced by the relatively small size of the test area and its proximity to major public transport hubs.

Further work is therefore required to validate the IoP surfaces. Firstly, where possible, non-resident walk-abouts should be carried out in Somers Town later in the day, at the end of the school day or after dark. Secondly, the creation and validation of the Index of Permeability should be extended to other, larger and more varied, areas. Currently data has been captured for Finsbury Park and Greenwich. Thirdly, additional work is required to explore the mechanisms used to identify and map 'fear' – could this, for example, be interpreted from observations of avoidance behaviour, through on-street or online surveys, through Participatory GIS or by interpreting comments noted during a walk-about around the area for the focus groups described above?

5. Acknowledgements

We would like to thank Dr Nastaran Azmin-Fouladi, Dr Jo Foord and Professor Graeme Evans, and the Cities Institute for their help in developing this tool and Antje Witting for help collecting data and organising the various focus groups. We would also like to thank the AUNT-SUE consortium for funding and Camden Council and Ordnance Survey for provision of datasets.

6. References

AUNT-SUE (2007), *About the Study*, [online], available from: http://www.aunt-sue.info/about.html, [accessed on 14th December 2007]

Azmin-Fouladi, N. (2007), *Inclusive Journey Environment – Street Auditing, analysis and findings in Somerstown and Elm Village: an Urban Design Approach,* Cities Institute, London Metropolitan University

Crime Stoppers (2008) - *Crime prevention through environmental design (CPTED)* [online], Available from: http://www.crimestoppers-uk.org/crime-prevention/helping-prevent-crime/community/crime-prevention-through-environmental-design-cpted [Accessed 10th December 2008]

Davies C, Mora R and Peebles D (2006) Isovists for Orientation: Can space syntax help us predict directional confusion? *Proceedings of the 'Space Syntax and Spatial Cognition' workshop*, Spatial Cognition 2006, Bremen, Germany, 24 September 2006

Ellul C and Calnan B, Modelling Street Safety in an Urban Environment, *Proceedings of GISRUK 2008,* Manchester, 2-4 April

Nasar J L and Jones K M (1997), Landscape of Fear and Stress (Fear Of Crime), *Environment and Behaviour*, Vol 29, No.3, pp. 291-323

ODPM (2004) *Safer Places - The Planning System and Crime Prevention* – Office of the Deputy Prime Minister, [online], Available from: http://www.securedbydesign.com/pdfs/safer_places.pdf [Accessed 10th December 2009]

Town S, Davey C L and Wootton A (2003) *Design Against Crime – Secure Urban Environments By Design – guidance for the design of residential areas,* University of Salford, Design and Innovation Research Group

8. Biography

Ben Calnan is a researcher at the Cities Institute at London Metropolitan University where he is working on urban perception modelling for the AUNT-SUE consortium (Accessibility and User Needs in Transport for Sustainable Urban Environments). Before this he completed a Masters in GIS at University College London in 2005.

Breaking Down the Silos

Niall Carter and Bruce Gittings
Edinburgh Earth Observatory
School of GeoSciences,
University of Edinburgh
Drummond Street
Edinburgh EH8 9XP
bruce@ed.ac.uk

Keywords: Joined-Up Government, Middleware, Information Silos, SDI.

1. Introduction

With the ever-increasing volumes of digital information and the ability to distribute through the rise of web services there is a growing need for a reassessment in the paradigm of system integration. While numerous information rich web sites are being developed by government, these exist individually in a vacuum, with bridges between them being rare. This paper argues for a single light-weight service capable of connecting these 'Information Silos' and breakdown the mentality of proprietary digital information. It goes on to describe the implementation of such a system.

We demonstrate work implemented at the University of Edinburgh in collaboration with key Scottish information providers to allow a bottom-up approach to accessing information. The 'Connecting Your Geographies' (CYGnus) system allows decentralised information to be delivered, from a single source, directly to users. No changes to underlying infrastructure are required for the information providers and they retain absolute control at all times as to the level, amount and quality of information returned. Delivery adheres to a recognised international standards and so cooperation has already seen seven Scottish information providers conform to the same method of delivering previously proprietary information to the public. By querying a selected location, thus treating Geography as the 'glue' through which information can be related, the user is able to have provided to them a standardised container for the information provided which can be either instantly visualised or readily consumed within a local client.

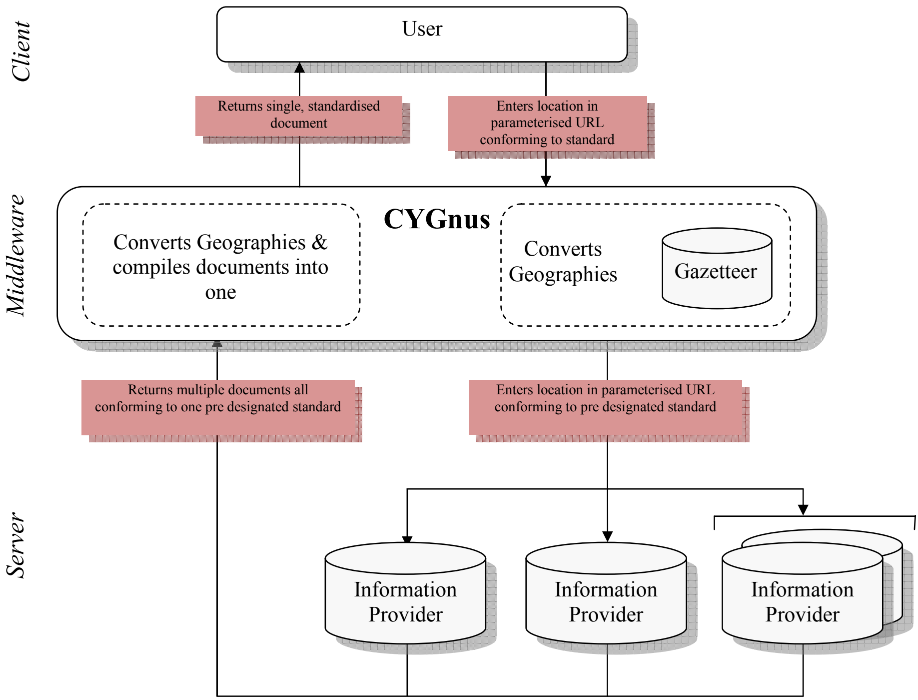
2. Origins of the project

The CYGnus system is a response to a seeming plethora of independent information resources and seeks to provide a means of integrating these at two different levels; namely URL sharing and data delivery. It takes some key cultural and general information resources from collaborating partners including the Royal Commission on the Ancient and Historical Monuments of Scotland, Historic Scotland, South Ayrshire Council, the Gazetteer for Scotland and the Scottish Government. It provides part of a Scottish SDI, itself a component of a geographical information strategy for Scotland (Scottish Executive, 2004). The aim of the project was to provide a *middleware* solution, which would lie between the end-user application and the information provider. The middleware is able to interpret a range of methods to describe geographical location and use it to return an aggregation of records from the information provider based upon that location. This approach has considerable appeal, because it avoids the need for the end-user to install specialist software or for the information provider to expend effort in a customised solution for a particular requirement. Whilst previous systems have been described as *data discovery* systems with only information pertaining to stored datasets being returned, CYGnus can be accurately described as a *data delivery* system (as advocated by Arsanjani et. al., 2003) with the distinction being that the user is be able to download *in a standardised format* the information held within the proprietary information provider's own databases. This was to be a lightweight solution so that different interfaces can be developed and different programs can consume CYGnus.

3. Aims

The service has several specific aims. First is to provide a data delivery service allowing access to detailed textual information held by information providers. This service must be based on established standards and be capable of sitting alongside, and integrating with, services such as OGC WMS and WFS. A further key aim was to develop functionality within a lightweight and scalable framework to remove barriers and encourage uptake. Lastly, the service should provide access to information through a geographical location and be independent of different geographies (such as projections). In the same vein be able to understand and deliver information in a range of geographical representations such as points, lines and polygons.

4. Enabling a range of geographies



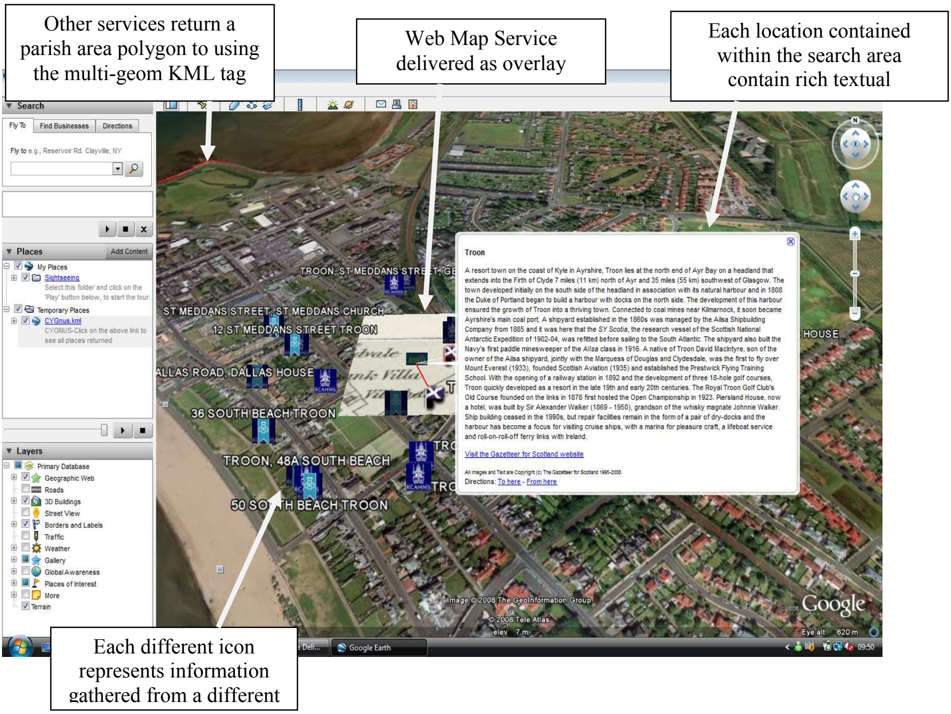
While there is a shared geographical context, the specific geographies used by partner organisations will naturally provide different. CYGnus is a middleware application accessed through a URL API, designed to be compatible with OGC web services. This allows CYGnus to be consumed in a number of different ways which could include a web page interface or from within another application such as ArcGIS Explorer. Geographical location is provided either as a place-name (via a short form gazetteer) or as an X and Y coordinate pair in either latitude/longitude or easting/northing. Support for areal units such as council areas, counties and parishes can potentially be added easily through use of a web service or an extension to the middleware application.

5. Standards

CYGnus uses the location entered as the basis for compiling further URLs for each information provider. Each information provider hosts a very simple interface to interpret these URLs, search their datasets based on the parameters provided and return a Keyhole Mark-up Language (KML) document back to CYGnus. KML was chosen because it is an industry standard which is relatively lightweight, able to support multiple representations of geography (point, line and polygon) but crucially also including text-based (including long-form or descriptive gazetteer) information which is not supported by OGC standards such as WMS and WFS. It also has the distinct advantage of being able to be readily consumed within widely available geo-Browsers such as Google Earth together with several GIS clients such as ESRI ArcGIS Explorer.

The method of retrieving the parameters and searching the dataset is completely controlled by the provider within their own system and so requires very little overhead to implement. The information returned is also completely controlled by the provider, which acts as a further security measure to prevent large volumes (data scraping) or sensitive information from being published.

The information returned by the service also involves KML, but extended using a GML snippit to encode coordinates which cannot be provided in latitude-longitude (eg. data referenced to the British National Grid). Harvesting each returned KML document in turn, CYGnus is able to convert geographical coordinates from such exotic coordinate systems into to WGS84 required for KML standards compliance. CYGnus then compiles all returned documents into a single, structured and internationally recognised standards compliant document which is delivered to the user for consumption.



. Conclusions and future directions

CYGnus provides a mechanism for a broad range of users (from professionals to the general public) to gain access to integrated information that has been previously held and offered solely through information silos. The solution is lightweight and standards compliant. It is has already been consumed in a variety of ways and provides proof of concept that a new paradigm of system integration can be developed which is not reliant on a heavy, costly top down approach.

CYGnus has successfully provided a single service which gives access to multiple information silos. It allows the enforcing of standards across several Scottish information providers by utilising Keyhole Mark-up Language to transport geographical information across the Internet. Further, it is implemented as a lightweight solution which is able to be consumed by a variety of client interfaces and so offer flexibility in its use and branding. Critically, CYGnus illustrates the ability of a grass roots approach to system integration can work and does not need to rely on a large public sector spending, coordinated infrastructure change and heavy-handed top-down managerial approaches (Groot, 1997).

Although initially a point-only service, CYGnus has already grown in terms of its functionality, a tribute to the ease of development which KML affords, to now allow a polygon service (providing polygons depicting the 1951 parish for the location entered) as well as an OGC WMS image overlay drawn from the National Library of Scotland. CYGnus does not host data, instead offering a access through simple spatial queries to a realm of data connected by geography.

Being middleware, this can be exploited by a range of different user interfaces. From a 'branded' web site with embedded mapping to a range of clients from geo-browsers such as GoogleEarth and Whirlwind to GIS clients which can read either KML or the XML feed, to other web services which can consume these formats. Other formats can be added with relative ease.

CYGnus provides a proof of concept which confirms that, with regard to technology, joining disparate information silos is not a difficult feat. It has demonstrated an international standard that can contain both points and polygons as well as offer web services (WMS) directly to clients. Just as CYGnus is able to convert between British National Grid and WGS 1984 projections it could also provide point-in-polygon operations and other lightweight geospatial processing tasks. Other web services such as WFS can be integrated to provide further added value.

CYGnus itself comprises only around 500 lines of code, with provider-services extending to only around 50 lines; a testament to the fact that a lightweight solution is available. Technology is not a factor in the ability to share data; rather it is the attitudes which govern the use of the data which are key. Without overcoming political barriers and thus encourage further uptake, CYGnus will not demonstrate anything other than a technological proof of concept. Already with seven information providers connected and providing data in a recognised international standard method of transport, the CYGnus model has provided an impetus which is worthy of both encouragement and support from across the public sector in Scotland

References

Arsanjani, *A.,* B. Hailpern*, J. M*artin and P. Tarr, “Web Services Promises and Compromises”, *ACM Queue*, 1(1), March 2003

Groot, R. (1997). "Spatial Data Infrastructure (SDI) for Sustainable Land Management", *ITC Journal* 3(4): 287-294.

Scottish Executive (2004) "One Scotland - One Geography: A Geographical Information Strategy for Scotland" GI Service, Scottish Executive.

Biographies

Niall Carter completed his M.Sc. in GIS in September 2008 and now works for ESRI-UK.

Bruce Gittings is a member of the Geographical Information Science group within the Edinburgh Earth Observatory. He teaches on the Edinburgh MSc in GIS and is currently supervising two PhD students. He chairs the National Steering Committee of GISRUK, maintains close links with industry through AGI, was a founder member of AGI-Scotland, and is Vice Chairman of the Royal Scottish Geographical Society.

Modeling Rules for Integrating Heterogeneous Geographic Datasets

Gobe Hobona[1], Mike Jackson[1], Suchith Anand[1], Stefania de Zorzi[2], Didier Leibovici[1]

[1]Centre for Geospatial Science, University of Nottingham, Nottingham, NG7 2RD, UK

email: gobe.hobona at nottingham.ac.uk   http://www.nottingham.ac.uk/cgs

Tel. +44 115 84 66058   Fax. +44 115 951 5249

[2]CORILA, S. Marco 2847, Venice, Italy

KEYWORDS: Integration, Fusion, INSPIRE, Ontology

Introduction

Cross border initiatives such as disaster response and infrastructure development require the integration of geographic datasets from various national mapping agencies. With the recent publishing of the INSPIRE directive, data models have been developed by INSPIRE Thematic Working Groups defining common feature types found in European datasets (EC, 2007-2008). The INSPIRE data models will help various geographic information communities within Europe integrate heterogeneous datasets into a common data model. To facilitate the integration of various heterogeneous geographic datasets, it is necessary to ensure that there is semantic interoperability between source and target data models. Rules have been identified as a possible mechanism for facilitating semantic interoperability (Visser et al, 2002; Klien, 2007). This paper proposes a model for expressing rules supporting the automated integration of geographic datasets.

Establishing Requirements for Merging Rules

One of the earliest studies into geospatial data fusion is by Devogele et al. (1998), where they offered a discussion of essential considerations for data integration. They highlighted that it is not just objects and their attributes that define a schema but also the composition, association and generalisation relationships between the objects. Regarding attribute mapping, they observed that thematic and geometric attributes from different objects may be mapped directly or through a function. They provided a detailed analysis of how aggregation conflicts between instances may be solved. They concluded that even from different viewpoints and scales, an integrated schema should be informed by the original data structures and preserve the semantics of the original schemas while maintaining relationships between objects. We observe that there are three main types of transformation functions that may be applied to feature types and attributes, namely conversion, aggregation and fragmentation. Therefore, our first requirement is that: a language for modelling rules for geographic data integration should be able to abstract functions for conversion, aggregation or fragmentation of feature types or attributes.

Duckham and Worboys (2005) proposed an approach that considers instance-level information in contrast to schema-level mappings. Their approach assumes that there is a function that maps the properties of features at a location to the concepts in a taxonomy. They provided a basis for mapping geometries to thematic attributes through a function. They observed that slight changes in the footprint of features resulted in very large changes in the product taxonomy, hence they proposed establishing thresholds to constrain the algorithm. Consequently, our second requirement is that data integration rules should allow for transformation of both text and geometries.

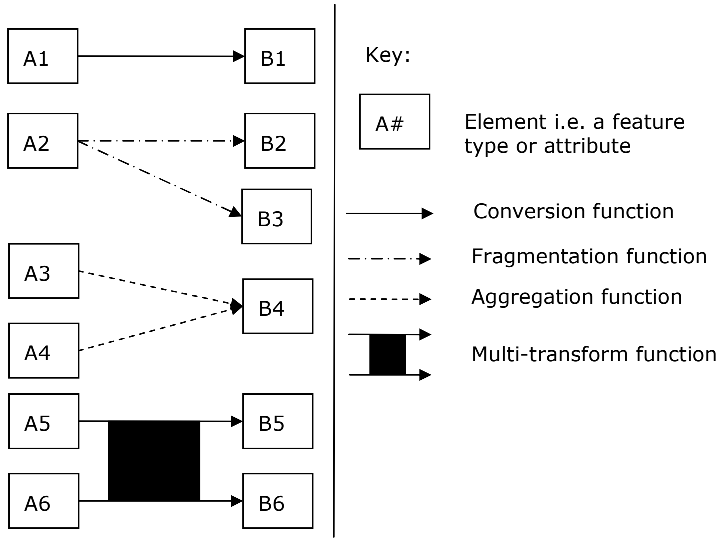


Figure 1 Conceptual View of Integration Functions

The conversion, aggregation and fragmentation functions are illustrated in *Figure 1*. The function mapping A1 to B1 depicts a conversion function. The function mapping A2 to B2 and B3 depicts a fragmentation as A2 is split into two other elements. The function mapping A3 and A4 to B4 depicts an aggregation as two separate elements are merged into a single element. Within a relational context, a conversion function is a one-to-one relation; a fragmentation function is a one-to-many relation; an aggregation function is a many-to-one relation. We define an additional function called a 'multi-transform' that is a collection of at least two transformation functions with at least two different input and output concepts. It is illustrated by the function mapping A5 and A6 to B5 and B6. The transformations within a multi-transform do not have to be the same but have to be applied simultaneously. A multi-transform is therefore a grouping of one or more conversion, aggregation or fragmentation functions. Within a relational context, a multi-transform function is a many-to-many relation. The following section proposes an approach that could be used to specify whether to clone, aggregate or fragment elements based on the semantics of feature types or attributes.

Applying Semantics to the Integration Process

When integrating heterogeneous datasets, the semantics of feature types and attributes are an important factor in deciding whether to aggregate or fragment instances of data (Fonseca et al., 2002, Bishr et al., 1999). Some of the feature types or attributes may be semantically different though lexically similar; for example the word ‘Orange’ denotes both a colour and a fruit. Representation of concepts is presented in an ontology – a shared conceptualisation of an application domain. Semantics may be expressed using an ontology language such as the Web Ontology Language (OWL)(W3C, 1999-2008) which is encoded in the W3C standard Resource Description Framework (RDF). An OWL knowledge base contains definitions of concepts and relationships between them; it can be flattened into a relational table with triples of 'conceptA-relation-conceptB'. This allows for a reasoner to infer assertions based of relationships between concepts. In a related study Pundt and Bishr (2002) suggested the use of RDF for the purpose of documenting semantics used by geographic information communities. Since then other studies have demonstrated the value of using OWL and RDF for this purpose(Raskin and Pan, 2005, Visser et al., 2002).

Proposed Approach

Bishr et al (1999) proposed an architecture that includes a ‘semantic mapper’ that provides a common ontology and interface to heterogeneous data sources. Klien (2007) proposed the use of annotation and the Semantic Web Rule Language (SWRL) for distinguishing between geographic features. Visser et al (2002) proposed the use of ‘semantic translators’ that re-classify a concept term based on semantics described in source and target ontologies. In this section we propose an approach that extends the role of the semantic translator and mapper to schema and geometry transformation during the integration of heterogeneous geographic datasets. Our approach defines a relational model required for integrating heterogeneous geographic datasets: the model includes a Semantic Schema Annotation (SSA) equivalent to a triple *(className, attributeName, owlReference)* and a Semantic Schema Transformation (SST) equivalent to a triple *(owlReference, transformationFunction, owlReference)*. A class model of the integration rules is presented in Figure 2 in Unified Modeling Language (UML).

SSA annotates structural definitions of feature types and attributes with references to OWL concepts; thereby providing the meanings of the feature types and attributes. For annotating feature types(i.e. class names), the attribute name entry in SSA is set to *null*. SST specifies how fields should be modified based on what the meanings of the fields are. Similarly, SST also specifies transformation for complete feature types. SST differs from XML Stylesheet Language Transformations (XSLT) through the selection of elements based on semantic definitions rather than lexical equivalents. We define a transformation of an attribute from one feature type to the other as the function resulting from the relational join of an SSA with the join of an SST with a different SSA. For example: where the symbol ∩ denotes a relational join, given R to be the SSA for an input schema, Q to be the SSA for an output schema and T to be the SST; then the relation $(R \cap T) \cap Q$ would define a complete semantic transformation from one attribute or feature type to another.

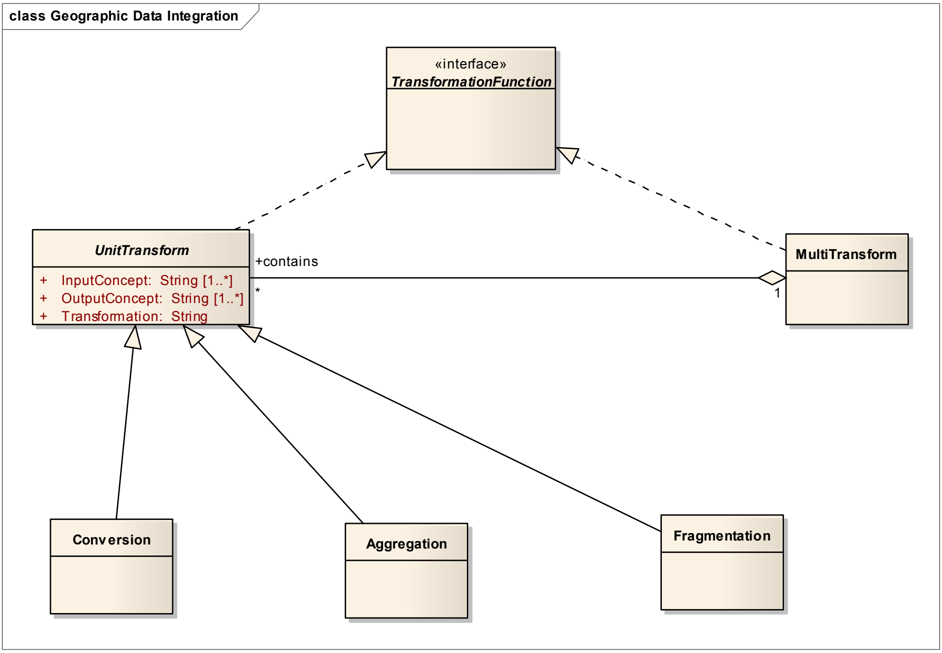


Figure 2 Class model of integration rules presented in UML

Some example rules of SST are presented in Table 1. Rules 1 and 7 are examples of conversion rules as the functions have a single output for each input. Rules 2a and 2b jointly illustrate a fragmentation rule as the single input is modified to produce two separate outputs. Rules 3a and 3b illustrate an aggregation rule applied to geometry, whereas Rule 5a and 5b illustrate an aggregation rule applied to an alphanumeric attribute. It should be noted that Rules 4a and 4b jointly illustrate another fragmentation on geometry, whereas Rules 6a and 6b illustrate a fragmentation on an alphanumeric attribute. Rules 8a, 8b and 8c illustrate a multi-transform as there are multiple types of inputs and outputs.

ID	Input Concept	Function	Output Concept
1	County	copy[Geometry]	AdministrativeUnit
2a	Parcel	withinFloodPlain[Geometry]	HighFloodRiskParcel
2b	Parcel	beyondFloodPlain[Geometry]	LowFloodRiskParcel
3a	Country	clip[Geometry]	Counties
3b	Administrative Boundaries	clip[Geometry]	Counties
4a	Road	buffer[Geometry]	RoadReserve
4b	Road	copy[Geometry]	Road
5a	City	concatenate[String]	Region
5b	Country	concatenate[String]	Region
6a	FullName	split[String]	FirstName
6b	FullName	split[String]	LastName
7	Country	reprojection[Geometry]	Country
8a	City	copy[String]	PlaceName
8b	County	copy[String]	Region
8c	Administrative Boundaries	centroid[Geometry]	Location

Table 1 Example Rules in a Semantic Schema Transformation file

Conclusion and Future Work

This paper has proposed an approach for modelling rules for integrating heterogeneous geographic datasets. The proposed approach adopts a relational model and is thus transferable to RDF, XML and relational databases. Whereas conventional GIS software offers only attribute mapping, our proposed approach also offers a methodology for representing aggregation and fragmentation of attributes. Future work will include the development of a Web Processing Service (WPS) that can apply SST rules to feature collections provided by Web Feature Services (WFS). The conference presentation will demonstrate application of the proposed approach to operational datasets used in the GIS4EU project.

Acknowledgements

This study is supported by the GIS4EU project which is funded by the European Commission through the eContentPlus Programme.

References

Bishr, Y., Pundt, H. and Rüther, C. (1999) 'Proceeding on the Road of Semantic Interoperability - Design of a Semantic Mapper Based on a Case Study from Transportation', *Lecture Notes In Computer Science,* 1580, pp. 203-215.

Devogele, T., Parent, C. and Spaccapietra, S. (1998) 'On Spatial Database Integration', *International Journal of Geographical Information Science,* 12, (4), pp. 335-352.

Duckham, M. and Worboys, M. F. (2005) 'An algebraic approach to automated geospatial information fusion', *International Journal of Geographical Information Science,* 19, (5), pp. 537-557.

EC. (2007-2008) *INSPIRE Directive and Data Models*. [Online]. Available at: http://inspire.jrc.ec.europa.eu (

Fonseca, F. T., Egenhofer, M. J., Agouris, P. and Camara, G. (2002) 'Using Ontologies for Integrated Geographic Information Systems', *Transactions in GIS,* 6, (3), pp. 231-257.

Klien, E. (2007) "A Rule-Based Strategy for the Semantic Annotation of Geodata," Transactions in GIS, vol. 11, pp. 437–452

Pundt, H. and Bishr, Y. (2002) 'Domain ontologies for data sharing-an example from environmental monitoring using field GIS', *Computers & Geosciences,* 28, (1), pp. 95-102.

Raskin, R. G. and Pan, M. J. (2005) 'Knowledge representation in the semantic web for Earth and environmental terminology (SWEET)', *Computer & Geosciences,* 31, (9), pp. 1119-1125.

Visser, U., Stuckenschmidt, H., Schuster, G. and Vögele, T. (2002) 'Ontologies for geographic information processing', *Computer & Geosciences,* 28, (1), pp. 103-117.

W3C (1999-2008) *World Wide Web Consortium Specifications*. Available at: http://www.w3.org (Accessed: 02/02/2006).

Biographies

Gobe Hobona is a postdoctoral research fellow in the Centre for Geospatial Science(CGS) at the University of Nottingham. He is currently a Work Package leader in the GIS4EU project and a consultant for the Open Geospatial Consortium (OGC) on the GIGAS project.

Mike Jackson is a professor of Geospatial Science at the University of Nottingham and Director of the CGS. He is a non-executive director of the Open Geospatial Consortium (OGC).

Suchith Anand is a postdoctoral research fellow in the Centre for Geospatial Science(CGS) at the University of Nottingham. He leads schematization research and Ordnance Survey-related research at CGS.

Stefania de Zorzi is the project coordinator for GIS4EU and the Information Systems Manager at the CORILA Research Institute in Venice, Italy. Since 2002 she is the Scientific Coordinator of Research at CORILA, in areas of data management for modelling and analysis of environmental data, hydrodynamic and sediment transport.

Didier Leibovici is a postdoctoral research fellow in the Centre for Geospatial Science(CGS) at the University of Nottingham. He leads CGS activities on the EU-funded eSoter project.

Historical Analysis of Habitat Associations with Intra-Guild Richness Hotspots for Farmland Birds: Clues for the Successful Deployment of Agri-Environment Schemes.

S.J. Dugdale[1], A.A. Lovett[1], A.R. Watkinson[1] and P.W. Atkinson[2]

[1] School of Environmental Sciences, University of East Anglia, Norwich, Norfolk, UK, NR4 7TJ.
Telephone: (+44) (0) 1603 592542 Fax: (+44) (0) 1603 591327
E-mail: s.dugdale@uea.ac.uk
[2] British Trust for Ornithology, The Nunnery, Thetford, Norfolk, UK, IP24 2PU.

KEYWORDS: farmland birds, guild, hotspot, agri-environment

1. Introduction

Birds are a good indicator of the general health of common and widespread wildlife in the countryside (Anon., 1999a), as they are widespread and tend to be near the top of the food chain (Gregory, Noble & Custance, 2004). Since 1966, an index of farmland bird abundance has shown a decline of 48% although the change since 1996 has been negligible (BTO, 2006). The UK government has adopted a Farmland Bird Index (FBI) as one of its Headline Indicators of Sustainable Development (Anon., 1999b) and made a Public Service Agreement to reverse the decline in farmland birds by 2020 (Grice et al., 2004).

Agri-environment schemes (AES) are a key mechanism for reversing these declines, but despite the huge sums of money invested in them, research into their effectiveness has been critical of the levels of success that they have achieved, e.g. (Kleijn & Sutherland, 2003).

Given that AES operate at a landscape scale and are the main delivery mechanism by which the declines in farmland birds can be reversed (DEFRA, 2008), it is critical that associations between birds and their preferred habitats are accurately defined. A guild is a group of species that use similar resources in a similar way. Bishop & Myers (2005) suggested that clustered areas of high intra-guild richness ("hotspots") should possess good quality and a high quantity of the resources required by species in a guild, as resources in those areas must be plentiful to enable so many different species competing for similar resources to co-exist. However, construction of bird-habitat association models using contemporary data may be confounded (Opdam & Wiens, 2002) since current habitats have already become degraded and are potentially sub-optimal.

2. Research Overview

This study has used data for 62 farmland bird species from two Breeding Bird Atlases for the whole of England and Wales. The first survey conducted 1968-1972 (Sharrock, 1976) sets a baseline at a time before farmland bird populations had begun to seriously decline. This is then compared with data collected 1988-1991 using the same methodology (Gibbons, Reid & Chapman, 1993) when the declines had begun. Survey methods were defined by the British Trust for Ornithology and involved every 10km square in the study area being surveyed between 1st April and 31st July. Analysis has been conducted using a guild based approach which encompass key life-history variables; diet during the breeding season, nest placement and preferred habitat type. Each of these three guild categories was sub-divided into guilds (Table 1). All species were allocated to three guilds, one in each guild category.

Table 1. Guilds within Each of the Three Guild Categories

Diet (Breeding Season)	Nest Location	Preferred Habitat Type
Invertebrates	Building	Mostly Covered
Plant Material	Canopy Layer	Open
Seeds	Cavity or Ledge	Semi-open
Seeds & Invertebrates	Ground	
Vertebrates	Shrub Layer	

Hotspots in the period 1968-1972 were then identified. A hotspot was defined as the 5% of the study area with the highest number of species of the same guild present (intra-guild richness). These hotspots should therefore correspond to areas of high habitat quality for the guild in question. Changes in the spatial relationships of the hotspots during the period of population decline have been assessed by undertaking a similar analysis of species data collected in 1988-1991. Any shift of hotspots over time may give some insights into the impact that the social and economic factors affecting farming have had on farmland birds. Large scale environmental factors such as climate change might be expected to cause the distributions of some species to move northwards and eastwards which would also be identifiable from such an analysis.

ArcGIS 9.2 (ESRI, 2006) was used to analyse the temporal and spatial relationships in the bird species data using raster files at a 10km resolution. Raster files representing the species distributions were overlayed to identify areas of high intra-guild richness in each time period. Where appropriate, a neighbourhood analysis was conducted using a focal mean operation on a 3x3 cell area to refine the analysis and identify the top 5% of grid cells as hotspots. Raster data of habitat and environmental factors were then created and relationships between guilds of birds and the underlying habitat were examined for locations identified as hotspots. By doing this spatially and temporally, it is possible to identify factors which have changed during the period of population decline. This provides valuable information at a landscape scale and could provide clues for the successful deployment of AES in the future.

3. Example of Hotspot Analysis

The hotspots of intra-guild richness in 1968-1972 for the 19 species in the Seeds and Invertebrates guild show a striking similarity to the distribution of chalkland in the UK (Figure 1a). By 1988-1991 however, this relationship appeared to be breaking down, with many of the 1968-1972 hotspots in both the North and South Downs being lost by 1988-1991 (Figure 1b). There are also many losses of hotspots in the more central chalkland regions such as the Chilterns.

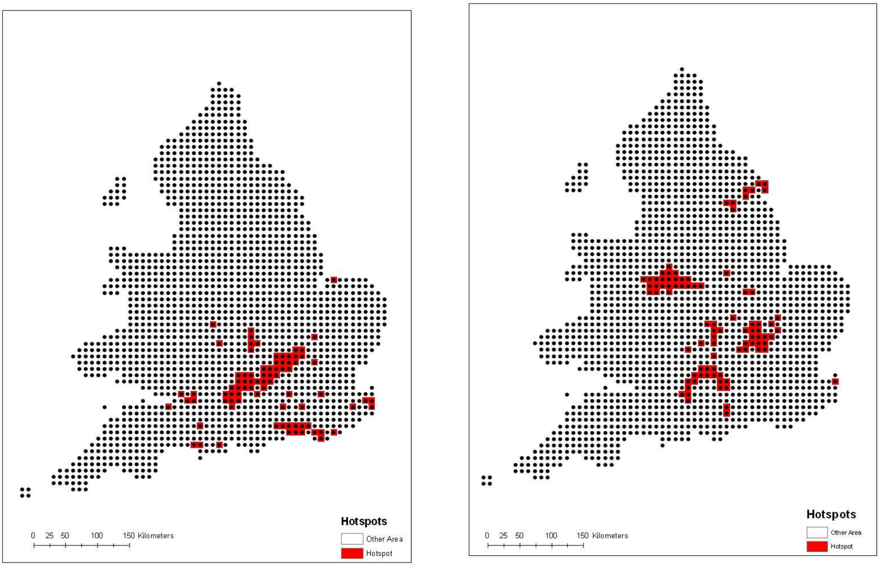


Figure 1: Change in hotspot location between (a) 1968-1972 and (b) 1988-1991 for the 19 species in the Seeds and Invertebrate breeding season diet guild.

4. Relationships between Hotspots and Environmental Variables

Having identified the guild hotspots, an analysis is being conducted of the habitat and environmental variables that were present in these locations at the same time as the bird survey data was collected. This will indicate the factors that constitute high habitat quality for each guild. Agricultural Census data relating to crop types, livestock density and farm classification and size are being analysed for hotspots areas. Where possible, additional factors being examined include soil type, elevation, climate, number of insects present, hedgerows and soil moisture content. Relationships and patterns of change in bird distributions and habitat change that occurred between the dates of the bird surveys are being explored. This will identify habitat associations at a landscape scale.

5. Conclusion

By identifying factors at a landscape scale that equated to high quality habitat for guilds of farmland bird species at a time before populations had declined, this study will provide some clues about what changes may be required to AES prescriptions or their configuration in the landscape. This has the potential to deploy AES more effectively, benefitting farmland bird populations whilst offering better value for money to the taxpayers.

6. Acknowledgements

This paper forms part of a Ph.D funded by the National Environment Research Council at the University of East Anglia, UK. The authors would like to thank the British Trust for Ornithology for providing the data.

References

Anon. (1999a). A better quality of life: a strategy for sustainable development for the UK (Cm 4345). The Stationery Office, London.

Anon. (1999b). Quality of Life Counts. Indicators for a Strategy for Sustainable Development for the United Kingdom: a Baseline Assessment. Department of the Environment Transport and Regions, London.

Bishop, J.A. & Myers, W.L. (2005) Associations between avian functional guild response and regional landscape properties for conservation planning. *Ecological Indicators*, **5**(1), 33-48.

BTO. (2006). The 2006 farmland bird Public Service Agreement (PSA) indicator for England, 1966-2005. http://www.bto.org/research/indicators/farmland_psa.htm. Date Accessed 6th June 2008.

DEFRA. (2008). Environmental Stewardship: Review of Progress.

ESRI. (2006). ArcGIS version 9.2. Environmental Systems Research Institutes, Redlands, California.

Gibbons, D.W., Reid, J.B. & Chapman, R.A. (1993) *The New atlas of breeding birds in Britain and Ireland : 1988-1991* T. & A.D. Poyser, London.

Gregory, R.D., Noble, D.G. & Custance, J. (2004) The state of play of farmland birds: population trends and conservation status of lowland farmland birds in the United Kingdom. *Ibis*, **146**(s2), 1-13.

Grice, P., Evans, A., Osmond, J. & Brand-Hardy, R. (2004) Science into policy: the role of research in the development of a recovery plan for farmland birds in England. *Ibis*, **146**(s2), 239-49.

Kleijn, D. & Sutherland, W.J. (2003) How effective are European agri-environment schemes in conserving and promoting biodiversity? *Journal of Applied Ecology*, **40**(6), 947-69.

Opdam, P. & Wiens, J.A. (2002). Fragmentation, habitat loss and landscape management. In *Conserving Bird Diversity, General Principles and Their Application* (eds K. Norris & D.J. Pain), pp. 202-23. Cambridge University Press Cambridge, UK.

Sharrock, J.T.R. (1976) *The Atlas of Breeding Birds in Britain and Ireland* T. & A.D. Poyser, Calton, England.

Biography

Steve completed an Ecology degree as a mature student in 2006 and is now a 3rd year PhD student at the University of East Anglia. He is particularly interested in the causes of avian population declines and the application of GIS to study this problem.

The Fluvial Information System

Carbonneau, P.E.[1], Dugdale, S.J.[2*], Clough, S.[2]

[1:] Lecturer in Physical Geography, Geography Department, Durham University, Durham, UK, email: patrice.carbonneau@durham.ac.uk

[2:] APEM Ltd. A17 Embankment Business Park, Heaton Mersey, Stockport, UK

* : Presenting Author.

KEYWORDS: Fluvial Remote Sensing, Salmon Habitat, River Science, River Management, Water Framework Directive

River ecologists have long been aware that our understanding of lotic ecology is severely limited by our lack of methods and concepts applicable to catchment scale processes (Fausch et al 2002; Wiens 2002). Furthermore, in terms of environmental management, recent EU legislation on freshwater management, the Water Framework Directive, now explicitly states that surface waters must be managed at catchment scales. This new awareness has created a need for an innovative approach to high resolution, catchment scale data collection and management in fluvial environments. In response to this need, remote sensing has been the focus of increasing interest in river sciences. Following an intensive development phase, it is now possible to map parameters such as water depth, grain size and habitat type with sub-decimetric resolutions over very large areas (Marcus and Fonstad 2008). Therefore this innovative usage of remote sensing is yielding unprecedented amounts of information about river systems. With such levels of information, crucial research questions about catchment scale ecology and river management can now be addressed. However, this data intensive approach produces vast amounts of raster data thus leading to significant issues in terms of spatially explicit data management. For example, Carbonneau et al (2004, 2005) use an image database comprised of 5556 images each having 6 million pixels. Extracting meaningful, spatially explicit, information from such large image databases poses a significant challenge which must be resolved if fluvial remote sensing methods are to deliver their full potential.

GIS has already been very successfully applied to manage remotely sensed data in a fully spatialised context. Unfortunately, when applied to modern fluvial remote sensing raster data, traditional GIS appears limited and unsuited to the specific tasks required by river scientists and managers. Indeed, widely used vector oriented GIS packages such as ArcMap and MapInfo have developed sophisticated modules for raster operations which allow for very high levels of analysis when working on a limited number of raster layers. However, experience with these packages has shown that they rapidly become overwhelmed and awkward when faced with gigabyte datasets distributed into thousands of rasters. Another fundamental issue with traditional GIS packages is the use of established Cartesian map projection systems. Given that rivers are curvilinear entities, the use of Cartesian, orthogonal grid, map projections is mismatched and curvilinear coordinate systems, unique to each river, will be required (Legleiter and Kyriakidis 2006).

This paper introduces the 'Fluvial Information System' (F.I.S.), a raster based GIS-type system designed to manage fluvial remote sensing data and automatically extract meaningful information. The F.I.S. rests on a 2D river coordinate system. Modelled after the orthogonal curvilinear system presented by Legleiter and Kyriakidis 2006, the downstream axis of this system follows the curvilinear river path as modelled by cubic splines whilst the cross-stream direction is locally orthogonal to the main axis. This river coordinate system is generated automatically. First, the high resolution images of the river channel are georeferenced by an automated co-registration process which matches the high resolution image to a previously referenced image having a lower resolution (See Zitova and Fluser 2003 for a review of automated image registration). Then, image classification is used to identify the river channel which allows the channel midpoints to be automatically indentified. The end result is a coordinate system which allows for a unique spatial localization of each high resolution image pixel and, crucially, for an accurate determination of the inter-image spatial relationships. Figure 1 shows an example of an automatically generated river centreline which acts as the basis for a river coordinate system on the river Wyre, Lancashire, UK.

This adaptation of GIS to fluvial systems is a significant innovation with important and significant applications to both fundamental river science and river management. With the F.I.S., managers can make an effective use of the rich information contained in high resolution imagery. Tools implemented in the F.I.S. interface allow the user to interrogate imagery at discrete intervals along the river centreline, yielding valuable information about fluvial processes. Previously published techniques for deriving grain size and depth statistics from 2D imagery (eg. Carbonneau *et al.*, 2005, 2006) are used in tandem with the river coordinate system to generate long-profile plots of substrate composition and bathymetric trends. A similar tool capable of plotting cross sections orthogonal to the river centreline is also useful for visualising downstream trends in bed topography, and has potential for incorporation into river flow models. Through the fusion of image-derived grain size and bathymetry maps with known salmonid habitat preferences (eg. Armstrong *et al.,* 2003), the FIS is also able to generate catchment-wide maps showing salmonid habitat extent and location. Such information can now allow us to quantify the available habitat for important species, and can be used in a river management decision support perspective where the estimation of stock survival is crucially dependent on available habitat area. Additionally, this information can be used in fundamental investigations on large scale patterns of habitat distribution and spatial ecology. Crucially, all this information is referenced in a coordinate system which is fitted to each river. Therefore, the Fluvial Information System offers a unique tool which promises to modernise our understanding of lotic ecology and our ability to manage surface water systems.

Acknowledgements

This project receives financial support from the Knowledge Transfer Partnerships programme (KTP). KTP aims to help businesses to improve their competitiveness and productivity through the better use of knowledge, technology and skills that reside within the UK Knowledge Base. KTP is funded by the Technology Strategy Board along with the other government funding organisations.

References

Armstrong, J.D., Kempa, P.S., Kennedy, G.J.A., Ladle, M., Milner, N.J. Habitat requirements of Atlantic salmon and brown trout in rivers and streams. Fisheries Research, 62, 143-170, 2003

Carbonneau P.E., Lane, S.N., Bergeron, N. Feature based image processing methods applied to bathymetric measurements from airborne remote sensing in fluvial environments. Earth Surface Processes and Landforms, 31, 1413-1423, 2006

Carbonneau P.E., Bergeron N.E. and Lane S.N. Automated grain size measurements from airborne remote sensing with an application to granulometric long profile measurements in fluvial environments. Water Resources Research, 41, W11426, doi:10.1029/2005WR003994, 2005

Carbonneau P.E., Lane S.N. and Bergeron N.E., Catchment-scale mapping of surface grain size in gravel-bed rivers using airborne digital imagery, Water Resources Research, 40, no 7, W07202, doi:10.1029/2003WR002759, 2004

Fausch K.D., Torgesen C.E., Baxter C.V. and Li. H.W. Landscapes to Riverscapes: Bridging the Gap between Research and Conservation of stream fishes. BioScience, 52(6), 483-498, 2002

Legleiter C.J. and Kyriakidis P.C. Forward and Inverse Transformations between Cartesian and Channel-fitted coordinate Systems for Meandering Rivers. Mathematical Geology, 38, No. 8, 927-958, 2006

Marcus W.A. and Fonstad M.A. Optical remote mapping of rivers at sub-meter resolutions and watershed extents. ESP&L, 33, doi:10.1002/esp.1637, 2008

Wiens J.A. Riverine landscapes: taking landscape ecology into the water, Freshwater Biology, 47, 501–515, 2002.

Zitova B. and Fluser J. Image registration methods: a survey. Image and Vision Computing, 21, 977–1000, 2003.

Interpolating land use data to hydrological units: methods and implications for diffuse pollution modelling

Paulette Posen[1], Michael Hutchins[2], Andrew Lovett[1], Helen Davies[2]

[1] School of Environmental Sciences, University of East Anglia, Norwich, Norfolk NR4 7TJ, UK
Tel. (+ 44 1603 592547) Fax (+ 44 1603 591327)
p.posen@uea.ac.uk; a.lovett@uea.ac.uk
[2] Centre for Ecology and Hydrology Wallingford, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK
Tel. (+ 44 1491 692478) Fax (+ 44 1491 692528)
mihu@ceh.ac.uk; hnd@ceh.ac.uk

KEYWORDS: agricultural census, diffuse pollution modelling, land use, nitrates, Water Framework Directive

1. Introduction

The EU Water Framework Directive (WFD) (CEC, 2000) calls for a major restructuring of European water management, with the aim of achieving “good ecological status” in all water bodies by 2015. One of the aims of the Catchment Hydrology, Resources, Economics and Management (ChREAM) study (Bateman et al, 2006), at the University of East Anglia, is to assess likely impacts of WFD implementation on agricultural land use, and consequent implications for water quality and farm incomes. An element of this analysis involves combining land use data with hydrological models to determine how agricultural inputs translate into concentrations of agrichemical compounds in water bodies.

Combining land use data with hydrological spatial units can involve a number of problems arising from the integration of a variety of data formats at a range of spatial and temporal resolutions, and the aggregation of source data over different spatial extents (Moxey et al, 1995; Moxey and Allanson, 1994; Geddes et al, 2003; Defra, 2006; Aalders and Aitkenhead, 2006; Huby et al., 2007). This paper sets out to identify the range of spatial resolutions at which reliable estimations of agricultural land use can be made, and scrutinises the ability to confidently predict the possible outcomes of future, policy-driven, land use change.

2. Background

The location chosen for detailed analysis was the River Derwent catchment in North Yorkshire, which covers an area of 1600 km^2, comprising 282 hydrological response units (HRUs) corresponding to areas of land over which surface water drains to discrete river stretches. The catchment encompasses a wide range of topography and land use types, ranging from grazed uplands to lowland arable and small urban areas (Figure 1) and, as urban land use occupies less than 8% of the catchment, most of the observed nitrate in rivers can be attributed to diffuse sources.

The study assessed four different methods of interpolating agricultural census data to HRUs and compared the results for each method, in terms of areas of agricultural land, at four different spatial scales ranging from a small hydrological unit to the entire Derwent catchment. As there is no true land use information against which to assess the accuracy of this areal interpolation, the comparison tested the robustness of each method and examined which, if any, showed particular sensitivity to catchment size.

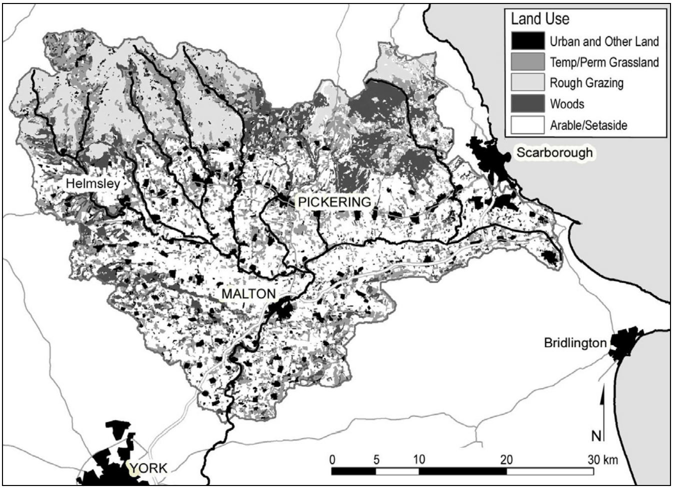


Figure 1. The Yorkshire Derwent catchment

3. Areal interpolation of agricultural census data to HRUs

An existing hydrological model, CASCADE (CAtchment SCAle DElivery), developed by hydrologists at the Centre for Ecology and Hydrology (CEH), was used to estimate nitrate inputs to the river network. CASCADE is a spatially distributed water quality model operating at daily time-step, which combines parameters such as land use, soil type, geology and rainfall, to predict nutrient deposition into waterways (Hutchins et al., 2006; Naden et al., 2001).

It was necessary to update the land use area inputs to the model; however, as there is no single data source that provides information at the ideal spatial resolution and attribute detail required for the hydrological units used in CASCADE, data from three sources had to be combined. These were (i) Land Cover Map 2000 (LCM 2000) (Fuller et al., 2002) which was reclassified into six broad land use categories used in CASCADE; (ii) Ordnance Survey digital Meridian 2 Developed Land Use Area (DLUA) boundaries (http://edina.ac.uk/digimap/description/products/meridian.shtml), to update the spatial extent of built-up areas; and (iii) 2 km grid resolution Agcensus 2004 data (EDINA, http://edina.ac.uk/agcensus/desription.shtml). The last of these datasets was used to distribute 14 arable crop types used in CASCADE within the areas defined as 'arable' and 'setaside' in LCM 2000.

The resolution of Agcensus data is coarse with respect to the HRU; therefore, as an initial step towards improving the spatial fit of the two datasets, the HRUs were aggregated into 61 larger spatial units used by CEH for modelling the effects of in-river processes on agricultural pollutants. Four different methods were then used to interpolate the Agcensus data to these aggregated units, and the resulting areal values were compared and appraised in order to assess what level of detail was required to give the most representative land use profiles for input to the CASCADE model. The interpolation methods (broadly based on a method used by CEH in the early 1990s to combine MAFF small area statistics with the LCM GB 1991 dataset) were:

1. 'Point in polygon' method to match 2 km Agcensus grid squares to aggregated HRUs. This

method interpolates data from all 2 km grid centroids falling within respective aggregated HRU boundaries.

2. Division of 2 km Agcensus grid squares into four equal 1 km grid squares, followed by a 'point in polygon' match of the 1 km squares to aggregated HRUs. This method includes data from all 'within-boundary' 1 km grid centroids.

3. Areal interpolation of the 1 km squares to obtain a proportional land use profile for each aggregated HRU. This method includes proportional data from all 1 km grid squares intersecting individual aggregated HRUs, regardless of whether the respective 1 km centroids fall within the hydrological boundaries.

4. Proportional interpolation as in Method 3, after initial forcing of agricultural land to those areas defined as agricultural by LCM 2000, rather than assuming that the Agcensus data are evenly distributed across each 2 km grid square.

The resulting land use datasets (representing absolute areal values for the various Agcensus crop and grassland categories) were evaluated against the available grassland, arable and setaside areas defined by LCM 2000. 'Scaling factors', each representing amounts by which the interpolated data had to be scaled to fit available LCM 2000 areas, were used as a comparative measure between the methods. (The extent of scaling required was likely to have been influenced by many different factors, including changes in land use over time, misclassification of land cover types, and methods used in the compilation of agricultural census data). The results were also compared across a range of spatial scales, defined by sub-catchment area (shown in Figure 2) in an attempt to identify the range of spatial resolutions at which reliable estimations of agricultural land use could be made.

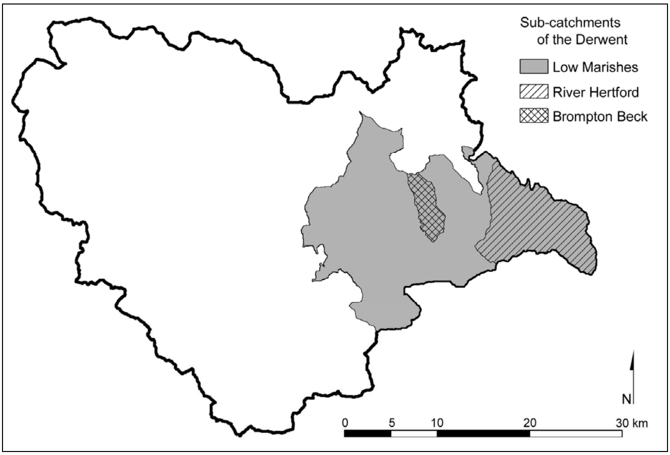


Figure 2. The River Derwent catchment, Yorkshire, UK, and three of its sub-catchments at different spatial scales. The dark outer boundary represents the entire Derwent catchment.

4. Comparison of results across a range of spatial scales

The scaling factors for each method for (i) interpolated grassland values, and (ii) interpolated arable values, for the River Derwent catchment and three of its sub-catchments, are given in Tables 1(a) and

1(b), respectively.

Table 1(a). Scaling factors representing amounts by which the interpolated grassland data had to be scaled to fit available LCM 2000 areas

Sub-catchment	Sub-catchment area (ha)	LCM 2000 area (ha) available for grassland	Scaling factor required to fit Agcensus data to available LCM 2000 space			
			2 km grid	1 km grid	proportional[a] unforced	proportional[a] forced
River Derwent	159480	35501	0.92	0.76	0.95	0.95
Low Marishes	36233	6423	0.82	0.63	0.83	0.83
River Hertford	8423	1132	0.52	0.51	0.54	0.54
Brompton Beck	1584	198	0.57	0.38	0.44	0.44

[a] Forcing the Agcensus data to the LCM 2000 agricultural areas before interpolation (Method 4) had negligible effect on the calculated grassland values and, consequently, showed no discernible difference in scaling factor at the sub-catchment scale

Table 1(b). Scaling factors representing amounts by which the interpolated arable data had to be scaled to fit available LCM 2000 areas

Sub-catchment	Sub-catchment area (ha)	LCM 2000 area (ha) available for arable land	Scaling factor required to fit Agcensus data to available LCM 2000 space			
			2 km grid	1 km grid	proportional unforced	proportional forced
River Derwent	159480	65453	1.00	0.79	1.06	1.02
Low Marishes	36233	21449	0.98	0.72	1.01	0.98
River Hertford	8423	5444	1.00	0.95	1.08	1.04
Brompton Beck	1584	1115	1.42	0.80	1.14	1.10

The most notable results from Tables 1(a) and 1(b) can be summarised as follows:

- The scaling factors for grassland data are all less than 1, indicating larger Agcensus grassland values than available LCM areas, whereas the opposite is true for the majority of arable values, whose scaling factors are close to, or greater than 1. The interpretation of grassland classes in LCM 2000, in which problems exist in the distinction between some of the unmanaged, semi-natural and improved grasslands may, in part, account for the better fit obtained for arable than for grassland values.

- As would be expected, scaling factors for grassland, across all interpolation methods, depart further from the value of 1 as catchment size decreases. However, the arable scaling factors show no such consistent behaviour, and may be influenced by methods used to preserve anonymity in the compilation of Agcensus data.

- Forcing Agcensus data to those areas defined as agricultural by LCM 2000 does not appear to offer any real benefit in respect of the amount of scaling required.

- Of particular interest is the finding that, in almost every case, the 1 km grid interpolation method gives the poorest fit over the entire range of catchment scales. This is thought to be an artefact of the way in which the Agcensus grid squares intersect the catchment boundaries (Figure 3) and is most problematic in those catchments which draw their land use values from a large number of ‘out-of-catchment’ grid squares.

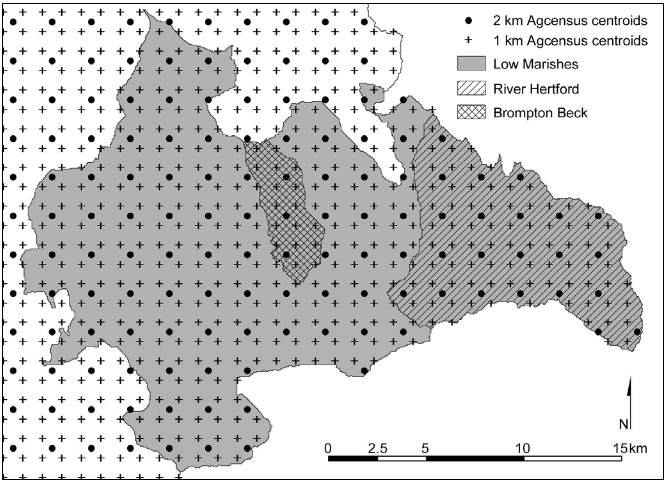


Figure 3. Placement of 2 km and 1 km grid square centroids in relation to the Brompton Beck and River Hertford catchment boundaries

This final issue is confirmed by the numbers of Agcensus cells contributing to the interpolated values at each catchment size, and the corresponding perimeter/area ratio of each sub-catchment, which influences the spatial extent over which relevant data are amassed (Table 2). In the current study, the comparatively high perimeter/area ratio of the Brompton Beck sub-catchment, combined with the involvement of a disproportionate number of ‘out-of-catchment’ Agcensus cells in the 1 km method, leads to the greatest variability in results at this catchment size.

Table 2. Perimeter/area ratios for each sub-catchment, and number of Agcensus cells contributing to land use data for each interpolation method

Sub-catchment	Perimeter/ area ratio	Number of contributing Agcensus cells			Comparative ratios between contributing Agcensus cells	
		2 km grid	1 km grid	proportional	proportional/ 2 km grid	1 km grid/ 2 km grid[b]
River Derwent	1.985	411	1977	471	1.14	4.81
Low Marishes	4.843	92	488	130	1.41	5.30
River Hertford	7.408	22	96	38	1.73	4.36
Brompton Beck	16.225	3	24	11	3.67	8.00

[b] One would expect the number of contributing cells in the 1 km interpolation to be roughly four times that of the 2 km method, but in the case of Brompton Beck, this value is doubled due to the spatial relationship between the catchment boundary and the Agcensus grid squares

5. Application to nitrate modelling

In terms of modelling mean nitrate levels (Table 3), the different interpolation methods do not appear to produce marked differences in output within individual catchments, and the only exceedance of the drinking water nitrate-N limit (11.3 mg l^{-1}) is seen at the smallest catchment scale, in Brompton Beck.

Table 3. Nitrate-N results obtained for interpolated land use values, using the different interpolation methods at a range of spatial scales. (The results for both the 'unforced' and 'forced' proportional methods (Methods 3 and 4, respectively) were identical.)

Sub-catchment	Sub-catchment area (ha)	Mean nitrate-N level (mg l^{-1}) modelled from land use values derived from three different interpolation methods		
		2 km grid	1 km grid	proportional
River Derwent	159480	6.365	6.372	6.396
Low Marishes	36233	9.897	10.010	9.943
River Hertford	8423	10.477	10.642	10.498
Brompton Beck	1584	10.990	11.482	11.135

A better indication of the variation in modelled nitrate outputs arising from the use of different interpolation methods might be conveyed by the fraction of total time that nitrate concentrations would exceed the drinking water limit. As shown in Table 4, this particular parameter of nitrate level (in contrast to mean concentration) seems to be more sensitive to choice of method, at least in the higher nitrate sub-catchments (i.e. Low Marishes, River Hertford and Brompton Beck).

Table 4. Fraction of time modelled nitrate-N levels would be above drinking water limit.

Sub-catchment	Sub-catchment area (ha)	Fraction of total time that nitrate-N levels would be above drinking water limit (11.3 mg l^{-1})		
		2 km grid	1 km grid	proportional
River Derwent	159480	0.000	0.000	0.000
Low Marishes	36233	0.119	0.155	0.127
River Hertford	8423	0.327	0.363	0.334
Brompton Beck	1584	0.405	0.649	0.497

6. Conclusions and recommendations

This analysis highlights some of the problems associated with assigning agricultural land use data to hydrological units. It is apparent that there is considerable variability between results derived for small catchments, relating both to the way in which source data are compiled and manipulated, and to differences in size, shape and location of the spatial units involved in the interpolation process. The indication is that derived land use profiles are more reliable at the scale of a large river catchment such as the Derwent, but that uncertainty increases in smaller hydrological units, with catchments at the scale of Brompton Beck being most problematic. The current study suggests that reasonable estimations of land use may be made at the scale of the River Hertford catchment (i.e. above 8000 ha), but that factors such as catchment shape and spatial relationships between input data are important considerations when assessing the reliability of such estimations, and that sophisticated interpolation procedures do not necessarily improve the outcomes. This has implications for diffuse pollution modelling in small-scale catchments, and suggests that caution should be exercised in interpreting the results.

7. Acknowledgements

EDINA at Edinburgh University Data Library and Defra are acknowledged as the sources of the Agricultural Census data. This analysis forms part of ongoing work for the Catchment Hydrology, Resources, Economics and Management (ChREAM) project, funded under the joint ESRC, BBSRC and NERC Rural Economy and Land Use (RELU) programme (award number RES-227-25-0024).

References

Aalders, IH and Aitkenhead, MJ (2006). Agricultural census data and land use modelling. *Computers, Environment and Urban Systems* **30** pp799-814

Bateman IJ, Brouwer R, Davies H, Day BH, Deflandre A, Di Falco S, Georgiou S, Hadley D, Hutchins M, Jones AP, Kay D, Leeks G, Lewis M, Lovett AA, Neal C, Posen P, Rigby D and Turner RK (2006). Analysing the agricultural costs and non-market benefits of implementing the Water Framework Directive. *Journal of Agricultural Economics* **57** pp221-237

CEC (2000). Council Directive 2000/60/EC of 23 October 2000 establishing a framework for Community action in the field of water policy. *Council of the European Communities Report No 2000/60/EC*

Defra (2006). Combining Defra June Census and IACS data to construct the first high resolution national 1km^2 agricultural land cover map. *Defra Agricultural Change and Environment Observatory Research Report No. 04,* Department for Environment, Food and Rural Affairs.

Fuller RM, Smith GM, Sanderson JM, Hill RA and Thompson AG (2002). The UK Land Cover Map 2000: Construction of a parcel-based vector map from satellite images. *The Cartographic Journal*, **39** pp15-25

Geddes A, Gimona A and Elston DA (2003). Estimating local variations in land use statistics. *International Journal of Geographical Information Science* **17** pp299-319

Huby M, Owen A and Cinderby S (2007). Reconciling socio-economic and environmental data in a GIS context: An example from rural England. *Applied Geography* **27** pp1-13

Hutchins MG, Deflandre A and Boorman DB (2006). Performance benchmarking linked diffuse pollution and in-stream water quality models. *Large Rivers* **17(1-2)** Volume 161 pp133-154

Moxey A and Allanson P (1994). Areal interpolation of spatially extensive variables: A comparison of alternative techniques. *International Journal of Geographical Information Science* **8** pp479-487

Moxey A, McClean C and Allanson P (1995). Transforming the spatial basis of agricultural census cover data. *Soil Use and Management* **11** pp21-25

Naden PS, Cooper DM and Boorman DB (2001). Modelling large-scale river basins. In *Land-ocean interaction: measuring and modelling fluxes from river basins to coastal seas*, DA Huntley, GJL Leeks and DE Walling (Eds), IWA Publishing, London.

Biography

Paulette Posen is a Senior Research Associate in the School of Environmental Sciences at the University of East Anglia. Her research interests include the application of GIS to land use/water quality relationships, and her current post involves the integrated modelling of EU Water Framework Directive impacts on rural land use and farm incomes.

Novel and disappearing climates in UK protected areas and their connectivity

Colin J. McClean

Environment Department, University of York, Heslington, York, YO10 5DD, UK.
Tel: +44 (0)1904 432995
Fax: +44 (0)1904 432998
E-mail: cjm8@york.ac.uk

KEYWORDS: climate change, conservation, functional connectivity, accessibility

1. Introduction

There are many different types of protected area across the UK, offering conservation to a wide variety of habitats. More than half of the species of conservation concern in Britain have been found to be highly dependent on protected areas for their persistence (Jackson & Gaston, 2008). Much recent research has focused on how appropriate the current protected area network is at protecting biodiversity (Gaston et al, 2006). One concern over effectiveness is how connected the protected area network is in relation to the individual species making use of them. Species may be at risk if habitats are too fragmented. The concept of functional connectivity is used by ecologists to ascertain how connected habitat within a landscape is for a species (Moilanen & Nieminen, 2002). However, most studies of functional connectivity have not begun to consider how changing climate might affect connectivity. The work presented here examines the connectivity of the protected area network in Britain in terms of how well connected its component parts are in relation to areas of similar climate, both now and under predicted future climate.

Many studies in ecology have used measures of functional connectivity (Moilanen & Nieminen, 2002). They can be grouped into three major types of methods: nearest neighbour; buffer measures; inter-patch distance matrix approaches. Interestingly, all such measures are independently used in quantitative geography, the latter group having long been referred to as accessibility indices (Guy, 1983). All can be translated into relatively simple algorithms formed by a number of functions available in proprietary and open source GIS software.

A recent approach to spatially characterising changing climate has been proposed by Williams *et al.* (2007). They use simple multivariate scaling techniques to identify areas of the Earth's terrestrial surface that are predicted to experience novel and disappearing climates by the year 2100. In effect, their methods are nearest neighbour connectivity measures in climate space. Here, similar measures of novel and disappearing climates within the protected areas of Britain are built into accessibility measures to assess their changing connectivity in the face of climate change.

2. Methods

The World Database on Protected Areas has now collated maps of protected areas, globally, for the first time (WDPA, 2008). The protected area network for Britain has been extracted from this database. The protected area network was converted to a 1 km resolution raster representing the presence or absence of protected areas in each cell. Conveniently, the proportion of land protected within Britain is close to the global proportion (Jackson & Gaston, 2008). Climate data are taken from the UK climate impacts programme (Hulme *et al.,* 2002). From the monthly average minimum temperature, maximum temperature and precipitation data available on a 5 km grid, the variables annual growing degree days above 5 degrees C, absolute minimum temperature and a moisture index have been calculated. These three variables are consistent with those used both by Williams *et al.* (2007) and in other work on climate change impacts (Prentice *et al.,* 1992; Pearson *et al.*, 2002). Four climate periods are used: a current climate period 1961-1990; a period centred on 2020; a period centred on 2050; a period centred on 2080. Future climate projections are based on the HADCM3 global circulation model results for four different emissions scenarios from low to high.

Measures of novel and disappearing climates where obtained by applying the approach proposed by Williams *et al.* (2007) to climate data for all protected area cells. The standardised Euclidean distance (SED) can be calculated between two pixels as follows:

$$SED_{ij} = \sqrt{\sum_{k=1}^{3} \frac{(b_{ki} - a_{kj})^2}{s_k}} \tag{1}$$

where a_{kj} is the value of a climate variable k at pixel j, b_{ki} is the value of climate variable k at pixel i and s is the standard deviation of the climate variable. When considering two time periods, a and b are the same climate variable in different time periods. Williams *et al.* (2007) used the variance of the climate variable at a pixel over a time series of values to standardise, but in this study space is substituted for time and the standard deviation of the variable across the study area is used. The minimum SED for each pixel to all other pixels is then found. If minimum SED is found from pixels in the current climate period to pixels in the future climate period, standardised by the variation of future climate variables, a measure of disappearing climate is obtained. Alternatively, minimum SED can be found from pixels in the future climate period to pixels in the current climate period, standardised by the variation of the current climate variables, giving a measure of novel climate. Minimum SED was also measured for each cell in the current period to all other cells in the current period to give a measure of how isolated current pixels are in terms of current climate.

The following functional connectivity measure, S, was then calculated for protected area cells (following Moilanen & Nieminen, 2002):

$$S_i = \sum_{j \neq i} \exp(-\alpha d_{ij}) A_j \tag{2}$$

where d is the distance between pixels i and j. A in a standard connectivity formulation is habitat patch area of the site j; in accessibility measures in Geography this is commonly the attractiveness of the site for some function. The parameter α gives the shape of the distance decay function in Geography, but in Ecology $1/\alpha$ is typically the average migration distance of a species in a metapopulation model framework. Here, A is the reciprocal of one of the minimum SED surfaces calculated above, either the current climate isolation surface or the novel or disappearing climate surfaces.

3. Results

Example maps of connectivity incorporating changes predicted for climate in 2080 are given in figure 1. The connectivity relative to novel climate shows reduced connectivity in southern protected areas and around the edges of upland protected areas further north. In relation to disappearing climate upland and northern areas show reduced connectivity.

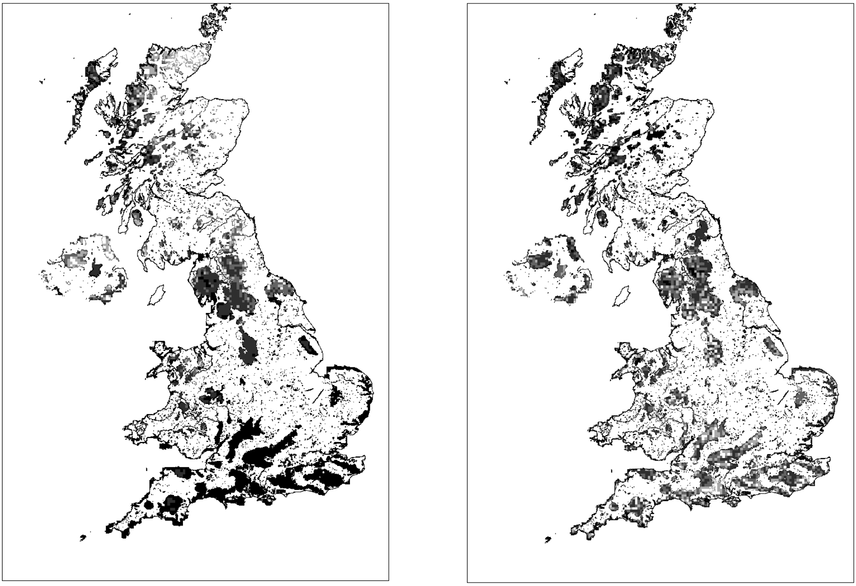


Figure 1. Left panel – connectivity by 2080 given high emission scenario and novel climate. Right panel – connectivity by 2080 given high emissions scenario and disappearing climate. Dark shades are lower connectivity.

The decline in connectivity for the entire study area by 2080, due to “novel” climate, in comparison to the baseline climate, is between a 60% and 80%, depending on climate scenario. The range in decline due to “disappearing” climate ranges from 61% to 66%. Over the two periods 2020 to 2050, and 2050 to 2080, the rate of decline appears constant and much reduced from the baseline to 2020 decline. Although, as expected, higher dispersal distances modelled show higher connectivity, percentage losses in connectivity across time are very similar for each of the dispersal distances used (1, 2, 4, 8 and 10 km). The results indicate that protected areas with “disappearing” climate will be impacted quickly, but further climate change impacts will be felt as “novel” climate conditions are experienced in the UK. The more detailed spatial structure of these patterns will be examined along with the policy implications of augmenting the protected area network in order to maintain connectivity under climate change.

4. References

Gaston, K.J., Charman, K., Jackson, S.F., Armsworth, P.R., Bonn, A., Briers, R.A., Callaghan, C.S.Q., Catchpole, R., Hopkins, J., Kunin, W.E., Latham, J., Opdam, P., Stoneman, R., Stroud, D.A. & Tratt, R. (2006) The ecological effectiveness of protected areas: The United Kingdom. *Biological Conservation*, 132, 76-87.

Guy, C.M (1983) The assessment of access to local shopping opportunities: a comparison of accessibility measures. *Environment and Planning B*, 10, 219-238.

Hulme, M., Jenkins, G.J., Lu, X., Turnpenny, J.R., Mitchell, T.D., Jones, R.G., Lowe, J., Murphy, J.M., Hassell, D., Boorman, P., McDonald, R. & Hill, S. (2002) *Climate change scenarios for the United Kingdom: The UKCIP02 Scientific Report*. Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich, UK, pp. 1–120.

Jackson, S.F. & Gaston, K.J. (2008) Land use change and the dependence of national priority species on protected areas. *Global Change Biology,* 14, 2132–2138.

Moilanen, A. & Neiminen, M. (2002) Simple connectivity measures in spatial ecology. *Ecology*, 83, 1131-1145.

Pearson, R.G., Dawson, T.P., Berry, P.M., & Harrison, P.A. (2002) SPECIES: a spatial evaluation of climate impact on the envelope of species. *Ecological Modelling,* 154, 289-300.

Prentice, I.C., Kramer, W., Harrison, S.P., Leemans, R., Monserud, R.A. & Solomon, A.M. (1992) A global biome model based on plant physiology and dominance, soil properties and climate. Journal of Biogeography, 19,117-134.

WDPA (2008) http://www.wdpa.org/

Williams, J.W., Jackson, S.T. & Kutzbach, J.E. (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the USA,* 104, 5738-5742.

Biography

Colin McClean is a senior lecturer with 20 years experience in applying spatial analysis to research fields including ecology, environmental economics, geomorphology and hydrology. Recent research has assessed the potential impacts of climate and land-use changes on the British and African floras.

Estimating domestic water demand using a scenario-based spatial microsimulation approach

Eran Md Sadek[1], Linda See[1], Rizwan Nawaz[2] and Gordon Mitchell[1]
[1]School of Geography, University of Leeds, Leeds LS2 9JT, UK
Tel. (+44113 343 3336), Fax (+44113 343 3308)
gyessm@leeds.ac.uk, l.m.see@leeds.ac.uk, g.mitchell@leeds.ac.uk
[2]Faber Maunsell Ltd., 5th Floor, 2 City Walk, Leeds, LS11 9AR, UK
rizwan.nawaz@fabermaunsell.com
Tel. (+44113 3012407)

KEYWORDS: domestic water demand, spatial microsimulation, water use behaviour

1. Introduction

The UK government has set tough domestic water demand targets for 2030 in order to achieve a drastic reduction in per capita consumption (PCC). The reduction is from the current average of 150 litres/person/day (*l/p/d*) to 120 *l/p/d*. Although the reduction may seem relatively small, it equates to a large savings in water use. The urgency for reducing water demand has been highlighted in the latest guidelines published by the Environment Agency (see Environment Agency, 2008) in response to the 2004-2006 droughts. To meet these targets, a reliable approach to estimate domestic water demand is required. Such an approach must be able to take water use behaviour, household size, population, technology (e.g. water efficient products), pricing and metering, education and awareness, and climate change into account. At present, most water companies in the UK have adopted a micro-component approach to forecast demand. This approach breaks down domestic water use into discrete components such as clothes washing, dish washing, personal washing and garden watering (Herrington, 1996, Gardiner and Herrington, 1986). Although this approach has proved to be successful, there are some limitations including lack of spatial detail, prediction of aggregate distribution rather than at the individual level and inadequate information on the behavioural elements of water use. To address these limitations agent-based modelling (ABM) and microsimulation can be used. This paper reports on ongoing progress in building a scenario-based microsimulation model to estimate domestic water demand in Leeds.

2. Previous work

There are many approaches to domestic water demand estimation and forecasting (see Gardiner and Herrington, 1986, Memon and Butler, 2006, Smith, 1986). Parson *et al.*(2007) provide an extensive review, dividing the approaches into four main categories: time series methods, causal methods, judgemental methods and experimental methods. The industry standard micro-component approach is an example of a causal method while ABM and microsimulation can be grouped into experimental methods. Unlike the micro-component approach, both ABM and microsimulation can provide individual information on water demand. ABM has been used in several studies of water demand (see Athanasiadis *et al.*, 2005, Downing *et al.*, 2003, RixonMoglia and Burn, 2007). However, most of these studies were carried out on a hypothetical gridded area.

Clarke *et al.* (1997) introduced a spatial microsimulation model in the estimation of small area water demand at Enumeration District (ED) level in Leeds. Microsimulation is a process to create a large-scale micro-level population so that the analysis of policy impacts can be carried out at the micro-level. The authors used data from the 1999 census combined with household water consumption data in the Yorkshire Water area. Data from the census were used to derive a synthetic micro-level population using *conditional probability* and *Monte Carlo sampling*. The variables used were number of people in the household, number of rooms, tenure type, property type and socio-economic characteristics. These variables provide an estimation of household water demand by determining the ownership of water-use appliances and the frequency of use. A revised version of the model included improvements in terms of modelling techniques, data used, projected water consumption, analysis of alternative future scenarios and results (WilliamsonMitchell and McDonald, 2002). The model was able to explain 44% of the variation in household water demand. The rest of the variation might be from unobserved factors such as garden size, water use behaviour and the effects of climate variability. Used in prediction of future water demand, the most optimistic results showed that there will be a 30% increase in water demand in Yorkshire Water area from the period 1991 to 2025.

Similar to the previous version of the model, several weaknesses can be identified. First is the lack of water behavioural information and second is validation of the model. One method of validation is by comparing the results with real water consumption data. This could be done by aggregating the simulated water consumption and comparing it with district meter area data. Alternatively, the prediction from a well established method such as micro-components could also be used as a comparison. A third weakness is that the scenarios used in the forecast have not taken issues of uncertainty into account. It is possible to model uncertainty utilising the probabilistic nature of the microsimulation model to produce confidence intervals or use a bootstrapping approach to produce a range of results. A sensitivity analysis could also be undertaken to determine the effect of different scenarios. Finally, other potential scenarios such as pricing and metering, introduction of an education and awareness campaign, and future climate variability could be included. This study will attempt to address some of these weaknesses.

3. Methodology

The study area is the city of Leeds, which is located in the region of Yorkshire Water Services. Leeds has a population of more than 700,000 in 300,000 households. Census Area Statistics (CAS) and Domestic Consumption Monitors (DCM) were used in a spatial microsimulation model. Figure 1 illustrates the process. In the first phase, a correlation analysis was carried out to determine which variables had the strongest correlation with water demand. As found by Williamson *et al.* (2002), the number of households, the number of bedrooms, the tenure, the property types, and washing machine and dishwasher ownership have the strongest correlation. These variables were used as constraints to generate a synthetic micro-level household population (see Phase 2 in Figure 1). In addition, ACORN geodemographic data were used because this relates water consumption to the socio-economic profile of a household. The synthetic micro-level household population was created using *deterministic reweighting* which was introduced by Ballas *et al.* (2005) and then further modified by Smith *et al.* (2007). The reweighting is calculated as follows:

$$n_i = w_i * s_{ij} / m_{ij} \qquad (1)$$

where:

n_i = the new weight of a household *i*

w_i = the old weight of a household *i*

s_{ij} = the element s of the particular area statistics table for *i* individual and attribute *j*

m_{ij}= the element m of the survey data table for *i* individual and attribute *j*

The microsimulation is carried out using the Flexible Modelling Framework (FMF) that has been developed in-house at the University of Leeds using java (Harland and Stillwell, 2007). The generated synthetic household population is validated by aggregating the simulated household into Census Output Areas (OAs) and compares the value of the constrained and unconstrained variables with known values in the census. This provides a model goodness-of-fit and determines whether the model is able to produce good estimates of baseline water demand for the area. Individual water consumption data for 2001 are used to correspond to the 2001 Census. The baseline water demand for the household population was derived by attaching the water consumption data in the DCM to the specific types of household in the OA (refer to Phase 3 in Figure 1). The sum of these values is the total water consumption for the whole OA as either an annual average demand or for the peak season e.g. water demand in the summer. A crude estimation of per household consumption (PHC) and per capita consumption (PCC) was derived by dividing the total water consumption for an OA by the number of households or persons living in that OA. Figure 2a shows the distribution of water demand in 2001 for the whole OA, Figure 2b for PHC and Figure 2c for PCC produced by an initial run of the microsimulation model and was produced using the ArcGIS software.

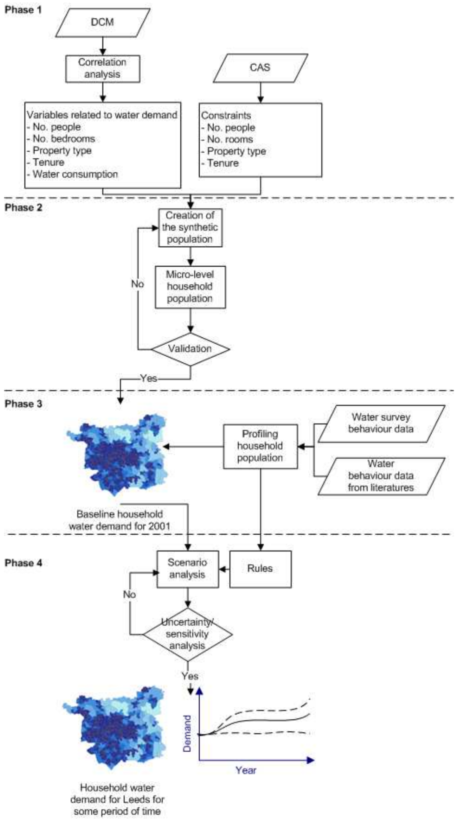
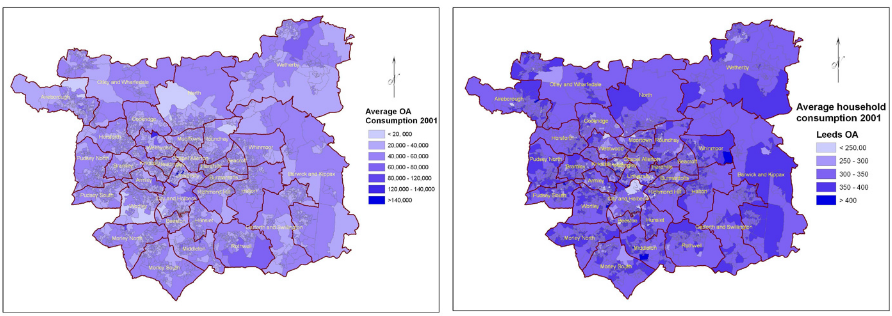
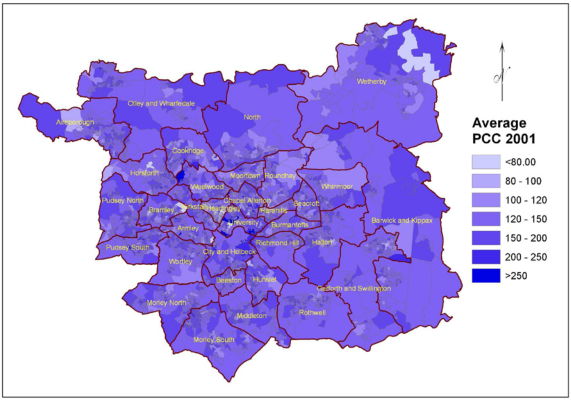


Figure 1. Microsimulation methodology to create estimates of water demand



(a) (b)



(c)

Figure 2. Estimated water demand of 2001 for (a) OA (b) PHC (c) PCC

The next step in the research is to include water use behaviour in the model (see Phase 3 in Figure 1). Because of the limited information on water use behaviour in the literature, this study has developed an online water use and conservation survey to collect as much data as possible about issues of water awareness and behaviour. In addition, a traditional paper-based survey also will be conducted to minimise the bias from the online sample, especially to types of people that have more limited access to the internet. The data collected from the survey will be used to categorise households into different types and form rules of behaviour. For example, household type "A" may have the highest degree of awareness, a certain socio-economic class and the lowest level of water consumption. Type "B" may be intermediate while type "C" may have the lowest degree of awareness, etc.

The final phase in this study (see Phase 4 in Figure 1) will be the application of scenarios to produce water demand projections. These scenarios will include potential changes in population and household size, new or changing policy and regulation such as pricing and metering, the effects of an education and awareness campaign, and future climate variability. The scenario analysis will measure how different types of household in the above categories will respond towards the scenario implemented in the model. Rules will be formulated from a combination of the survey data and from findings in the literature.

4. Conclusion

This paper reports initial results from an ongoing study of domestic water demand using spatial microsimulation. Future work includes adding several different DCM datasets from other water companies. This will expand the capability of the model to be applied to different areas. At the moment this model is only applicable to Leeds. It is hoped that when the model is fully functional, it will be possible to understand the effects of different drivers on domestic water consumption, with the ability to make recommendations about what policies may be effective in reducing water demand in the future.

5. Acknowledgements

The main author is grateful to the Malaysian Government for funding this study. The DCM data used in this work have been obtained courtesy of the WaND project, in particular Prof Adrian McDonald and Patrick Sim.

References

ATHANASIADIS, I. N., A. K. MENTES, P. A. MITKAS & Y. A. MYLOPOULOS. 2005. A hybrid agent-based model for estimating residential water demand. *Simulation,* **81** (3), pp.175-187.

BALLAS, D., G. CLARKE, D. DORLING, H. EYRE, B. THOMAS & D. ROSSITER. 2005. SimBritain: A spatial microsimulation approach to population dynamics. *Population, Space and Place,* **11** (1), pp.13-34.

CLARKE, G. P., A. KASHTI, A. MCDONALD & P. WILLIAMSON. 1997. Estimating Small Area Demand for Water: A New Methodology. **11** (3), pp.186-192.

DOWNING, T. E., R. E. BUTTERFIELD, B. EDMONDS, J. W. KNOX, S. MOSS, B. S. PIPER, E. K. WEATHERHEAD & (THE CCDEW PROJECT TEAM). 2003. *Climate change and the demand for water.* Oxford: Stockholm Environement Institute Oxford Office.

ENVIRONMENT AGENCY. 2008. *Future Water - The goverment's water strategy for England.* Bristol: Environment Agency.

GARDINER, V. & P. HERRINGTON. 1986. The basis and practise of water demand forecasting. *In:* V. GARDINER & P. HERRINGTON, eds. *Water demand forecasting.* Norwich: Geo,

HARLAND, K. & J. STILLWELL. 2007. *'Using PLASC data to identify patterns of commuting to school, residential migration and movement between schools in Leeds'. Working Paper 07/02,* [online]. [Accessed 24 September 2008]. Available from World Wide Web: <http://www.geog.leeds.ac.uk/wpapers/PLASC.pdf>

HERRINGTON, P. 1996. *Climate change and demand for water.* London: HMSO.

MEMON, F. A. & D. BUTLER. 2006. Water consumption trends and demand forecasting techniques. *In:* D. BUTLER & F. A. MEMON, eds. *Water Demand Management.* London IWA Publishing, pp.1-26.

PARSON, J., P. REES, P. SIM & A. MCDONALD. 2007. *MACROWater: a top-down, policy driven model for forecasting domestic water demand. Working Paper 07/1* [online]. [Accessed 28 April 2008]. Available from World Wide Web: <http://www.geog.leeds.ac.uk/wpapers/MACROWaterWorkingPaper07_1.pdf>

RIXON, A., M. MOGLIA & S. BURN. 2007. Topics on system analysis and integrated water resources management. *In:* C. ANDREA & R. SONCINI SESSA, eds. *Topics on System Analysis and Integrated Water Resources Management.* Oxford: Elsevier,

SMITH, D. M., K. HARLAND & G. P. CLARKE. 2007. *SimHealth: estimating small area populations using deterministic spatial microsimulation in Leeds and Bradford. Working Paper 07/06* [online]. [Accessed 30 April 2008].

SMITH, R. 1986. Forecasting Industrial demand for water. *In:* V. GARDINER & P. HERRINGTON, eds. *Water demand forecasting.* Norwich: Geo,

WILLIAMSON, P., G. MITCHELL & A. T. MCDONALD. 2002. Domestic water demand forecasting: A static microsimulation approach. **16** (4), pp.243-248.

Biography

Eran Md. Sadek is a postgraduate student in the School of Geography, University of Leeds. He is sponsored by the Malaysian Government and is employed in the Department of Surveying Science and Geomatics, Universiti Teknologi MARA, Malaysia.

Dr Linda See is a Senior Lecturer in GIS and Computational Geography. Her research interests include soft computing techniques (fuzzy logic, neural networks and genetic algorithms) and agent-based modelling as applied to a range of spatial problems: flood forecasting, geodemographics, crime reduction, etc.

Dr Rizwan Nawaz works as a consultant for Faber Maunsell Ltd. His expertise is primarily related to investigating the effects of climate change on flood and drought risk. He has undertaken project work both in the UK and overseas, especially in developing countries..

Dr Gordon Mitchell is a Senior Lecturer in the School of Geography and Institute for Transport Studies, University of Leeds. His research interests are in environmental planning and assessment including sustainability appraisal, natural resource management and decision support.

Re-Drawing the World: An Approach towards a gridded World Population Cartogram

Benjamin D Hennig[1], Danny Dorling[1], Mark Ramsden[1]

[1]Department of Geography
The University of Sheffield
SHEFFIELD S10 2TN
United Kingdom
Tel. +44 114 222 7900 Fax +44 114 279 7912
b.hennig / danny.dorling / m.ramsden [@sheffield.ac.uk]
http://www.sasi.group.shef.ac.uk/

KEYWORDS: griddata, worldmapper, cartogram, population, geographic visualisation

1. Introduction

This work builds upon the worldmapper project which utilises cartograms that preserve the shapes of countries while distorting the area according to a specific variable. These maps are open to potential criticism when it comes to their informative value. One such criticism is the variation of the depicted topic within the territorial borders is not taken into consideration. A possible solution to this problem is presented here with a redrawn edition of worldmapper's world population cartogram calculated with grid-based data rather than using a single population figure for each country.

1.1 Worldmapper and its World Population Cartogram

In the first stage of the worldmapper project, a wide range of maps depicting various human dimension of the world have been published (http://www.worldmapper.org). Since the publication of the first new world population cartogram in 2006 (Webb 2006) nearly 600 maps have been produced, covering topics such as education, poverty, and pollution (Dorling, Barford & Newman 2006).
The worldmapper cartograms show the data for 200 territories, thus making this new view on the world to some extent an arbitrary view: territorial borders are artificial. The world population cartogram was used to test different ways to calculate new more detailed grid-based cartograms beyond the territorial borders.

2. Data and Cartogram Calculation

Data used in this work were derived from the Socioeconomic Data and Applications Center (SEDAC) of Columbia University, New York. The Gridded Population of the World (GPW) database contains the distribution of world's population on a gridded base (http://sedac.ciesin.columbia.edu/gpw/), including population data and estimates from 1990 to 2015. These data are available in resolutions of up to 2.5 arc minutes leading to a population grid of 8640x3432 pixels. Data from the year 2000 have been used to make results comparable to the original worldmapper population cartogram.
These raster format data were imported to ESRI's ArcGIS, converted to polygons and combined with further metadata (e.g. country labels) to match gridcells for further visualisation tasks. The cartogram script (see below) uses a 4096x2048 pixel-sized lattice for its map results.
The cartogram itself is calculated by using the ArcScript *Cartogram Geoprocessing Tool* by Tom Gross (http://arcscripts.esri.com/details.asp?dbid=15384) which uses Newman and Gastner's density-equalizing method methodology (Gastner & Newman 2004). Unlike the worldmapper cartograms that distort an initial projection of the boundaries of the territories, each population grid is treated as a separate part for the calculation, not taking any territorial information of borders into account. Thus each grid cell marks a border so that distinct shapes of countries are intentionally of no interest in the

calculation.
Changes in the distortion of the resulting cartogram thus are only possible by adjusting the factor to smooth the original density. In addition, data from the United Kingdom have been extracted from the 2.4 arc minutes population grid and are calculated separately in the same way to produce a more detailed view of the resulting grid and its interval variation.

3. Results

The resulting cartograms require some final visualisation steps to adapt them to appear similar to the original worldmapper cartograms. The polygons of the calculated world population cartogram are dissolved according to their affiliation to the worldmapper territories and coloured according to the distinctive worldmapper colour scheme. The gridlines in the UK cartogram are conserved to show the degree of distortion within the grid.

3.1 A redrawn World Population Cartogram

Compared to its predecessor (Figure 1), the redrawn World Population Cartogram (Figure 2) shows considerable differences. For example, in China the sparely populated Himalayan regions can be distinguished from the densely populated eastern coastal regions. Internal variation within the United States and Mexico can also be recognized. Somewhat harder to identify but still evident are North-South differences in Great Britain and West-East differences in Germany. Hence, our goal to take the varying distribution of population on a sub-national level and make them visible on a global view has been achieved. However, sub-national variation can be difficult to analyze in more detail because the gridcells are eliminated to sustain the view on the global scale. In addition, more distinctive national shapes are far more distorted than in the original cartogram, which for some users might appear odd when interpreting such maps.

3.2 Down to Earth: A Population Cartogram of the United Kingdom

To counter the loss of familiar national boundary shapes a separate population cartogram is produced for the United Kingdom (Figure 3). The shape of the cartogram has more detail compared to the shape of the UK on the world population cartogram. This is because more gridcells are used in the calculation of the cartogram and no other polygons (e.g. from the European continent) influence the calculation. The different scale also allows the visualisation of each gridcell so that sub-national variation can be recognized. An “original” map of Britain with its familiar shape is shaded-in underneath the grid to aid interpretation.
This visualisation on a different scale is an improvement in the visualisation that goes far beyond the current capabilities of the worldmapper project by using gridded base data to allow not only a different view on population distribution worldwide but also within separate regions. By using cartogram techniques, a different view on the regional variations of the human geography is created which can hardly be achieved with traditional mapping techniques.

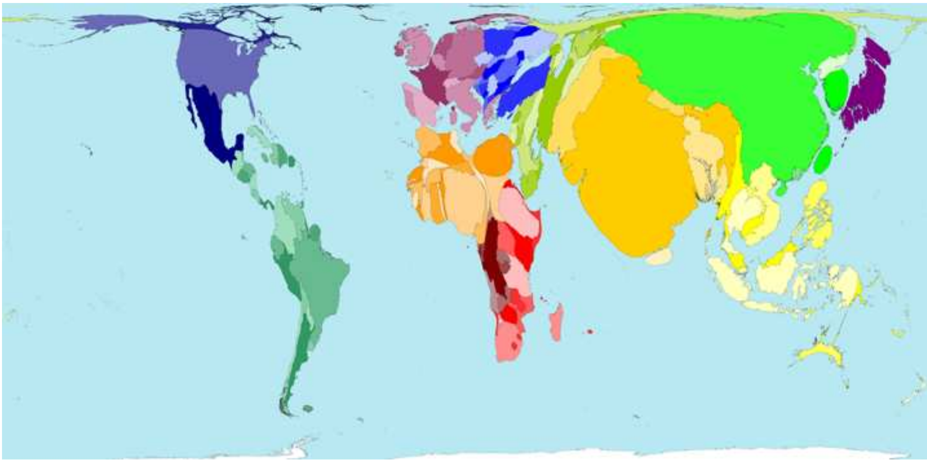


Figure 1. Worldmapper Population Cartogram
(Source: http://www.worldmapper.org/images/largepng/2.png)

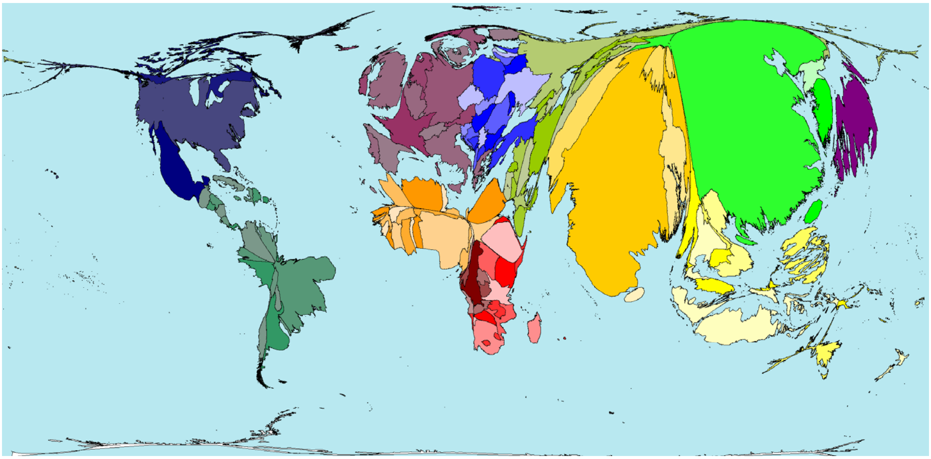


Figure 2. Grid-based World Population Cartogram (2000)

Figure 3. Grid-based Population Cartogram of the United Kingdom (2000)

4. Outlook

The most significant obstacle to the realisation of gridded depiction for worldmapper will be the vast quantity of the different topics covered and availability of data. Reliable gridded socio-economic data for the whole world are barely available and rarely of such good quality as the population data. The estimation of missing national data for some topics has already been a serious matter in the existing worldmapper cartograms (Dorling, Barford & Newman 2006; Dorling 2007). Such estimations will not meet the demands of gridded datasets, so new ways of data estimation are needed (e.g. Gaffin et al 2004; Hay, Graham & Rogers 2006).

Revised gridded cartograms offer great potential to enhance the variety of worldmapper's visualisation capabilities. A different view of the "real" location of the depicted topic can present a better understanding of the human geography of our planet. However, distortions associated with the gridded method can potentially undermine the purpose of the used algorithm to preserve country shapes. Therefore, the potential of the gridded approach and the desire to preserve familiar shapes must be carefully balanced. Nevertheless, much potential lies in adding more user-interactivity and detail to worldmapper. Grid-based cartograms have the advantage of allowing a user to zoom in to view national and regional details, within a global context. Finally, a transfer to popular digital globes can thus easily be realised, allowing viewers to identify the regional dimension of a subject. Separate regional editions of gridded population cartograms can be generated to visualise the regional variation of population distribution.

5. Acknowledgements

We would like to thank our colleague John Pritchard for all his help and the Leverhulme Trust for funding the Worldmapper project. The trust played no role in the submission or preparation of this article.

References

Dorling, D (2007), Worldmapper: the Human Anatomy of a Small Planet, *PLoS Medicine* **4(1)**, pp13-18.

Dorling, D, Barford, A, and Newman, M E J (2006), Worldmapper: the world as you've never seen it before, *IEEE Transactions on Visualisation and Computer Graphics* **12 (5)**, pp757-764.

Gastner, M T, and Newman, M E J (2004), Diffusion-based method for producing density equalizing maps, *Proc. Natl. Acad. Sci. USA* **101,** pp7499-7504.

Gaffin, S R, et al. (2004), Downscaling and geo-spatial gridding of socio-economic projections from the IPCC Special Report on Emissions Scenarios (SRES), *Global Environmental Change* **14(2)**, pp105-123.

Hay, S I, Graham, A, and Rogers D J (2006), *Global mapping of infectious diseases: methods, examples and emerging applications*. Academic Press, London.

Webb, R (2006), Cartography: a popular perspective, *Nature* **439**, p800.

Biographies

Benjamin Hennig completed his Master degree in 2005, working with remote sensing applications and geographical information systems. He joined the worldmapper team at Sheffield University in 2008 where he currently works on a PhD project to improve worldmapper's geo-visualisation capabilities.

Danny Dorling worked on children's play schemes in the late 1980s, but has been trapped in universities since then. He has worked with several colleagues on a number of books, papers and reports. He is currently professor of human geography at Sheffield University.

Mark Ramsden is lecturer in Social and Spatial Inequalities at the Geography Department, University of Sheffield. Mark's interests focus on spatial expressions of social inequality. Currently he is working on a project investigating gentrification, ethnicity and schooling in East London.

Automated and Subjective Terrain Feature Extraction: A Comparative Analysis

Delroy Brown, David Mountain, Jo Wood

giCentre, City University London, EC1V 0HB
www.gicentre.org
Email:brown.delroy@gmail.com

Available online: www.gicentre.org/papers/gisruk09/BrownD.pdf

KEYWORDS: terrain feature extraction; digital elevation modeling; Ethnophysiography

1. Introduction

Ethnophysiography, proposed by Mark and Turk (2003), examines people's conceptualizations of landscapes through interviews and description. This paper investigates one aspect of this concept, comparing subjective perception of terrain features with automated extraction techniques, for an area in the English Lake District. The primary research objective is to determine if there is a scale of analysis for automated terrain feature classification which coincides with human perception of the same features in the landscape.

This analysis required subjective tests to occur in rugged terrain where large numbers of candidate features exist and frequented by people. The area chosen, the Old Man of Coniston and its environs, is a region of dramatic relief, of which the summit is a popular destination for hikers.

The remainder of this paper is in two sections: method and discussion. The method, describes a subjective study that identified objects in the landscape that human subjects considered representative of different classes of terrain features; and an automated process which extracts terrain features from a digital elevation model across a range of scales. In the discussion, these datasets are compared to detect the scale of analysis for automated feature extraction processes that coincides with the scale at which people perceive different classes of terrain features.

2. Method and results

2.1 Subjective terrain feature identification

The first dataset created represents a subjective assessment of prominent features in the study area. It was generated by interviewing one hundred hill-walkers from the summit of the Old Man of Coniston between July 25–27 2008. Participants were asked to identify what they considered the first and second most *prominent* of four classes of morphometric features - peaks, channels, ridges and passes - from the observation point. The word prominent used in the interviews as suggested by Greatbatch *et al* (2007) was intended to convey to participants that their choice need not be constrained to the largest or closest feature. Respondents were guided by the schematic feature representations in Figure 1.

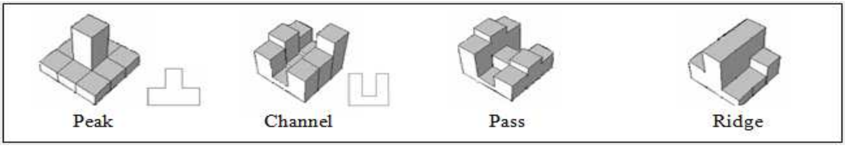


Figure 1. Schematic Morphometric Features (Adopted & modified after Wood, 1996)

One potential source of error was a lack of synchrony between the feature identified by the participant and that recorded by the interviewer. To minimize this error, the 1:25000 Ordnance Survey (OS) map of the subject area was draped over a three dimensional image of the terrain, and a paper copy of this scene annotated in the field to confirm the features identified. Respondents highlighted summits on the OS explorer map to record peaks, while an approximation of the centre of channels and ridges and their lengths were recorded with lines. Shapefiles were created for each feature type by manual digitizing inESRI ArcGIS, then imported to the LandSerfGI System (Wood, 2008).

In all, nine peaks, three ridges, four channels and six passes were identified over the course of 100 interviews (Figure 2). While all respondents were able to identify peaks, ridges and channels, only 43% were able to identify passes.

Initial analysis allowed the following observations to be made:

1. Changes in visibility over the period had no obvious impact on the results.

2. Features identified as being prominent tended to be within a radius of approximately 4 kilometers from the observation point.

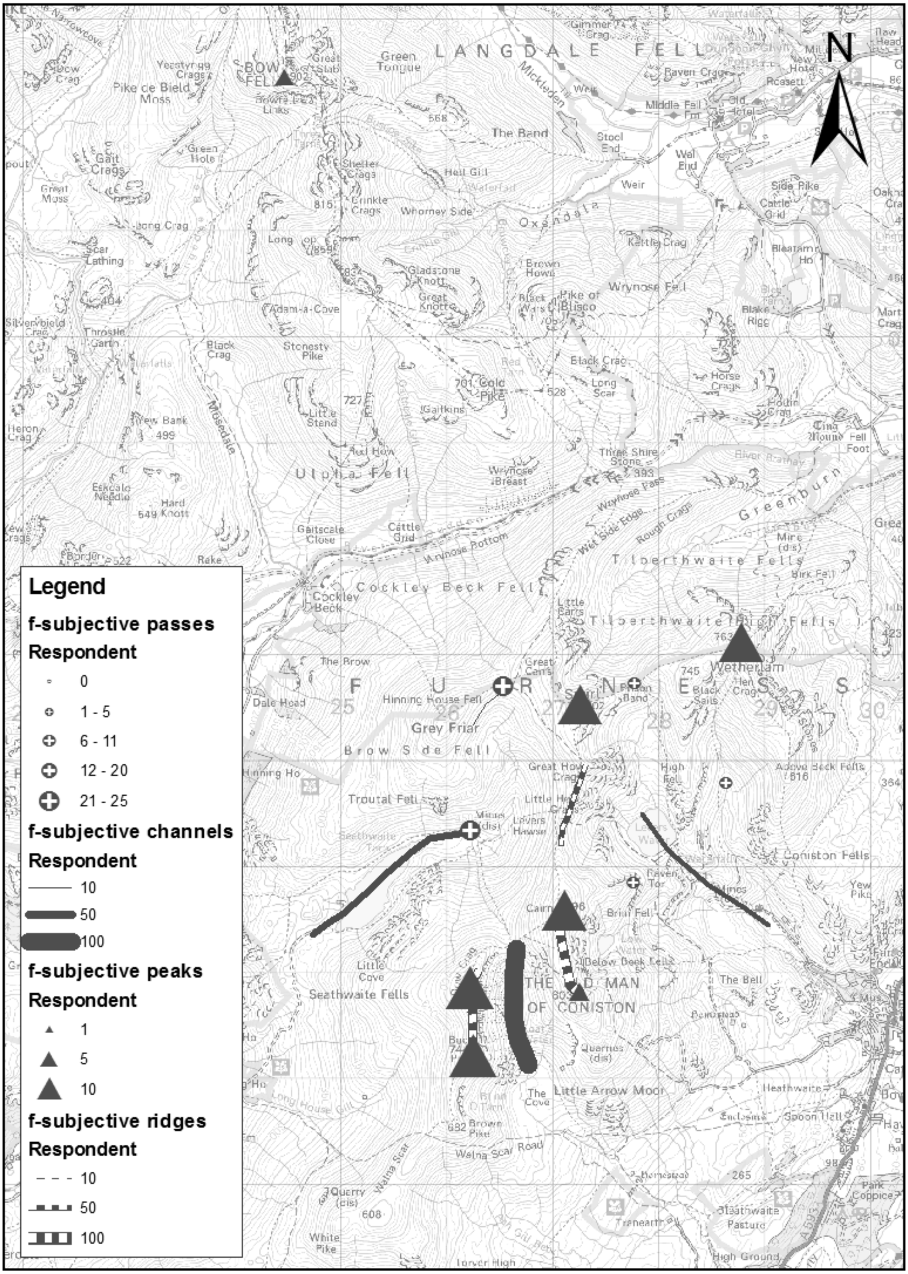


Figure 2: Terrain features identified by hill walkers from the Old Man of Coniston
Crown Copyright/database right 2008. An Ordnance Survey/EDINA supplied service

2.2 Automated terrain feature extraction

One of the most widely used set of automated morphometric classifications partitions a landscape into distinct terrain features (Wood 1996). Automated feature extraction algorithms, such as Fowler and Little, assigns the individual cells of a Digital Elevation Model (DEM) a single feature class by examination under a 3x3 kernel. Each of the eight neighbors surrounding the mid cell are assigned a positive or negative value depending on their elevation relative to the mid cell. The pattern of these neighbours is then used to define which feature class the cell belongs to: for example, the mid cell would be defined as a peak if all neighbours are lower (Figure 3).

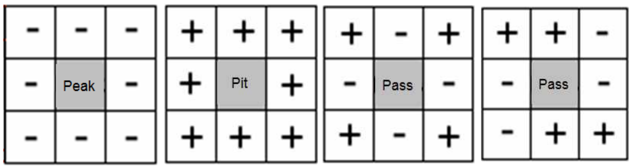


Figure 3: Cell Classification (Fowler and Little Algorithm (Adopted from Wood 2007)

Terrain features were extracted using LandSerf at various scales of analysis using the OS 50m DEM of the study area. This was achieved using the Feature Network tool, which maintains topological integrity of features. A distance decay of 0 was used, assigning equal importance to cells. The output of this process is a continuous surface with cells assigned to one of six candidate feature classes: peak, channel, ridge, pass, pit or plane (Wood, 1996). Figure 4 shows some results of the process.

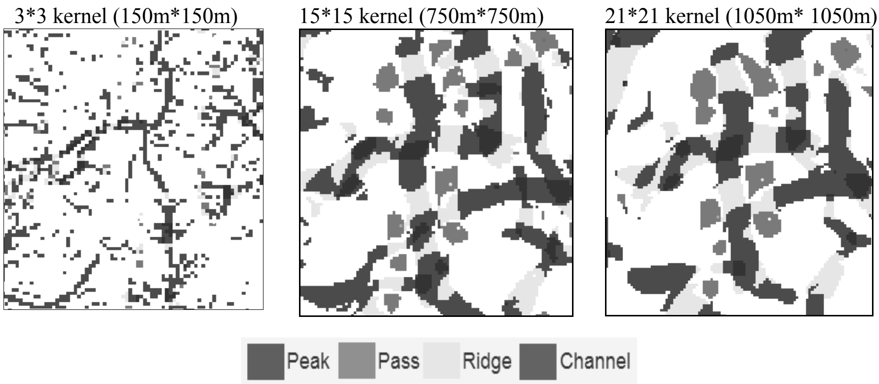


Figure 4: Effect of scale on automated feature classification

3. Discussion: Comparative Analysis

The vector dataset representing subjective terrain features was overlaid and compared against the automatically generated raster datasets. This procedure was repeated for four feature types, peaks, ridges, channels and passes. This section assesses the results of the subjective classification prior to attempting to establish the scale of analysis at which human participants identified terrain features.

3.1 Peaks

Figure 5 shows the peaks identified by the subjective classification, against the number of responses, and figure 6, the relationship between elevation, distance from observation point and number of responses. This demonstrates a relationship between distance and features identified as being prominent. Beyond 3.6 kilometers, very few features were identified.

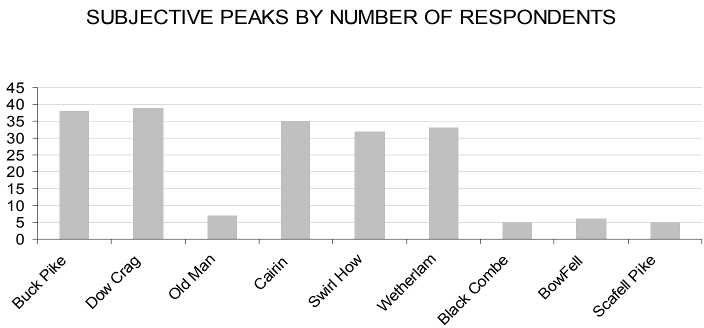


Figure 5: Frequency of specific peaks, from subjective analysis

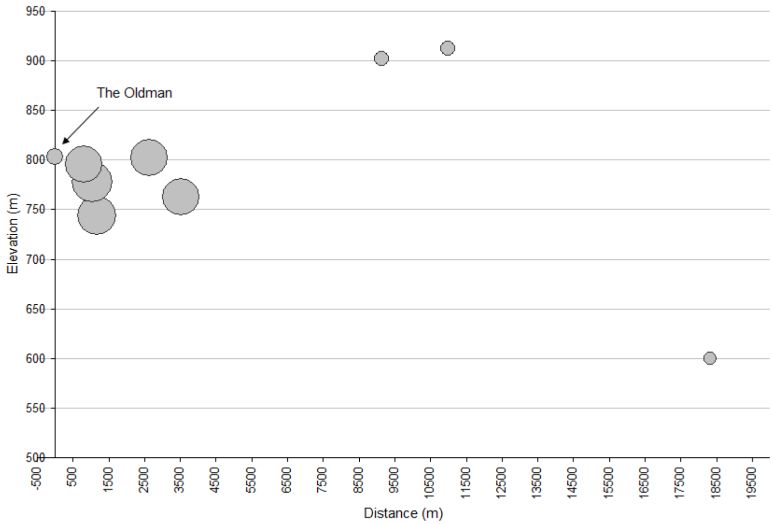


Figure 6: Identified Peaks: elevation (y-axis) vs distance from observer (x axis).
Dot sizes represents number of feature responses

The peaks identified subjectively display the closest alignment with the automated results using kernel sizes between 750m and 1650m (15x15 to 33x33 cell kernels), as in Figure 7. At finer scales, there are many peaks which were not identified by participants, suggesting that features must exceed this threshold before being considered a peak. At wider scales of analysis, what participants considered to be peaks tended to be identified as other features by the automated extraction process. This confirms a known issue in the identification of mountain features: whether a particular feature is a peak on its own or part of a ridge.

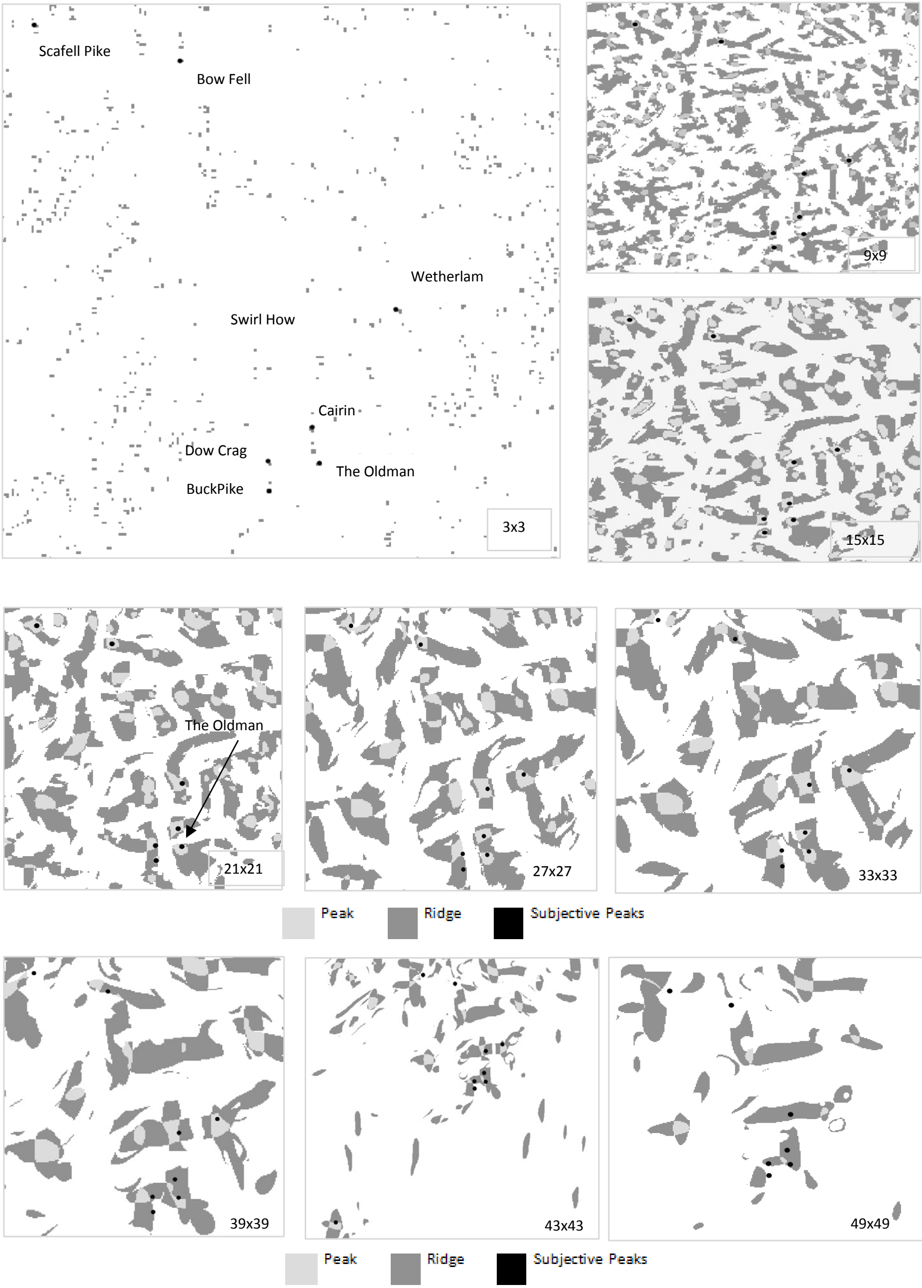


Figure 7: Subjective and Automated Feature Overlay (Peaks)

3.2Ridges

Participants identified fewer ridges than peaks (just 3), all of which were within 4km of the observation point. Ridges A and B had an almost equal number of responses totaling 78.8 percent, and were both connected to features that were identified as peaks in the subjective feature classification (see Figure 8). At smaller window sizes subjective ridges were a composite of multiple peaks and ridges in a linear manner, and were within 0.9m close proximity to the observation point. At window sizes 9x9 and 21x21, subjective ridges were classified as passes, and a higher degree of synchrony with automated results with sizes 27x27, 33x33, and 39x39.

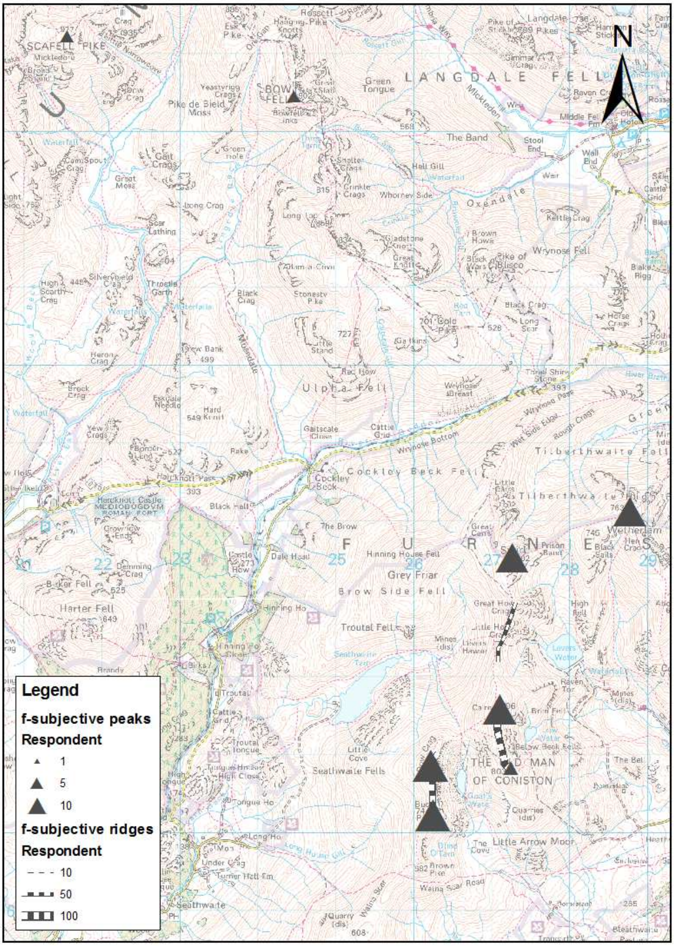


Figure 8: Locations of Subjective Ridges and Peaks
Crown Copyright/database right 2008. An Ordnance Survey/EDINA supplied service

It was observed that there are similarities in classification with the subjective views using kernel sizes of 1350m – 1950m (27*27 – 39*39 cell kernels), as shown in Figure 9. Features tend to decompose when smaller kernels were used.

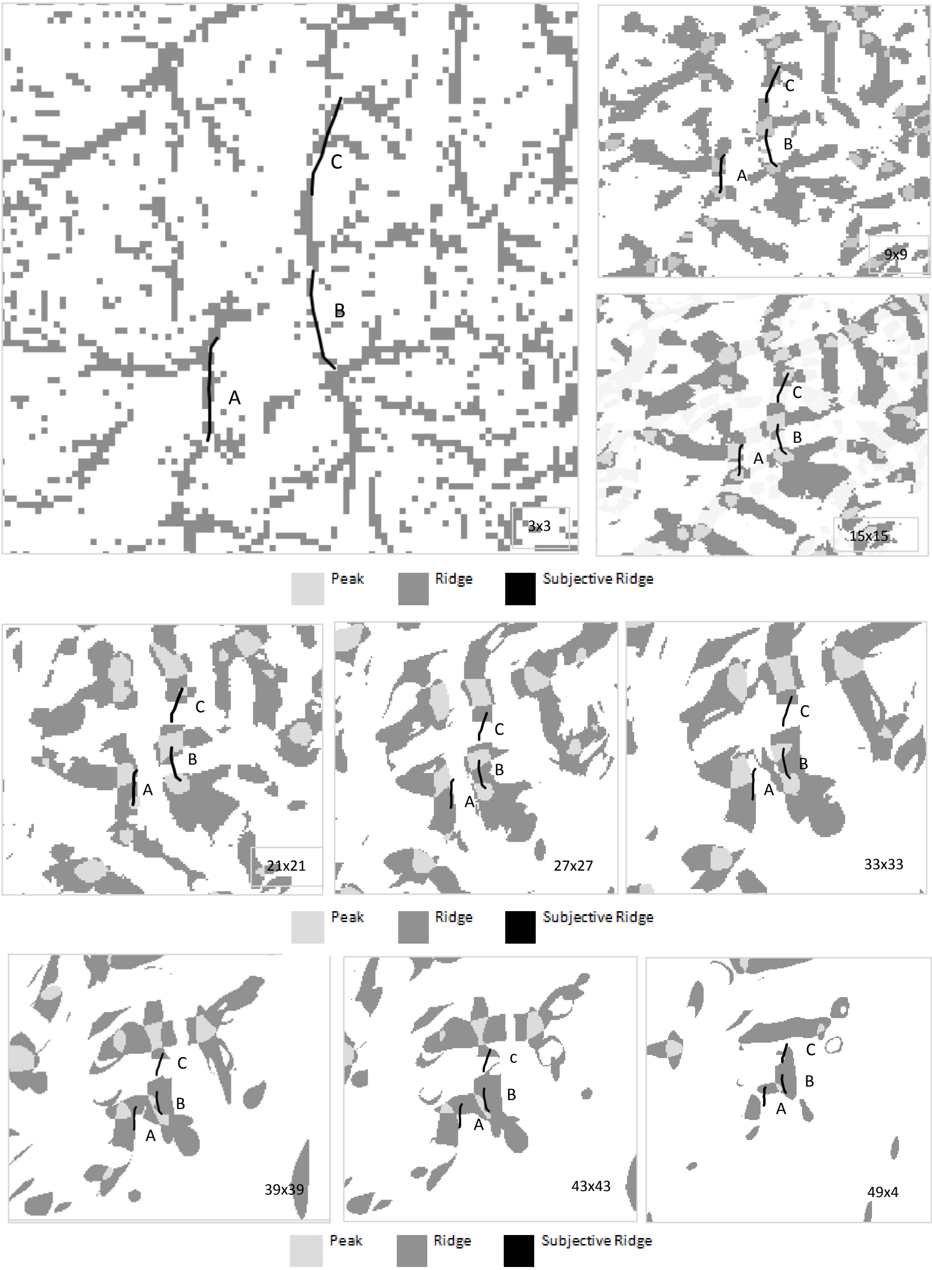


Figure 9: Comparison of subjectively and automatically identified ridges

3.3 Channels

Ninety six percent of the participants identified the same three features (A, B and D) as channels; all of which were within a horizontal distance 3km of the observation point. Figure 10, shows the locations of these subjective channels in relation to peaks. All identified channels were within 3km of the observation point, suggesting that it is harder to identify distant channels than peaks.

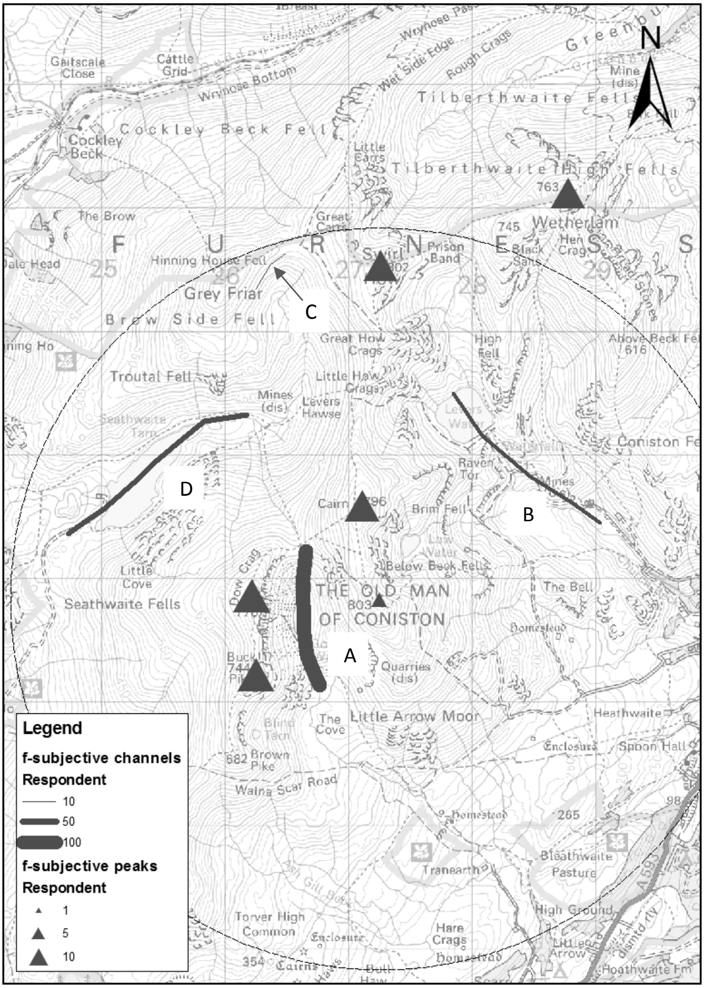


Figure 10: Subjective Channels Shaded-represents Peaks-Threshold-600m

People's perception of channels was consistent with those extracted automatically using kernels between 1050m – 1650m (21*21- 33x33 cell kernels), although there is some discrepancy between human and automated classification across the scales.

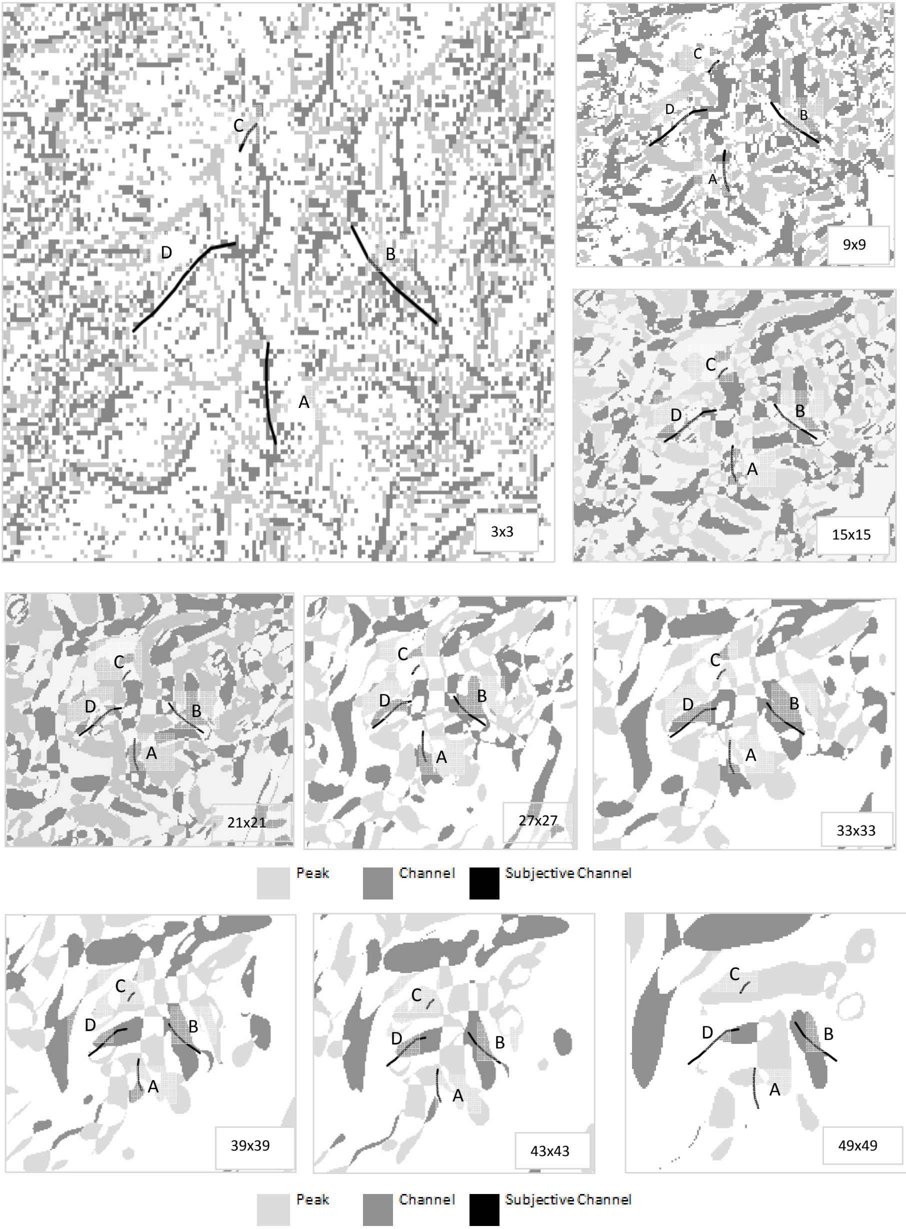


Figure 11: Comparison of subjectively and automatically identified channels

3.4 Passes
Five of six passes were within 3.6 kilometers of the observation point. The data suggests features C, E and F are more prominent examples of, passes with a total of seventy two percent of the responses (Figures 12 and 13). Like the location characteristics of peaks, the passes identified were similar, with a minimum and maximum distance from each subjective peak of 0.6 and 1.9 kilometers.

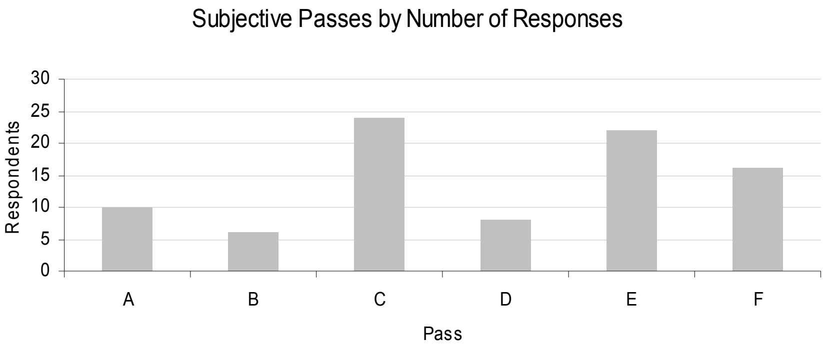


Figure 12: Subjective Passes and Number of Responses

Of the subjective passes, F, E and C remained across the automated scales of analysis from 150m (3*3) – 2150m (43*43), suggesting they are the least scale dependent features. There is consistency at these scales with both results. This suggests that passes are recognized within the range 3x3–33x33.

Figure 14 shows the agreement between terrain features identified by participants in the survey (subjective features), and those calculated using LandSerf's surface feature classification algorithm over a range of scales. These values were calculated by sampling the pixels in the feature classification surfaces which intersected the locations at which participants identified terrain features: these locations were points for passes and peaks, and lines for channels and ridges. Higher values suggest greater agreement between the subjective and algorithmic approaches. For example, for peaks at a scale of analysis of 450m, the value of 1 indicates that all those subjectively identified peaks coincided with pixels classified as peaks by the feature classification algorithm.

It can be seen that for peaks there is greatest agreement between the subjective approach with algorithmic analysis conducted at scales of 500-1250 metres. This agreement drops sharply for scales of analysis above 1500m. Closer inspection of figure 7 suggests, that this is due to the algorithm classifying the locations identified as peaks by participants, as being part of a ridge for these coarser scales of analysis.

The profile of passes is similar to that for peaks, with greatest similarity between the subjective identification and algorithmic classification at a scale of roughly 1000m. Significantly, there is far greater overall agreement in the subjective and algorithmic approaches for peaks than for passes, suggesting there may be less ambiguity about what form a peak takes.

The profiles for channels and ridges follow a similar trend: the greatest agreement occurs at scales of around 2000m, but both have secondary peaks at fine scales of analysis. For channels, there is a strong trend for less agreement for fine scale analysis (<750m) and coarse scale analysis (>2250m). The trend for ridges is less clear, however when compared to peaks and passes, the trend for these linear features (ridges and channels) is for greater correspondence at coarser scales of analysis.

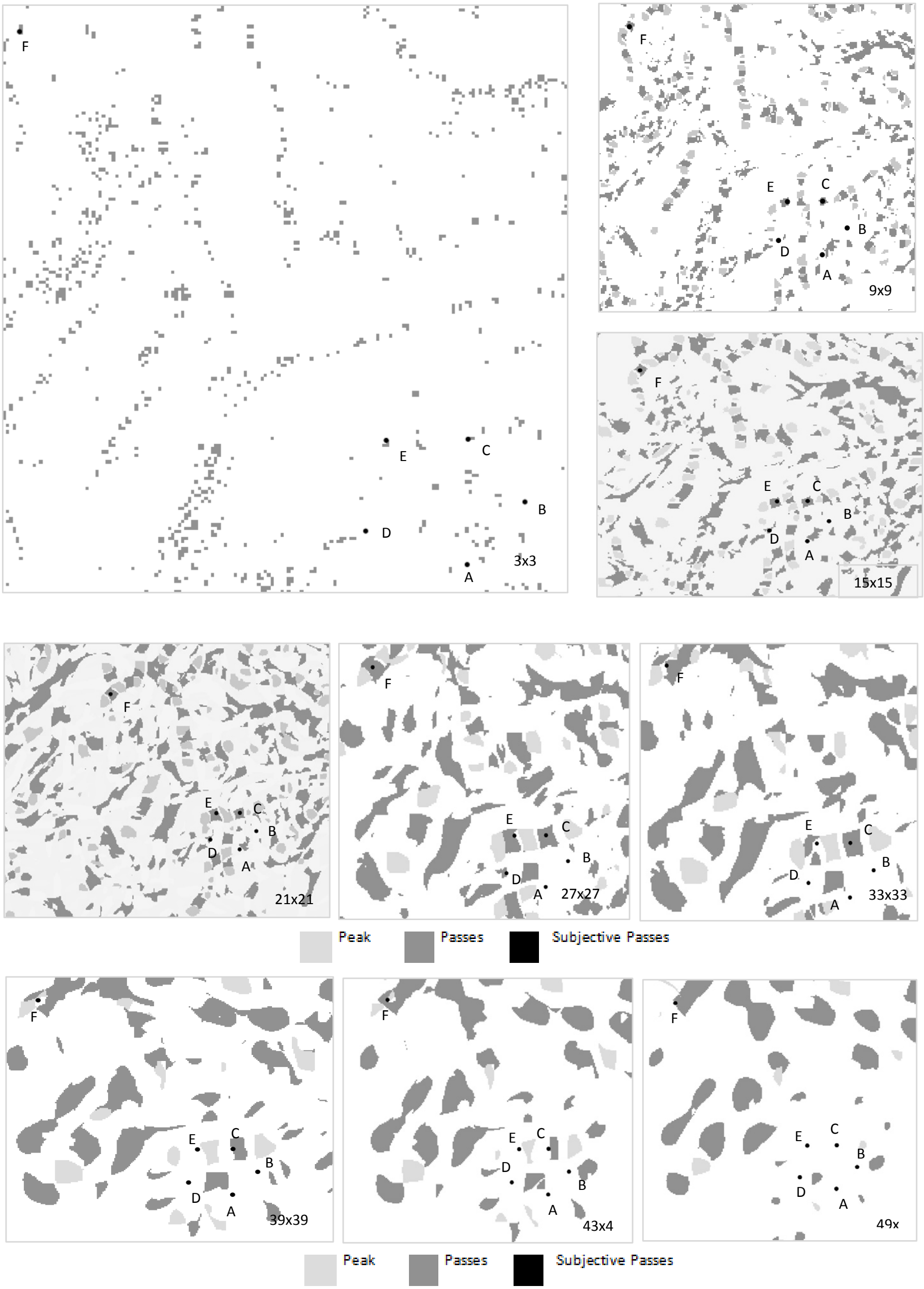


Figure 13: Comparison of subjectively and automatically identified passes

4. Agreement between subjective and algorithmic approaches

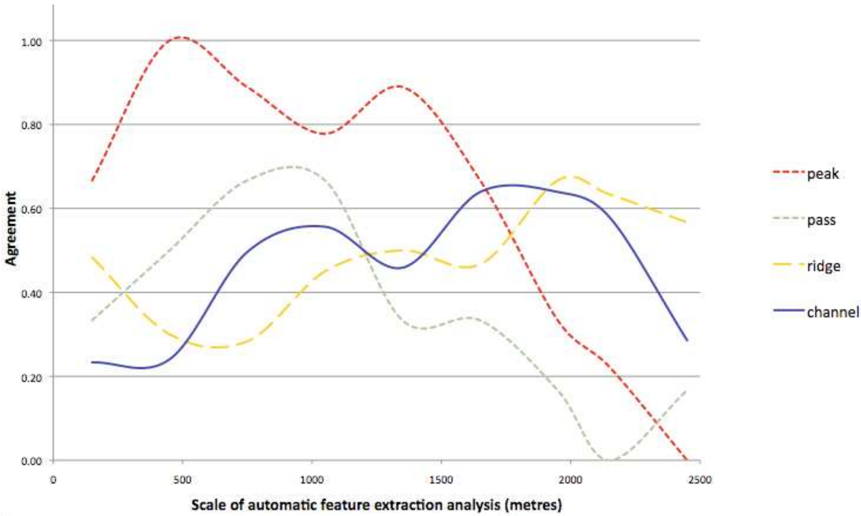


Figure 14: Agreement between subjective and algorithmic feature classification vs scale of analysis

These findings provide some evidence for the scale at which to conduct automatic feature classification, in order to correspond with scale at which human's perceive those features in the landscape.

5. Conclusions

This study indicates that for all features people identify specific classes over an identifiable scale range: features with the same morphology at larger or smaller scales may not be perceived as the same class of feature. For this particular study, using a single observation point, the scale of analysis at which there was greatest agreement between the automated feature class extraction and the human perception is indicated below:

1. Peaks ≈ 500m
2. Passes ≈ 1000m
3. Ridges: ≈ 2000m
4. Channels ≈ 1800m

Further research in this area may advance the objectives of ethnophysiography and provide empirical evidence for geographic ontologies. This research may also provide guidance to scales of analysis for researchers use in automatic terrain feature classification, to coincide with human scales of perception.

References

Greatbatch, I. Wood, J. and Fisher, P. 2007. Using the Google API to compare the cultural and physical prominence of mountain features, in Winstanley AC (Ed) Proceedings of Geographical Information Science Research UK (GISRUK) 2007, 11th-13th April Maynooth Ireland

Mark, D., & Turk, A., 2003, "Ethnophysiography" *Conference on Spatial Informatin Theory (COSIT03) Ellicottville, NY.*

Smith, B, & Mark, D 2001, "Geographical Categories: An Ontological Investigation"

Wood, J., 1996. "The GeomorphologicalCharacterisation of Digital Elevation Models", *Ph.D. thesis, Department of Geography, University of Leicester, Leicester*.

Wood, J (2007) "Modelling Surfaces" Geographic Information Science, City University London

Wood, J. (2008) LandSerf GIS, www.landserf.org

Biography

Delroy Brown undertook this work as part of his MSc in GIS at City University London. He has a background in urban planning and interest in spatial analysis and mobile GIS.

Jo Wood is a senior lecturer in geographic information at the giCentre, City University London with interests in geovisualization and terrain analysis.

David Mountain is a lecturer in geographic information at the giCentre, City University London with interests in GI Science and location-aware computing.

The estimation of the socio-economic impacts of machiya (traditional wooden townhouse) demolitions: A dynamic spatial microsimulation approach

Kazumasa Hanaoka[1]

[1]Ritsumeikan University, 58 Komatsubara-kitamachi, Kitaku, Kyoto, Japan, 603-8341
Tel. +44-(0)75-8801 Fax +44-(0)75-8825
Email kht27176@fc.ritsumei.ac.jp, HP http://www.ritsumei.ac.jp/~kht27176/

KEYWORDS: spatial microsimulation model, simulated annealing, synthetic microdata set, object oriented, Machiya (traditional wooden townhouse)

1. Introduction

Machiya is a traditional wooden townhouse mainly built during the period between the Meiji era (1868-1912) and World War II. These houses are regarded as a core element of the historical landscape of Kyoto City, Japan. However, the number of machiyas has been falling last fifty years. According to the surveys by Yano *et al*. (2006), the number of machiyas decreased from 28,000 in 1998 to 24,000 in 2004 in the central part of Kyoto. Thus, protecting machiya is an urgent policy task.

In order to analyse machiya demolitions, a spatial microsimulation is effective in two ways:

First, individual-based simulation approaches such as microsimulation is useful to understand complex and dynamic socio-economic phenomena (Gilbert and Troitzsch, 1999). For example, Ballas *et al*. (2006) estimates job losses and analyses the related or multiplier effects on socio-economic status of population by using a spatial microsimulation. It is known that machiya demolitions were occurred in relation with (1) deterioration of the exterior conditions of machiyas, (2) demands for high-rise buildings, (3) deterioration of neighbourhood environments for living in a machiya along with the demolitions, and (4) change of generation because most of machiyas are reaching a time of replacement and the structure and interior are less favourable for younger generations (Hashimoto *et al*., 2001; Hanaoka, in press). These factors contribute to machiya demolitions individually while they interact with each other. Therefore, use of spatial microsimulation approach allows us to model behaviours and interactions of each of machiyas, persons and households specifically.

Second, microsimulation can create a synthetic microdata set by combining multiple aggregated datasets (Williamson *et al*., 1998). Since there is no microdata set on machiya residents available, microdata creation allows understanding both their demographic attributes and locations in details.

Therefore, the aim of this article is to construct a dynamic spatial microsimulation model called MachiyaSim to estimate machiya demolitions in the next 15 years in order to provide quantitative information about the number of machiya survived and socio-economic impacts on machiya residents by machiya conservation policy.

2. Specification of MachiyaSim

The datasets used for MachiyaSim are the Population Census of Japan in 2000, the Machiya Survey in 1995-98 and 2003-04. The surveys consist of a machiya exterior survey and the resident surveys. In the exterior survey, a location, type, traditional elements, physical condition and façade condition were examined. In the resident survey, both face data (# of household members, age of household head, and household type) and attitude toward machiya preservation (wants of rebuild, # of repairs in the last 5 years, a type of building after demolition) were surveyed. Samples of the Person-Trip Survey (PT) in 2000 are also used to create a baseline population dataset of machiya residents.

The procedure to construct MachiyaSim consists of four major steps as follows: (1) construction of a synthetic microdata set of households living in machiyas, (2) design of simulation components, (3) model validation, and (4) application of MachiyaSim for evaluating alternatives of machiya conservation policies.

2.1 Construction of a synthetic microdata set

A synthetic microdata set is created by combining multiple tables and an existing microdata set when an appropriate microdata set is not available for analysis. In MachiyaSim, a procedure of constructing a synthetic microdata set consists of three major steps: (a) construction of a synthetic microdata set of households living in Kyoto City, (b) estimation of a household head from household members, and (c) determination of whether they live in a machiya (Figure 1).

First, a simulated annealing which is one of combinatorial optimisation algorithms is used to construct a synthetic microdata set of households. The method allows finding the most fitted set of samples from the PT survey against census table at a small area level. Three census tables: (a) households by family type, (b) households by household size, and (c) population by age group and industry at a small area level are included as a constraint.

Second, a household head is selected from members of a household by using a Monte-Carlo simulation. Conditional probabilities of being a household head for each member are estimated from a census table of households by household head and family type.

Third, to determine whether a household lives in a machiya or not, probabilities of living in machiyas are calculated by age group of household heads. Thereafter, a Monte-Carlo simulation is applied to allocating households to each machiya.

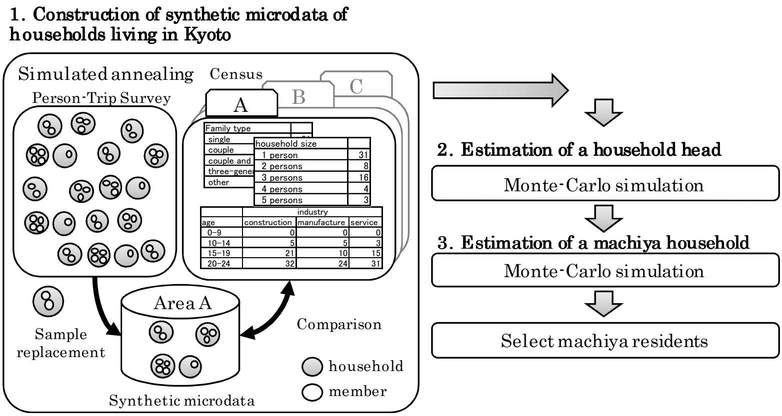


Figure 1. Procedure of constructing a synthetic microdata set

2.2 Design of simulation components

The simulation components of MachiyaSim are constructed by using AnyLogic5.5, which is an integrated development environment for an agent-based modelling developed by XJ Technologies Company LTD. In MachiyaSim, persons, households and machiyas are defined as objects. Properties and methods of the objects are presented along with a relationship among the objects in Figure 2.

Demographic events such as aging, birth, death, leaving home and marriage for a person object, move-out/in and separation of members for a household object are considered. The parameters of the demographic changes are estimated from Vital Statistic in Japan. Transitions of physical and façade conditions of machiyas depend on time since the last repair and/or vacancy. Machiya demolitions are estimated by a logistic regression model which includes a physical condition, traditional façade condition, height zoning, distance from the nearest road, total area of machiyas within 50m radius as independent variables. Total area of machiya within 50m is included to consider neighbourhood effects on machiya demolitions identified in Hanaoka (in press). Details of the simulation components are discussed in Hanaoka (in review).

A graphical user interface is incorporated in MachiyaSim to display simulations in real time (Figure 3). A spatial distribution of machiyas painted according to a physical condition is shown in a map at the center. Bar and line graphs display population transitions and compositions.

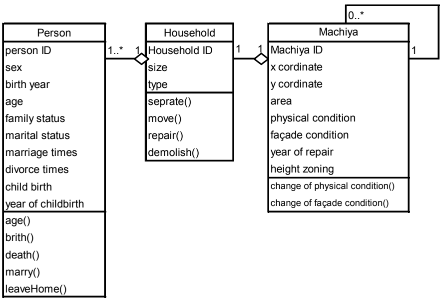


Figure 2. Class diagram of MachiyaSim

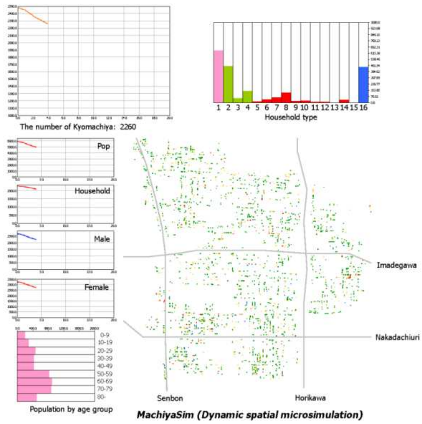


Figure 3. GUI for MachiyaSim

3. Simulations

3.1 Study area

Nishijin district is selected as study area (Figure 4) since computation resources for simulating whole machiya residents are not readily available. Nishijin is one of the historical centers in Kyoto and transformations from machiyas to high-rise accommodation were observed (Fujitsuka, 2004). There were 2,400 machiyas survived in 2004.

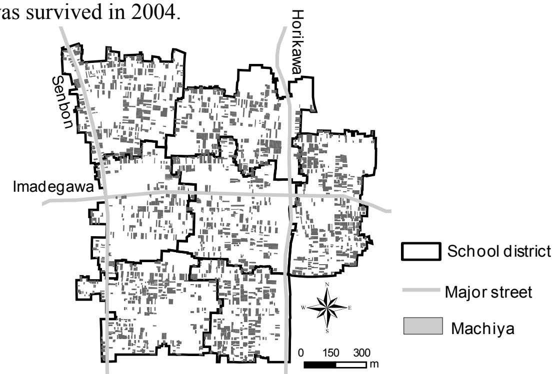


Figure 4. Study area (the spatial distribution of machiyas in 2004)

3.2 Model validation

Using a dataset of machiyas identified in the first survey and the same parameter setting mentioned in Section 2.2, the demolitions of machiyas from 1998 to 2004 are simulated for model validation. The average numbers of survived machiyas by school district are estimated based on 100 simulation runs of MachiyaSim. Table 1 shows average of error ratios are 2.3% and a coefficient of correlation between the observed and estimated frequencies of the machiyas by school district is 0.976. Therefore, MachiyaSim can reproduce a similar spatial pattern as an observed one.

Table 1. The observed and estimated number of the machiyas survived by school district

school district	observed (year 2004) [a]	estimated[b]	error ratio
Karaku	252 (89.0%)	251	0.00
Kenryu	444 (88.8%)	439	-0.01
Ogawa	350 (88.4%)	350	0.00
Seishin	425 (86.7%)	425	0.00
Nishijin	351 (91.9%)	334	-0.05
Touen	295 (81.3%)	316	0.07
Juraku	362 (90.5%)	357	-0.01
			average of |error ratio|
total	2478(88.1%)	2,471	0.02

values in brackets: # of machiyas in 2004 / # of machiyas in 1998 * 100
error ratio = (b - a) / a

3.3 Scenario analysis

MachiyaSim is run 100 times each under different conservation policy settings below based on 2,478 machiyas in 2004 to understand the effects of the policies quantitatively. Parameters for each scenario are estimated from the machiya survey.

A: keeping the status quo
B: increase of repair ratio by a financial support (low ratio)
C: increase of repair ratio by a financial support (high ratio)
D: increase of move-in ratio by a support for searching new residents (low ratio)
E: increase of move-in ratio by a support for searching new residents (high ratio)
F: tighter height zoning
G: combination of C, E and F

Under Scenario A, the current status of demolition ratio is preserved. Under Scenario B and C, more financial supports for repair are introduced. Thus, probabilities of improving machiya exterior conditions within 5 years increase. These probabilities are decided based on a question about willingness to repair in the machiya resident survey. They are 20.2% for Scenario C and 38.9% for Scenario D, respectively. Assumption of Scenario D and E is based on a new policy which supports for searching new residents. These policies increase the probabilities of move-in after a machiya is vacant within 5 years. These parameters are estimated based on the results of machiya resident survey. They are 43.3% for Scenario D and 59.4% for Scenario E. Under Scenario F, the tightest height zoning in status quo applies to whole study area. All parameters of height zoning in a logistic regression model are changed to one of the tightest zoning. Under Scenario G, all assumptions of Scenario C, E and F are simply applied.

Table 2 shows the average numbers of machiyas survived in 100 simulation runs along with ratio of survived machiyas and ratio against Scenario A (the status quo). Under the condition of scenario A. MachiyaSim estimates that, in the next 15 years, 67.3% of machiyas are survived and the number of machiyas decreased from 2478 to 1667. On the contrary, with a financial support for repair, ratios of survived machiyas are increased to 70.1% (Scenario B) and 74.4% (Scenario C) respectively. Further, with a support for searching new residents, the ratios are slightly improved to 67.7% (Scenario D) and 68.1% (Scenario E). Tighter height zoning has also reduced the number of machiya demolitions and the ratio of survived machiyas is 76.1%. Finally, under Scenario G, demolitions of machiyas are the most strongly restricted and the ratio of survived machiyas is 76.1%. Comparing with the number of machiyas (2,478) in 2004, if the comprehensive policy will be implemented, 82.5% of machiyas can be protected on the contrary to 67.3% if the status quo will be preserved.

Table 2. Comparison of simulation results by scenario

scenario	# of machiya	ratio of survived machiya	ratio against scenario A
A: keeping status quo	1,667	67.3%	100.0%
B: increase of repair ratio (low ratio)	1,738	70.1%	104.2%
C: increase of repair ratio (high ratio)	1,845	74.4%	110.7%
D: increase of move-in ratio (low ratio)	1,679	67.7%	100.7%
E: increase of move-in ratio (high ratio)	1,687	68.1%	101.2%
F: tighter height zoning	1,886	76.1%	113.1%
G: combination of C, E and F	2,045	82.5%	122.7%

ratio of survived machiya= # of machiya in 15 years / # of machiya in 2004 * 100

Furthermore, MachiyaSim can provide detailed machiya and population transitions by re-aggregating the output of the microdata set. Figure 5 shows the number of survived machiyas decreased at the same time conditions of machiya and neighbourhood machiya (a ratio of machiyas' area within 50m radius) are changed. The changes of a physical condition are slower under Scenario G compared with Scenario A. Figure 6 displays a population transition under Scenario A. The result shows that aging of machiya resident continues and the population composition is changed drastically in coming 15 years. As such, MachiyaSim can provide useful quantitative information flexibly.

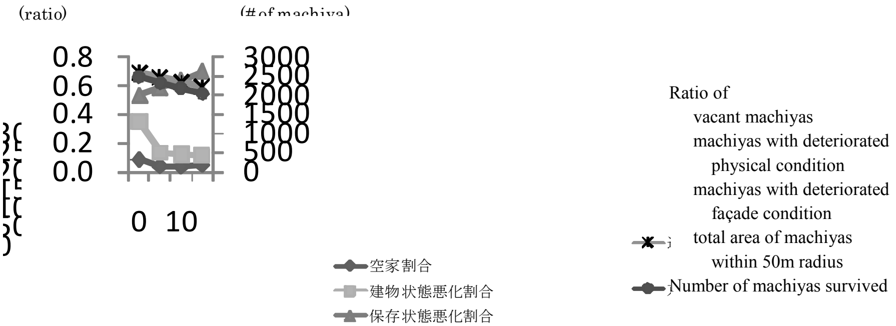


Figure 5. Machiya transition by condition (Scenario A & G)

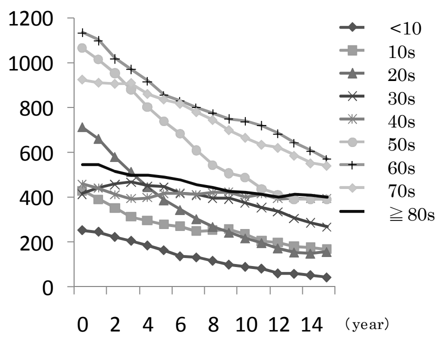


Figure 6. Population change by age group (Scenario A)

4. Conclusions

The results are summarized as follows:

(1) A synthetic microdata set of machiya residents were constructed by combining multiple existing datasets in the manner which the sums of the synthetic microdata set agree with those of the census datasets. The use of the synthetic microdata allows us to analyse detailed household demographics and the process of machiya demolitions at the small area level.

(2) Decision making units such as individuals, households and machiyas can be modeled in the same way they exist and behave in the real world by using object-oriented modelling. A merit of this approach is that objects are able to be updated dynamically in a relation with other agents.

(3) The results of simulations show that, in the next 15 years, only 67.3% of machiyas will be preserved and aging of the residents will continue. On the other hand, when a comprehensive preservation policy is implemented, 82.5% of machiyas can be protected. The application of MachiyaSim is useful to understand the process of machiya transformations. Furthermore, what-if simulations based on machiya preservation policies help evaluating the effectiveness of policies.

(4) The results of simulations such as changes of machiyas and residents are displayed in the forms of maps and charts in real-time via a graphical user interface. Such visual presentations would allow city planners and citizens to understand the results easily so as to encourage them to participate in forming preservation policies of machiyas.

Acknowledgement

I thank to Urban Improvement Section, Planning Division, City Planning Bureau of Kyoto City, Kyomachiya Saisei Kenkyukai, Kyoto Center for Community Collaboration and Kyoto Prefecture for allowing to using their datasets. This study is a part of the Academic Frontier Project (Establishment of academic principles for the protection of cultural heritage and artistic works from natural disasters) and International Training Program supported by Japan Society for the Promotion of Science (JSPS).

References

Ballas D, Clarke P and Dewhurst J (2006) Modelling the Socio-economic Impacts of Major Job Loss or Gain at the Local Level: a Spatial Microsimulation Framework *Spatial Economic Analysis*, **1(1)**, pp127-146

Fujitsuka Y (2004) Gentrification and neighborhood dynamics in Japan: the case of Kyoto. in the *New Urban Colonialism: in Gentrification in a Global Context* edited by R. Atkinson and G. Bridge, Routledge, pp137-150

Gilbert N, Troitzsch K G (1999) *Simulation for the Social Scientist*. Buckingham: Open University Press

Hanaoka K (in press) An analysis on the determinants of Kyomachiya (traditional wooden townhouses) demolitions in Nishijin district, Kyoto City *Geographical review of Japan*.

Hanaoka K (in review) A dynamic microsimulation model for predicting Kyomachiya (traditional wooden townhouse) transformation *Journal of Geography* (in Japanese)

Hashimoto S, Tohiguchi M and Muneta Y (2001) Traditional timber building stock and its characteristics in the central area of Kyoto City *Journal of Architecture and planning* **542** pp183-189

Williamson P, Birkin M and Rees P. H. (1998) The estimation of population microdata by using data from small area statistics and samples of anoymised records. *Environment and Planning A*, **30**, pp785-816.

Yano K, Isoda Y, Nakaya T, Kawasumi T, Matsuoka K, Takase Y, Kawahara D, Inoue M, Tsukamoto A and Kirimura T (2006) Creating virtual time-space of Kyoto *E-Journal GEO* **0(0)** pp12-21

Biography

Kazumasa Hanaoka is a post-doctoral fellow in Ritsumeikan University with interests in a spatial microsimulation modelling for retailing, disaster mitigation and, urban land use change.

Framing the structure of spatial literacy using an empirical method

Jim Nixon, Robert J. Abrahart

Centre for Geospatial Science, University of Nottingham, UK, NG7 2RD
Tel. (044) 115 84 15445 Fax (044) 115 95 15249
jim.nixon@nottingham.ac.uk, www.nottingham.ac.uk/cgs/

KEYWORDS: Spatial Literacy, Nottingham Button Model, Nottingham Sputnik Model, Cluster Analysis, Spatial Skills Test

1. Introduction

Last year at this conference we suggested that the AAG spatial skills test (SST) could form the basis of a useful and robust tool for the assessment of spatial literacy. We suggested a variety of changes, based on student feedback and our own critical analysis that could be made to improve the test. These changes included question re-wording and alterations in graphics or interaction. These changes have now been made and full electronic delivery of the test is possible at the University of Nottingham.

We also proposed the so-called 'Nottingham Sputnik Model' (NSM) to account for the three major areas of ability that could be used in order to infer high or low levels of spatial literacy (Nixon et al., 2008). We proposed that spatial literacy is a latent concept comprised of three broad areas. Three spokes of interest are included: spatial knowledge comprises declarative knowledge or geographical facts. Spatial inference is the ability to interpret and make inferences from spatial data. Finally spatial operations and function can be thought of as the cognitive part of spatial literacy; the ability to manipulate and transform space in the mind. Each group of distinct skills and knowledge are tested in the SST hence their inclusion in our conceptual model. The NSM suggests that spatial literacy is composed of these three 'spokes' which when converged form the raw ability that allows the individual to manage and interpret geospatial data and materials.

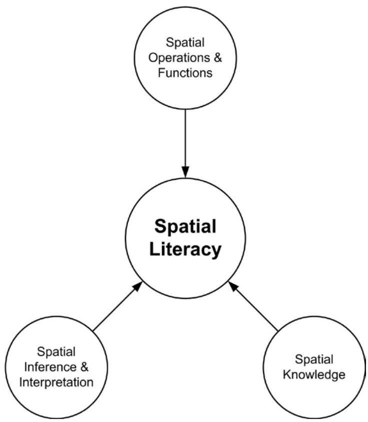


Figure 1 The Nottingham Sputnik Model (NSM)

In this paper, we present an empirical analysis of the SST, categorising questions into groups. Analysis of the patterns of responses in the test uncovered groups of abilities which can be applied to specific spatial problems. Examination of these groups suggests latent parts of spatial literacy which

underlie groups of questions in the test. Similar performance across certain groups of questions may indicate a common spatial skill or theme which can then be separated or investigated further.
As long as the test is viewed as a measure of spatial literacy (see Nixon et al., 2008 for discussion of this point), then responses on these underlying groups could inform teaching or suggest where increased support of skills may be required. At the end of this paper we present a revised model to reflect the groupings of the questions found in the test.

2. Method

2.1 Participants

The test was delivered to forty-two, second-year undergraduates enrolled in an advanced GIS module. All students had studied GIS for a year long module the previous year.

2.2 Materials

The SST was delivered to students using Perception 4.1. The software automatically delivered the questions and scored the responses. Questions were scored correct or incorrect and a total score in addition to the individual question scores were recorded for further analysis. The time taken for completion of the test was recorded. The software randomised the order of item presentation for each individual student to mitigate the possibility of two or more students comparing answers. Students were unable to return to questions that they had answered previously. Questions consisted of the full set from the AAG SST. These included wayfinding questions, visual manipulation questions, abstract stimuli questions, inference questions and analytical questions:

2.3 Wayfinding questions

These questions presented students with a list of operations that were to be followed to move from an origin pint to a destination point on the map (Figure 2). New features developed for the Nottingham version of this test included a larger, more prominent North arrow and removal of the ambiguous description 'block' in the wayfinding instructions.

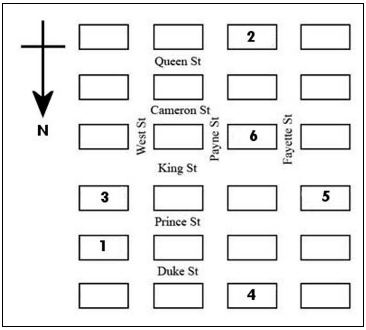


Figure 2 An example of the stimuli used in wayfinding questions

Visual manipulation

These types of questions required students to visualise scenes from different viewpoints (Figure 3). In the example shown, a location on the 3D terrain model must be matched to the correct location on an aerial view of the area. Other questions required judgements formed from visualising contours and matching to slope profiles. Another set of questions required visualisation of a variable across space and their subsequent representation on a two-dimensional graph.

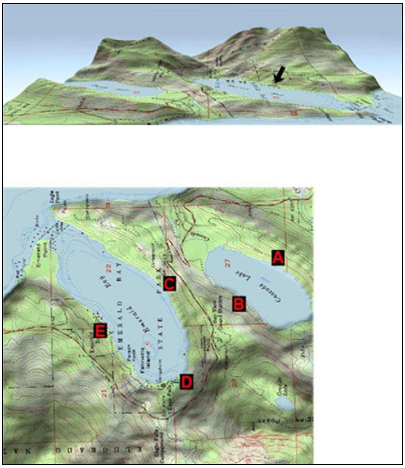


Figure 3 An example of the stimuli used in visual manipulation questions

Abstract Stimuli

The abstract questions required students to manipulate shapes according to a range of logical operators. Questions asked the student to correctly identify the operator (Figure 4) given the result, or to select an answer from a range of results given the operator.

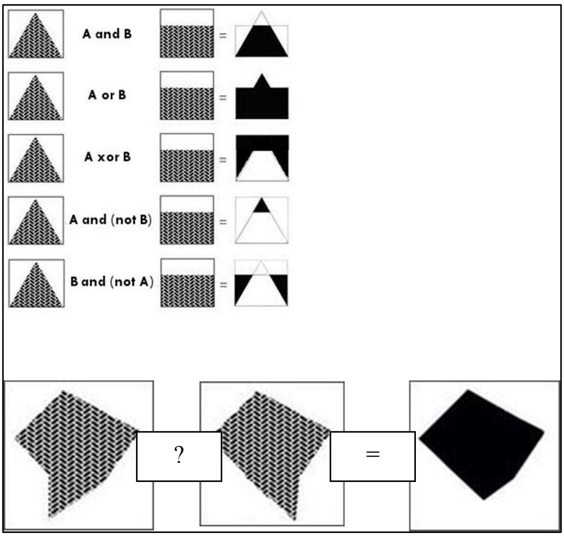


Figure 4 An example of the stimuli used in abstract questions

Inference questions

These types of questions presented spatial information in various forms to students. Students were then required make a judgement based on some or all of the information presented. In the example shown in Figure 5, students had to select the best location for a new coffee shop based on information about the local area and the requirements of the business.

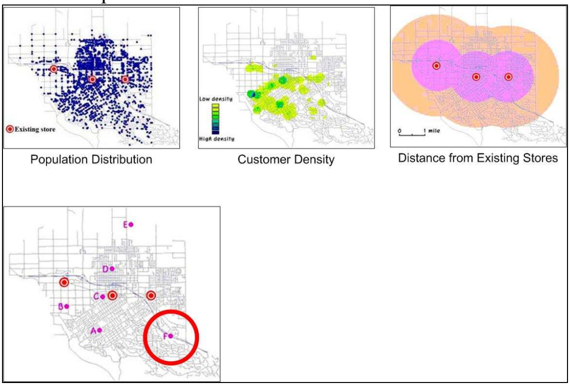


Figure 5 An example of the stimuli used in inference questions

Analytical

These questions all involved correlation of values which had a spatial component (Figure 6). Questions required students to select the map that shared an attribute, positive or negative correlation, with a target map.

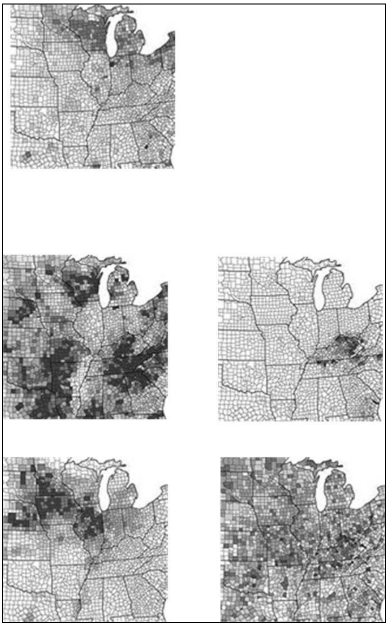


Figure 6 An example of the stimuli used in analytical questions

3. Results and Discussion

3.1 SST Metrics

Analysis indicates that test scores for this group are not normally distributed (Figure 7). The mean test score is 12.4 (SD = 2.3) which equates to an average score of 77%. This high score coupled with the low variance indicates that test may not have been challenging enough for students working at this level. Scores show negative skew indicating that many students did not find the items particularly challenging. This may indicate either an inherently high level of spatial literacy or experience in manipulating and using the skills required of the test due to their training and education in geography. With the notable exception of one student who obtained a near perfect score in a very short time, test timings are normally distributed (Figure 8) with a median time of 822.0 seconds (SD = 179.7 seconds). A weak, but significant negative correlation indicated that students who performed better in the test tended to complete the test in less time than those students with lower test scores ($r_s = -0.31$, $N = 42$, $p < 0.05$) indicating no speed-accuracy trade off in the test results.

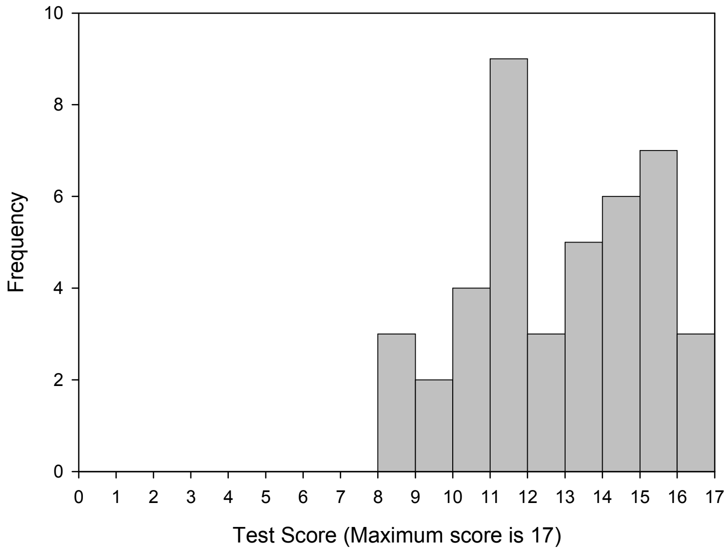


Figure 7 Distribution of test scores across the group

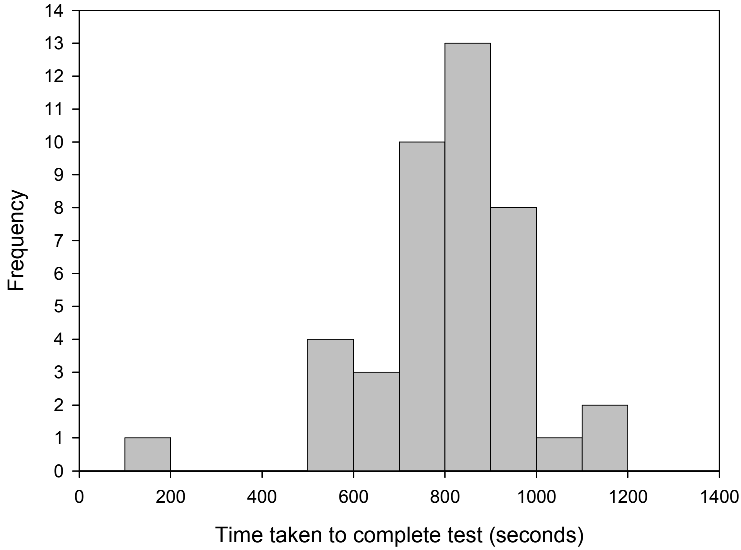


Figure 8 Distribution of test completion times for the group

3.2 Hierarchical Cluster Analysis

Approach

A binary cluster analysis was performed on the test data. Participant scores were of the form 0 for an incorrect response to an item, 1 for a correct response. Cluster analysis proceeds in two major steps. Firstly, a distance measure is specified in order to produce a similarity matrix. This similarity matrix is then entered into a clustering algorithm to determine cluster structure and membership. Many different distance measures and algorithms are available to the researcher. Finch (2005) reviews a variety of distance measures for use with dichotomous data. The Jaccard measure when used with Ward's clustering algorithms delivers reliable cluster structures when analysing dichotomous data and will be used to analyse the results from the SST in this paper. Figure 9 shows the results of the analysis. The partition line divides the questions into three distinct clusters and one outlying question (Q3) which is not associated with any major cluster.

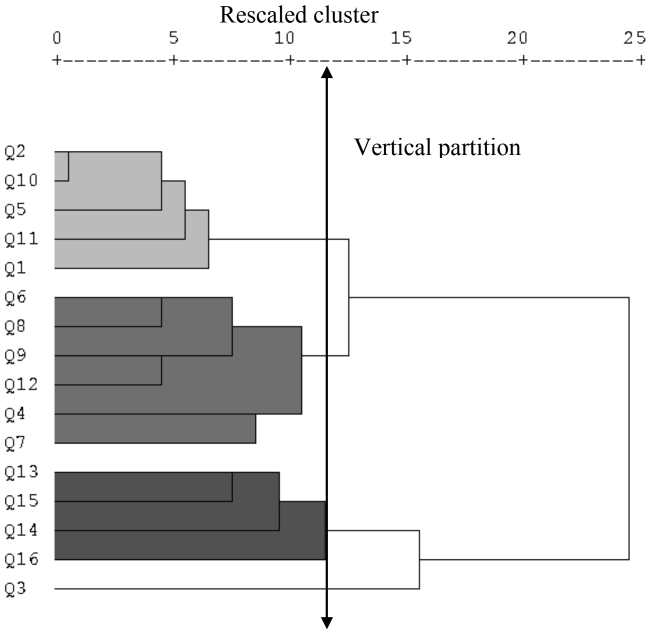


Figure 9 Dendrogram showing cluster membership for each question

Cluster Membership

The first cluster contains questions that do not require a change on viewpoint to answer successfully, both target and question stimuli are presented in the same viewpoint. The wayfinding and analytical questions are members of this cluster (Q1,2,10,11). A second cluster contains the questions which require a change in view-point to answer successfully, the visual manipulation questions (Q6,7,8,9,12). Examination of the third cluster reveals a strong grouping of all questions involving the abstract stimuli (Q13,14,15,16). All Students tended to find these questions more difficult given their lack of familiarity with logical operators applied to spatial problems. As such, a similar pattern of responses has emerged for this group of questions.

The spatial inference questions (Q4, Q5) are split across clusters one and two. At first glance this result is inconsistent since given the NSM, these questions would be expected to form a cluster of their own: spatial inference. This is an example of the joint limitations of binary data and cluster analysis in deriving these groups. The analysis examines patterns of response rather than any qualitative aspect of the question and this should be remembered when interpreting the analysis. Closer examination of the two inference questions shows that the cluster membership is consistent with the matched/un-matched views explanation proposed above. In Q4 (shown in Figure 5), students are required to manage a number of different variables having a spatial component, customer and population density have both position and value. These variables must then be matched to a location. The operation is fundamentally mulitivariate: a number of different or unmatched variables must be considered by the student in order to answer the question. This spatial operation is reflected in Q4's membership of the unmatched views cluster. In Q5 (not shown) students are asked to give the best location for a flood management facility. The constraints in this question are matched with the requirements of the answer and are all spatial in nature. For example, one constraint is 'must be

within 60ft from an electric line'. The answer is completely matched with the information given in order to infer the location. Only one variable is considered: distance.
Q3 is not associated with any pattern of response. One use of this technique is to single out problem questions and further examination of Q3 reveals potential confusion which explains why no clear cluster membership has evolved. Figure 10 shows Q3. The expected membership for this cluster would be cluster 2, unmatched views. The majority of students (54%) proposed the second graph (linear, positive) which was recorded as correct. A large proportion of students (27%) proposed the first graph (curvilinear, positive) as correct.

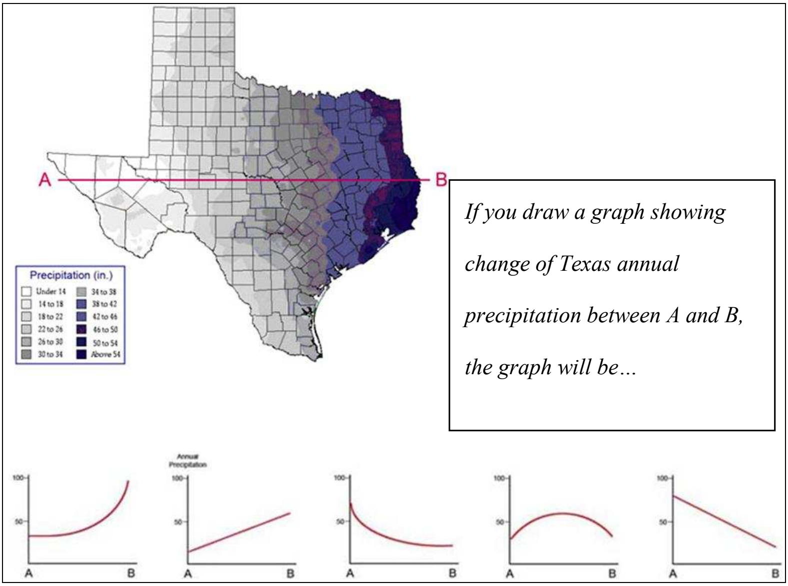


Figure 10 Stimuli and responses to Q3

One explanation for this result may lie in the wording of the question. Indeed, the authors could not agree on whether the question demanded the rate of change in rainfall or the absolute change in rainfall across the area. Certainly in the former case, a curvilinear relationship seems more likely. Given this ambiguity, the lack of a clear response pattern means that membership within a main cluster could not be assigned.

Figure 11 shows percentage success on questions grouped into their clusters. Clearly, all students performed well on all questions. The best performance is when response-stimuli views or variable are matched, below average performance is associated with the abstract stimuli.

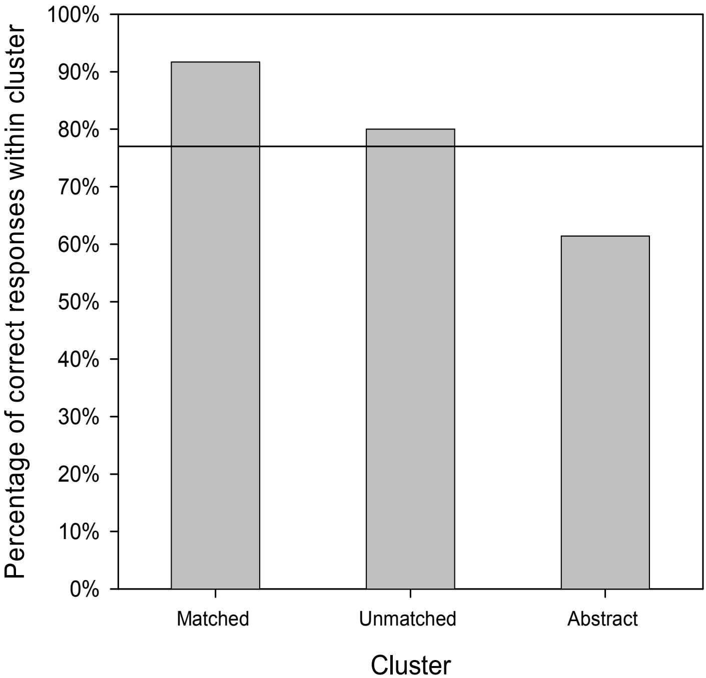


Figure 11 Percentage of correctly answered questions summed for each cluster (reference line shows average overall test score)

4. Conclusion

This basic analysis of the SST has indicated that three distinct groups of questions are present in the test: matched, unmatched and abstract. Students tended to find the unmatched questions more challenging but all students found the abstract questions the most challenging of all. Of particular interest was the grouping of the inference questions. In the NSM model, these questions would be grouped under the spatial inference spoke. The data suggests that the questions are distinguished by the type of operation required for a correct answer. A direct link between inference and spatial operation and function is implied but absent in the original NSM.

The difficulty that students experienced when answering the abstract questions may indicate a lack of experience using logical operators or an absence of context increasing the difficulty of the question. Fortunately, GIS packages are very good at performing large numbers of these computational problems leaving our students free to concentrate on the more interesting and often important aspects of GIS; deciding on the type of analysis to perform to solve spatial problems and interpreting the results.

Given these results, our original model may have been too rigid, excluding the independent information flows between the various 'spokes'. Separating inference from operations and knowledge is not supported by the data presented in this paper. In light of these results we propose a revised model of spatial literacy, the 'Nottingham Button Model' (NBM) shown in Figure 12. In this model information flows seamlessly between the three functions proposed in the original model. Spatial literacy is the over-arching concept that connects all of the essential function together. Our results suggest that inference, interpretation, operation and function all interact and that the process of interaction in order to generate an answer to a spatial question is mediated by spatial literacy. Those students with higher spatial literacy can engage these different processes and abilities more easily.

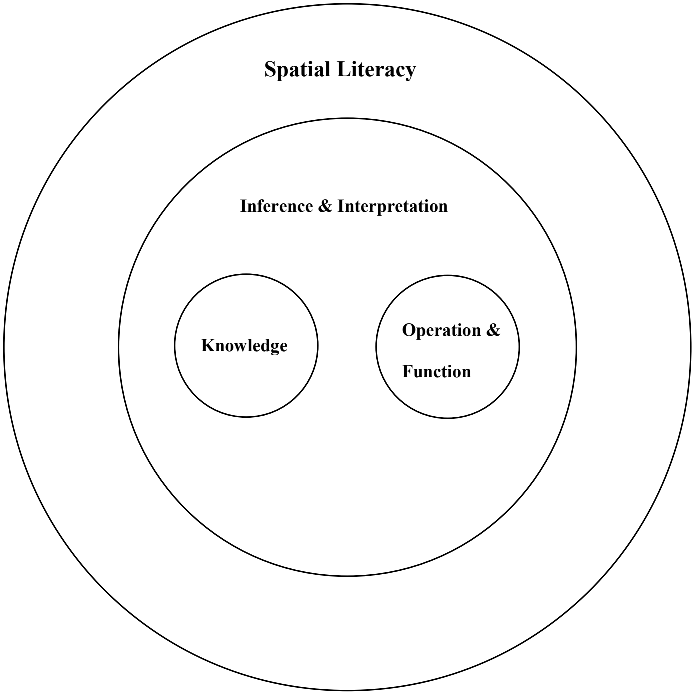


Figure 12 The Nottingham Button Model (NBM)

This short study demonstrates the potential power of testing in order to generate data about a concept that is difficult to define and/or measure like spatial literacy. Of course the technique and results are bound by the quality and types of questions asked. A wide variety of questions are included in the spatial skills test and we hope that at the very least, this empirical examination will inspire others to enlarge and define the concept of spatial literacy and examine its impact on students in higher education.

Acknowledgements

Extra special thanks to all the second-year GIS students who participated in this study. Thanks also to the model naming committee: Iain Cross, Oliver Dunnet, Lauren Gough, and especially to Lucy Veale who made the winning suggestion.

References

Finch, F. (2005) Comparison of distance measures in cluster analysis with dichotomous data. Journal of Data Science, **3**(1), pp85 - 100.

Nixon, J., Abrahart, R. J. & Chambers, C. (2008) In Search of Spatial Literacy: The

Nottingham Sputnik Model. IN LAMBRICK, D. (Ed.) Proceedings of Geographical Information Science UK (GISRUK) 2008, pp94-102. 2nd - 4th April Manchester Manchester Metropolitan University

Biography

Jim Nixon has recently been awarded his PhD in human factors and spatial representation. He is currently working as a chartered psychologist at National Air Traffic Services Ltd (NATS).

DEM creation for delineation of lahar inundation hazard zones

Amii Darnell[1], Andrew Lovett[1], Jenni Barclay[1] and Richard Herd[1]

[1]School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ UK
Tel. +44 (0)1603 591346, Fax. +44 (0)1603 591327
a.darnell@uea.ac.uk

KEYWORDS: Digital Elevation Models, data fusion, hazard mapping, fitness for use

1. Introduction

A lahar is a rapidly moving mixture of volcanic rock debris and water. The raster data structure in a GIS permits neighbourhood calculations that can simplify lahar movement from a volcano over a digital elevation model (DEM). However, the DEM must be carefully constructed for use in volcanic hazard assessment, where the adverse consequences of using unfit data can be particularly extreme. LAHARZ, implemented in a GIS, is a common method for delineation of lahar hazard zones (Schilling, 1998). This paper focuses on DEM preparation for LAHARZ and provides an approach for long-term hazard mapping.

2. Study area

The Soufriere Hills Volcano, Montserrat (West Indies) has been active for over 13 years, intermittently ejecting ash and other volcanic debris. The tropical environment, with its frequent heavy rainfall, continually mobilises this material into lahars. The only major drainage that could channel lahars from the volcano to inhabited areas is the Belham River Valley (Figure 1).

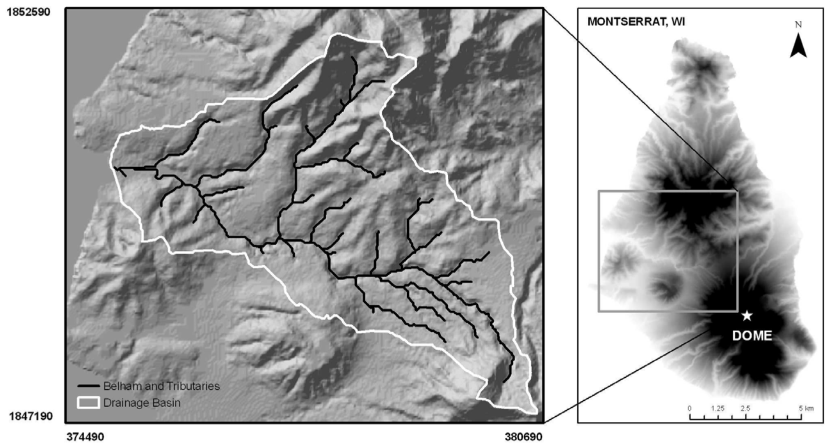
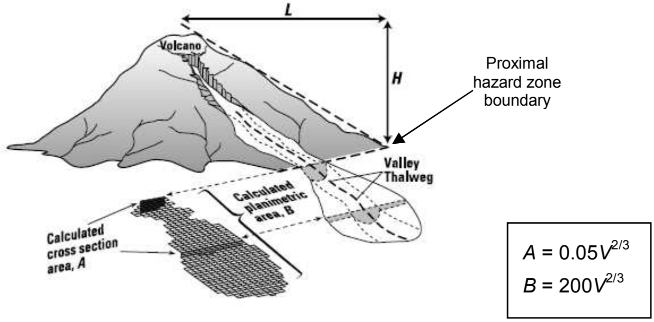


Figure 1. Montserrat and Belham Valley study area
(Montserrat National Grid)

3. LAHARZ

The distal inundation area for a lahar can be estimated from the volume of material leaving the proximal hazard zone. The proximal hazard zone boundary marks the transition from erosive to depositional behaviour and is defined by the geometric relationship between the horizontal runout (L) and vertical descent (H) (Figure 2). Iverson et al (1998) developed semi-empirical equations that were used to predict the valley cross-sectional area (A) and planimetric area (B) inundated by a lahar with volume, V, leaving the proximal hazard zone. These equations approximate lahar movement to a waveform, with constant velocity and volume. LAHARZ user input consists of a DEM, derived supplementary grids and a specified H/L value (to define the start cell).



$$A = 0.05V^{2/3}$$
$$B = 200V^{2/3}$$

Figure 2. An idealised lahar path and geometric relationships between H and L, which describe the extent of the proximal hazard zone, and A and B, which describe the extent of the distal lahar inundation zone (After Iverson et al, 1998).

4. Updating the digital elevation model

The extent of a DEM used by LAHARZ must encompass the volcano and the downstream extent of the drainage (Schilling, 1998). Regional data were acquired as a 10 m resolution island-wide DEM, originally derived from contour lines. A study area was extracted that included the energy cone apex (volcano summit) and the Belham drainage basin (Figure 1). However, the dynamic nature of the volcanic system induces geomorphological changes to the valley. These adjustments are discernible on an annual/ semi-annual scale, rendering digital representations of the valley rapidly outdated. Additional elevation data were gathered in the field through a roving Global Positioning System (GPS). Coverage of most of the valley floor was attained, although data were irregularly distributed and didn't encompass steep escarpments and valley sides (where satellite communication was lost) (Figure 3a). Only points resolved to an accuracy of 0.1 m were retained for further analysis.

GPS points were interpolated as mass points to create a Triangular Irregular Network (TIN), and the Belham thalweg was imbedded in the TIN fabric as a hard breakline. The TIN was then converted to raster format for the ease of neighbourhood calculations, common in the derivation of hydrologic parameters. Fitness of this DEM for hydrological analyses was investigated using a steepest decent flow model. This application-driven approach was used to reconstruct the DEM more reliably, testing uncertainty in the output flow paths by individually varying model inputs (e.g. resolution, magnitude and spatial structure of errors) (Darnell et al, In Review). As a consequence it was essential that the DEM be further augmented with vegetation information digitised from aerial photographs and field notes.

4.1 Data fusion

Two overlapping regular gridded matrices of elevation values were thus available. First, the severity of differences were considered e.g. spatial resolution, vertical datum, horizontal and vertical shifts, production errors, type of terrain model (surface model or bare earth etc.) and spatial distribution of errors (Reuter et al 2007). Comparison of the overlapping cells revealed the GPS-derived DEM generally gave a greater elevation (mean change over time = + 4.16 m), however the standard deviation was high due to great spatial variability (*sd* = 4.82) and some cells experienced an apparent elevation decrease. It was concluded recorded changes were greater than expected due to error and resolution effects alone. Furthermore, photographic evidence supported channel change due to erosion and deposition, especially in areas of extreme calculated change. Due to the temporal difference of the datasets (on a decadal scale) it was not plausible to amalgamate all points, and an exploratory analysis of a simple ‘find and replace’ algorithm for overlapping cells created sinks at the valley sides. Two techniques were tested further:

1) Void, fill and feather. The ‘fill and feather’ method, an interpolation method common for correcting voids in Shuttle Radar Topography Mission data (e.g. Reuter et al, 2007), was adapted to smooth the transition boundary of the two datasets. The DEMs were first superimposed and where cells overlapped the elevation values of the contour-derived DEM were replaced (filled) with those from the GPS-derived DEM. An artificial void was then created by extruding a buffer zone external to the GPS-derived DEM. The buffer width was determined by distance between the break in slope (contour-derived DEM) and the closest GPS point. These were important features needed to maintain integrity of the surface. Finally, a low pass 5 × 5 filter (moving average window) was applied to diffuse, or feather, any abrupt change.

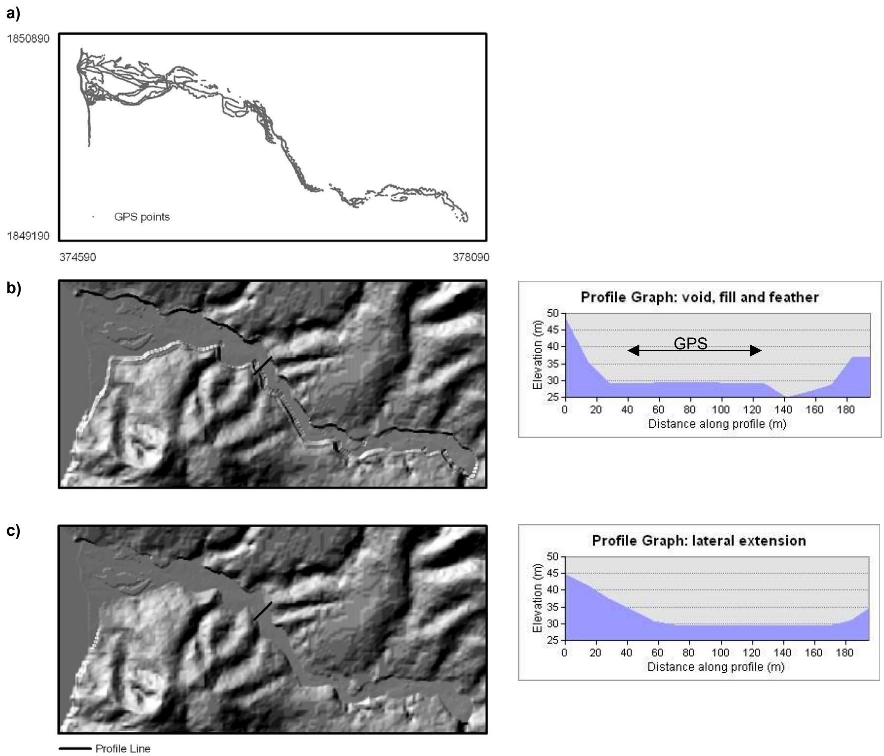


Figure 3. a) distribution of GPS points, and planimetric view of the merged section of the study area for b) void, fill and feather, and c) lateral extension method. Profile graphs show the resultant cross-sections; coverage by the GPS-derived DEM is also indicated.

2) Lateral extension of recent elevation data. The second approach also prioritised conformity to the GPS-derived topography. For this method, the GPS-derived DEM was extended laterally by a neighbourhood average elevation. A moving window computed elevation values in the NoData pixels based on the local average of the neighbouring elevation values (5 × 5 pixel window). A number of iterations were performed until the dimensions of the valley floor were surpassed (as determined from a mask digitised using aerial photos). A differential surface was then created by subtracting the contour-derived DEM. Negative values were then reclassified to zero and the result added to the contour-derived DEM, generating a DEM showing only positive elevation change (inferring deposition). To allow negative elevation change within the channel, only negative values from the differential surface were extracted using the valley floor mask. Finally, this erosive surface was added to the deposition-only DEM.

4.2 Discussion and fitness of fused surface

The contour-derived DEM was unsuitable for predicting future flows as it was outdated. Preliminary analyses suggested the GPS-derived DEM was fit for use in this context (Darnell et al, In Review) but this DEM didn't cover the extent of the study area. Differences between the two fused methods are visible in Figure 3. The fill and feather process corrupted the contour-derived DEM, creating artificial slopes in the data that neither conformed to the original surface nor represented the true ground (a common problem with this technique (Grohman et al, 2006)). Furthermore, the topographic variance of the two surfaces was too high to match the surfaces over such a distance (Figure 3b). Various buffers and filter neighbourhoods were trialled, resulting in minimal or no improvement. An improvement on the 'fill and feather' method was perceptible in the fused surface from the lateral extension method (Figure 3c). Profile graphs show the same channel cross-section as viewed by the different methods. The difference in slope is clearly visible and this will affect flow direction.

5. Hazard zone delineation

The distal area inundated by a single flow event can be output from LAHARZ in the form of a binary grid. Changing volume magnitude and overlaying resultant binary grids can provide information on likely hazard zones (e.g. Figure 4). However, uncertainty in model inputs will propagate to the model output and thus error should be incorporated in a probabilistic assessment. Furthermore, successive lahars will change the morphology of the underlying topography, altering future flow paths and inundation areas. Evolving terrain should be considered for mid- to long-term hazard management. Change in elevation cannot be extracted directly from LAHARZ, but altering the AML code enables maximum inundation height to be obtained per cross-section. Deposition can be elicited from maximum inundation depth if deposit thickness correlates with flow depth for maximum discharge. This new approach requires validation of simplifying assumptions and preliminary results are being tested.

6. Summary

Data fitness is especially relevant when modelling hazards. This paper has demonstrated an original approach for creating and updating DEMs. Modelling geomorphologic change for long-term hazard mapping is in progress.

7. Acknowledgements

The first author is in receipt of an ESRC/NERC studentship PTA-036-2006-00016. We would like to acknowledge the help of the Montserrat Volcano Observatory together with the input of Horatio Tuitt (Director of the Disaster Management Coordination Agency, Montserrat).

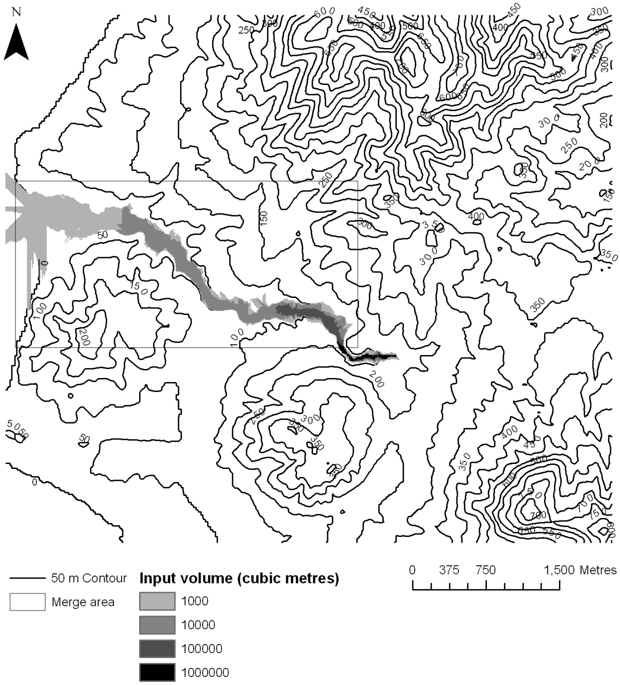


Figure 4. Inundation areas corresponding to different input volumes run over the DEM merged by lateral extension.
(The merge area has been provided for orientation with Figure 3).

References

Darnell AR, Lovett AA, Barclay J and Herd RA (In Review) An application-driven approach to terrain model construction. *International Journal of Geographical Information Science*

Grohman G, Kroenung G and Strebeck J (2006) Filling SRTM voids: the delta surface fill method. *Photogrammetric Engineering & Remote Sensing* March pp213-216

Iverson RM, Schilling SP, Vallance JW (1998) Objective delineation of lahar inundation hazard zones. *Geological Society of America Bulletin* **110** pp972-984.

Reuter HI, Nelson A and Jarvis A (2007) An evaluation of void-filling interpolation methods for SRTM data. *International Journal of Geographical Information Science* **21(9)** pp983-1008

Schilling SP (1998) LAHARZ: GIS programs for automated mapping of lahar-inundation hazard zones *U.S. Geological Survey, Open File Report* pp93-638.

Biography

The first author is a PhD student researching the use of GIS for volcanic hazard assessment. She is currently in her third year of research under the supervision of Andrew Lovett, Jenni Barclay and Richard Herd.

Developing Topographic Descriptors to study Orographic Processes under a Changing Climate

Emma Ferranti[1], Duncan Whyatt[1], Roger Timmis[2]

[1]Department of Geography, Lancaster University, Lancaster. LA1 4YQ
Tel. 44 1524 510246 Fax 44 1524 510269
e.ferranti@lancs.ac.uk
[2]Environment Agency, Lancaster University, Lancaster. LA1 4YQ

KEYWORDS: precipitation, rainshadow, uplands, Cumbria

1. Introduction

The seasonality and characteristics of rainfall across the UK are altering as a result of climate change (Jenkins *et al.*, 2007). The amount and intensity of rainfall is generally decreasing in summer but increasing in winter, particularly in northern and western areas (Maraun *et al.*, 2008: Osborn *et al.*, 2000). Recent analyses suggest the greatest changes in winter rainfall are occurring in upland regions and adjacent rainshadow zones (Burt & Ferranti, *[in review]:* Malby *et al.*, 2007). Understanding spatial and temporal changes in rainfall in such regions is important for the future management of water resources and ecosystems (Fowler *et al.*, 2007: Orr *et al.*, 2008). The current generation of regional climate models simulate observed precipitation poorly in such areas of complex topography (Fowler *et al.*, 2005), casting doubt over future predictions. Furthermore, the mechanisms controlling rainfall processes in leeside (rainshadow) zones are poorly understood (Roe, 2005).

This paper presents a novel approach to analysing rainfall patterns and processes under a changing climate. A database of meteorological and topographical parameters was compiled for Cumbria to enable analysis of rainfall under different conditions. Cumbria was selected as an example of a topographically diverse mid-latitude region that has a predominately maritime and westerly defined climate. The region has a dense network of rain gauges and the database contains monthly, daily, hourly and sub-hourly rainfall records from more than 400 stations that operational for a period between 1900 and present day (BADC, 2008) (Figure 1). Other local meteorological observations such as temperature, wind speed and wind direction are included, together with Lamb-Jenkinson Weather Types (LJWT) that provide a daily snapshot of air mass origins and the synoptic weather conditions over the British Isles (Jones *et al.*, 1993). Used in combination, these variables allow rainfall patterns observed in Cumbria to be 'typed' by air mass origin and/or local meteorology to enable greater understanding of the processes delivering rainfall (Figure 2). Although previous studies have produced precipitation maps and climatologies from observational data (e.g. Frei & Schar, 1998: Perry & Hollis, 2005) none have used directional analysis to understand the rainfall patterns and processes associated with different synoptic conditions, nor reviewed how these patterns and processes are affected by a changing climate.

Rainfall at a location depends not only on synoptic and local meteorology but also on the surrounding orography (Salter, 1918) and distance to sea (Hobbs *et al.*, 1975). As such the database also incorporates a series of topographic descriptors unique to each station defined using GIS techniques (Figure 2). Currently under development, these descriptors will inform the relationship between rainfall and topographic setting and help understand how this may have changed over time. Although previous studies have used GIS techniques to develop similar topographic descriptors (e.g. Marquínez *et al.*, 2003), none have used the descriptors to gain a greater understanding of orographic processes or applied these in a climate change context.

This paper presents a series of topographic descriptors developed using the GIS software ArcGIS and gives a preliminary analysis into the relationship between the descriptors and rainfall. An example of database application is given, and the direction of future research summarised.

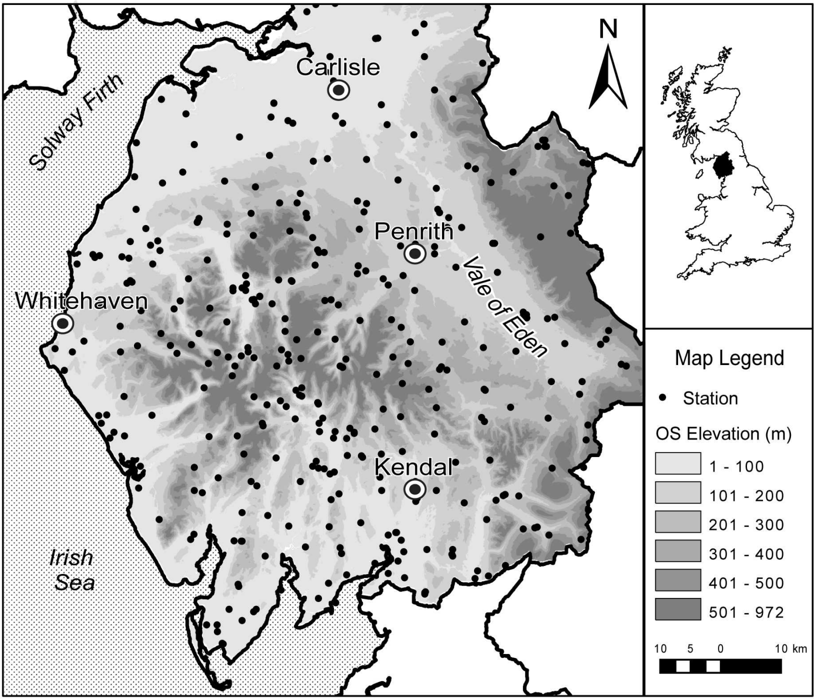


Figure 1. Data from over 400 rainfall Stations in Cumbria are stored within the database

2. Topographic Descriptors

Topographic descriptors can be used to quantify the relationship between local meteorological conditions (e.g. precipitation, temperature, wind strength) and the orographic setting of the station. Understanding this relationship will provide greater insight into the mechanisms controlling rainfall in areas of complex topography, and how these are affected by a changing climate. A better understanding of the topographic controls on rainfall will help inform the climate modelling process, and ultimately lead to better simulation of rainfall in areas of complex topography. Given that orographic setting and distance to sea have greatest influence on rainfall (Marquínez *et al.*, 2003) the following topographic descriptors have been defined using ArcGIS for each of the 400+ weather stations:

Station Elevation: the elevation of the station as given by 50 m resolution Ordnance Survey Panorama data.
Zonal Mean Elevation and *Zonal Max Elevation*: the mean elevation and max elevation within a specified buffer (1-10 km) derived to characterise the topographic setting.
Distance to Coast: the distance from each station to the nearest stretch of coast as calculated using a Euclidian distance function to investigate the influence of open water on rainfall.
Distance to West (southwest): the distance to coast along a westerly (south-westerly) ray bearing 225° (270°). Each ray was clipped to the convex hull of the Cumbrian coastline to minimise the impact of minor bays. This provides an estimate of orographic processing of an air parcel prior to arrival at the

station and gives the distance to the moisture source for W or SW air masses.
Absolute Elevation Change: the absolute change in elevation along a westerly (south-westerly) trajectory from station to coast was summed to estimate the prior orographic processing of the air mass.
Aspect: the down-slope direction of maximum change in elevation was calculated at different resolutions (50 m – 1 km) to determine the windward/leeward position of the site.
Slope: of the pixel containing the station was derived at different resolutions (50 m – 1 km) to typify the surrounding topography (e.g. flat or undulating).

These topographic descriptors will also be used to classify stations into categories such as *coastal*, *lowland* or *upland* and *windward* or *leeward* position of each station. As previous studies (Burt & Ferranti, *[in review]:* Malby *et al.*, 2007) suggest upland and leeward regions are undergoing the greatest increase in winter rainfall amount and intensity, defining these positional characteristics and investigating how their relationship with rainfall changes under different weather conditions over time is paramount to the success of this study.

Methodology

1. Select the synoptic conditions under study from the database, *e.g. weather originating from SW.*
2. Select the local meteorological conditions under study, *e.g. rainfall, temperature at a station.*
3. Analyse how rainfall patterns have changed in time for one or more sites, *e.g. has the number of rainfall days associated with SW weather increased 1959-2004?*
4. Use topographic descriptors to establish which areas have seen change and to further understand rainfall processes, *e.g. where has the rainfall associated with SW weather increased 1959-2004?*

This analysis will provide greater understanding of the relationship between rainfall and topography under a changing climate in mid-latitude uplands.

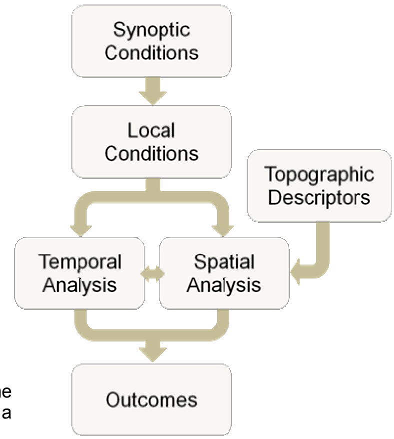


Figure 2. Meteorological observations and topographic descriptors stored within the database can be used to understand rainfall in areas of complex topography under a changing climate

3. Preliminary Analysis of Topographic Descriptors

This section presents a preliminary investigation into the relationship between rainfall and 3 topographic descriptors that describe elevation (*Station Elevation*, *Zonal Mean Elevation*, *Zonal Max Elevation*). The descriptor with the strongest relationship can be used to classify the stations as upland or lowland, using a boundary of 200m (after Minder *et al.*, 2008). The descriptors were compared with daily precipitation associated with winter air masses of SW origin (calculated as a 45-year mean); these systems are frequent and wet (Kelly *et al.*, 1997), and hence were selected to give a strong signal. Daily averages were selected in preference to annual totals to allow data from all weather stations, regardless of their measurement duration to be included.

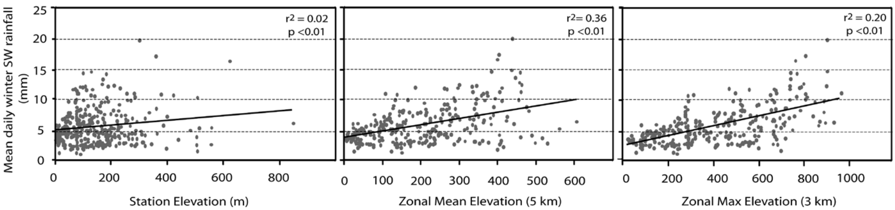


Figure 3. The relationship between mean daily winter rainfall and (a) Station Elevation, (b) Zonal Mean Elevation (5km) and (c) Zonal Max Elevation (3km).

Figure 3 shows the relationship between *Station Elevation*, *Zonal Mean Elevation (5 km)*, *Zonal Max Elevation (3 km)* and rainfall. Testing revealed these radii had the strongest relationship with rainfall for the particular topographic descriptor. Preliminarily analysis suggests that *Zonal Max Elevation (3 km)* has the strongest relationship with mean daily winter rainfall ($r^2 = 0.36$; $p<0.01$) and is therefore the best descriptor to classify the weather stations as lowland or upland.

4. Sample Application: change in SW winter rainfall

Zonal Max Elevation (3 km) was used to classify the weather stations as lowland or upland using a boundary of 200m. Figure 4 shows the average winter rainfall associated with SW air masses for nine 5-year periods between 1960 and 2004 for lowland (<200 m) and upland (> 200 m) stations. Although not significant at 5-year resolution, the data suggests upland winter rainfall associated with SW air masses has increased relative to lowland stations. Moreover, there is an increasing divergence between the amount of rainfall experienced in upland and lowland areas ($r^2 = 0.71$; $p<0.01$).

Rainfall for these 5-year periods can also be plotted in ArcGIS by interpolating between stations and Figure 5 shows mean daily winter rainfall maps for 3 of the 5 year periods shown in Figure 4. In agreement with Figure 4 the central uplands appear to have experienced a relatively greater increase in rainfall compared to lowland and coastal regions. The rainshadow spanning the Eden Valley appears to have weakened in recent years, changing the pattern of rainfall, and suggesting a relatively greater increase in rainfall has also taken place on leeside slopes. These findings are consistent with (Malby *et al.*, 2007) and this changing ratio of windward/leeward rainfall will be investigated further using the topographic descriptors defined above.

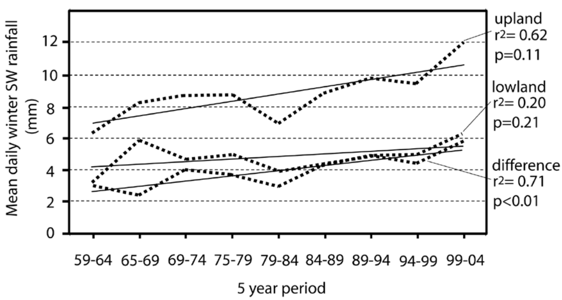


Figure 4. Upland rainfall has increased relative to lowland rainfall 1959-2004.

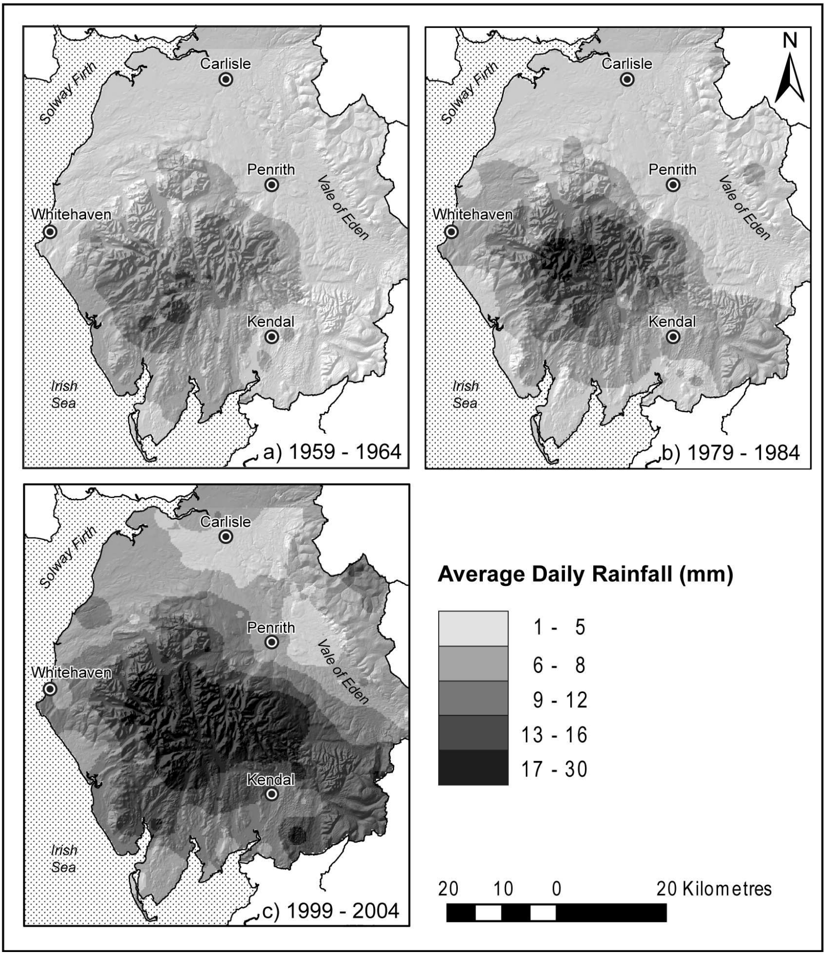


Figure 5. The average winter daily rainfall associated with SW weather types for (a) 1959-1964, (b) 1979-1984, and (c) 1999-2004.

5. Conclusions

A high resolution database of meteorological observations has been constructed for Cumbria. Synoptic and local site characteristics allow the rainfall associated with specific conditions (e.g. SW winter winds) to be defined and analysed over different time periods. In addition, a series of topographic descriptors are being developed to provide greater understanding of the relationship between rainfall and topography. The changing pattern of SW winter rainfall is presented in this paper to show how the database can be applied to provide greater understanding of changing rainfall patterns and processes. Topographic descriptors have been used to establish the regions undergoing greatest change although the full benefits of the descriptors have yet to be fully exploited. It is hoped that this novel approach will provide greater insight into the changing mechanisms of upland rainfall and will help inform the climate modelling process. This will ultimately lead to better simulation of rainfall in areas of complex topography, and provide stronger support for future water and ecosystem management decisions.

6. Acknowledgements
This worked was undertaken as part of a NERC studentship NE/F007434/1 supported in part by the Environment Agency for England and Wales. All rainfall data were provided by the British Atmospheric Data Centre and are available online at http://badc.nerc.ac.uk/data/ukmo-midas.

References
BADC, 2008: UK Meteorological Office. MIDAS Land Surface Stations data (1853-current). [Available online from http://badc.nerc.ac.uk/data/ukmo-midas.].
Burt, T. and E. Ferranti, [in review]: Changing patterns of heavy rainfall in upland areas: a case study from northern England.
Fowler, H. J., C. G. Kilsby, and J. Stunell, 2007: Modelling the impacts of projected future climate change on water resources in north-west England. Hydrology and Earth System Sciences, **11,** 1115-1124.
Fowler, H. J., M. Ekström, C. G. Kilsby, and P. D. Jones, 2005: New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. Assessment of control climate. Journal of Hydrology, **300,** 212-233.
Frei, C. and C. Schar, 1998: A precipitation climatology of the Alps from high-resolution rain-gauge observations. International Journal of Climatology, **18,** 873-900.
Hobbs, P. V., R. A. Houze, and T. J. Matejka, 1975: Dynamical and Microphysical Structure of an Occluded Frontal System and Its Modification by Orography. Journal of the Atmospheric Sciences, **32,** 1542-1562.
Jenkins, G. J., M. C. Perry, and M. J. Prior, 2007: The climate of the United Kingdom and recent trends. Met Office Hadley Centre, Exeter, UK.
Kelly, P. M., P. D. Jones, and K. Briffa, 1997: Classifying the winds and weather. Climates of the British Isles: Present, Past and Future, M. Hulme and E. Barrow, Eds., Routledge, 454.
Malby, A. R., J. D. Whyatt, R. J. Timmis, R. L. Wilby, and H. G. Orr, 2007: Long-term variations in orographic rainfall: analysis and implications for upland catchments. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, **52,** 276-291.
Maraun, D., T. J. Osborn, and N. P. Gillett, 2008: United Kingdom daily precipitation intensity: improved early data, error estimates and an update from 2000 to 2006. International Journal of Climatology, **28,** 833-842.
Marquínez, J., J. Lastra, and P. García, 2003: Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis. Journal of Hydrology, **270,** 1-11.
Minder, J. R., D. R. Durran, G. H. Roe, and A. M. Anders, 2008: The climatology of small-scale orographic precipitation over the Olympic Mountains: Patterns and processes. Quarterly Journal of the Royal Meteorological Society, **134,** 817-839.
Orr, H. G., R. L. Wilby, M. McKenzie Hedger, and I. Brown, 2008: Climate change in the uplands: a UK perspective on safeguarding regulatory ecosystem services. Climate Research, **37,** 77-98.
Osborn, T. J., M. Hulme, P. D. Jones, and T. A. Basnett, 2000: Observed trends in the daily intensity of United Kingdom precipitation. International Journal of Climatology, **20,** 347-364.
Perry, M. and D. Hollis, 2005: The generation of monthly gridded datasets for a range of climatic variables over the UK. International Journal of Climatology, **25,** 1041-1054.
Roe, G. H., 2005: Orographic precipitation. Annual Review of Earth and Planetary Sciences, **33,** 645-671.
Salter, M. d. C. S., 1918: The relation of rainfall to configuration. British Rainfall, **58,** 40-56.

Biography
Emma Ferranti is a 2nd year NERC-funded PhD student studying upland rainfall processes and patterns under a changing climate.
Duncan Whyatt is a Lecturer in GIS with interests in air pollution modelling at regional and local scales.
Roger Timmis is Air Science Manager for the Environment Agency with interests in the detection and dispersion of air pollution.

A sense of place in Virtual Environments

Stuart Ashfield and Claire Jarvis

University of Leicester, University Road, Leicester, LE1 7RH
Tel. (0044 116 252 2522) Fax (0044 116 252 2200)
sa295@le.ac.uk, chj2@le.ac.uk

KEYWORDS: Virtual Environments, Place, Game Engines, Virtual Space, Presence

1. Introduction

It is reasonable to suppose that virtual worlds (and virtual space) can contain virtual places. Simply put, place has been defined as a particular space that is overlaid with meaning (Harrison & Dourish, 1996). However, little research has so far been directed at understanding place and sense of place within a virtual world, and there is much to explore concerning the degree to which the literature on place is relevant. This is important given that data visualisation and geographical analysis software increasingly display data in three dimensional virtual worlds.

This research seeks to identify participants reported experience of place within a desktop 3D virtual environment (VE). In particular, this study focuses its research questions explicitly into the use of games engines as tools for visualisation, and explores user sense of place within a recent example of this technology.

2. Literature Review

2.1 Place

One of the pioneering works in the field, Place and Placeslessness (Relph, 1976), identifies three main dimensions of place identity; the *physical setting*, the *activities* and the *meanings*. Tuan (1977) concluded that spaces need to become meaningful in order to eventually become places. To Harrison and Dourish (1996), places derive much of their meaning from their own spatiality. There is an evident consensus within the literature highlighting place as an area or part of space that has meaning, or life.

Turning to a sense of place in a virtual world, established approaches from the sociological literature would appear to have important limitations (Turner & Turner, 2006). For example, real places are richer and more sensuous than anything that technology can currently produce, while place studies often focus on longer term use of a place, something which is difficult to measure in a VE. Similarly, looking at some of the ways in which a sense of place is developed, an emphasis in the real world place literature on body orientation and movement through the environment (Zacharias 2006) does not translate to desktop displays that allow for only limited immersion through traditional control measures. The validity of previous work on place is therefore questionable when we consider place in a VE.

2.2 Presence

Many authors point out that presence is an essential, defining aspect of virtual experience (Gaggioli et al, 2003). According to some researchers it is impossible to examine the concepts of virtual space without considering the related themes of presence and immersion (Randall, 2006). Building on earlier work by Slater et al (1994), Witmer et al (2005) put forward a model of presence in immersive VEs using principal component-analysis. They found that a 4-factor model comprising the categories *involvement, immersion, sensory fidelity and interface quality* provided a best fit to their data. Involvement in this sense relates to how focussed the participant was on the task at hand, with immersion relating to the specific sense of “being there” (Whitelock & Jelfs, 2005).

It was thus decided to incorporate participant feelings of presence into the original concept of place within VEs. Desk-top game engine environments are not renowned for being immersive experiences; however, presence in them has not been examined to a large extent. Whether a sense of place in a VE can be identified without feeling present therefore becomes one of the main concerns of this work.

3. Methodology

3.1 Software Environment

Game engines have been extensively tested, work on standard systems and can be easily disseminated (Trenholme & Smith, 2008). Overall, we concur with Corbett & Wade (2004, pp114) as follows:

'The modifiability of the virtual environment and game code components and the robustness of the rendering engines and networking code all contribute to the potential of game engines as an interactive and/or immersive 3D cartographic visualization tool.'

This research made use of the CryEngine, created by software developers Crytek. Like many modern engines it has a modular structure and can be reused (Lewis & Jacobson, 2002). A virtual island was created, based on imagery from Google Earth.

3.2 Exploration & Interview Design

14 volunteers took part in the study. The participants were situated at a desk and were allowed to become familiar with standard mouse and keyboard controls. Participants were shown an overview map with 6 locations marked and instructed to move through the environment between markers taking them through a range of different land cover types, including water. At each marker they were asked to stop and survey the scene before moving on.

Following exploration of the VE, detailed interviews were held with participants in order to examine a sense of place, alongside how beneficial they felt this to be for their own navigation and experience. Questions were designed to relate to the main themes of both place (Section A) and presence (Section B) taken from the literature. Participants were asked questions about their experience and their responses, alongside more specific questions relating to their opinions on virtual place and the related concept of presence.

3.3 Analysis

Analysis of the data set was conducted in the following way.

1. Familiarisation with the audio recordings and resulting transcriptions including notes made from any additional comments provided by participants
2. Identification of similar fragments of data corresponding with a sense of place and presence based on categories taken from the literature and subsequent identification of any relevant themes which were different or outstanding

   Section A: Place (Turner & Turner, 2006)

- *Physical Attributes* – Identification and description of features in the environment, sights and sounds etc.
- *Personal meanings associated with the environment* – Being reminded of other places, positive and negative emotions etc.
- *Use of environment* – Any references made to actual or desired activity in the environment.
- *Interaction* – References to how the environment and its contents were interacted with.

  Section B: Presence (Witmer et al, 2005)

- *Involvement* – Aiming to discover how involved a participant became in the VE

- *Immersion* – Aiming to discover how immersed a participant felt in the VE
- *Sensory fidelity*
- *Interface quality*

3. Review of the categories and identification of other themes
4. Comparison of resultant themes with previous literature

4. Results

The interview data was collated and transcribed. Following this, it was coded to allow responses to be assigned to categories involving a sense of place and feeling towards levels of presence in the VE. Table 1 shows responses given by participants.

Table 1. Summary of Responses Mentioned by Participants

Categories	*Frequency mentioned*	*% Positive response*	*% Negative response*
Physical Attributes	68	32	3
Meaning	69	25	9
Use of Environment	35	6	9
Interaction	39	8	72
Involvement	60	13	58
Immersion	24	8	63
Sensory Fidelity	28	7	75
Interface Quality	27	59	26

4.1.1 Physical Attributes

All participants made some reference to visual features of the environment. These were typically related to beach/sand, trees and plants, water and cliffs/rocks. Some participants spoke of wishing to hear/smell things such as the noise/smell of the sea. The physical nature of the environment seemed of interest to all participants. Sensory input other than by sight was mentioned infrequently, notably by its absence: *'It doesn't smell like I'm standing in the sea'*.

4.1.2 Meaning

People expressed positive and negative emotions towards the environments. It was described as lovely as well as lonely. People seemed to associate the environment with other types of media, magazines, television, films and computer games. People described the experience as relaxing, peaceful and tranquil.

4.1.3 Use of the environment

People noted a frustration at being unable to move easily. People stated a wide range of activities to do in the environment should they be there "for real". Participants noted only basic movements such as walking and swimming and exploring their surroundings. In the VE people still seemed to think of the space as lonely or isolated: *'I think I would probably sit in the shade or sunbathe… there wouldn't be any point in doing them here'*.

4.1.4 Interaction

A lack of interaction was noted by nearly all participants. Comments focussed on what technology cannot yet provide: *'I couldn't pick up a stone and skim it or anything; it's not flexible like, my decisions are different because I can't do some things'.*

5. Discussion

5.1 A Sense of Place

Physical representations in virtual space can be seen, visually at least, to offer participants an attractive environment to explore. The VE made people think of similar places that they had seen or been to. The literature tells us that the ways in which people construct knowledge about a VE is different from that of the real world. Is it fair to compare a sense of place in two environments when one is experienced only for a short period of time and the knowledge of both are constructed via different means?

How a place is used relates to the sense of place felt, and this is true in VEs, however the uses may differ between the real and the virtual, rendering comparisons difficult at best. Further work on virtual place must also take into account specific patterns of activity and use associated with VEs.

There were very low feelings of presence felt by participants. This is to be expected in an environment that is not thought of as being particularly immersive (Slater, 1999). Randall (2006) has argued that presence is necessary to look at virtual place, but this might not actually be true. Low levels of presence did not stop participants from registering some recognition of the VE as a place.

One consideration must be that all places are individual, so a study into place in one environment will not necessarily translate into another environment. Turner and Turner (2006) suggest that in order to further understand the concept of virtual place, we must select the aspects of place that can best survive their translation into virtual reality. This suggests a new chapter in research into places within VEs.

6. Conclusions

Presence levels in a 3D VE created in a standard game engine are low; however basic concepts relating to a sense of place are still apparent. The visual dominance of this type of VE requires consideration to enable environments to be created where a sense of place can successfully be felt. Further research must explore the differences between virtual and real places, as well as the similarities.

References

Gaggioli, A., Bassi, M. & Delle Fave, A. (2003) Quality of experience in virtual environments. In, *Being There: Concepts, effects and measurement of user presence in synthetic environments* G. Riva, F. Davide, W.A IJsselsteijn (Eds.)Ios Press, 2003, Amsterdam, The Netherlands

Harrison, S. & Dourish, P. (1996) Re-*place-ing space: The roles of space and place in collaborative systems*. Procedues CSCW'96 ACM Press

Lewis, M. & Jacobson, J. (2002*)* Game Engines in Scientific Research. *Communications of the ACM*: 45 (1), 27 – 31

Randall, R. (2006) Presence, Space and Place in Virtual Environments in Quality and Impact of

Qualitative Research In (Ruth, A.) (Ed) *Quality and Impact of qualitative research. 3rd annual QualIT conference, Brisbane, Institute for Integrated and Intelligent Systems*, Griffith University, 94 – 103

Relph, E. (1976) *Place and placelesness*. London, Pion Books

Slater, M., Usoh, M. & Steed, A. (1994) Depth of presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*: 3 (2), 130 -144

Slater, M. (1999) Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoperators & Virtual Environments*: 8 (5), 560 – 565

Trenholme, D. & Smith, S. P. (2008) Computer game engines for developing first-person virtual environments; in *Virtual Reality*, (2008) In Press Received 23/07/07 Accepted 12/02/08

Tuan, Y. F. (1977) *Space and place*. Minneapolis: University of Minnesota Press.

Turner, P & Turner, S. (2006) Place, Sense of Place, and Presence. *Presence: Teleoperators & Virtual Environments* 15 (2), 204–217.

Whitelock, D. & Jelfs, A. (2005) Would you rather collect data in the rain or attend a virtual field trip? Findings from a series of virtual science field studies. *Int. J. Cont. Engineering Education and Lifelong Learning*, Vol. 15, (1), 121-131

Witmer, B. G., Jerome, C. J. & Singer, M. J. (2005) The Factor Structure of the Presence Questionnaire. *Presence: Teleoperators & Virtual Environments* 14 (3), 298–312.

Zacharias, J. (2006*)* Exploratory spatial behaviour in real and virtual environments. *Landscape and Urban Planning:* 78, 1 – 13

Biography

Stuart Ashfield completed my MSc in GIS at the University of Leicester, where the above work formed part of my dissertation. He is currently working for SPLINT, conducting research into spatial literacy. Research interests include geo-visualisation and representation, virtual environments, games and learning, and associated technologies.

Virtual archives: a geo-located perspective

Rob Millman, Claire Jarvis, Jing Li

Department of Geography, University of Leicester, University Road, Leicester LE 1 7RH
Tel. (+44) 0116 252 3808 Fax (+44) 0116 252 3856
r_millman@hotmail.com, chj2@le.ac.uk, jl185@le.ac.uk

KEYWORDS: Local history, digital archives, virtual reality

1. Introduction

Over the last 10-15 years historians have come to embrace the spatiality of history to an increasing degree. Many research projects have looked at ways in which GIS can help visualise and explore historical data (e.g. Siebert 2000; Mosaker 2001; Weaver *et al*. 2007) and more recently to provide new avenues for analysing the spatiality of historical data (Owens 2007; Knowles 2008).

The main focus of this particular project was to examine how we might geo-contextualise representation of history; how, for example, might we digitally represent newspaper clippings, contextualise photographs or incorporate news bulletins of actions relating to campus life but not necessarily being performed within it? In particular the study aimed to produce a working real-time virtual environment showing the space-time development of the Leicester University campus, in a manner that incorporated archival data in the form of geo-located media.

2. Approach

2.1 Archival data

The University's virtual reality model of the campus as it was in 2004 provided a robust starting point for this research.

Historical data used in this work came from several different sources. Photographs were largely sourced from the university's official archives, which started around the same time as the university itself. One group of photographs was selected for the purpose of re-texturing buildings and ground cover according to period; further selections were made for direct user interaction, based on the degree to which they gave an insight into life at the university from a student perspective. Video files were obtained from the Media Archive for Central England (MACE). Additionally, historical maps were sourced from the Ordnance Survey historical map collection. Sound files for the project were recorded specifically for this project from newspaper records and letters.

2.2 Modelling Approach

2.2.1 Representation of time

The rate at which University buildings developed across the campus has not followed a linear time sequence but rather has been a function of the economic and social setting of the period. The general strategy for selecting the time slices was to use at least one iteration per decade, assuming that some significant change such as the erection of a new building had occurred. Overall, the period of largest scale building development in the University's history was between the mid- 1960s to 1970. The final, modelled time slices in real-estate terms were: 1921, 1951, 1958, 1963, 1970, 1977, 1985 and 1994. Each of these models contains social media and reportage pertaining to the intervening period.

Modelling and viewing transitions over time is an important issue within the virtual environment. Our goal was to allow the user to progress through time *at any location they are standing in* within the environment, with a view to maximising presence and inclusion within the immersive scene. The buildings themselves simply go from present to absent within the VR scene.

2.2.2 Spatial Context

Study of the maps showed that the current university campus grew from a smaller number of buildings; land on which some of the university's buildings are currently situated did not always belong to the institution. With this in mind, only the extent of university property at any time slice was displayed, avoiding assumptions based on partial information regarding what may have been located on that land before the university took ownership.

2.2.3 Representation of Archival Data

Within the environment, audio, video and photo materials were visualised in different ways. Amongst the archive material were several newspaper cuttings and some diary entries written by a past student of the university. A selection of stories were chosen and recorded as a mock 'Newscast' which is played at relevant points within the environment. Videos were added in a way that allowed them to be viewed on a drive-in style cinema screen placed within the virtual campus, while the video continually loops so that the user will always be able to see the full film. Photos were added to the visualisation software in groups as slideshows which appear on screen as the user travels around the environment.

2.3 Software & Hardware Environment

The visualisation of the 3D environment took place in the 120^{o} field of view (FOV) stereo virtual reality theatre in the geography department at the University of Leicester. Participants in the evaluative study were navigated or "flown" around structures of significance by a free roving pilot in real time, who was also able to respond to specific requests for movements over space-time.

Plenoptics™ Creator was the primary program for modelling the campus whilst another of Plenoptics' software solutions, Vega Prime, was used for real-time visualisation of the environment. Sound recording was done using Audacity with video encoding completed using Adobe Premier Pro.

2.3.1 Modelling

Using the contemporary 3D model of the campus as a starting point, each building and its surrounding grounds were divided into separate units. Having completed this separation a time slice model was then created by adding each of the components needed to portray the university as it would have been at that time. In doing this, gaps became evident where current buildings are; these were replaced with a grassy texture.

A further modelling consideration was the design and positioning of the cinema board on which movies could be played. This model was taken from an existing 3D library and adapted in Creator to suit the needs of the project. The board was positioned within the environment so that it remained in a prominent place and was not obscured by any of the other buildings on campus.

Historical Ordnance Survey maps were exported from ArcScene, with base heights, as VRML files. These VRML files were converted to OpenFlight format using Polytrans software and then opened in Creator.

2.3.2 Real-Time Visualisation

Real-time visualisation of the environment was carried out using Vega Prime. The programme allows the selection of models to be displayed in a scene, adds environmental effects, and dictates how the user can move through the environment. Two dimensional overlays can be added for the display of photos and text. Vega Prime also allows the addition of sound.

Using the standard Vega Prime interface, objects added to the scene cannot be changed at runtime. Additionally, once audio sounds and photo overlays are enabled, they remain active throughout a user's interaction with the visualisation. Enhanced functionality was therefore required, implemented in C++ via the Vega Prime development kit, in order to:

- Enable the real-time transition of models from one time period to the next;
- Enable the real-time addition of photos based on the proximity to a specific building (within 15m) using a Line of Sight (LOS) vector;
- Enable the addition of sound based on the proximity to a specific object (within 15m), again via a LOS vector;
- Enable the viewing of video material within the environment, via the use of real-time object texturing.

3. Results & discussion

The outcome of the study was the production of a virtual reality model of the University of Leicester campus which included geographically located historical media. A selection of interested parties, including archivists and local historians, were given the opportunity to use the tool.

The model goes through 8 stages of time from 1921 to 1994 showing the progress of building work on the campus (Figure 1 and Figure 2).

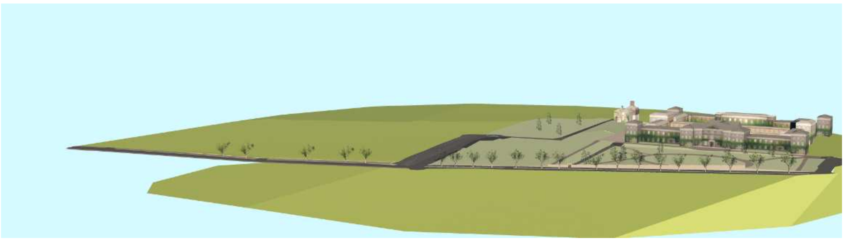


Figure 1. Overview of campus in 1921

Figure 2. Overview of campus in 1994

As Figure 3 demonstrates, photo overlays for separate buildings showed extra detail regarding the life of the students present at that time.

Figure 3. Image showing tool running in the background with photo overlay in bottom left hand corner and the year of the time slice in the top left.

Figure 4 demonstrates an example of a historical video embedded within the fourth time slice model, displayed on a cinema screen within the virtual environment. The LOS method used allowed the observer to focus particularly on proximal material, and also to geographically contextualise the photos and audio. The use of the LOS also meant that the user maintained a generally clear field of view throughout their progress through the visualisation.

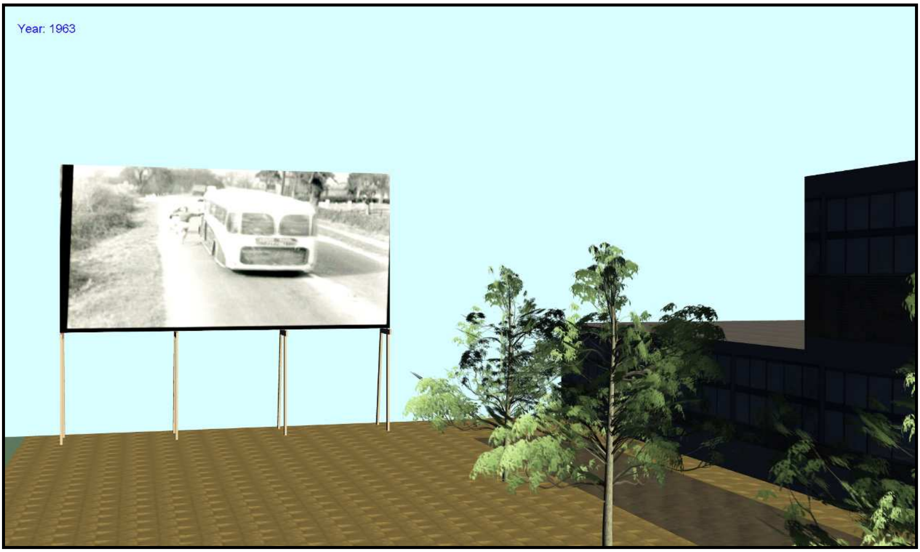


Figure 4. Image showing tool running with cinema screen in the environment.

The use of the outdoor cinema screen and its ability to "include items beyond the extent of the GIS" but still relevant to the university population was considered a particular success by the user test group. The novel newscasts also were well regarded, their appeal to an extra sense other than sight thought to provide a "broader learning experience and allows the user to be more engaged". Visually adding these textual sources to the environment would have been unsatisfactory; used as the basis for recorded sounds, they allowed for the inclusion of a wide ranging source of material that would otherwise be lost.

Further, more detailed, evaluation is required, together with an exploration of appropriate navigation approaches for stand-alone users interacting with such an environment in different contexts. Meanwhile, the commentary of the expert group suggests that the geo-located embedding of archival data and materials in a VR environment provides a powerful, immersive experience with considerable general potential for visualising histories; the prototype demonstrator triggered discussion of a number of future developments and directions. These included both the "bringing alive" of oral histories to the general public and new strategies for archiving materials for research access. The multimedia approaches used were considered an appropriate match for bringing together physical and social histories, but clearly more focused applications for directed tasks will require detailed user requirements analyses from which a variety of tailored interfaces & functions will emerge.

4. Conclusions

This work illustrates a virtual model representing temporal progression of the campus at the University of Leicester, in which not only are archival data used to support the urban modelling but where the approach also incorporates geo-located historical multimedia. The use of audio to encode historical newsclippings as geo-located newscasts offered a novel means of placing and expressing significant historical events place. A static movie screen was used to play video clips relevant to particular time periods but where the events were located off-campus. The work extends the typical role of historical 3D visualisation, in which the primary goal is for the user to experience the developing area from multiple perspectives and vantage points, to more fully integrate a range of contextual or source materials.

References

Knowles, A.K. (2008) *Placing History: How Maps, Spatial Data and GIS are Changing Historical Scholarship*, ESRI Press, CA.

Mosaker, L. (2001) Visualising historical knowledge using virtual reality technology, *Digital Creativity* **12** pp. 15-25.

Owens, J.B. (2007) Towards a geographically-integrated, connected world history: Employing geographical information systems (GIS). *History Compass* **5/6**, 2014-2040.

Siebert, L. (2000) Using GIS to Document, Visualize and Interpret Tokyo's Spatial History, *Social Science History* **24**, pp. 537-74.

Weaver, C., Fyfe, D., Robinson, A., Holdsworth, D., Peuquet, D., MacEachren A. (2007) Visual exploration and analysis of historic hotel visits, *Information Visualisation* **6**, pp. 89-103.

Biography

This research was completed by Robert Millman as part of his MSc in GIS (Distinction) at the University of Leicester, 2007/8. Rob, who graduated in Geography at Leicester in 2007, is now working for infoterra.

Visualizing Public Transport Quality of Service

Adam Winstanley, Bashir Shalaik, Jianghua Zheng and Rebekah Burke

Department of Computer Science, National University of Ireland Maynooth, Co.Kildare, Ireland.
Tel: +353 1 708 3853; Fax: +353 1 708 3848
Email: adam.winstanley@nuim.ie

KEYWORDS: vehicle tracking, GPS, GPRS, transport, QoS

1. Introduction

The recent advances of geo-positioning hardware, computer software and mobile communications have combined to offer new opportunities for improved public transport services. Today many transportation companies are using the Global Positioning System (GPS) and wireless communication systems (e.g. radio data systems or GSM/GPRS) for communicating their vehicle location information and other details to a central server (Predic et. al 2007) (Kane, Verma and Jain 2008). By tracking their bus fleet in real-time, operators can monitor schedule adherence and service efficiency, give better operational support and provide users with real-time service information. There are several bespoke systems commercially available that do this.

In order to improve services, as well as providing real-time information, these systems build up an archive of data that can be analysed and mined to explore and show behaviour of the transport system over time, indicating problems such as vehicle bunching and delays due to congestion. In addition, to qualify for public subsidies, operators must report Quality of Service metrics to government. These are usually calculated manually but the existence of a full archive of data gives the potential for automation.

Quality of Service (QoS) in public transport is a set of metrics used to measure the reliability and punctuality of bus services (Transport for London 2007). For a high frequency route (a route with five or more buses per hour) it is important for buses to run at evenly spaced intervals and not bunch. A standard metric, the Excess Waiting Time (EWT), is used to measure the additional wait experienced by passengers due to the irregular spacing of buses or those that failed to run. For low-frequency bus routes (a route with four or less buses per hour) passengers using the service tend to rely on the timetable. It is more important that services run as close as possible to the time specified on timetables. A metric that measures any deviation is required.

In a joint project between NUI Maynooth and Blackpool Transport methods are being explored to visualise the behaviour of vehicles in ways to allow the operator to better assess and improve the quality of their public service. The system uses off-the-shelf GPS/GPRS integrated units programmed to transmit location at regular intervals (45 seconds approximately) while the vehicle is in motion. The data is stored on a server and can be visualised through a standard web browser to show views representing current locations of vehicles in close-to-real-time. In addition tools are provided to visualise vehicle behaviour over time and to calculate metrics and summaries. The system uses web technologies such as JavaScript, MySql, XML, PHP and Ajax. In addition there is a public interface that can display and update vehicle locations in Microsoft Virtual Earth (bustracking.co.uk).

2. Visualization of real-time information

One of the most important challenges in visualization research is to determine how best to depict a set of data so that the information it represents can be accurately and efficiently understood. Both design and evaluation have key roles to play in this process. Some recent advanced public transportation systems (APTS) make use of various data visualization tools in a GIS context (Yu, Mishra and Lin 2006) (Hoar 2008) (Maclean and Dailey 2001). These tools concentrate on information for passengers and the real-time display of bus locations. In this paper we present a computer visualization prototype which is more aimed at public transport operators. It includes real-time tracking but also allows operators to view performance over periods of time and to help in decision making.

Visualization is a process of transforming data into a visual form enabling the user to better understand and extract the information contained. In transportation services, vehicle information can be visualized in an on-line or off-line environment through tables, maps, data plots and other graphical outputs (Zolfaghari, Jaber and Azizi 2001) (Okunieff, P.et al. 1997). To visualize real-time information, such as the current/last vehicle location, time at location, this is integrated with real-time data sources from the vehicles. The system displays real-time locations of buses pictorially, textually and, using the facilities provided by the Microsoft Virtual Earth API, with 2D and 3D map visualizations (figure 1). The displays can show adherence to the published timetable through colour coding (figure 2).

Figure 1. The public interface showing updating textual display plus moving locations on Microsoft Virtual Earth.

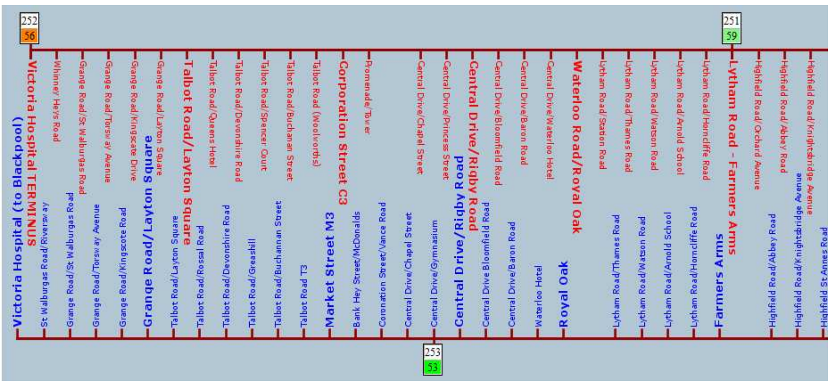


Figure 2. Route diagram visualisation, colours indicating adherence to timetable.

3. Visualization of vehicle behaviour

Historical as well as current vehicle behaviour is required for an assessment and improvement in service quality. The system can visualize the behaviour of vehicles in different easy-to-understand ways. The daily spatial-temporal behaviour of vehicles (vehicle locations in term of bus stops passed against time) is displayed as a series of line-graphs. As well as showing the regularity of the service, poor service phenomena such as vehicle-bunching and erratic service intervals are easily detected (figure 3). Similarly the scheduled timetable can be plotted in a similar way providing comparisons between advertised and scheduled services (figure 4) and clearly indicating when a bus is ahead or behind. This can be summarised through the calculation of rms values between the two curves to give an overall measure of timetable adherence (table 1). The plots can also be used to display another relevant variable through the thickness of the plotted line, such as in figure 5. The on-vehicle units have a serial interface allowing it to monitor and transmit data from other devices such as ticket-issuing machines, motor controllers and other sensors.

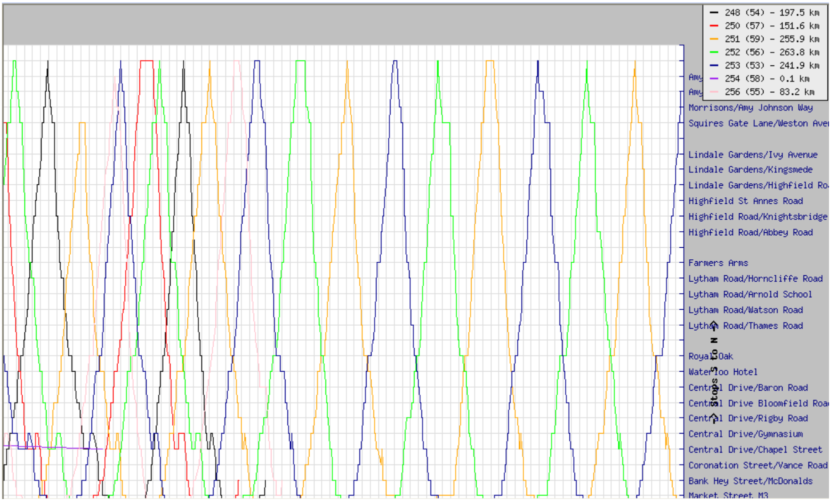


Figure 3. Extract of plot of vehicle locations, in terms of bus-stops passed, against time. Bunching of vehicles is easily detected.

4 Visualizing Quality of Service indicators (QoS)

Quality of service indicators are metrics that are used in evaluating public transit performance. These provide passengers and operators a measure of how reliable services are and help operators to improve schedule adherence and service efficiency. Similar the regulatory authorities usually require reporting of QoS metrics to comply with licensing rules and the conditions for operating subsidies. QoS is defined as the "overall measured or perceived performance of a transit service from the *passengers'* point of view" (Transport for London 2007).

With respect to QoS, frequency of service can be divided into two categories, *high* and *low* depending on the number of vehicles serving an individual route. For low frequency routes, defined as those with four or less vehicles per hour, it is important that the service runs exactly to the time specified on timetable and QoS is specified as the mean deviation of buses from their scheduled time. On high frequency routes (with five or more buses per hour), passengers tend to arrive at stops without consulting a timetable because they expect buses are running at evenly spaced headways. QoS is measured by calculating the average Excess Waiting Time (EWT) that passengers have waited above the theoretical waiting time given by the service interval.

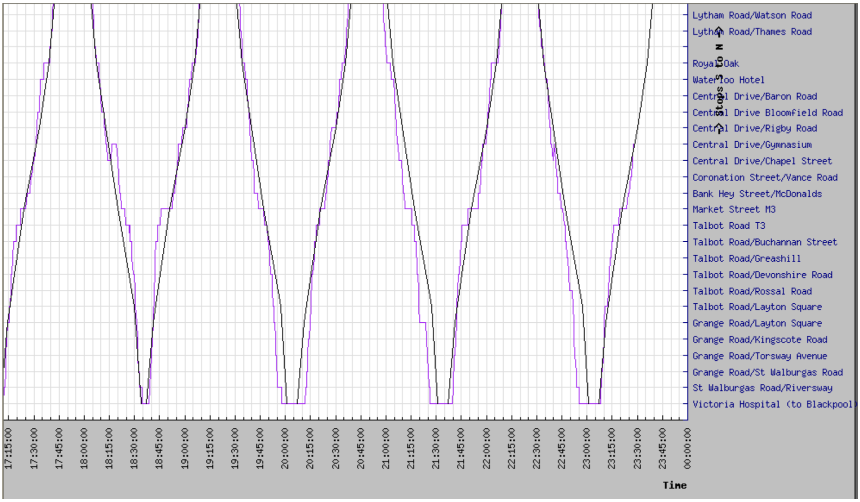


Figure 4. Extract of plot of vehicle location in comparison to scheduled timetable.

Date: 2008-08-01

Imei No	Bus	Route	rms
357541000328460	256	53	1265
358279000289809	250	51	2546
358279000289825	251	57	74
358279000290989	253	59	1745
358279000298263	257	0	1035

Table 1. RMS values in seconds for different vehicles against their timetable

4.1 Arrival Time Prediction for Low Frequency Routes

An algorithm has been implemented to estimate bus arrival time. This algorithm involves using historical operational data on the route collected over several months and averaging bus travelling time (between bus stops) for each hour of each day of the week with the underlying assumption that buses will be running in same operational conditions at the same time each week. This averaged data is used to predict how long it will take an individual bus to get from where it is to a particular stop. The system also uses the travel time of the preceding buses, weighted by how recent they are, to factor in conditions that may be prevailing on that particular day. The effectiveness of this algorithm is currently being evaluated and will be presented. The system is demonstrated by implementing a web-based next-bus indicator.

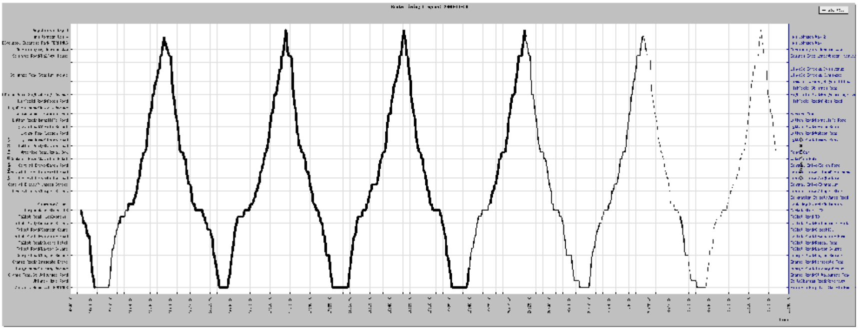


Figure 5. The thickness of the line on the plot can indicate variables such as battery voltage or passenger loadings by location and time.

Figure 6. Predicted arrival time based on historic data.

4.2 Excess Waiting Time (EWT) for High Frequency Route

A standard metric, the Excess Waiting Time (EWT) is commonly used to measure the quality of service. It can be defined as the average additional wait experienced by passengers due to the irregular spacing of buses or those that failed to run (Transport for London 2007). This is a key performance indicator since it denotes how much time passengers had actually to wait in excess of what we would have expected them to if the service were perfect. EWT is calculated by subtracting Scheduled Waiting Time (SWT) from Average Waiting Time (AWT) and it is this which is used as the measure of reliability. The greater the EWT is, the less reliable the service (Okunieff, P. et al. 1997).

$$EWT=AWT-SWT \qquad (1)$$

where AWT is the average time that passengers actually waited and SWT is the time a passenger would wait, on average, if the services ran exactly as planned during the periods observed. The system can automatically generate daily, weekly, monthly and annual reports of EWT for any stop. We are currently investigating ways to visualise its variation over time.

5. Conclusions

The current system provides several tools for visualising vehicle behaviour and calculating metrics for service quality. They provide views designed to allow easy detection of many of the symptoms of a poor bus service – bunching, late-running, missed services. They also automatically calculate metrics for reporting and comparing QoS. However, evaluation of how effective they are in a commercial operational environment is on-going and will be used as input to a revision of the system. Other planned work involves developing 3D visualisations and adapting techniques used in computer network QoS visualisation for this domain.

Acknowledgements

Thanks are due to Blackpool Transport Ltd for facilitating this project and particularly to Oliver Howarth for his comments and feedback. This project was partly supported by the StratAG project, funded under the SFI Strategic Research Cluster Programme (07/SRC/I1168). One author is supported by a PhD studentship from the Libyan Ministry of Education.

Data for **Market Street M3** on **20/05/2008** from 8am - 6pm.

(a)	(b)	(c)	(d)
08:06:00			
08:35:00	29	14.5	420.5
08:46:00	11	5.5	60.5
09:21:00	35	17.5	612.5
09:37:00	16	8	128
10:04:00	27	13.5	364.5
10:14:00	10	5	50
10:53:00	39	19.5	760.5
11:11:00	18	9	162
11:35:00	24	12	288
11:42:00	7	3.5	24.5
12:23:00	41	20.5	840.5
12:37:00	14	7	98
13:05:00	28	14	392
13:13:00	8	4	32
13:54:00	41	20.5	840.5
14:03:00	9	4.5	40.5
14:32:00	29	14.5	420.5
14:43:00	11	5.5	60.5
15:23:00	40	20	800
15:36:00	13	6.5	84.5
16:04:00	28	14	392
16:20:00	16	8	128
16:54:00	34	17	578
17:05:00	11	5.5	60.5
17:37:00	32	16	512
17:48:00	11	5.5	60.5

Total (Sum of square of headway/2) 8211
Time between 1st and last observed bus is 582
Average Waiting time in minutes 14.1082474227

Number of Buses scheduled per hour 6
Scheduled Waiting times (SWT) in minutes 5

Excess Waiting Time= 9.10824742268

Figure 7. Automatic calculation of Excess Waiting Time.

References

Predic, B., Rancic, D. et. al (2007) Automatic Vehicle Location in a Public Bus Transportation System, *11th WSEAS International Conference on Computers*, Crete, Greece, 675-680, 2007.

Kane, L., Verma, B. and Jain, S. (2008), Vehicle tracking in public transport domain and associated spatio-temporal query processing, *Computer Communications*, 31, 2862-2869, 2008.

Transport for London (2007) *London Bus Quality of Service Indicators*, Transport for London, 2007.

Yu, D., Mishra, S., and Lin, J. (2006): Visualization of Bus Adherence Using GIS, *Applications of Advanced Technology in Transportation*, 159-164, 2006.

Hoar, R. (2008), 'Visualizing Transit through a Web Based Geographic Information System' *Proceedings of World Academy of Science, Engineering and Technology*, 36, 180-185, December 2008.

Maclean, SD, and Dailey, DJ (2001) 'Busview: A Graphical Transit Information System, *IEEE Intelligent Transportation System Conference*, 1073-1078 2001.

Zolfaghari, S., Jaber, M.Y and Azizi, M. (2001): A Multi-Attribute Performance Measurement Model for Advanced Public Transit system, *Intelligent Transportation Systems*, 7:295-314, 2001

Okunieff, P.et al. (1997), *Synthesis of Transit Practice 24: AVL systems for Bus Transit*, Transit Cooperative Research Program, Transportation Research Board, National Academy Press, 1997.

Biographies

Dr. Adam Winstanley is the Head of Department of Computer Science, National University of Ireland Maynooth. His research interests lie in several fields, including Location Based Services; transport information systems; passenger information systems; graphics recognition using shape, context, and Statistical Language Modelling for recognition and validation of graphical objects and electric vehicle control systems.

Bashir Shalaik is a postgraduate research student at National University of Ireland, Maynooth. He did his Masters degree in computer science at NUIM. He is interested in IT in Intelligent Public

Transportation and GPS-based bus tracking systems.

Dr.Jianghua Zheng is a postdoctoral researcher in Department of Computer Science, National University of Ireland Maynooth. He previously was associate professor in Xinjiang University China. His interests include GIS for transportation, emergency management and Location Based Services.

Rebekah Burke is a final year undergraduate studying Computer Science and Software Engineering at NUI Maynooth. She implemented some of the work described here as part of the third year internship programme.

Flow Trees for Exploring Spatial Trajectories

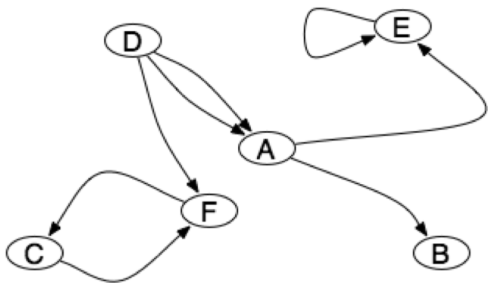
Jo Wood, Jason Dykes, Aidan Slingsby and Robert Radburn

giCentre, School of Informatics, City University London EC1V 0HB
Tel.+44 (0)20 7040 0146 Fax +44 (0)20 7040 8584
{jwo | jad7|a.slingsby|sbbd476} @soi.city.ac.uk, www.soi.city.ac.uk/~ {jwo | jad7|sbbb717|sbbd476}

KEYWORDS: trajectories, flows, OD matrices, treemaps, commuting, visualisation

1. Introduction

A trajectory is a directed path that defines a link between two spatial locations. That path may be as simple as the Euclidean shortest distance between a start and end point, or may involve a more complex traversal through time and space to travel from start to end. Within GI analysis, trajectories are used to represent phenomena such as movement of people as migration and commuting, goods and information. Trajectories are commonly represented as *Origin-Destination Matrices* (OD matrices) (Voorhees, 1955) that record the numbers of directed links between a set of start and end points (see Figure 1).



		Destination					
		A	B	C	D	E	F
Origin	A	0	1	0	0	1	0
	B	0	0	0	0	0	0
	C	0	0	0	0	0	1
	D	2	0	0	0	0	1
	E	0	0	0	0	1	0
	F	0	0	1	0	0	0

Figure 1. Simple trajectory network and its origin-destination matrix. Note how spatial information is lost in the matrix view.

This paper presents a novel alternative representation of origin-destination topology that makes it more amenable to the visualization of structure and spatial organisation of trajectories.

2. Challenges for Visualizing Trajectories

Trajectories can be numerous and dense in space making their direct visualization as GIS vector lines problematic. Spatial heterogeneity of density (e.g. commuting flows become denser towards the centre of a city) or constrained flow paths (e.g. along transport networks) can further complicate direct visualisation of all trajectories. While there have been proposals for generalising high density network paths for visualization (e.g. Cui *et al* 2008), some of the network structure and spatial arrangement can be lost in the process. Rae (2009) explores some of the ways in which the spatial structure of migration trajectories can be preserved using contemporary GI System tools.

As an alternative approach, Marble *et al* (1997) proposed visualization of the OD matrix directly rather than vectors of each trajectory. This declutters the visual display since only the frequency of origin and destination connections is shown, not the geometry of the paths that connect them. Visualising this matrix as a surface projected in 3-dimensional space, they argued that insight into the topology is gained by clustering OD cells of similar value.

We would argue that there are two significant problems with this approach. Firstly, there seems little benefit in projecting what is a 2-dimensional raster of comparatively low resolution into a 3-dimensional continuous space. This leads to problems of occlusion, dominance of ‘spikes’ and difficulty in navigation and orientation. Secondly, and more significantly, spatial relationships are lost when constructing an OD matrix – 2-dimensional spatial location is collapsed along 1 dimension of each matrix axis. This can result in spurious lineations that are a function of the arbitrary ordering of locations within the matrix (see Figure 2, right).

We propose an alternative reordering of the OD matrix, here termed a *flow tree*, that preserves the spatial arrangement of both origin and destination locations while maintaining the topological structure of the OD matrix. Consider a rectangular space partitioned into *n* by *n* grid cells. Each cell represents the origin of trajectories that start within that cell’s bounding box. Next, partition each cell into *n* by *n* sub-cells, where each of these smaller child cells represents the location of the destination of any trajectories that originated from their parent cell’s origin location. This is illustrated in Figure 2 where 500 Gaussian random trajectories are partitioned into 5x5 origin cells and then a further 5x5 destination cells inside each origin. This produces a set of 25 small multiples of trajectory destinations, each a copy of the entire region, but filtered and ordered by their origin location. In other words, every origin cell has its own destination map embedded within it.

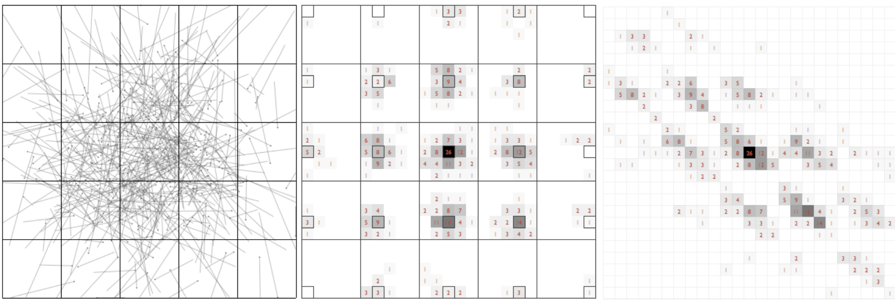


Figure 2. 500 random trajectories shown as (left) a conventional trajectory map of origin to destination vectors; (middle) a flow tree; (right) an OD matrix. Numbers of trajectories are indicated as grey shading and numeric symbols. Within the flow tree, locations of trajectories that share the same origin and destination cell are highlighted.

The flow tree contains the same set of cell values as the OD matrix, but more usefully ordered to show spatial relationships. This can be regarded as a special case of the *spatial treemap* (Wood and Dykes, 2008a, 2008b) where space is partitioned regularly and the hierarchy is two levels deep.

The benefits of the flow tree representation can be seen in Figure 3, which shows 500 random trajectories with significant directional bias. In the conventional trajectory map, this is not visually obvious in the denser central part of the region, but the flow tree shows this clearly, with no destination cells shaded to the east (right) of their origin location.

For many real distributions, short flows are likely to be much more frequent than longer ones. Figure 4 shows the flow distance frequency distributions for commuter flows within Ohio and within Leicestershire and Rutland. In both cases, a log-type scaling of grey levels would be required in order to show any variation beyond the most local flows.

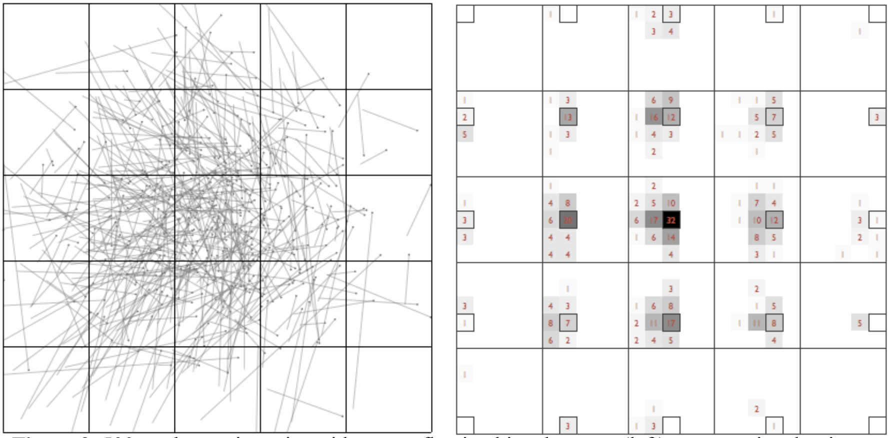


Figure 3. 500 random trajectories with a west-flowing bias shown as (left) a conventional trajectory map and (right) a flow tree.

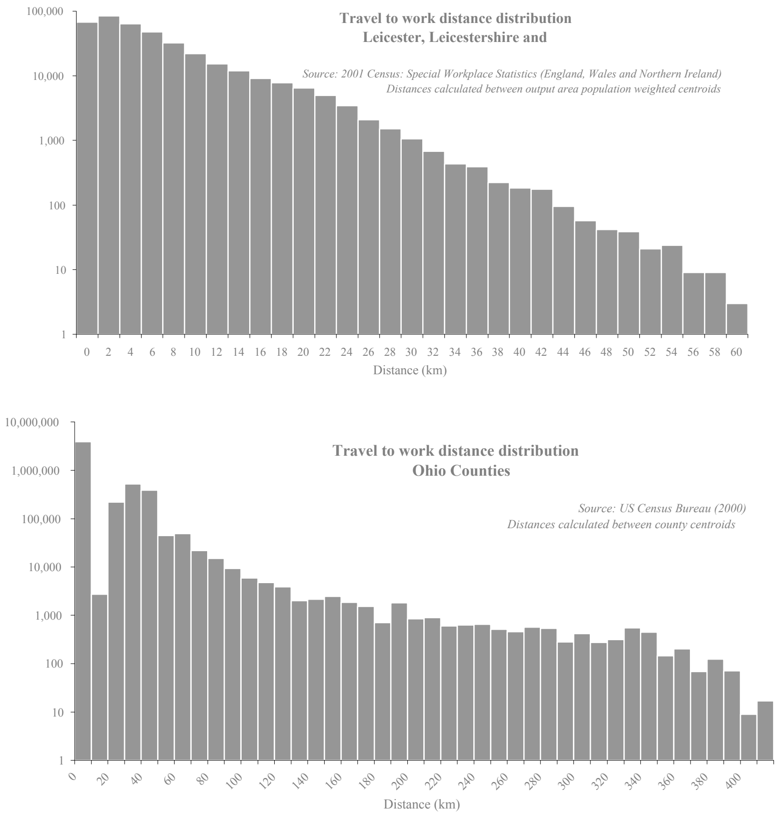


Figure 4. Distance frequency histograms for travel to work commuting in Leicestershire and Ohio.

3. Partitioning of Space

The form of the flow tree is dependent partly on the way in which space is partitioned into regions. An important design decision in the flow tree layout is that destination regions are partitioned in exactly the same way as origin regions, thus ensuring the small multiples of the region. To achieve this, the partition must result in self-similar sub-regions that tessellate within their parent region. The simplest of these is to partition into *N* squares, where *N* is perfect square.

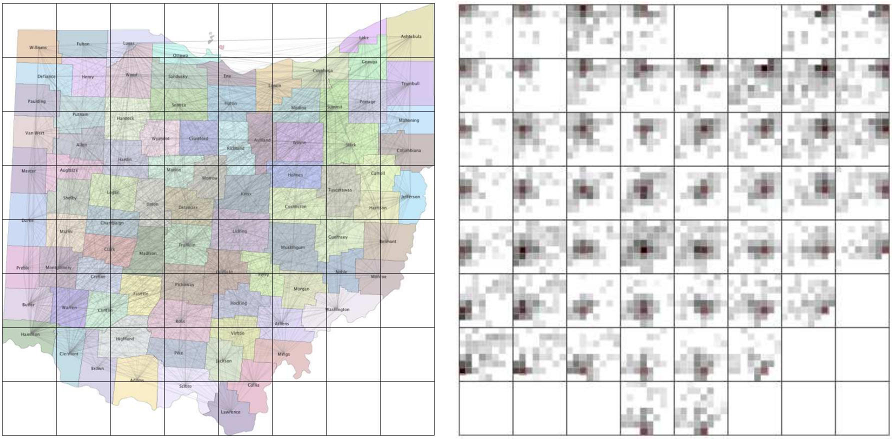


Figure 5. Square regions for Ohio county-to-county travel to work trajectories and flow tree.

The size of each grid cell should reflect the approximate scale of interest, so for example Figure 5 shows *N*=64 for Ohio, where each cell is approximately the average size of the counties being examined. However, this can lead to aliasing problems where counties are split between adjacent cells, or inconsistently aggregated within a single cell.

Aggregating at a much coarser scale than the trajectory origin and destination regions can reduce aliasing. For example, Figure 6 shows partition of census output areas (approximately 100 households) into non-square rectangles of 10x8km that reflect the aspect ratio of the entire region.

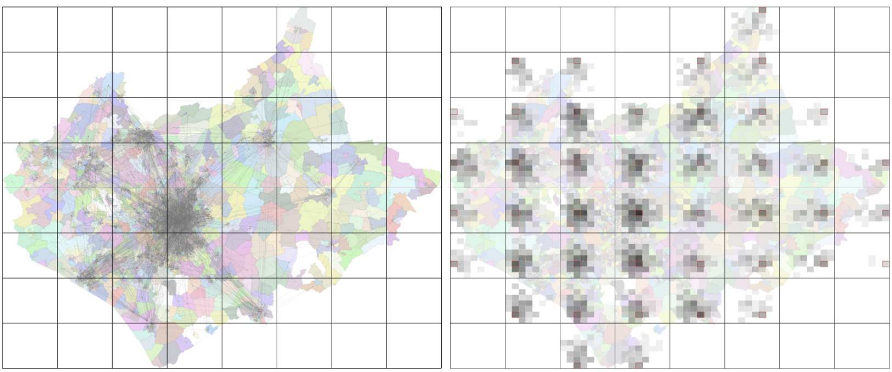


Figure 6. Rectangular regions used for Leicestershire Output Area travel to work trajectories and flow tree representation. Boundary data Crown Copyright, 2009.

Alternatively, if data resolution does not permit aggregation into coarser grid squares, a rectangular spatial treemap tessellation (Wood and Dykes, 2008a) may be applied avoiding arbitrary gridding. Extra ‘dummy nodes’ may have to be inserted into the tree to ensure the number of nodes is a perfect square and thus preserve small multiples (see Figure 7).

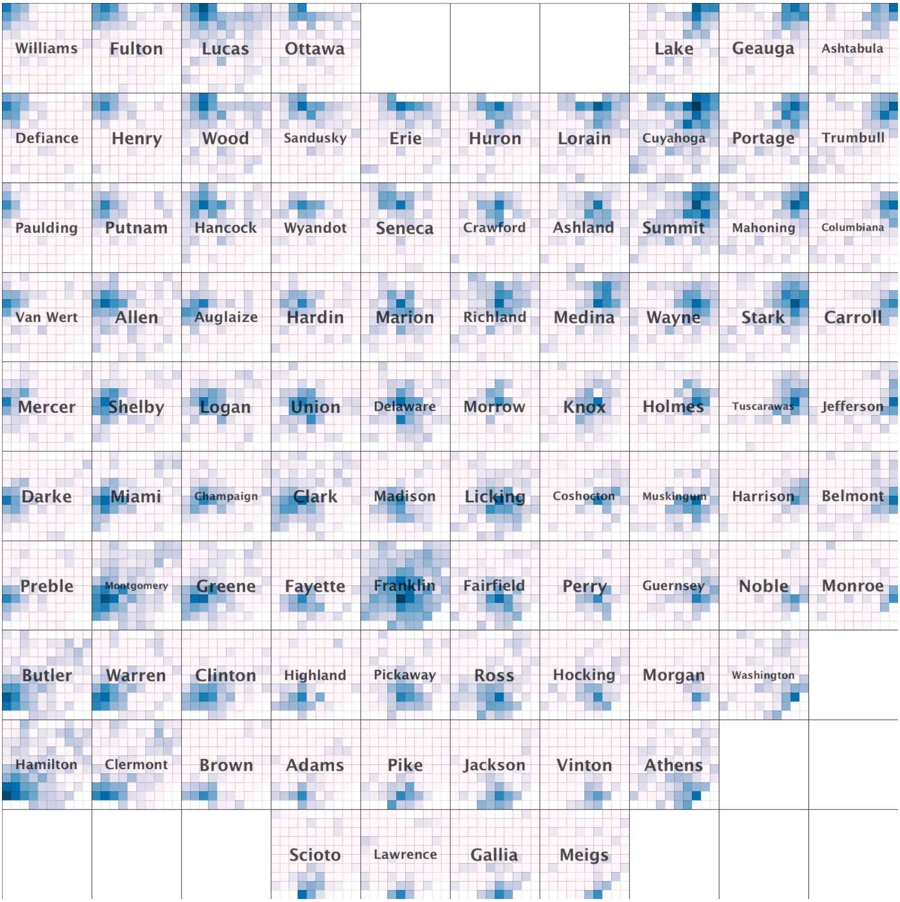


Figure 7. Flow tree for Ohio county travel to work using spatial treemap tessellation of counties. 12 extra dummy nodes (white) are added to preserve small multiples of destination cells.

4. Conclusions and Further Work

Visual representation of sets of flows in space is challenging due to localised high density of flows typical in many geographic phenomena. Topological OD matrix views can help to simplify networks, but can also lose important spatial information. The flow trees presented here preserve both topological and spatial structure by using a two-level spatial hierarchy.

Ongoing work is examining how interactive software can allow spatial portioning to be varied and to use brushing and selection to show geometric paths of trajectories projected onto this two-level hierarchy. We are also exploring how background context mapping can be used to ease the cognition of the spatial transformation involved in projecting a flow tree. Finally, we are investigating non-linear projections to show the spatial heterogeneity of flows in typical urban-rural environments.

Acknowledgements

Leicestershire travel to work data derived from Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Output Area boundary data provided through EDINA UKBORDERS with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown.

References

Cui, W., Zhou, H., Qu, H., Wong, P. and Li, X. (2008) Geometry-based edge clustering for graph visualization. *Transactions on Visualization and Computer Graphics*, 14(6), 1227-1284.

Marble, D., Gou, Z. and Saunders, J. (1997) Recent advances in the exploratory analysis of interregional flows in space and time. In **Kemp, Z.** (ed), *Innovations in GIS 4*. London: Taylor & Francis, 75-88.

Rae, A. (2009) From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census. *Computers, Environment and Urban Systems*, *in press* DOI:10.1016/j.compenvurbsys.20009.01.007

Voorhees, A. M. (1955) A general theory of traffic movement. *Institute of Traffic Engineers Past Presidents' Award Paper,* New Haven: ITE.

Wood, J. and Dykes, J. (2008a) From slice and dice to hierarchical cartograms: Spatial referencing of treemaps. In Lambick, D. (ed.) *Proceedings of the GIS Research UK 16th National Conference, Manchester Metropolitan University*, 1-8.

Wood, J. and Dykes, J. (2008b) Spatially ordered treemaps. *IEEE Transactions on Visualization and Computer Graphics*, 14(6) 1348-1355.

Biography

Jo Wood is a Reader and Jason Dykes a senior lecturer in geographic information at the giCentre, City University London with interests in developing and using interactive graphics. Aidan Slingsby is a Willis Research Fellow at the giCentre with interests in spatial modelling and geovisualization with for applications relevant to the global insurance industry Robert Radburn holds an ESRC UPTAP research fellowship at City University London and is a Senior Research Officer at Leicestershire County Council.

Vehicle routing problem and travel time prediction

Ivana Ćavar

Faculty of Transport and Traffic Sciences, University of Zagreb, Vukeliceva 4, HR-10000 Zagreb
Tel. +38512380225 Fax +38512314415
ivana.cavar@fpz.hr

KEYWORDS: VRP, VRPTW, travel time prediction, Geographic Information System (GIS), Global Positioning System (GPS)

1. Introduction

The Vehicle Routing Problem (VRP) arises from determining optimal routes used by a group of vehicles in the course of serving a group of customers. The solution to the problem is a set of routes with the first and last stops in the depot of the corresponding vehicle. All customer requests should be fulfilled, all constraints satisfied and each customer served only once.
In practice, the basic VRP is extended with constraints (Figure 1). The objective of solving the VRP is to minimize total transportation costs and upgrade the level of service provided by the transportation company.

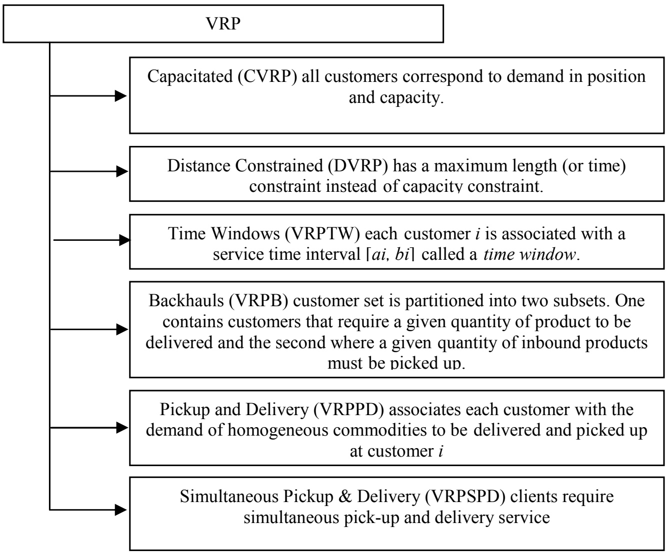


Figure 1 VRP Classification

In cases when only one vehicle serves customers with no additional constraints, the VRP is reduced to the Travelling Salesman Problem (TSP). Both problems are NP hard, which means that in standard cases the exact solution cannot be found in polynomial time. For instance, to solve the TSP problem with 10, 18 and 20 customers, a single contemporary computer with processing power 10^9 operations/sec, needs less than 1 second, 74 days and 77 years respectively in order to accomplish the task. Thus, to solve the VRP problem, the following exact, heuristic and meta-heuristic algorithms are used (Table 1).

Table 1 Solutions for exact, heuristic and meta-heuristic algorithms on Solomon benchmark problems

	Exact		*Heuristic*		*λ interchange*	
Problem	**Vehicles**	**Distance**	**Vehicles**	**Distance**	**Vehicles**	**Distance**
C101	10	827.3	10	828.94	10	828.94
C102	10	827.3	10	828.94	10	978.6
C103	10	826.3	10	828.06	11	1174.38
C104	10	822.9	10	824.78	11	1112.89
C105	10	827.3	10	828.94	10	828.94
C106	10	827.3	10	828.94	10	844.58
C107	10	827.3	10	828.94	10	828.94
C108	10	827.3	10	828.94	10	867.86

2. VRP Case Studies

To solve a practical VRP problem, attention must be paid to a number of specific elements:

- A digital map must be obtained with traffic layers including all important elements, such as street directions and prohibited turns (Figure 2)

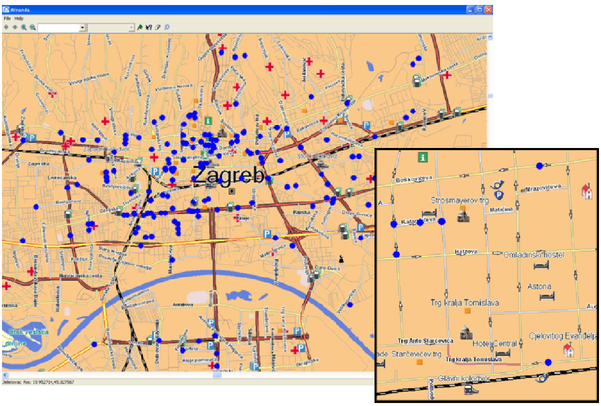


Figure 2 Digital Map (Mireo d.o.o.)

- Due to the above mentioned, the matrix of minimal distances between customers is asymmetric (Figure 3)

	0	1	2	3	4	5	6	7	8	9	10
0	0	1133	1502	1502	1876	1163	1537	964	1170	2095	1118
1	1331	0	338	338	712	594	373	602	601	2892	1915
2	1457	1235	0	0	$\boxed{1413}$	500	279	508	507	2388	1411
3	1457	1235	0	0	1413	500	279	508	507	2388	1411
4	2356	1559	$\boxed{1168}$	1168	0	1424	1203	1432	1431	3917	2335
5	1103	1356	1160	1160	1534	0	1195	154	153	2034	1057
6	1177	955	759	759	1133	1015	0	1023	1022	2728	1751
7	949	1202	1006	1006	1380	667	1041	0	674	1882	905
8	951	1204	1008	1008	1382	669	1043	243	0	1882	905
9	1203	1932	1736	1736	2110	1397	1771	1198	1404	0	988
10	1214	1943	1747	1747	2121	1408	1782	1209	1415	1301	0

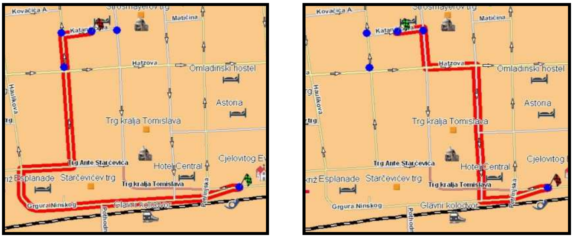


Figure 3 Matrix of minimal distances between customers

- A GPS record of referent/initial solution should be obtained if it exists
- The possibility of exporting the optimized route to a digital map should be examined

The overall procedure of solving practical VRP problems is shown in Figure 4.

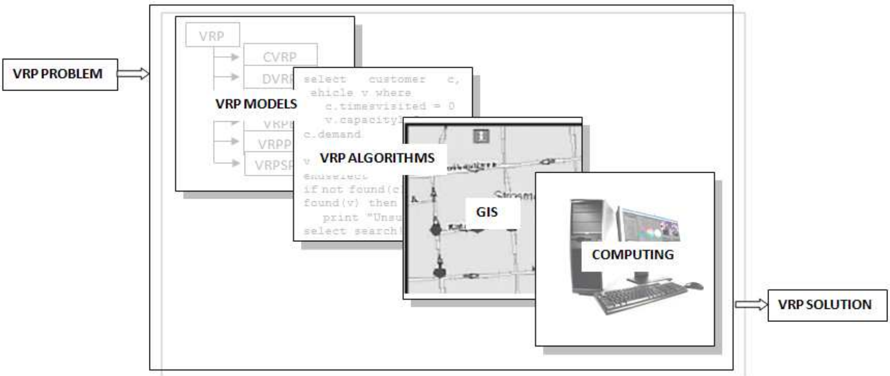


Figure 4 Solving a VRP problem

Optimized routes have a significant impact on transport organization by reducing costs and travelling duration. Practical experiences from literature show that up to 20% in transportation cost reductions may be achieved. Main users are carriers who perform delivery and collection, e.g. transport companies, the postal service, etc.

1.1 CVRP

Waste collection in the city of Zagreb (the capital of Croatia) represents a solved practical CVRP problem for which substantial savings have been achieved compared to the standard experience based routes. The utility company *Cistoca* collects waste at 156 locations, using seven trucks and covering a path with total length 140.93 km. The results of applied algorithms are presented in Table 2.

Table 2 Solutions for practical CVRP problem obtained by Simulated Annealing (SA) and λ interchange (First Best – FB, Global Best – GB) algorithms

ALGORITHM	COMPUTING TIME [s]	VEHICLES	COST [km]
SA	355.27	6	109.632
λ - FB	98.40	6	119.567
λ - GB	222.65	6	110.416

1.2 VRPTW

The second case study involves the national post operator *Hrvatska Posta*. Initial practical solutions were manually designed routes that were not sensitive to variations in the volume of consignments. The geographical locations of the customers and the vehicle routes for delivery and pick-up problems were recorded by a GPS device, and a simulation of the routes was done as reference data for further analysis.

For the delivery of consignments, the postal operator uses 16 vehicles with total route length 240.786 km. Optimization of the delivery problem was performed and the results are presented in Table 3.

Table 3 Solutions for practical VRPTW delivery problem obtained by SA and λ interchange algorithms

ALGORITHM	COMPUTING TIME [s]	VEHICLES	COST [km]
SA	1034	11	230.94
λ - FB	480	11	231.91
λ - GB	1118	11	233.11

Some of the new proposed routes have been tested and recorded with the aim of proving the feasibility of the new routes.

The results show the following:

a) An experienced dispatcher with a good knowledge of urban area routing should be involved in the process of final route construction. For example, in the case of a delivery, if two customers are physically close but on the opposite side of the street and there is an obstacle (prohibited turn) that prohibits crossing the street with the vehicle, a dispatcher can change the route as shown in Figure 5.

Figure 5 Generated route that can be improved by expert knowledge

b) VRPTW solutions greatly depend on vehicle travelling times through an urban area. Travelling times differ for morning delivery and afternoon pick-up due to different network conditions and a travel time prediction model should be made.

3. Travel time prediction model

The travel time prediction model will be developed on the basis of data collected from 297 vehicles of different categories that performed their usual duties in the urban area of the city of Zagreb.
Data collecting was performed during a time period of thirteen months by GPS and GPRS devices that recorded *log time*, *vehicle ID*, *X coordinate*, *Y coordinate*, *speed*, *course*, *GPS status* and *engine status*. These data were cleansed and a map-matching procedure was carried out.
In the GIS database roads are defined at sublink level (link is a road arc from one intersection to another and the sublink is a one-way arc connecting two intersections with respect to the exiting link) so that each road direction can be analyzed separately.
Furthermore, a travel time pattern for work days and at weekends has been analysed to define specific peak hours.
The idea is to identify the main elements that affect travel time and congestion in urban areas so that they can be built into the model to achieve a satisfying level of prediction accuracy.
Data collected for model development from the Meteorological and Hydrological Service for the same area and time period have also been included to enhance model quality. Therefore, data fusion has been done.
GPS data are statistically analyzed to classify roads into five categories and to apply various data mining methods to each category so that the results of the methods can be compared based on road categories.
So far the kNN and ARIMA methods have been considered, but the possibility of applying other methodologies is open.
The following objectives should be observed:

- Investigating the differences between estimation of travel time duration in urban networks as opposed to highways and motorways.
- Studying different available data mining models for travel time estimation and comparing them.
- Finding the most suitable models for travel time estimation in urban traffic networks based on data collected with GPS equipped vehicles.
- Investigating the applicability of the selected models for:
    - The improvement of : - real-time methods and models for VRP algorithms,
- VRPTW algorithms,
- congestion managing systems
    - The application of decision-making processes
- Proposing the most appropriate model for travel time estimation in urban networks based on the performed data mining analysis and performed validation.

The following software tools have been considered as primary support:

- StatSoft Statistica and Data miner,
- Microsoft SQL Server 2005,
- Miranda (GIS),
- VISUM & VISSIM (PTV).

4. Acknowledgements

This work has been realized with the data and digital map provided by *Mireo d.o.o.* Their support is gratefully acknowledged.

References

Caric T. (2004) Improvement in Transport Organization by Usage of Heuristic Methods. *Ph.D. Thesis, Faculty of Transport and Traffic Engineering, University of Zagreb. (in Croatian)*

Cavar I, Galić A, Fosin J, Caric T and Gold H (2004) Development of Distributed System for Practical Vehicle Routing Problems, *International Conference on Software, Telecommunications and Computer Networks, Croatia and Italy,* pp. 99

Cavar I, Gold H and Caric T (2005) Assessment of Heuristic Algorithms for Solving Real Capacitated Vehicle Routing Problems by Analytic Hierarchy Process, *12th World Congress on ITS, San Francisco, USA*

Galic A, Caric T and Gold H (2005) Mars programming language for solving vehicle routing problems, *City Logistics IV, Langkawi, Malaysia*

Gold H, Caric T, Galic A and Cavar I (2005) Vehicle Routing Problem Solving System as CRO-GRID Application, *MIPRO, Croatia*

Biography

Ivana Ćavar received a B.Sc. degree from the Faculty of Transport and Traffic Sciences, at the University of Zagreb, Croatia. Currently, she is a Ph.D. student working at the same Faculty as an assistant. Her interests include combinatorial optimization, VRP, transport planning, GIS, data mining and travel time prediction.

Incorporating Egocentric Routing Preferences into Pedestrian Navigation Devices

Luke Studden[1], Gary Priestnall[2], Mike Jackson[3]

[1]Atkins Limited, Intelligent Space, Euston Tower, 286 Euston Road, London, NW1 3AT, UK
[2]School of Geography/Centre for Geospatial Science, The University of Nottingham, Nottingham, NG7 2RD, UK
[3]Centre for Geospatial Science, The University of Nottingham, Nottingham, NG7 2RD, UK

KEYWORDS: Pedestrian Navigation Device, Egocentric Routing, Mobile Computing, Spatially-Aware.

1. Introduction

Mobile Location Based Services (MLBS) are more widely typified by the mass use and uptake of “In Vehicle Navigation Systems” (IVNS) colloquially termed “Sat-Nav”. For example, 2006 saw a 269% growth in car navigation system ownership compared to 2005 (Poropudas, 2007). The success of IVNS and the consequent increase in user familiarity with mobile positioning technology is now leading to strong interest in the extension of navigation systems to also cover multi-modal travel and pedestrian way-finding. The MLBS sector of pedestrian-oriented navigation services is also encouraged by the development of mobile phones capable of supporting such applications (Brimicombe and Li, 2006; Garmash et al 2001) and GPS-enabled mobiles with assisted GPS services are now part and parcel of standard mobile phone contracts where the device is supplied as a standard component of the service contract.

However to create an “effective” Pedestrian Navigation Device (PND), the navigation industry and data suppliers need to address the issue of “user-appropriate” route generation, or more specifically, the representation of “egocentric walking preferences”. The underlying aim of egocentric routing is for the PND to have the capability of offering a user a tailored route based on their walking preferences. Such preferences may vary widely but might, for example, include considering specific path gradients and the types of terrain which is acceptable to traverse. This capability would enable user groups who were previously overlooked by developers to take advantage of MLBS. User groups include people with reduced mobility such as those confined to wheelchairs or walking aids. The ability to query and return a spatial route through the environment, appropriate to an individuals walking ability or conversely, disability, would promote environmental accessibility and mobility.

At the conceptual level a pedestrian navigation service is a simple extension of the “sat nav” concept and part of a conceived future service proposition that would offer door to door routing and navigation to cover all circumstances. At the implementational level, however, it raises difficulties, many of which are as yet unresolved. These difficulties exist at multiple levels within the service proposition. Road data for car navigation is, in the main, well defined and comprehensively mapped together with associated meta-data that is applicable for routing such as one-way roads. Because of the ability to link the in-car navigation system to a vehicle’s controls limited periods of poor GPS reception can be compensated for based on dead-reckoning calculations and/or the limited route options the road network offers. Snapping a GPS position to a road segment is likewise algorithmically relatively straightforward. The options for routing and the richness of data that is appropriate to be displayed for car navigation is also constrained enabling the device user interface to be kept simple.

When one considers a parallel routing service for the pedestrian, however, many of the natural constraints of road navigation are absent. The pedestrian is not limited in the selection of routes to a well defined network. Whilst features such as pavements (side-walks) may be generalised as a linear network there will be many situations such as parks, large civic squares and countryside which the individual might navigate in multiple ways. The routes are much more likely to pass through areas of limited or absent GPS coverage, dead-reckoning add-ons become more difficult to design in a manner which gives the necessary performance, the richness of meta-data that becomes relevant escalates and the user interface issues correspondingly multiply.

These research challenges are being tackled on multiple fronts but this study is concerned with “egocentrism” and the selection of datasets which will support egocentric pedestrian navigation. The paper describes an approach to achieving this objective with its aim of developing a schema capable of quantifying egocentric walking preferences for integration within conventional and future PND network datasets such as the Ordnance Survey’s pedestrian orientated Integrated Transport Network 2 (ITN2) pilot dataset.

Five objectives are identified in order to achieve this;

Objective 1: *Construct a methodology enabling the capture of real-time pedestrian spatiotemporal behaviour.*

Objective 2: *Using the spatiotemporal data obtained, construct a schema allowing egocentric preference values to be ascertained for individuals.*

Objective 3: *Conduct statistical analysis determining if there is a difference between the "assumed walking preference" (shortest path, quickest time) of PNDs and the collected egocentric preference data.*

Objective 4: *Conduct exploratory analysis to determine if categorisation can be achieved from the collected data.*

Objective 5: *Suggest a data-framework enabling future PND network datasets to incorporate egocentric preferences in route calculation.*

2. Methodology

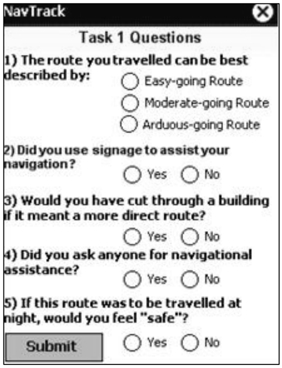
A pilot study consisting of navigational tasks involving four participants over a period of six hours, determined whether participants familiar or unfamiliar with the campus be used in the main experiment. Unfamiliar participants consulted a large proportion of navigational cues such as signage and predominantly followed routes on pavement proving to be considerably longer in length (533.7m further) as opposed to the same tasks navigated by those familiar with the campus. The familiar group used no signage and took considerably shorter routes, traversing over a range of surfaces that can be considered as their personal optima due to possessing an established "mental map" (Raubal and Egenhofer, 1998). It is this established routing which the study aimed to extract and analyse, thus twenty participants familiar with the study area, The University of Nottingham University Park Campus, took part in the main experiment to facilitate the extraction of walking preferences.

A methodology was constructed and executed as advised by Millonig and Gartner (2007), utilising an across method triangulation approach of various quantitative data capture techniques. At the heart of this objective was NavTrack, a custom written PDA application serving as an interactive platform for recording user "trip diaries", recording spatiotemporal movements and administering eight diverse navigational tasks in order to capture walking preferences. NavTrack interfaced with a Holux GPSlim 236 High-sensitivity Global Positioning System (HSGPS) module (Figure 1a) enabling spatiotemporal behaviour to be logged for post-processing (example screenshots are shown in Figure 1 b, c and d).

Figure 1 (a) Dell Axim x51V PDA with NavTrack installed linked via bluetooth™ to a Holux GPSlim 236 HSGS module.

(b) NavTrack Initial task commencement screen.

(c) Screen displayed to the user whilst navigating. Note that NavTrack simply records spatiotemporal tracks and does not offer routing guidance.



(d) Trip Diary questionnaire page allows the user to respond qualitatively to the route they have just travelled in order to build up an overriding user profile and route hierarchy.

NavTrack was written utilising the Visual Studio .NET Compact Framework programming language and served three main functions;

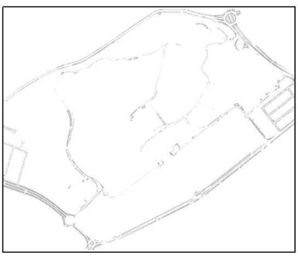
1) To provide an easy to use mobile interface, capable of administering eight navigational tasks.
2) To track via HSGPS, the routes traversed by each participant.
3) To record a “trip diary” in the form of a questionnaire at the end of each navigational task for a “reflective” task summary.

Each participant was tested sequentially using two phases of testing. Phase one involved a “route preference questionnaire” aimed at determining baseline route preferences. Phase two involved a daytime navigation experiment lasting approximately 1hour 30 minutes, conducted during periods of sustained dry weather, representing an individual’s “ideal routes”. Participants were also given a map of the study area with task locations marked as a reference.

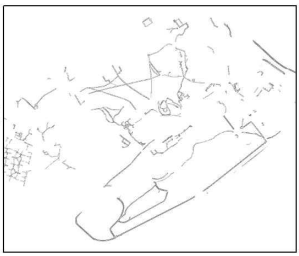
Each task involved navigating to a specific building or feature on campus. After each of the eight tasks, a questionnaire termed by Millonig and Gartner, (2007, p.31) as a “trip diary”, was automatically deployed by NavTrack, recording the participant’s reflection of the route travelled in terms of physical exertion, use of navigational assistance (such as physical signage), navigational preferences and, if travelled at night, perceived safety. To avoid experimental bias, participants were not followed or observed during the experiment but were instructed on what the task required.

3. Quantification of Walking Preferences - A Three Phased Schema

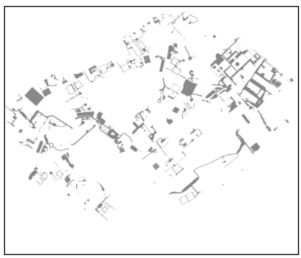
A three level schema was developed capable of describing an individual’s walking preference by combining their spatiotemporal tracks with weighted “pedestrian preference” raster layers. These were constructed from a fusion of OS Master Map data, LiDAR-derived slopes, and land cover data derived from aerial imagery (figure 2). The schema quantifies an individuals walking preference into an “Egocentric Preference Value” (EPV).



Pavement: Weighting 10



Paths: Weighting 9



Manmade Surfaces: Weighting 7

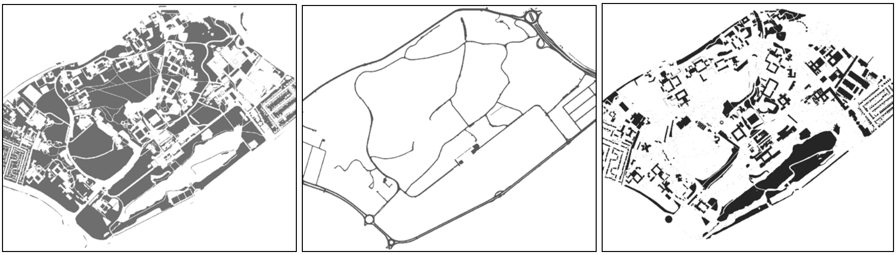


Figure 2. Overview - Feature components are extracted from a multitude of data sources, converted to raster and weighted. The layers are combined using a conventional weighted linear combination technique to form the basis of the pedestrian preference map. **Data Source:** Ordnance Survey Master Map

Phase I: HSGPS track-log download, semi-automated generalisation via map truthing utilising OS Master Map and 12.5cm resolution aerial imagery, and finally conversion into point files for use in phase III.

Phase II: The creation of weighted raster pedestrian preference maps. Data from OS Master Map relating to traversable surface types was classified, converted to raster datasets and weighted on a scale of 0 to 10 (0: least attractive to walk upon, 10: most attractive) so that each 25cm pixel related to an attractiveness for pedestrian movement as with military "*going"* maps. LiDAR-derived slope and additional land cover information from aerial imagery was also factored in using a similar 0 to 10 weighting (0 pertaining to a highly steep slope and 10 representing a completely flat extent of terrain), resulting in a final egocentric preference composite raster map.

The maps were summed via raster math and consequently range on a scale from 0 to 20 (Figure 3); 0 representing an extremely steep or impenetrable area and 20 representing a completely flat, highly attractive region (pavement) to walk upon.

Phase III: Derivation of EPV statistics for each participant by overlaying track-logs onto the weighted preference maps and conducting a raster-point intersection (figure 4). This enables the mean EPV to be determined for each participant by averaging the values resulting from overlaying track log points. EPV values fall in the range 0 - 20 (0 = "arduous exertion", preference to cut-across open spaces; 20 = "easy-going exertion", preference to remain on flatter, established walkways).

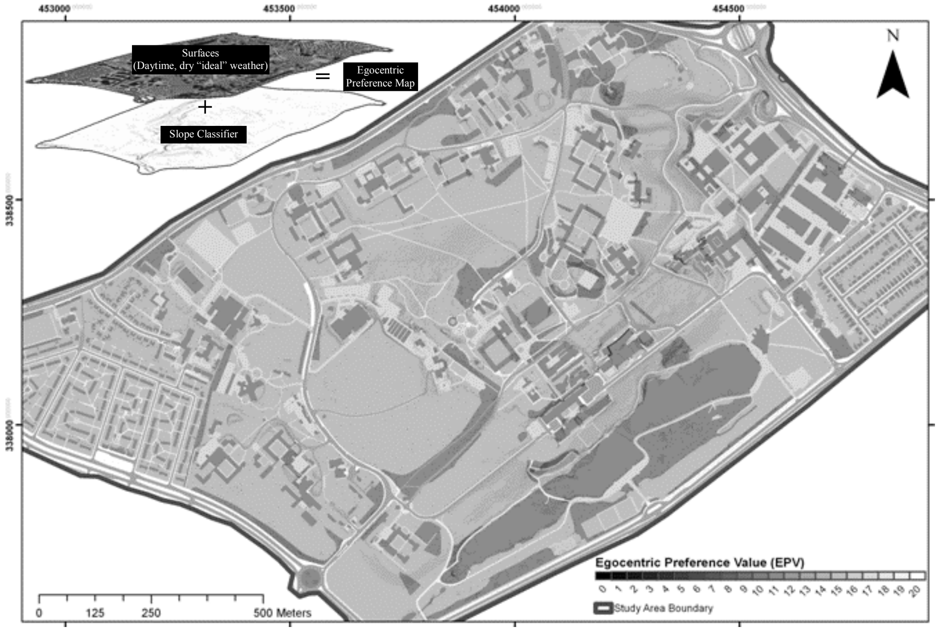


Figure 3. Compiled egocentric preference map. This map is the sum of the weighted pedestrian preference map and a LiDAR derived, weighted slope raster. **Data Sources:** Ordnance Survey Master Map, Infoterra Ltd.

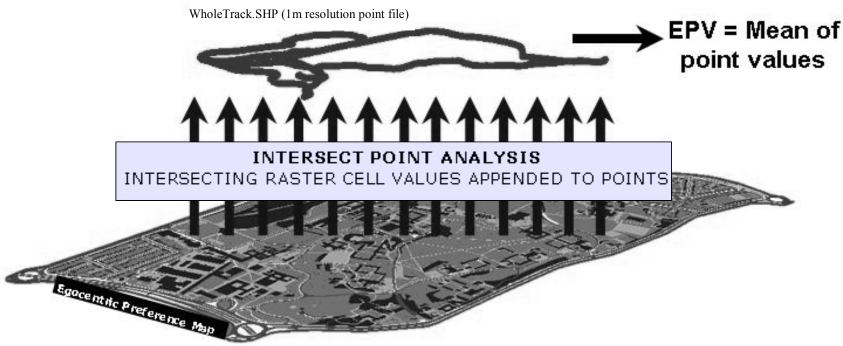


Figure 4. The raster EPV cell values are appended to the 1m resolution point track-logs derived from phase 1. The average point value represents the overall EPV for the participant. **Data Source:** Ordnance Survey Master Map

4. Results

The devised schema is successful in capturing egocentric walking preferences as exhibited by the range of EPVs captured (figure 5) ranging from 16.51 (figure 6) to 17.54 (figure 7), with a mean of 17.02 and standard deviation of 0.282, describing very different, distributed walking styles.

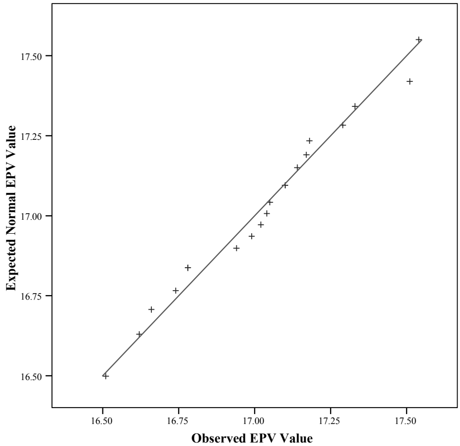


Figure 5. Normal Q –Q plot indicating the distribution of recorded EPV values.

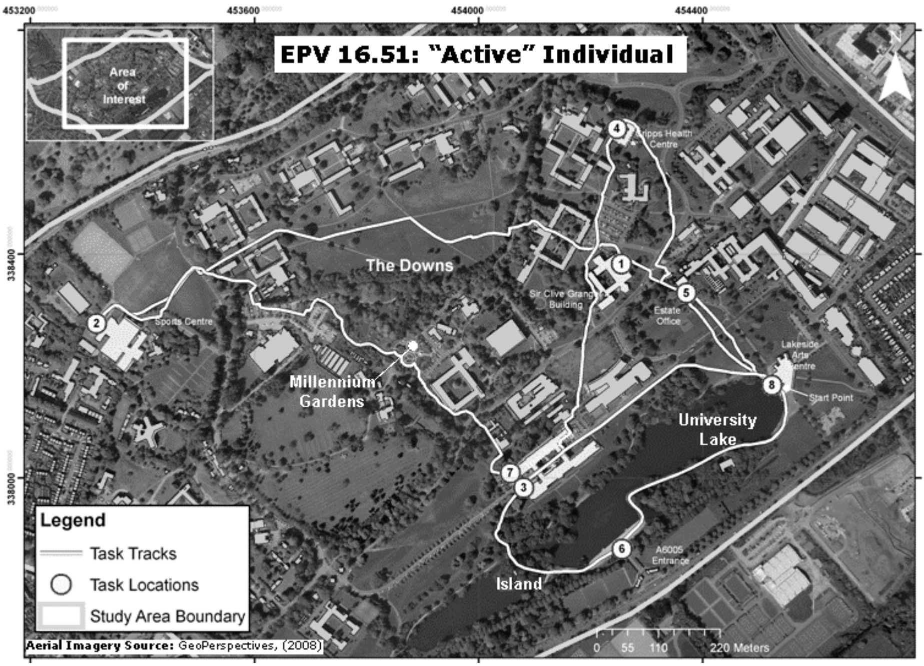


Figure 6. Track-log of a participant exhibiting an "active" walking preference, cutting across grass fields and traversing over steep gradients. The low EPV of 16.51 reflects this. **Data Source:** Ordnance Survey Master Map

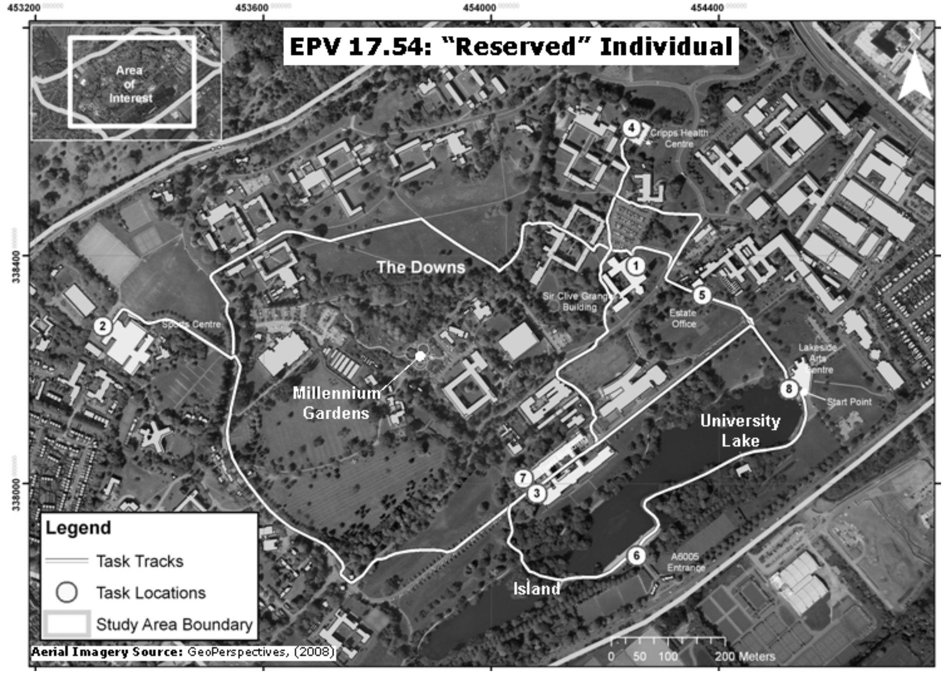


Figure 7. Track-log of a participant exhibiting a "reserved" walking preference, remaining on established walkways and traversing over flat gradients. The derived EPV of 17.54 reflects the difference between this and figure 6 accordingly. **Data Source:** Ordnance Survey Master Map

Whilst routes taken by individuals are based on a range of intrinsic criteria (familiarity with the area, visibility to the destination or aesthetics) they must be considered as the personal optima for the individual, given the temporal and environmental conditions at the time. Although a large proportion of these elements are subjective, they must also translate into a spatial route enabling the individual to convert their preference into a spatial journey. In essence, such criteria is reflected by the routes taken by the participants which the EPV considers. However, these separate entities represent an arm for future inclusion within the EPV schema.

4.1. Testing Conventional PND Assumptions

To place into context the requirement for egocentrism in PNDs, a one sample t-test was set up to determine if the PND assumption of "shortest path, quickest time" (equating to the recorded EPV 16.51) was the over-riding preference of all navigational trips. The null hypothesis was rejected at the 1% level with $p < 0.01$, meaning the alternative hypothesis that individuals had differing navigational preferences was therefore accepted.

4.2. Integration of EPVs into Conventional PND datasets – Egocentric Routing

A similar method emulating the existing distance and time accumulation method can be used to integrate EPV based routing into conventional PND datasets. Each line segment is appended by its EPV by means of a "line-raster intersection" as facilitated by Hawth's tools (Beyer, 2004). Line intersection appends the cell values of a specified raster (EPV raster) to the line segments overlaying it and calculates its mean. Therefore, each line segment can be appended by its EPV value (figure 8) and egocentric routing can commence by taking the mean EPV of all line segments in the routing solution and returning the most applicable journey. Limitations exist, however.

Existing datasets fail to represent pedestrian degrees of mobility and their traversable environment. In the example used above (OS ITN), road segments only are represented and thus the EPV pertaining to that particular surface will be calculated. The nature of pedestrian route preference is one that avoids roads and therefore this EPV augmented dataset is not reflective of the pedestrian environment. Ongoing research is developing "proof-of-concept" systems which utilise alternative data models to support free pedestrian movement.

Figure 8. OS ITN with EPV augmentation for each road segment. Note the differences in EPV values as dictated by the relative slope and pedestrian preference weightings. **Data Source:** Ordnance Survey. **Aerial Imagery:** GeoPerspectives, (2008).

5. Ongoing and Future Research:

In addition to more complete network-based data structures for pedestrian navigation e.g. OS ITN2, there is a requirement to explore a broader range of data types such as complex polygonal areas (which may incorporate time constraints in terms of access/passage), fuzzy zones which may be held in a full-topological model or incorporated as a probabilistic raster cost surface and 3D building structures. Vector data might, for example, represent pavements, roads, paths and tracks whilst raster data represents open areas such as fields and man made surfaces capable of being walked in any direction. The challenges lies in the development of computationally efficient algorithms which can fully utilise both attributed vector networks and raster cost surfaces both of which may hold qualifying attribution and meta-data that need to be incorporated into the routing calculations even in the 2 or 2.5D scenarios. The current imprecision in mobile-phone GPS and positioning technology where positional accuracy can vary significantly at the pedestrian scale over short periods of time, make the definition of an appropriate data model highly challenging at this time. However, as the merging of positioning technologies such as GNSS, other signals of convenience, dead-reckoning technologies and map/image matching improve, then the vision of a user-friendly device which can precisely position you and accurately navigate you from any location to any other based on your physical limitations and preferences becomes more achievable.

6. Conclusions

It was concluded that the EPV was able to represent an individual's walking preference, being able to distinguish a range of different walking types. A method to integrate the EPV for routing in conventional pedestrian navigation systems was introduced. Limitations were identified and a hybrid data model for pedestrian mobility was proposed.

Acknowledgements

Much of the work presented here formed part of an MSc in Geographical Information Science dissertation undertaken within the School of Geography, the University of Nottingham, supported by a NERC Postgraduate Training Grant. Thanks go to Jim Nixon in the Centre for Geospatial Science for advice on experimental design.

The authors are grateful for the support of Gemma Polmear for advice on the development of the mobile application and for the loan of the PDA and HSGPS from the SPLINT (SPatial Literacy IN Teaching) equipment pool.

Ordnance Survey Master Map and ITN data were downloaded via EDINA whilst ITN2 data was obtained through the Ordnance Survey MSc projects programme. Image data was obtained from GeoPerspectives and Lidar Digital Surface Model data from Infoterra Ltd.

References

Beyer, H.L. 2004, Hawth's Analysis Tools for ArcGIS. http://www.spatialecology.com/htools [Accessed on 12/08/07]

Brimicombe, A.J., Li, Y. 2006, Mobile space-time envelopes for location-based services. *Transactions in GIS*, **10**(1), pp.5-23.

Garmash, A., Virrantaus, K., Markkula, J., Terziyan, Y.V. 2001. Developing GIS-Supported Location- Based Services. In: *Proc. of WGIS'2001 – First International Workshop on Web Geographical Information Systems,* Kyoto, Japan, pp.423–432.

Millonig, A., Gartner, G. 2007, Monitoring Pedestrian Spatio-Temporal Behaviour. In *Proceedings of the 30th German Conference on Artificial Intelligence*, Osnabrück, Germany, pp.29-42.

Poropudas, T. 2007, *US GPS navigation market picks up speed.* 13th March, 2007. http://www3.mobilemonday.net/mm/story.php?id=5296 [Accessed 20/08/08]

Raubal, M., Egenhofer, M.J. 1998, Comparing the complexity of wayfinding tasks in built environments. *Environment & Planning B*, **25**, pp.895–913.

Biographies

Luke Studden graduated from the MSc GIS course at The University of Nottingham in September 2008 and presently works for Atkins Intelligent Space, specialising in pedestrian movement modelling and consulting. His research interests include, GNSS based routing, pedestrian navigation and ubiquitous positioning.

Gary Priestnall is an Associate Professor within the GIS research group, School of Geography, University of Nottingham, whose research interests lie in geographic representation, landscape visualisation, and spatially-aware mobile applications for geography fieldwork.

Mike Jackson is Professor of Geospatial Science and Director of the Centre for Geospatial Science (CGS). He was previously Director of the Space Division at QinetiQ and Head of Location Platform at Hutchison 3G ("3"). He is a non-executive director of the Open Geospatial Consortium Inc (OGC) and Chairman, Commission 5 (Networks) of EuroSDR.

The role of user generated spatial content in mapping agencies

Vyron Antoniou[1], Jeremy Morley[2], Muki Haklay[3]

University College London, Gower Street, London, WC1E 6BT
Tel. +44 (0)20 7679 2000
[1] v.antoniou@ucl.ac.uk
[2] jmorley@cege.ucl.ac.uk
[3] m.haklay@ucl.ac.uk

KEYWORDS: user generated spatial content, mapping agencies, Web 2.0, spatial usability

1. Introduction

Since its emergence, the World Wide Web (Web) served as a medium to deliver spatial information, and its impact on geographic information (GI) has been constantly growing. Today, this medium is going through a major transformation. The most striking element in the so called Web 2.0 is the new role of users. Users ceased being simply the consumers of information and instead are taking part in creating, sharing, consuming and disseminating information in the Web. Today's users populate the Web with many types of information from personal thoughts in their blogs, to videos broadcasted at YouTube, to articles in Wikipedia, with some of these integrated in high profile media outlets.

2. User Generated Spatial Content - UGSC

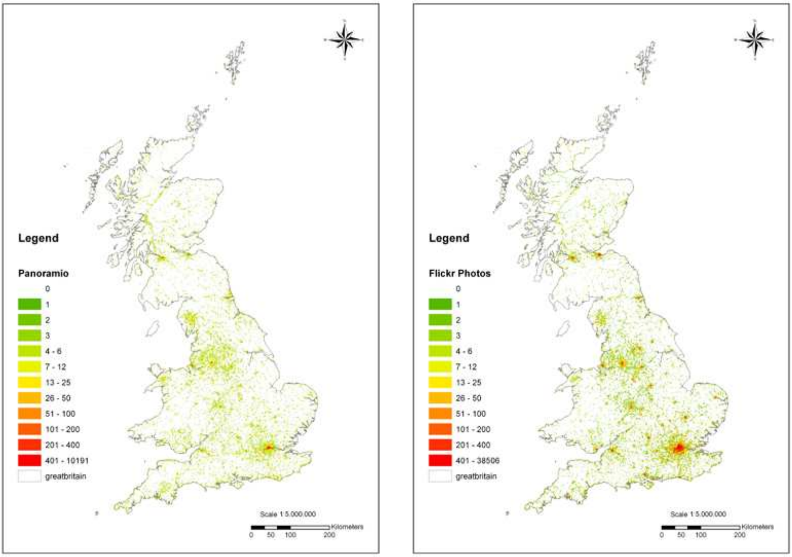
Spatial content has seen a major transition in Web 2.0. Numerous Web mapping applications have been created that allow users to upload, digitize, update or annotate spatial content. Google My Maps, Wikimapia or OpenStreetMap are just a few examples of a new reality in Web mapping described as neo-geography (Turner 2006).

Although researchers have acknowledged the power of user generated spatial content, termed by Goodchild (2007a) as "volunteered geographic information" or VGI, still there is scepticism regarding the phenomenon itself. Starting with the word "volunteered" scholars (Obermeyer 2007, Sieber 2007, Williams 2007, Elwood 2008b, Bishr and Mantelas 2008) support that it can be misleading regarding the possible uses of such data and the intentions of the data providers. Additionally, the quality of such data (Flanagin and Metzer 2008) and its fitness for purpose (Haklay 2008) has drawn research interest, revealing that the use of UGSC can find implementations into many applications. On the other hand though, it must be noted that when examining the social aspect of the phenomenon (Goodchild 2007a, Haklay 2008) it is evident that only people from the privileged side of the digital divide can participate. Still, the inability of traditional methods of spatial data collection to effectively capture and attribute data that are not detectable remotely considerably enhances the importance of the phenomenon (Goodchild 2007b).

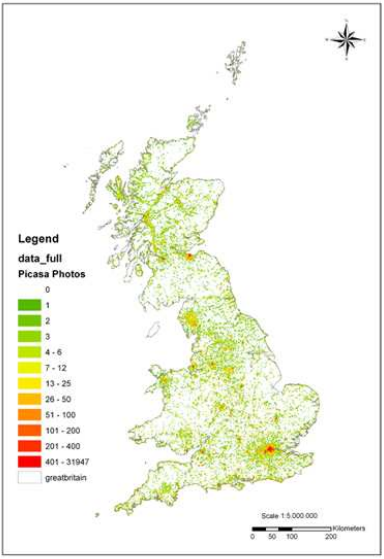
While some research was carried out about the quality of this data, an intriguing question is how traditional mapping organisations (such as national mapping agencies) can utilise this information. Do the procedures of building, updating and auditing spatial databases need to adapt to the new challenges presented by this evolution or should they stay solely in the hands of experts?

To answer these questions, we need to look at aspects such as user participation, bi-directional flow of information and other characteristics of UGSC and examine their impact and their importance in GI. The focus is on how a mapping agency can incorporate UGSC into well established mapping procedures such as data updating, change detection and map auditing.

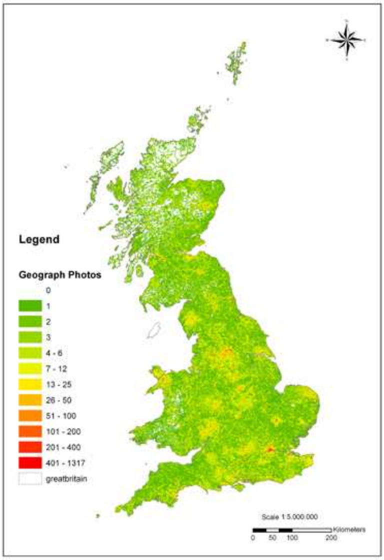
The first step is to examine the nature of UGSC sources. Preliminary work indicates that the ubiquity of spatially-related Web applications can be grouped into two broad families: spatially implicit and spatially explicit applications. The sources which do not directly engage their users into posting spatially-related content (e.g. Flickr or Piacasa Web) are considered as spatially implicit. In contrast, spatially explicit applications urge their users to upload or interact with spatial content (e.g. OpenStreetMaps or Geograph). The second step is to examine the nature of UGSC itself. Issues like spatial usability, distribution, intellectual property rights (IPRs) and data flow is examined. We define the spatial usability of content (picture) posted in a Web application as the ability to produce useful spatial knowledge and thus it should portray geographic features in a recognizable and conceivable way (e.g. a photo of a point of interest or a new building). For example, regarding the spatial distribution, Figures 1a, 1b and 1c shows the distribution in space of geotagged photos for UK from spatially implicit applications (Panoramio, Flickr and Picasa Web), in contrast with the distribution of a spatially explicit source of UGSC such as Geograph (www.geograph.org.uk) (Figure 1d). The findings show that spatially explicit applications can act as sources of GI with high and steady flow of information and with distribution that can cover the needs of a mapping agency.



(a) **(b)**



(c)



(d)

Figure 1. The spatial distribution of (a) Picasa Web geotagged photos and (b) photos from Geoograph (source: www.geograph.org.uk)

Even spatially implicit sources provide coverage that, in some cases, can be adequate (e.g. urban and touristic areas as shown in Figures 1a, 1b and 1c). However, to discover the usable content of significant spatial value a more ‘intelligent’ filtering is needed in order to improve the low signal to noise ratio that characterises implicit sources. In that context, only 18.4% of 500 random geotagged photos uploaded to Flickr are spatially usable. In contrast, when the filtering incorporates tag evaluation the usability increases to 72.6%.

Spatial usability has been examined also for vector encoded data. A comparison between an Ordnance Survey (OS) Master Map dataset and roads from OpenStreetMap has revealed the potentials of UGSC to help mapping agencies to improve their spatial databases if embodied into their mapping procedures. Figure 3 show that such comparisons can reveal missing or misclassified entities in the spatial database of the mapping agency.

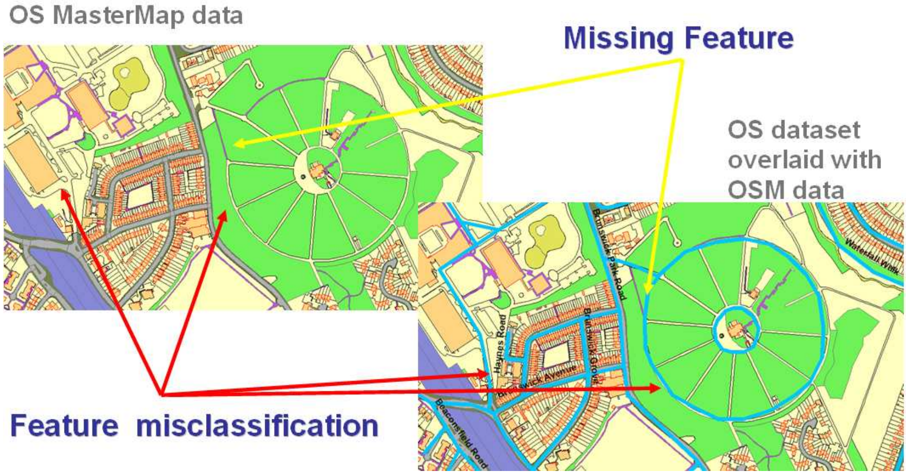


Figure 2. Comparison between Ordnance Survery MasterMap and OpenStreetMap road data

Ongoing research is focusing on the automation of harvesting and filtering of spatial content available on the Web and also on automatic comparison between UGSC and reference data from mapping agencies.

3. Mapping agencies and UGSC

In 1994, an interesting argument was raised by Estes and Mooneyhan (1994) stating that our world is far from being well mapped and that the popular notion supporting the opposite is mistaken. Despite the evolution of geospatial technology and space imagery, current observations (Goodchild 2007b) support the validity of this argument. In fact, mapping and map updating programs are experiencing serious delays in many countries. In that context the value of the UGSC phenomenon is something that mapping agencies can not afford to discard and thus questions can be raised about the best way forward in order to incorporate UGSC into their procedures. The gains of such an approach are that traditional mapping procedures can be further enhanced and existing mapping products can be enriched. For example, UGSC can enrich geographic products like gazetteers which traditionally were partially based on the local knowledge of individual (Goodchild 2008). Moreover, UGSC is expected to initiate the creation of a new breed of spatial data and knowledge taking advantage of the dynamism and temporal aspect of the phenomenon (Elwood 2008a). Finally, change detection which is a principal procedure of mapping agencies in their effort to keep up to date their spatial databases can be enhanced by UGSC. The argument is that either changes happen where people exist (e.g. urban areas) or first changes happen and then people's presence increases (e.g. the construction of a new recreation area). Either way, increased presence of people indicates that a change is likely to be recorded either from an explicit or implicit source. For the latter though, further research should be undertaken in order to examine if similarity exists between the spatial distribution of occurring changes and the clustered pattern of content from spatial implicit sources.

4. Conclusions

Although there are issues raised about the IPRs, quality and usability of spatial data available on the Web it is acknowledged that UGSC is a phenomenon with considerable impact on GI. Initial findings from the study of UGSC sources and the spatial content itself show both the value of the phenomenon and its potentials in assisting longstanding mapping procedures followed by mapping agencies in building more complete and innovative mapping products.

5. Acknowledgements

This work is funded by the Hellenic Ministry of Defence.

References

Bishr, M., Mantelas, L., 2008. A trust and reputation model for filtering and classification of knowledge about urban growth. *GeoJournal* 72:133–135

Elwood, S., 2008a. Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. *GeoJournal* 72:133–135

Elwood, S., 2008 b. Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal* 72:173–183

Estes, J.E. and Mooneyhan W., 1994. Of maps and myths. *Photogrammetric Engineering and Remote Sensing,* 60 (5), p. 517-524.

Flanagin A., Metzer M., 2008. The credibility of volunteered geographic information. *GeoJournal*,72: 72:137–148

Goodchild, M. F., 2007a. Citizens as sensors: The world of volunteered geography. *GeoJournal* 69(4) 211–221.

Goodchild, M. F., 2007b. Citizens as voluntary sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial data Infrastructure Research*, Vol. 2, 23-32.

Goodchild, M. F., 2008. Commentary : Wither VGI?. *GeoJournal* (2008) 72:239–244

Haklay, M., 2008. How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England. Retrieved 30 Oct 2008 from http://www.ucl.ac.uk/~ucfamha/OSMv20data%20analysis%20070808_web.pdf

Obermeyer, N., 2007. *Thoughts on volunteered (geo) slavery*. Retrieved March 13 2008 from http://www.ncgia.ucsb.edu/projects/vgi/participants.html.

Sieber, R., 2007. *Geoweb for social change*. Retrieved March 13, 2008 from http://www.ncgia.ucsb.edu/projects/vgi/supp.html.

Turner, A. J., 2006. *Introduction to neogeography*. Sebastopol, CA: O'Reilly Media Inc.

Williams, S., 2007. *Application for GIS specialist meeting*. Retrieved March 13, 2008, from http://www.ncgia.ucsb.edu/projects/vgi/participants.html.

Biography

Vyron Antoniou is a Captain in the Greek Army and since 1998 serves at Hellenic Military Geographical Service. Currently is an MPhil/PhD student at UCL in the Civil Environmental and Geomatics Engineering department. His research interest is in GIS and particularly in UGSC, spatial data transmission, spatial databases and web mapping.

Jeremy Morley is a lecturer in Geographic Information Systems in the Dept. Civil, Environmental & Geomatic Engineering at UCL. His research is in GIS, remote sensing and terrain modelling. Research aims include especially the greater integration GIS with geoscientific modelling, provision of GIS analysis via the Internet, planetary mapping, and remote sensing techniques, especially synthetic aperture radar.

Mordechai (Muki) Haklay is a senior lecturer in Geographical Information Science in the department of Civil, Environmental and Geomatic Engineering at UCL. His research interests are in public access to environmental information, Human-Computer Interaction (HCI) and Usability Engineering for GIS, and Societal aspects of GIS use. He received his PhD in Geography from UCL. He's a member of ACM, Fellow of the RGS and member of the AGI. Contact him at the Department of Civil, Environmental and Geomatic Engineering, UCL, Gower St. London, WC1E 6BT, United Kingdom, m.haklay@ucl.ac.uk <mailto:m.haklay@ucl.ac.uk>

Geographic Data Mining of Online Social Networks

Alex D Singleton[1]

[1]University College London, Department of Geography, Gower Street, London, WC1E 6BT
Tel.+44(0)20 7679 0510 Fax +44 (0)20 7679 0562
a.singleton@ucl.ac.uk, www.alex-singleton.com, www.spatial-literacy.com

KEYWORDS: Social Networks, Facebook, Social Capital, GIS

1. Introduction

The development and subsequent rise in popularity of websites such as Facebook, Beebo and Myspace offers an unprecedented ability to study the composition and structure of online social networks. Using the Facebook (www.facebook.com) application programming interface (API) this paper presents an application which enables users of Facebook to profile themselves, their neighbourhoods, and their "friends" network according to a series of light hearted metrics. This paper demonstrates how this information can be mined to examine real world geographic associations within online socially networked data.

2. Networks of Social Capital

Although the term "Social Network" (SN) has entered common parlance with the increasing popularity of websites such as Facebook, Beebo and Myspace; social network analysis (SNA) as an area of academic endeavour has an established history rooted in the early twentieth century Sociology and Anthropology (Scott, 2007). SNA is a method of expressing the social environment as "patterns or regularities in relationships among interacting units" (Wasserman, 1994:3). Units can apply to a variety of different objects ranging in scale from individuals to regions; or as organisational units, such as companies or activity based groups. In all social network analysis, actors possess relational data which link them together. The complexity of these relations ranges from a binary connection score (1-0), through coarse directional attribution (+ve, -ve) and finally to strength of association weightings (e.g. a continuous integer). Additionally, actors within SNs may possess attribute data; so in the case of individuals, this could include attributes such as their name, age or gender. SNA has been applied in a variety of areas such as health (Christakis and Fowler, 2007), education (Hawe and Ghali, 2007), crime(Calvó-Armengol and Zenou, 2004) and politics (Crossley, 2007). Within the discipline of Geography there has been limited interest in SNA (Butts, 2003) which presumably relates to the aspatial nature of much SN data, where emphasis in collection is placed on social over spatial relationships. However, this is not to say that Geographers have ignored networks, as they have long been used as a method of conceptualising space (Haggett and Chorley, 1969), utilising both topological and geometric relationships to solve a range of spatial problems. For example, the most efficient way a truck could be routed on a transport network between two locations given a series of pick up and drop off points; or, how a rainfall in a river basin may drain, given a network of tributaries, streams and rivers all with an associated capacity to carry water. SN are commonly visualised as a series of nodes (or vertices) connected by lines (or edges) (See Figure 1a), which can be used to build complex representations (See Figure 1b).

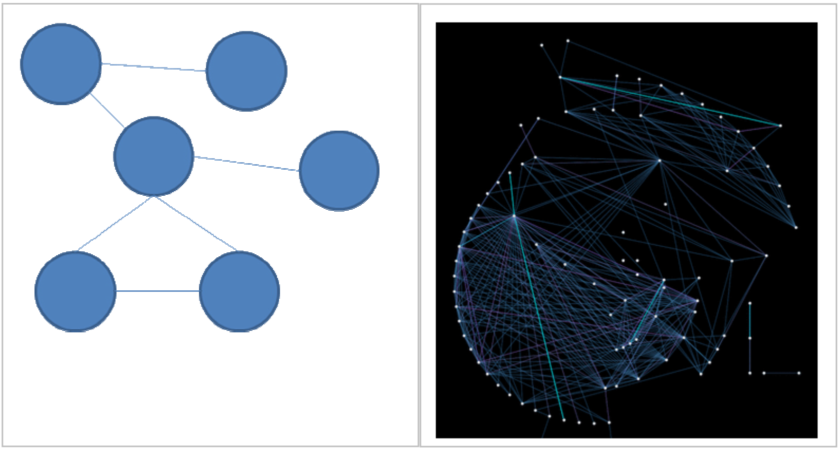


Figure 1. a) A basic SN graph made up of nodes and connecting lines, b) A more complex radial graph of the author's social network on Facebook, visualised using nexus.ludios.net/

A concept which can be related to SN is the idea of social capital (SC). There are multiple formulations of SC (Sampson et al., 1999) which broadly relate to those "features of social life – networks, norms, and trust – that enable participants to act together more effectively to pursue shared objectives" (Putnam, 1995:664-5), often having "forceful, even quantifiable effects on many different aspects of our lives" (Putnam, 2000:23). These quantifiable effects are elaborated by Bourdieu (1986) who describes how the social capital of individuals can be converted to other forms of capital (e.g. economic or cultural), and as such, be used to leverage relative social advantage. For example, parents applying for competitive state school places may learn through their SN of specific information which could improve the chances of a successful application relative to those parents external to the network and thus the privileged information.

3. Social Networking Online

Castells (2000:180) describes how"[n]etworks are the fundamental stuff of which new organizations are and will be made", and these changes are increasingly evident in the many new businesses that have been set up around services popularised online which link individuals together into virtual places through their personal relationships, or sharing of common interests. Recent usage statistics illustrate the success of these services, with Facebook (www.facebook.com) alone attracting 123.9 million unique visitors during May 2008 (McCarthy, 2008). Parallel to this high general demand there has been increased academic interest in social networking websites (Boyd and Ellison, 2008), specifically as evidence suggests that the composition of relationships in online SNs are similar to those offline (Lenhart and Madden, 2007). However, in isolation this statement is geographically naïve (in the UK at least), specifically given those social, demographic and economic disparities in access to new information communication technologies (Longley and Singleton, 2009a, Longley and Singleton, 2009b). As such, any analysis conducted using online SN are only representative of those who have access to the service.

The application presented in this paper uses the Facebook application programming interface (API). Applications created by the API have to be manually installed by a user, and by doing so, users agree to give the software privileged access to a range of information from their personal profile (e.g. age, gender, name etc), their list of friends, and a list of their friends' friends. Facebook profiles are not geocoded, however there are a range of possibilities through which location information can be derived. Firstly, when joining Facebook users affiliate themselves to a Network. There are fixed lists of networks[1] which are approved by Facebook and can be regions, schools or colleges. An installed

[1] http://www.facebook.com/networks/networks.php

application may access network affiliation data about both the user, and that of their friends. Other location information which could be mined includes hometown, school or college information. Finally, an application could request the location of a user.

4. Project Description

The substantive aim of this research is to create automatic methods of assessing the extent that Facebook SNs are constructed through real world spatial associations; and additionally, with reference to social capital, to examine clusters within the SN for their socio-economic status. In these analyses a number of measures of socio-economic status are derived. At the individual level a classification of surnames and forenames was created based on a relative likelihood that they are possessed by people living within areas typically inhabited by those of high socio-economic status. Although an ecological fallacy (Robinson, 1950), without access to individual level information on occupation this provides the closest surrogate for grouping names into a ranking of socioeconomic status. This information was calculated for all names with a frequency of 100 in a modified version of the 2001 Electoral Roll. At an aggregate spatial level where a unit postcode could be identified, the ACORN geodemographic system from CACI was used to infer relative socio-economic status.

The web application was created with PHP and links to the Facebook API. When a user loads the application for the first time they are requested to give their postcode; then the names of their friends within their SN are profiled to examine socio-economic status. An overall score is provided for the users own name, their SN and their neighbourhood (see Figure 2). For added interest a series of rankings compares their friend names with those of celebrities.

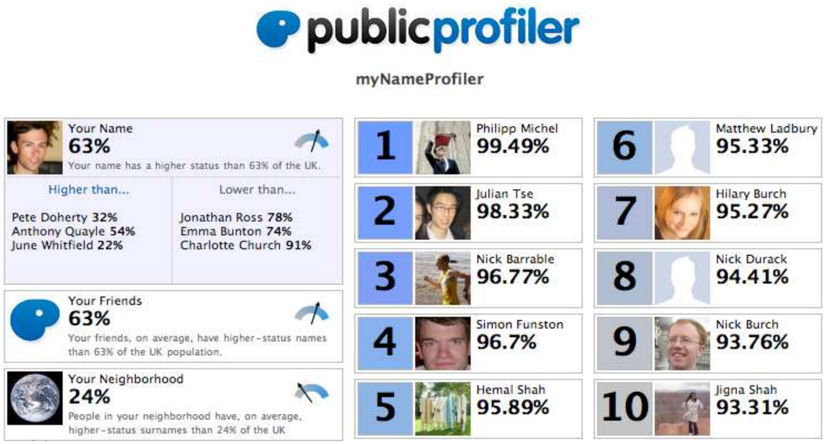


Figure 2. Facebook Social Profiler

The null hypothesis in this research is that there is no real world spatial association between members of online SN (Facebook). This hypothesis is tested by mining spatial location attributes from within a users SN. Two aspects of space are considered. Firstly, where the individuals currently lives (e.g. Home Town, Network Association); and a secondly, the extent that the relationships between individuals may have been formed as function of real world activity, and as such, of geographical significance (e.g. school, university).

The spatial location of members within a SN can be extracted from a series of variables. For each person, a rule based extraction is used to assign the highest geographic resolution reference available. This extraction occurs across attributes in order of: current postcode, current city, current regional network and finally current educational network. Some users do not complete all variables (specifically postcode), and additionally, some users when installing the application supply a postcode. A simple Euclidean distance metric is then calculated between the Facebook user and the location of members of their SN.

A second aim of this analysis is to examine how real world activities influence the formation of the Facebook SN. As such, a method of identifying and then geo-coding structural groups within the SN was required. The Newman community identification algorithm (Newman, 2004) was used to assign members of the SN into groups based on between member linkages. Once these groups were identified, the attributes of the members within each group were examined to extract information which indicated how the groups were formed. Although there are many different ways in which groups can form, two of the simplest to extract and classify are school and university as these activities are often detailed across a series of educational attributes on member profiles[2].

5. Preliminary Results

The Facebook application that collects the data associated with this research project has only just been released, and as such, it is too early to draw substantive conclusions on the relationship between space and the Facebook SN. However, using automated mining of the author's SN, two main clusters of friends were identified, one for university and one for school. Out of 90 friends, 63 friends (70%) were attributed geographic location, 30 friends were from university, 8 from school and 52 classified as other associations.

6. Conclusion

Although at a preliminary stage, the data mining exercise enabled by this research will provide key insight into the geographic processes which influence SN formation. A modified Newman community identification algorithm provided a method of identifying structure within a SN which through data mining could be assigned to real world spatial interaction.

References

BOURDIEU, P. (1986) The forms of capital. IN RICHARDSON, J. G. (Ed.) *The Handbook of Theory: Research for the Sociology of Education.* New York, Greenwood Press.

BOYD, D. M. & ELLISON, N. B. (2008) Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication,* 13**,** 210-230.

BUTTS, C. T. (2003) Predictability of Large-scale Spatially Embedded Networks. In Ronald Breiger, Kathleen Carley, and Philippa Pattison (eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, 313-323. Washington, D.C.: National Academies Press

CALVÓ-ARMENGOL, A. & ZENOU, Y. (2004) Social Networks and Crime Decisions: the role of social structure in facilitating delinquent behavior. *International Economic Review,* 45**,** 939-958.

CASTELLS, M. (2000) *Rise of the Network Society,* Oxford, MA.

CHRISTAKIS, N. A. & FOWLER, J. H. (2007) The Spread of Obesity in a Large Social Network over 32 Years. *N Engl J Med,* 357**,** 370-379.

CROSSLEY, N. (2007) Social Networks and Extraparliamentary Politics. *Sociology Compass,* 1**,** 222-236.

HAGGETT, P. & CHORLEY, R. J. (1969) *Network Analysis in Geography,* London, Edward Arnold.

HAWE, P. & GHALI, L. (2007) Use of social network analysis to map the social relationships of staff and teachers at school. *Health Educ. Res.***,** cyl162.

LENHART, A. & MADDEN, M. (2007) Social Networking Websites and Teens: An Overview. Washington, Pew Internet Project.

LONGLEY, P. & SINGLETON, A. (2009a) Classification Through Consultation: Public Views Of The Geography Of The E-Society. *International Journal of Geographical Information Science,* In Press.

[2] Facebook was started as a college network in the USA, and as such contains extensive options to add educational attributes.

LONGLEY, P. A. & SINGLETON, A. D. (2009b) Linking social deprivation and digital exclusion in England. *Urban Studies,* In Press.

MCCARTHY, C. (2008) ComScore: Facebook is beating MySpace worldwide.

NEWMAN, M. E. J. (2004) Fast algorithm for detecting community structure in networks. *Physical Review E,* 69.

PUTNAM, R. (2000) *Bowling Alone: The Collapse and Revival of American Community,* London, Simon & Schuster Ltd.

PUTNAM, R. D. (1995) Tuning In, Tuning Out: The Strange Disappearance of Social Capital in America. *PS: Political Science and Politics,* 28**,** 664-683.

ROBINSON, W. S. (1950) Ecological Correlations and the Behavior of Individuals

American Sociological Review, 15**,** 351–357.

SAMPSON, R. J., MORENOFF, J. D. & FELTON, E. (1999) Beyond Social Capital: Spatial Dynamics of Collective Efficacy for Children. *American sociological review,* 64**,** 633-660.

SCOTT, J. (2007) *Social Network Analysis,* London, Sage.

WASSERMAN, S., FAUST, K. (1994) *Social Network Analysis: Methods and Applications,* New York, Cambridge University Press.

Biography

Alex D Singleton completed his PhD in November 2007 and is now a Research Fellow in the Department of Geography and Centre for Advanced Spatial Analysis at University College London.

Crowdsourcing Spatial Surveys and Mapping

Andrew T. Crooks [1], Andrew Hudson-Smith, Richard Milton and Mike Batty

[1]Centre for Advanced Spatial Analysis
University College London, 1-19 Torrington Place, London, WC1E 7HB, England
Tel. +44(0)207 7679 1782 Fax +44(0)207 7813 2843
andrew.crooks@ucl.ac.uk, asmith@geog.ucl.ac.uk, richard.milton@ucl.ac.uk, m.batty@ucl.ac.uk
http://www.casa.ucl.ac.uk

KEYWORDS: GMap Creator, MapTube, Google, Geodemographics

1. Introduction

This paper presents the potential of linking the GMap Creator[1] software and the MapTube web service (www.maptube.org) to create near-real time spatial surveys. Three different surveys will be presented which map people's perceptions about certain questions, including the current financial crisis, anti-social behaviour and peoples thoughts on road pricing. Basic results will be highlighted for each and the geodemographic profiles of respondents will be explored.

2: Near Real-Time Spatial Surveys

MapTube allow people to share and view other people's maps as shown in Figure 1, but it can also be used in more innovative ways. For example, as web surveys are often aspatial (e.g. surveymonkey.com), the ability to combine GMap Creator and MapTube offers a simple solution to build spatial surveys for large areas. Figure 2 shows the process of creating the near real-time maps. Users are asked a series of questions and to enter their postcode so that the results can be geo-coded. This is then sent to a web server, time stamped and stored in a database. Every 30 minutes (however, this can be varied) a script is run to create a new shapefile, compiling all the results from a survey, aggregating them into a spatial units (in this case postcode sectors). The shapefile is then passed to GMap Creator along with an XML file containing information including: settings for colour thresholds, maximum level of zoom and the field name of the shapefile for which the map is to be created on. GMap Creator runs creates a series of image tiles which updates the map on MapTube which can then be served back over the internet.

What follows are three surveys which map people's perceptions about certain issues done in association various BBC organisations. For each survey no personal information was collected and participants were reassured that actual locations could not be identified. This was ensured through the use of postcode districts rather than the postcode unit or building address therefore preserving data confidentiality. Used in conjunction with MapTube, it allowed participants and other users to take other information and lay the maps on top of one other.

2.1. Mapping the Credit Crunch

A pilot study was carried out as an experiment to create a mood map of the credit crunch within the United Kingdom in conjunction with BBC Radio 4 iPM show[2] . Based on what is the "singly most significant factor hurting the person the most about the credit crunch", participants were asked to enter the first part of their postcode (postcode sector) so their responses could be geo-tagged along

[1] http://www.casa.ucl.ac.uk/software/gmapcreator.asp

[2] Link to the BBC Radio 4 show: http://www.bbc.co.uk/blogs/ipm/2008/04/mapping_the_credit_crunch.shtml

with one of six options to choose from: mortgage or rent, fuel, food prices, holidays, other, or the credit crunch is not affecting me.

Between 26th April and 29th June 2008 there were 23475 responses to the survey with 48.8% of response saying that fuel was most significant factor hurting the person the most about the credit crunch (Figure 3). However there was spatial variation around the country with more respondents within Greater London saying it was either mortgage or rent, or food as shown in Figure 4.

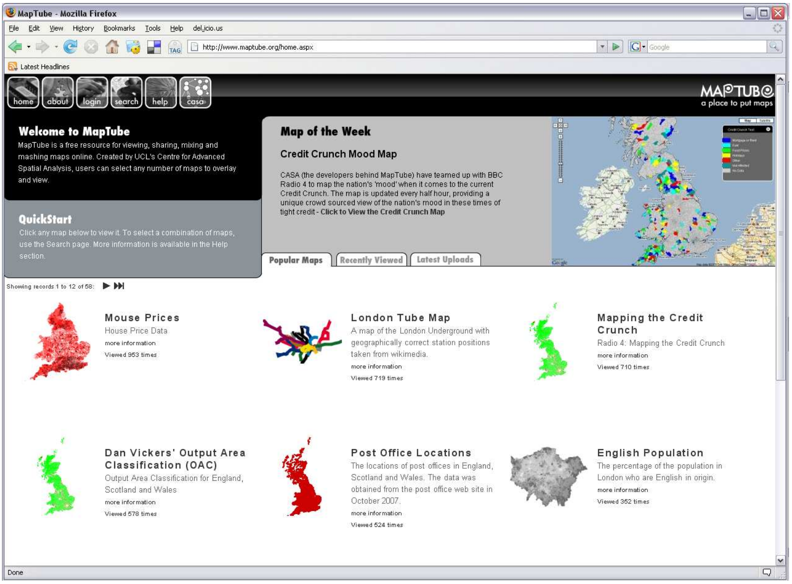


Figure 1. MapTube home page showing the most popular maps.

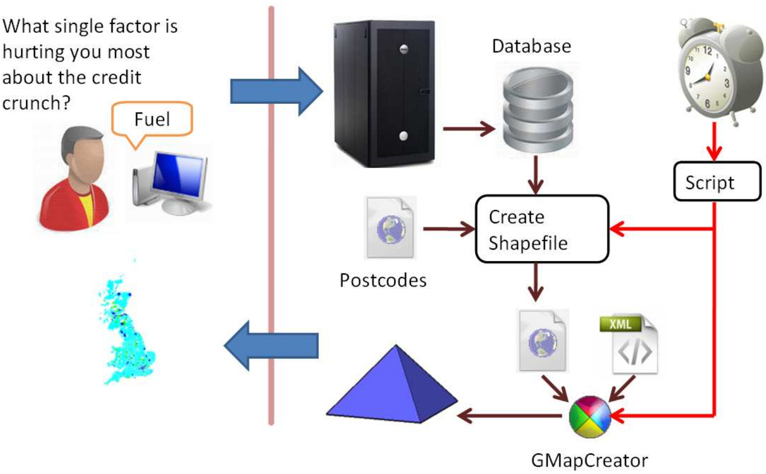


Figure 2. The process of gathering, storing and creation of maps

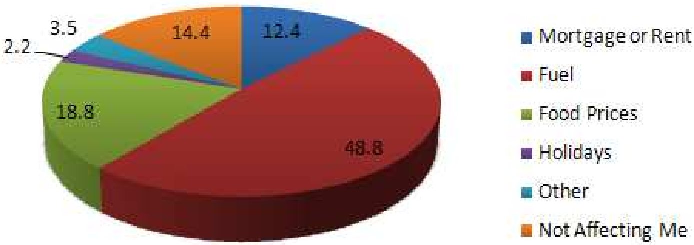


Figure 3. Overall percentages for the Credit Crunch Survey.

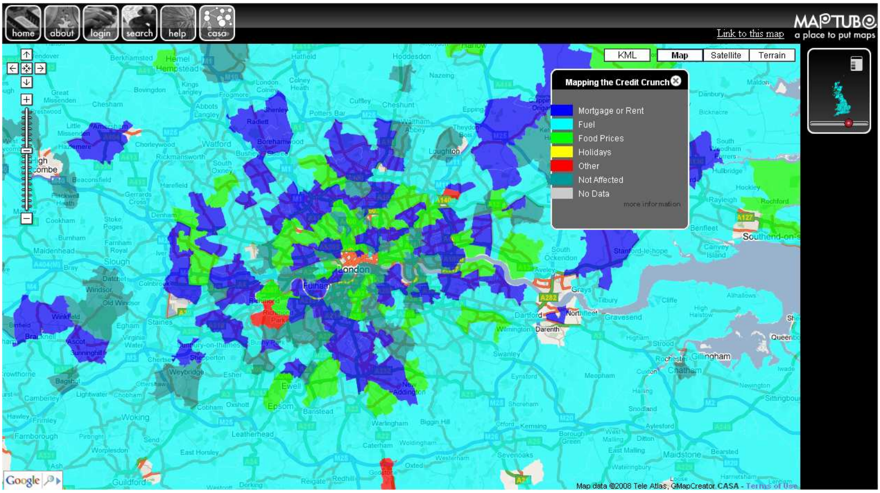


Figure 4. Results of the Credit Crunch Survey Focused Around London (Note: the Colour represents the Most Frequent Response in the Postcode Sector).

2.2. Anti-Social Behaviour in East Anglia

The Credit Crunch Map has since led to BBC Look East, using the system to map peoples perceptions of anti-social behaviour[3]. Each respondent was asked "what problems do you face where you live?" Respondents had five options: drunken youths, noisy neighbours, boy racers, no problems, great community and no problems. The survey ran between 4th July 2008 and 12th September 2008. During this time 6902 responses were received. Figure 5 shows the overall percentages, with 33.7% saying drunken youths with the other categories broken down relatively evenly between 14 to 18%. Figure 6 maps the responses with drunken youths clustering around urban areas such as Norwich and Newmarket.

---

[3] Link to the BBC Look East site
http://www.bbc.co.uk/lookeast/content/articles/2008/07/02/behaviour_feature.shtml

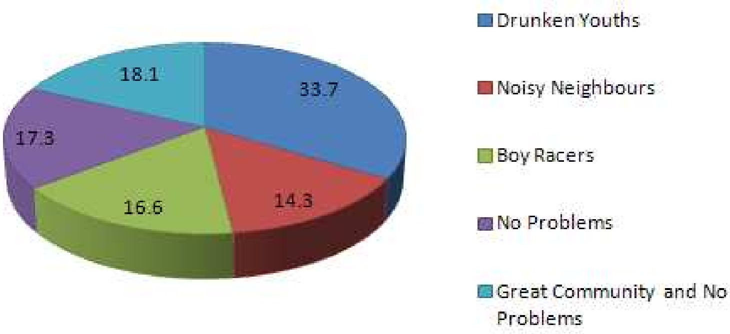


Figure 5. Overall Percentages for the Anti-Social Behaviour Survey.

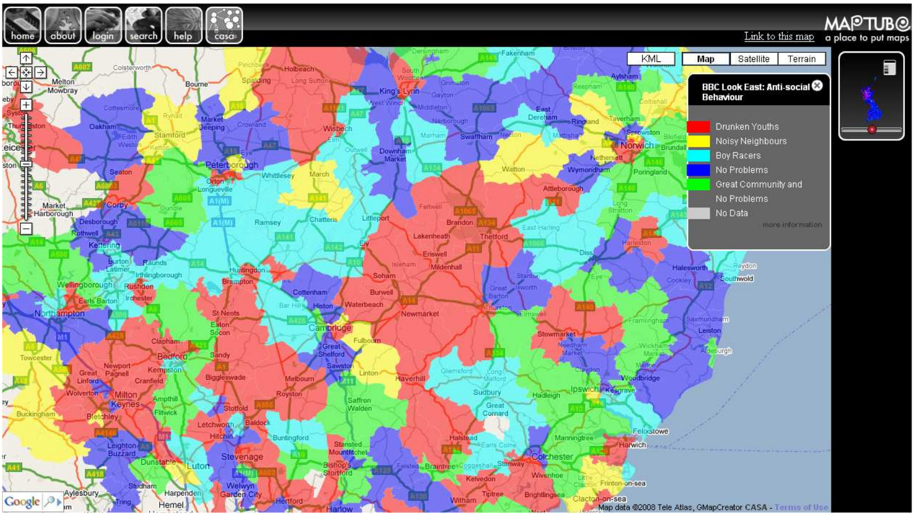


Figure 6. Results of the Anti-Social Behaviour Survey Focused Around East Anglia (Note: the Colour represents the Most Frequent Response in the Postcode Sector).

2.3. The Manchester Congestion Charge

At the time of writing, there was a proposal for Manchester in introduce a congestion charge zone motorists pay to drive in and out of the city at peak times. The BBC North West Tonight program[4] wanted people's reaction to the proposed Greater Manchester congestion charge, from within the city but also people who drive in from outside the region. As these people don't get a vote but may end up paying the charge (subsequently the people of Manchester said no).

People were asked the following question "If a congestion charge is introduced in Greater Manchester, along with significant investment in public transport, will you:" and then asked to select one of the following options: drive and pay the charge, drive at different times, use public

[4] Link to the BBC North West Tonight site
http://www.bbc.co.uk/northwesttonight/content/articles/2008/10/20/201008_congestion_feature.shtml

transport/motorbike/bicycle, work or shop elsewhere, or I am not affected by these changes. The survey began on 14th October 2008. By the 10th December 2008, there were 14933 responses with 46.8% saying they would work or shop elsewhere (Figure 7). This online collaboration provided a unique picture of how well the proposal was going down across the north west of England as the map is updated every day.

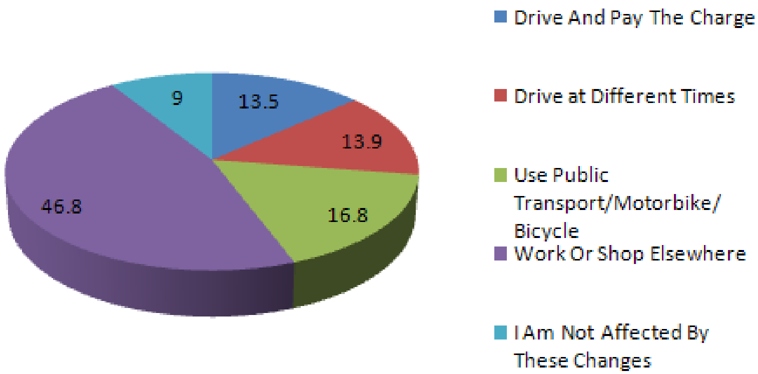


Figure 7. Overall percentages for the Manchester Congestion Survey.

3. Geodemographic Profiles of Respondents

While we only asked for respondents or their first part of their postcode, many entered their full postcode as can be seen in Table 1. We note that this in not a representative sample but it does provide an opportunity to further investigate who is responding to such surveys. To gain this understanding we use two geodemographic classification schemes. First, the Acorn classification from CACI which categorises neighbourhoods based on multidimensional socio-demographic attributes. The second being the e-Society geodemographic classification (Longley et al., 2008) which categorizes neighbourhoods based on their engagement with new information communication technologies.

For the analysis, index scores was calculated. An index score compares the over or under representation of a specific target variable against a base population (e.g. the national average). Where a score of 100 is the national average, 200 is double the national average and a score of 50 is 50% below the national average. From such analysis it is the middle and upper classes who are over-represented within the surveys as shown in Table 2, this potentially relates to demographics of the readers, listeners, and viewers Radio 4 and the BBC news. The over representation of E-business users in the E-society classification (Table 3) suggest many respondents are answering the questionnaire while at work. Furthermore the geodemographic profiles of responses to individual questions can also be explored as seen in Table 4. Across all demographic groups the biggest concern was fuel.

Table 1. Total Number of Respondents to Surveys and Number Who Entered Their Full Postcode.

	Credit Crunch	Anti-Social Behaviour	Congestion Charge
Total No. of Responses	23475	6902	14993
No. of Full Postcodes	15847 (68%)	2749 (40%)	13660 (91%)
No. Which Matched Acorn Classification	14698 (63%)	2109 (31%)	10354 (69%)
No. Which Martched E-Society Classification	13776 (59%)	1952 (28%)	9897 (66%)

Table 2. Index Scores of Respondents by Acorn Category Classification.

Acorn Category Classification	Index Scores		
	Credit Crunch	Anti-Social Behaviour	Congestion Charge
Wealthy Achievers	111	129	85
Urban Prosperity	121	35	64
Comfortably Off	116	136	142
Moderate Means	91	85	137
Hard-Pressed	49	67	75
Unclassified	116	60	83

Table 3. Index Scores of Respondents by E-Society Group Classification.

E-Society Group Classification	Index Scores		
	Credit Crunch	Anti-Social Behaviour	Congestion Charge
E-unengaged	82	104	111
E-marginalised	90	84	123
Becoming engaged	154	112	120
E for entertainment and shopping	138	150	166
E-independents	115	133	99
Instrumental E-users	160	169	118
E-business users	188	163	148
E- experts	196	107	131
Unclassified	35	20	23

Table 4. Percentage of Responses to the Credit Crunch Survey Broken Down by Acorn Category

Answer	Acorn Category					
	Wealthy Achievers	Urban Prosperity	Comfortably Off	Moderate Means	Hard-Pressed	Unclassified
Mortgage or Rent	10	19	13	13	13	11
Fuel	58	30	52	51	47	26
Food Prices	14	21	18	20	23	34
Holidays	2	3	2	2	3	4
Other	3	4	3	2	3	2
Not Affected	13	23	13	11	12	23
Total	100	100	100	100	100	100

4. Discussion

This paper has demonstrated the potential of using GMap Creator and MapTube for near-real time spatial survey thus providing a resource to map the nations opinions to specific questions over space and time both statistically and geographically. The potential of this approach for gathering spatial information is enormous. For example, it could easily be used to gather other information such as fear of household burglary, the quality of primary school education and so on. We consider this in many senses this to be Web 2.0 and Neogeography in action.

However, the geodemographics of the respondents shows there is an inherit bias in who is answering the questions and there is the question to whether or not respondents are influenced by the maps before answering the questions. Further work is to explore how the maps evolve over time, as each response is time stamped and how this relates to news headlines. Additionally, we are currently exploring the geodemographic profiles of each survey in more detail. We are currently re-running the credit crunch with the BBC[5] with slightly different options to the answer

5. Acknowledgements

[5] http://www.bbc.co.uk/blogs/ipm/2008/10/credit_crunch_the_return_of_th.shtml

The authors are grateful to the Economic and Social Research Council for funding through the National Centre for e-Social Science (NCeSS). The authors would also like to acknowledge the support of GLA Economics.

6. Reference

Longley PA, Webber R, Li C, (2008) The UK geography of the e-society: a national classification *Environment and Planning A* **40(2)** pp362-382.

Biography

Andrew Crooks is a GLA Economics Research Fellow in Urban Systems at the CASA.

Andrew Hudson-Smith is Senior Research Fellow at CASA. He led the NCeSS GeoVUE project.

Richard Milton is a research fellow at CASA specialising in computer programming, who along with Hudson-Smith initiated GMap Creator and MapTube.

Mike Batty is Bartlett Professor of Planning at UCL and the director of CASA.

Mapping Future Climate: A Case Study for the Deployment of the Open Source Geo-stack in Scalable Web-based Applications

Philip James[1], Simon Abele, David Alderson

[1] School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, NE1 7RU

Tel. (+44 (0)191 222 6445) Fax (++44 (0)191 222 6502)
philip.james@ncl.ac.uk, www.ceg.ncl.ac.uk

KEYWORDS: Climate, Open Source, Web GIS

1. Introduction

Dealing with the possible consequences of climate change depends on understanding predictions and taking action to mitigate against predicted changes, to adapt, or both. Deciding what action to take will require weighing up risks and benefits and evaluating alternative strategies. Decision makers will range from individuals, through local government, to national governments and intergovernmental negotiators, and in the public sector alone, cover a gamut of professions from engineers and educators to policy makers and scientists. The Department for Food and Rural Affairs (DEFRA) has funded, and continues to fund, projects which produce climate prediction data and scenarios and advice for the UK climate impacts community. The latest set of future climate statistics due for release in early 2009 involves terabytes of probabilistic climate data. The UK Climate Projections (UKCP) project will supply decision makers, through a web application, with access to this data. This paper provides an overview of the development of this web application, focussing on the geospatial tools used in its development and assesses the utility of the open-source geo-stack for scalable web applications.

The main UKCP datasets consist of probabilistic climate change data over land and sea. For the land based data the statistics have been calculated by the Met Office Hadley Centre (MOHC) using inputs from global and regional climate models. The computed probabilistic variables have been downsized to a 25km grid covering the UK land masses and some of the outlying islands. In addition to individual 25km grid square the MOHC has computed aggregate statistics for a number of administrative regions and river catchments. The probabilistic data is provided as monthly statistics or aggregates thereof (Jenkins, 2008). Those applications that require higher temporal resolution can make use of asynchronous execution of a Weather Generator (WG) program that utilises the climate change projections to simulate future weather variables at daily and sub daily levels (Kilsby, 2007). The WG data is provided for up to forty contiguous 5km grid squares based on user selections. The marine data is based on a set of marine polygons and 50km grids corresponding to the various marine climate regions around the UK.

2. Specifications

The specification for the UKCP User Interface (UKCP UI) stated that the application should provide a rich, interactive user experience including map based selections and outputs. In addition, the UKCP UI had to support a wide range of browsers and provide accessible functionality for both expert and novice users. The map based UCKCP UI requirements is listed in Table 1 below.

Table 1: Map specifications for UKCP UI

Specification	Type	Notes
25km grid selection	Input	Interactive map
Catchment selection	Input	Interactive map
River Catchment selection	Input	Interactive map
5km Grid selection	Inputs	Must be contiguous. Max 40. Interactive map
5km grid matching	Process	Drilldown lookup - IDs of underlying 25km climate grid
Geocoder	Input	UK postcode
Probabilistic output on 25km grid	Output	Interactive map
Marine area selection	Input	1x region/grid square only

It is almost expected that today web applications provide the rich, interactive and responsive map interfaces experience typified by popular commercial mapping clients and APIs such as Google Maps. It is through these interfaces that most non-specialist users interact with geospatial data on the Web. The UKCP UI shall undergo scrutiny from a selected panel of interested stakeholders and form the user testbed to offer initial advice on improvements to the UI experience. The specifications state that any implemented solution must allow a minimum number of simultaneous users (c. 1000) and be able to scale to meet possible future demand scenarios.

3. System Architecture

The system is comprised of a browser based client communicating with multiple server side data and processing servers. Spatial data is provided to the UI by a spatially enabled relational database accessed directly and via intermediate Web Map Servers (WMS). Client requests for data and derived outputs are handled by a suite Open Geospatial Consortium (OGC) Web Processing Framework (WPS) standard (OGC 2008) compliant (geo)processes. A combination of synchronous and asynchronous process calls are used to interact with short (e.g. GetPlot() which generates custom plots and/or charts based on user selections) and longer running (e.g. the Weather Generator service) processes respectively. These processes access data and process dependencies via a number of non-public APIs.

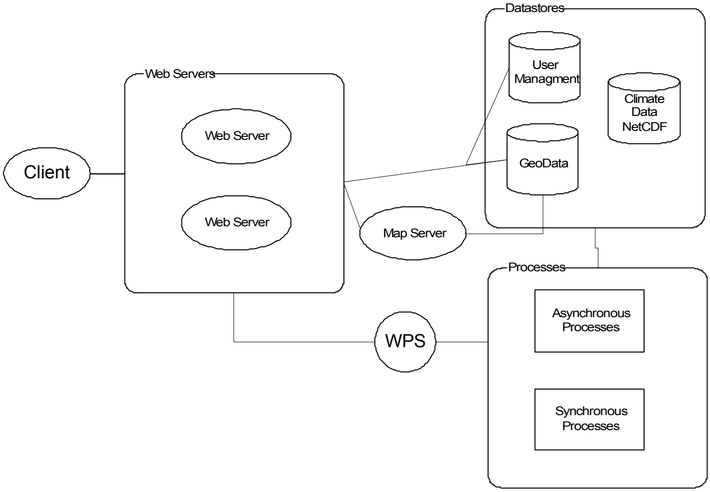


Figure 1: System Architecture Overview

The service and all supporting hardware and software resources employ a variety of techniques to ensure it is both fault tolerant and scalable. These include, but are not limited to; physical and virtual server replication for data and system state, load balancing and management software allow processes and input/output loads to be allocated and distributed across all available resources.

4. Geospatial Tools and Components

The following section outlines a few of the major pieces of geospatial open source technology used by the system. Multiple spatially-enabled instances of PostgreSQL (PgSQL) the open source Object-Relational DBMS are used to store and manage the majority of system data, ranging from user session and system state to geospatial data. Spatial enablement of PostgreSQL is provided by OGC compliant PostGIS spatial database extension which adds support for geographic objects to PostrgreSQL. PostGIS offers efficient spatial indexing (GiST) to enable rapid searching of spatial data, provides a range of spatial predicates (inside, contains, overlaps etc.) that allow for spatial querying and implements a set of spatial functions defined by the OGC Simple Features Profile for SQL (OGC, 2006).

GeoServer (GeoServer, 2008) provides web map serving capability to the system for both static and dynamic geospatial data. GeoServer is an open source software server written in Java designed to publish data from any major geospatial data source using open standards. GeoServer is the OGC reference implementation of the Web Feature Service (WFS) (OGC, 2006 (4)) and Web Coverage Service (WCS) standards and a certified compliant Web Map Service (WMS) (OGC, 2006 (3)).

TileCache, a popular open source tiling component written in Python, is used to provide fast response times for maps of static data (TileCache, 2008). A limitation of WMS which has been highlighted by the introduction of so called ‘slippy-map’ clients, stems from one of their key benefits, the ability to render raster map images of dynamic data for any area, at any scale, in any style, on demand. Most mapping clients render WMS data for every query, resulting in unnecessary and often unacceptable processing and response times. One solution popularised by commercial mapping clients (e.g. Google Maps, MS Virtual Earth) to overcome these problems has been to generate and save (cache) map tiles as they are requested. TileCache does this by acting as a proxy between mapping client (such as OpenLayers) and server (such as any WMS-compliant server). TileCache intercepts new requests for maps and returns pre-rendered tiles from disk if stored, or calls the source WMS server to render new tiles as necessary.

The client side UI uses a highly customised version of the open source map client OpenLayers which provides both data selection and data output viewing capability on the browser. The OpenLayers source code was modified to allow for constraining grid box selections and for handling climatic data.

5. Analysis and Discussion

Currently the UKCP UI is in private beta, with public release due in spring/early summer 2009. Initial tests suggest that the system and software architecture is stable and load testing has not yet yielded any serious issues. The open source geo-stack is now a robust platform and is able to deliver commercial standard, scalable web applications. The factors behind this, whilst not unique to the “geo” world perhaps suggest why certain open source developments provide successful implementations and others fall by the wayside. The initial factor is that the geo-stack is built on a solid platform of well tested and supported libraries that provided much of the core geo-related functionality e.g. GDAL, Proj4, Geotools, OGR, GEOS. Another influence is an effective standards body to push interoperability. The open source tools provide the benchmarks and reference implementations for compliance with international geo-standards. Standards compliance is an acknowledged advantage in some areas (e.g. public funded developments) and can also provide a ready made framework for application development, compressing development schedules. The nature of the types of open source developments also play a role in the robustness of the open source tools. An analysis of the open source geo software and the activity, contributors and market penetration suggests that initial “success” of open source projects is founded on the motivation behind the initial development.

1. Correct tools are not available e.g. GRASS (GRASS 2006)
2. The emerging standards are not well supported by proprietary software vendors e.g. GeoServer, Mapserver.
3. There is an identifiable gap in a software stack e.g. OpenLayers
4. Open source development is used to create indirect revenue e.g. PostGIS/OpenLayers

All of the open source tools used in this project have highly active development teams with a relatively small team of core developers (OHLOH, 2008 (1)). Many of these tools are crucial to the commercial wellbeing of companies founded to exploit the opportunities in supporting, developing and using open source tools (MetaCarta, Refractions Research, OpenGeo) and it is often from the ranks of these types of companies that the implementations of new open standards emerge.

Acknowledgements
The UKCP Data Delivery Program is funded by DEFRA under the CEER/06/07 contract. The project is led by the British Atmospheric Data Centre (BADC) part of the Science, Technology and Facilities Council (STFC).

Bibliography
Jenkins, 2008, Jenkins G., Perry M. and Prior J., The climate of the United Kingdom and recent trends, UKCIP 2007
Kilsby, 2007, Kilsby,C.G, P.D. Jones, A. Burton, A.C. Ford, H.J. Fowler, C. Harpham, P. James, A. Smith, R.L. Wilby, A daily weather generator for use in climate change studies, 2007,Environmental Modelling and Software, 22(12) 1705-1719
OGC, 2008, Open Geospatial Consortium, OGC Web Processing Standard, OGC 2008
Mornjian, 2000, Mornjian, B., How PostgreSQL Rose to Fame, 2000, http://www.oreillynet.com/pub/a/network/2000/06/16/magazine/postgresql_history.html?page=2, Visited 01 10 2008
OGC, 2006, OGC, OpenGIS® Implementation Specification for Geographic information - Simple feature access - Part 2: SQL option, 2006Open Geospatial Consortium
Stolze, 2003, Stolze, K., SQL/MM Spatial: The Standard to Manage Spatial Data, Relational Database Systems, 2003
OGC, 2006 (3), Open Geospatial Consortium, OGC Web Map Service, 2006 Open Geospatial Consortium,Visited 17 12 2008
OGC, 2006 (4), Open Geospatial Consortium, OGC Web Feature Service, 2006 Open Geospatial Consortium,Visited 17 12 2008
GeoServer, 2008, GeoServer, GeoServer Information, 2008, http://geoserver.org/display/GEOS/Welcome, Visited 17 12 2008
TileCache, 2008, MetaCarta, TileCache Information, 2008, http://www.tilecache.org, Visited 17 12 2008
GRASS, 2006, GRASS Development Team, Grass Home, 2006, http://grass.osgeo.org/devel/grasshist.html, Visited 12 06 2008
OHLOH, 2008 (1), Ohloh, Ohloh, 2008, http://www.ohloh.net/projects/, Visited 12 06 2008

About the Author
Phil James is a Senior Lecturer in Geographic Information Science at the School of Civil Engineering and Geosciences, Newcastle University. He heads the Ilab-Geo and leads the Geo-informatics research group. His work largely involves attempting to implement rather than just theorise, combining current and emerging technologies, standards and tools. Currently the main area of activities include sensor grids, web based services, grid computing and open source software development with applications in transport, carbon modelling, climate data management and deployment and sensor and grid interactions.

Wading Through Derwent Water: Taking Digital Terrains from the Real World to Second Life.

Nick Mount[1], Gary Priestnall[1], Dan Weaver[1], Andy Burton[2]

[1]University of Nottingham, School of Geography, University Park, Nottingham, UK, NG7 2RD
Tel. +44(0)115 951 5438 Fax +44 (0)115 951 5249
nick.mount@nottingham.ac.uk,
http://www.nottingham.ac.uk/~lgzwww/contacts/a-z/index.phtml?name=mount
[2]Institute of Engineering, Surveying and Space Geodesy, University of Nottingham
KEYWORDS: second life, neo-geography, visualisation, scale

1. Introduction

The increasingly-rapid development of Web 2.0 technologies is having a major impact on the ways in which geographic information is being presented and accessed. The dominance of desktop GIS experienced throughout the 1990's is now being replaced by new, freely available tools for viewing, creating, interacting with and publishing geographic information. Accompanying this is a change in the profile of the geographic information user together with the scope of audience that can be reached by them; and these users are unencumbered by the constraints of traditional geographical standards and methodologies. The result is a renaissance in geographic information (Hudson-Smith and Crooks, 2008) and the evolution of a 'neogeography' that does not necessarily operate within the constraints of professional geographers (Eisnor, 2006). Nowhere, perhaps, is the potential for neogeography greater than in 3D virtual worlds such as Second Life (SL) (Rymaszewski *et al.,* 2007).

2. Space and location in Second Life

Second Life is a virtual world, maintained through a combination of client software and hosting servers. Users access the world through avatars in a way that is similar to various multiplayer online games (Figure 1). However, it has the unusual quality that nearly all of the content is user-provided. To this end, the client software includes facilities for building and rendering 2D and 3D entities and this is beginning to be used as an opportunity for the presentation of interactive geographical data sets to a new, potentially enormous user group (SL currently has in excess of 4 million users).

Figure 1. The first author's SL avatar 'wading' through Derwent Water; part of a DSM covering the area around Cat Bells (background), UK Lake District.

In Second Life spatial referencing is not constrained by the fixed three-dimensional figure of an oblate sphere, rather the entire world is created on a planar surface referenced by the servers used to store the 3D entities which make up the world (Figure 2). Consequently, there is no horizon,no true concept of diurnal cycles nor a universal time.

The surface of SL is divided into regions, each 256 m x 256 m and extending to 1000 m elevation. Within each region location is given relative to a set of x,y,z Cartesian coordinates with the origin located at the bottom left of region. Regions are identified by a textual name which refers the SL client software to the server on which a region's entities are stored. Together the region name and x,y,z coordinates provide a SLURL (Second Life URL); a unique reference to space within the SL world. The result is that Tobler's Law breaks down at the region scale and traditional geographical relations between regions cease to exist.

This results in enormous challenges for spatial analysis, and the restrictions of the SL client software creates numerous technical issues that impact the selection of appropriate scales and resolution for geographic data. It is these latter issues that are the focus of this paper.

Figure 2. The Second Life 'world'. The world, also known as the 'grid' is composed of large numbers of tessellating land parcels known as regions. Each region is the equivalent of 256 m x 256 m in area.

3. Background to the study

The work reported here forms part of the JISC-funded Design of Learning Spaces in Virtual Worlds (DELVE) project (Minocha *et al*., 2008a; Minocha *et al*. 2008b, **)** (Open University and University of Nottingham) and HEFCE-funded SPLINT (SPatial Literacy IN Teaching) centre of excellence (University of Leicester, University of Nottingham and UCL). This work seeks to investigate how the levels of reality and immersion offered within SL can be used to design effective learning environments; involving comparisons of digital surface models (DSMs) within semi-immersive virtual-reality (VR) (see Burton et al., 2008) and SL environments. A $4km^2$ area around Cat Bells in the UK Lake District is used as a case study (Figure 3). Data for the DSM comprise the Next Map, 5m DSM and aerial photography at 12.5 cm resolution comprise the data for the image drape.

4. Building Digital Terrains in Second Life

All three-dimensional objects within SL are built from structures called primitives; a convolution of sweeps lofts and extrusions (Joy, 2008). Primitives are generally used to construct buildings and other physical entities within the SL environment. However, a variant called a sculpted prim (commonly known as a sculptie) exists that is a three-dimensional mesh whose surface morphology can be determined by texture files and to which drapes can be added. Consequently, digital surface models can be used to determine a sculptie's morphology following conversion to the necessary texture file format and resampling of original DSM and imagery (Figure 4).

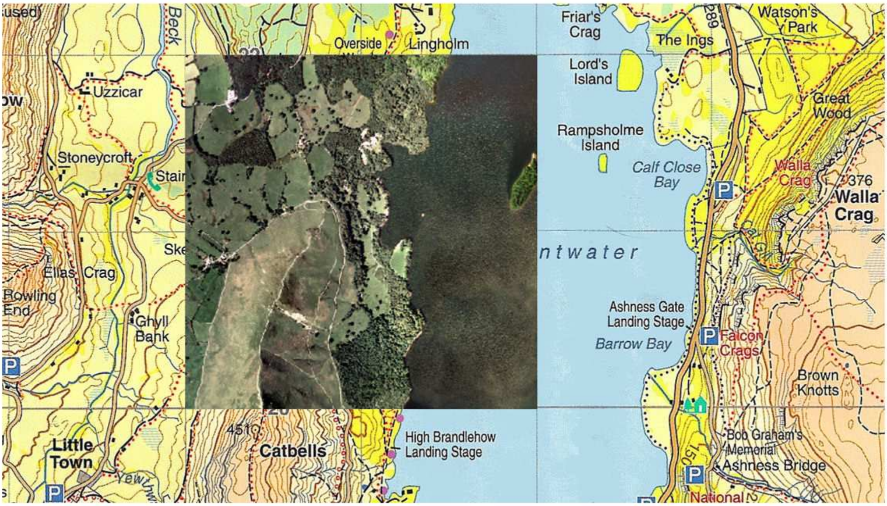


Figure 3. The area around Cat Bells in the UK Lake District used in the experiment.

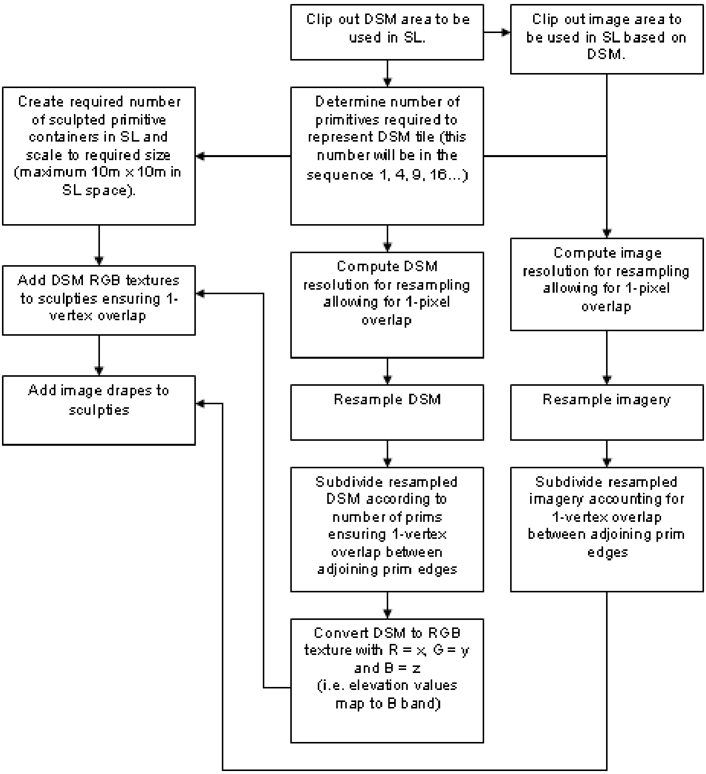


Figure 4. The work flow for creating a draped DSM in SL.

4.1. Limitations of sculpties

Whilst sculpties offer a means of integrating digital surface models into the SL environment they are highly restricted:

- A single sculptie can be any size but can not exceed 10m x 10m in SL area, although several, overlapping sculpties can be stitched together to make larger surfaces;
- Where multiple sculpties are used, an overlap of at lease 1 vertex must be maintained to avoid artefacts in the rendered model
- A sculptie's resolution is restricted to 64 x 64 vertices and can be draped with an image that does not exceed 1024 x 1024 pixels;
- Interpolation between vertices uses a variation of the b-spline algorithm, with no additional options available;
- The level of detail with which a sculptie is rendered depends on distance from the viewing location (usually an avatar or camera position). Proximal to the avatar sculpties will render at 64 x 64 vertices, decreasing to 32 x 32 and 16 x 16 vertices with increasing distance. Beyond about 100m in SL space, the sculptie will not render;
- Sculpties must be contained entirely within a single region.

As a consequence it is problematic to create anything close to 1:1 scale representations of terrain morphologies for anything but very small areas that do not exceed 1000m elevation range. Even if large numbers of sculpties are stitched together, the constraints of the regional architecture make it difficult to build 1:1 terrains greater than 0.125 km^2. Coupled with this, rendering restrictions mean that only proximal locations within the terrain are rendered, resulting in a highly restricted visualisation in which perspective is highly compromised. Consequently, scaled terrains represent the most feasible solutions, particularly where large digital surface models are required that can be rendered over large distances. In the case of the DELVE project, where the location of wind farms on a landscape was key, scaled digital terrain models, represented the only feasible approach for recreating the Cat Bells terrain within SL.

5. Scale and resolution of DSMs in SL

5.1. DSM Resolution in SL

The scale and resolution with which a digital surface and image drape data are presented to an avatar in the SL environment is dependent on the area covered by the original data, the area covered by the SL model, the vertex density of the sculpties used to construct it, the size of the overlap between adjoining sculpties and the level of detail provided by the SL rendering engine.

The DSM resolution of an SL model (DSM_{res}), relative to an avatar, can be estimated according to the following formulae for the simplest case of a square digital surface model tile with a single vertex overlap. Amendments can be made for rectangular models.

$$DSM_{res} = \left[\frac{\sqrt{A_{DSM}}}{\sqrt{n_v} + \left(\sqrt{n_p} - 1 \cdot \sqrt{n_v} - 1\right)} \right] * \sqrt{\frac{n_v}{LOD_{DSM}}} \tag{1}$$

Where;

DSM_{res} is the rendered resolution of the DSM in SL space (m)

A_{DSM} is the original area of the DSM tile (m^2)

n_v is the total number of vertices per sculpted prim (i.e. 64 x 64 vertices = n_v = 4096)

n_p is the total number of sculpted prims used to build the DSM in SL

LOD_{DSM} is the level of detail (in total number of vertices rendered per prim) with which the DSM is rendered in SL

A reduction in DSM resolution results in poorer representation of terrain morphology but the qualitative impact is relatively low. The b-spline interpolation used in SL results in general smoothing with loss of breaks in slope and degraded terrain definition (Figure 5). However, this loss of information is most acute proximal to the avatar position as reduced rendering resolution makes comparative difference less marked distally.

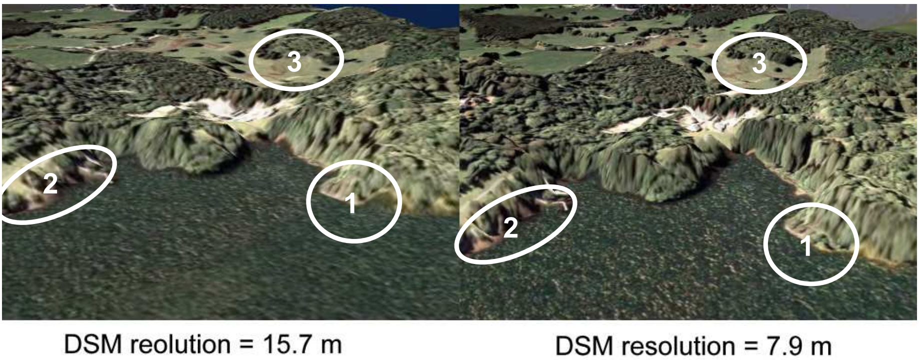


Figure 5. The effect of DSM resolution. Significant morphological detail is lost proximal to viewing location including the loss of low angle slopes at the land / water interface (1), smoothing of steep slope angles (2) and loss of definition of small topographic features such as the hill at (3).

5.2. DSM Scale in SL

The representative fraction of the DSM scale is the ratio between the original DSM size and its size in SL. However, the requirement for overlaps between multi-prim DSMs introduces complexity into the computation of this ratio and means that the representative fraction is seldom a simple multiple:

$$DSM_{scale} = \frac{1}{\sqrt{A_{DSM}} \Big/ \left(\sqrt{A_p \cdot n_p}\right) - \left(\sqrt{\frac{A_p}{n_v}} \cdot \sqrt{n_p} - 1\right)} \tag{2}$$

Where;

A_p is the size of each sculpted prim (m^2).

Scale affects the size of the DSM in SL relative to the avatar (Figure 6) and in general, large scale DSM require large numbers of prims. Where the area of a terrain to be built in SL is limited (as in the Cat Bells DSM) scales of up to 1:50 are quite possible using relatively low numbers of prims. However, where the area of terrains is larger and one wishes to preserve a full perspective view of the terrain, it is necessary to consider the rendering restrictions of SL. Where many prims are required to build a large-scale DSM, those prims distal to the avatar will not render, and in periods of high user activity, may render very slowly or at a reduced resolution. Therefore, the choice of DSM scale is a balancing act between scale of terrain and the rendering restrictions of SL.

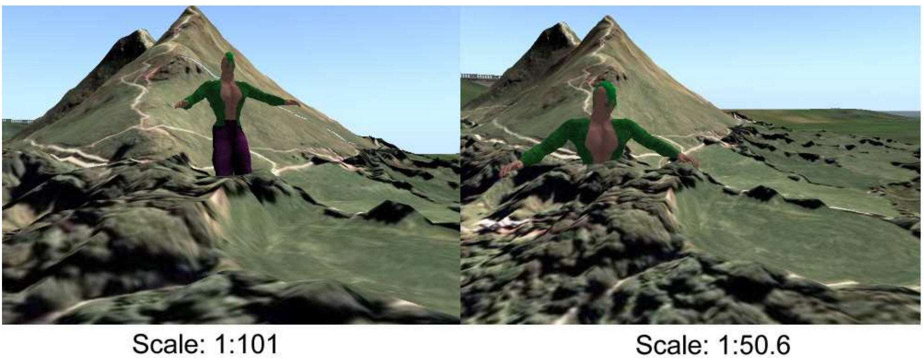


Figure 6. The effect of scale. Scale has an obvious impact on the size of the terrain relative to the avatar. The impact of the necessary prim overlaps can be seen in representative fractions that are not simple multiples.

5.3. Image drape resolution in SL

SL is capable of draping each prim in a DSM with any image of resolution up to 1024 x 1024 pixels, and adjoining prims do not need to have drapes of the same resolution (Figure 7). A version of Equation 1, modified to account for the number of pixels lost due to prim overlaps, is required to compute the resolution of the image drape (I_{res}):

$$I_{res} = \left[\frac{\sqrt{A_{DSM}}}{n_{pix} + \left(n_{pix} - \left(\sqrt{\frac{n_{pix}}{n_v}} \right) * \sqrt{n_p} - 1 \right)} \right] * \sqrt{\frac{n_{pix}}{LOD_I}} \tag{3}$$

Where;

n_{pix} is the total number of pixels in the drape assigned to each prim

LOD_I is the level of detail (in total number of pixels rendered per prim) with which the image drape is rendered in SL.

Figure 7. DSM draped with 0.98 m resolution imagery (right) and 0.49 m resolution imagery (left). Second Life's rendering engine is clearly capable of rendering the improved imagery detail.

6. Conclusions

Second Life and other neo-geographic worlds offer new opportunities for the presentation and integration of geographic data. However, the architectural constraints within the client software result in notions of space and time that do not necessarily conform to those in the real world. These architectural constraints also produce a number of technical difficulties when converting geographic information from GIS data formats to those that can be used and rendered properly in SL. Moreover, the neogeography focus on display loses the opportunity to investigate and create meaning (ie analysis), such that the professional spatial scientist (and desktop GIS) are far from redundant. Contrary to the views of some, in the evolving world of neo-geography, there remains a very clear role for the professional geographer.

7. Acknowledgements

This work has been funded by the Joint Information Systems Committee (JISC) Teaching and Learning Innovations Grant scheme and the HEFCE-funded Spatial Literacy in Teaching (SPLINT) centre of excellence. Digital data for the development of the virtual model were obtained from Intermap (the DSM) and Getmapping (aerial photography). The backdrop in figure 3 shows a portion of the Harvey's British Mountain Map: The Lake District (Reproduced with permission from scanned paper mapping, ©Harvey 2008).

References

Burton A, Priestnall G, Polmear G and Mount NJ (2008) Real-Time Landscape Visualisation: Experiences in a Teaching and Learning Context. *Proc. 16th Annual GISRUK Conference, Manchester Met. University*, 308 – 315.

Eisnor D (2006) Neogeography, http://www.platial.com, accessed 17th November 2008.

Hudson-Smith A and Crooks A (2008) The Renaissance of Geographic Information: Neogeography, Gaming and Second Life. *CASA Working Paper* **142**, *UCL, London.* 16 pp.

Joy B (2008) How SL Primitives [Really] Work, http://www.realityprime.com/articles/how-sl primitives-really-work, accessed 24th November 2008.

Minocha S, Kear K and Mount NJ (2008a) *Design of Virtual Learning Spaces in 3D Virtual Worlds: Project Plan*. Joint Information Systems Committee, 40 pp. Available Online: http://www.jisc.ac.uk/media/documents/programmes/elxinstit/ltigdelveppredact.pdf

Minocha, S, Kear, K Mount, NJ and Priestnall, G (2008b) Design of Learning Spaces in 3D Virtual Environments. *ReLive 2008*, Open University, UK, 20-21st November 2008.

Rymaszewski M, Wagner J, Wallace M, Winters C, Ondrejka C, Batstone-Cunningham B and Rosedale P (2007) *Second Life: The Official Guide*, Wiley, New Jersey. 334 pp.

Biography

Nick Mount is lecturer in Geographical Information Science at the University of Nottingham and winner of the Lord Dearing Award for his use of emerging technology to support teaching and learning activities.

Gary Priestnall is Associate Professor in Geographical Information Science and the University of Nottingham and director of the Nottingham arm of SPLINT. His research interests focus on approaches to the visualisation of geographical information.

Dan Weaver is Research Assistant on the DELVE project. His research interests include the development and adaptation of technology to support teaching and learning and the use of self-organising maps in spatio-temporal pattern analysis.

Andy Burton is Post-Doctoral Research Fellow on the SPLINT centre of excellence. His expertise include the production of semi-immersive virtual reality visualisations of geographical information.

Building bridges between methodological approaches: a meta-framework linking experiments and applied studies in 3D geovisualization research

Susanne Bleisch[1], Jason Dykes[2], Stephan Nebiker[1]

[1]FHNW University of Applied Sciences Northwestern Switzerland,
Institute of Geomatics Engineering, CH-4132 Muttenz, Switzerland
(susanne.bleisch, stephan.nebiker)@fhnw.ch
[2]City University London, giCentre, School of Informatics,
London, United Kingdom, jad7@soi.city.ac.uk

KEYWORDS: meta-framework, research methodology, 3D geovisualization, experiment, case study

1. Introduction

We propose a meta-framework that relates studies using a range of research methodologies as we investigate the use of abstract graphical representations of numeric data in 3D desktop based virtual environments. Selected approaches varying along a continuum from perceptual experiments to studies in applied settings provide the supports for the methodological bridge that we build by relating the design and findings of each. Doing so enables us to benefit from the advantages and overcome the limitations associated with studies conducted at any one stage. For example, we can evaluate the focused results of a controlled experiment in a more applied setting and thus address a key shortcoming of controlled experiments by taking into account contextual information.

2. The 'land'

At either end of the continuum we bridge are contrasting established approaches - the 'in vitro', quantitative and controlled environments used in psychophysical response studies and the 'in vivo', qualitative, case study research that evaluates applications 'in the wild' (Figure 1).

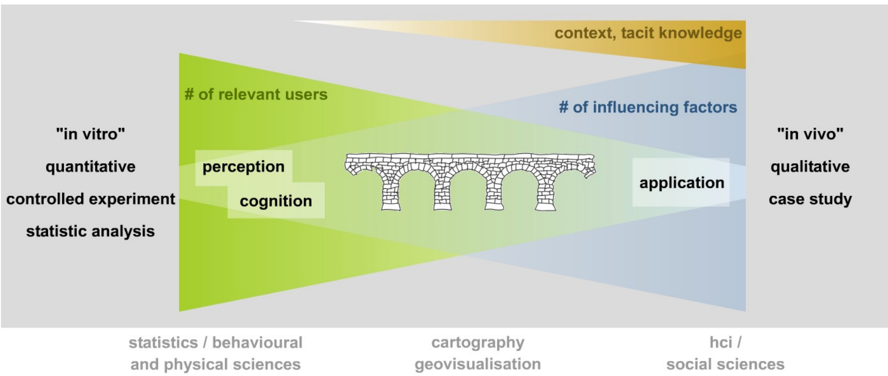


Figure 1. The meta-framework bridge between the two sides of 'in vitro' and 'in vivo' research

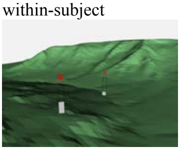
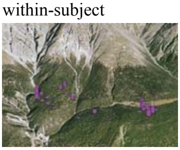
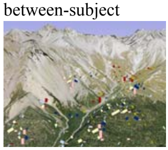
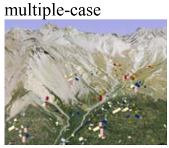
Activities on the experimental end are typical for the behavioural and physical sciences including cartography (Montello 2002). They usually involve large numbers of users, with little contextual information and thus no (or few controllable) influencing factors. On the applied side very few users are typically involved and qualitative approaches are employed. Much contextual and tacit knowledge influences and enriches these studies and many influencing factors exist that we cannot or do not want to control (Yin 2003). Research in geovisualization is done mostly towards the experimental side of the bridge (e.g. Bair and House 2007; Fabrikant, Montello et al. 2006). Case studies or applied settings are often used regarding implementation or usability issues (e.g. Brooks and Whalley 2008; Koua, MacEachren et al. 2006) and rarely evaluate the effectiveness of a visualization. Research along the bridge, for example, using experimental settings for the evaluation of different visualization types for easing understanding of and gaining insight into a dataset (Rester, Pohl et al. 2007) is rare.

3. The bridge

In the case of geovisualization we typically have a small number of experts who work with large amounts of data in an applied setting. They employ complex visual tasks to better understand and gain insight into the data. The complexity of such a setting with its many influencing factors, such as the tacit knowledge of the user, may be best researched 'in vivo'. However, certain more generic aspects of a geovisualization application can additionally be researched in a more controlled environment with a larger number of informed participants and thus underpin the 'in vivo' evaluations.

The four stages of our bridge are different methodical frameworks. We briefly describe those used in our evaluation of the graphical representation of numeric data through abstract symbols in 3D desktop virtual environments. Table 1 gives an overview of the specific or common characteristics of the different stages used when we employ this approach. The images in the table show how each visualisation might look like in the context of our particular research activity.

Table 1. Characteristics of each stage along the bridge

	stage I	**stage II**	**stage III**	**stage IV**
data type	synthetic/random	deer data SNP[1]	deer data SNP[1]	case data
variables	two values	# of visits per location	# of visits per time and location	many
symbol	single bars	single bars	bar charts	bar charts
setting	synthetic landscape	realistic landscape	realistic/real landscape	real landscape
# of tasks	two, pre-defined	seven, pre-defined	some, open, insight	open
task complexity	low	medium	medium/high	high
main sampling method	questionnaire (quantitative)	questionnaire (qualitative)	insight reports	observation, interview
participants	many (GI students)	many (GI students and staff)	some (GI students and staff)	few (data-experts)
study design	within-subject	within-subject	between-subject	multiple-case
example visualisation				

3.1 Stage I

Stage I consists of a controlled experimental framework that compares two different symbol types (bars with and without frames) in two different display settings (2D and 3D). The experiment participants complete two simple tasks: defining the higher of the two bars and judging the size of the smaller bar in comparison to the taller bar. The details of the framework and the results of this study, which enabled us to establish that task performance is no less effective but a little less efficient in the 3D setting, can be found in Bleisch et al. (2008). We hypothesise that the participants taking longer in 3D are also engaged in other processes such as assimilating and understanding the landscape. In the

[1] Swiss National Park

experimental setting of stage I these cause a longer task completion time only. The usefulness of such processes can only be explored in more context-rich settings where, for example, comprehension of the landscape is important for the analysis of the data displayed – this may be where geovisualizationis most useful.

3.2 Stage II

Stage II sets up an experimental framework using a within-subject design. The data consist of aggregations of deer sightings and thus have some context. As in the first experiment the data are displayed as single bars. But, given the knowledge that low-level tasks can be achieved in the 3D setting, participants undertake more complex tasks that require more cognitive processes than in stage I - such as finding and relating patterns of deer sightings with regard to location and altitude.

Initial results show that the times needed to perform the tasks have a high variance but in contrast to the experimental results of stage I efficiency is not significantly different between the interactive 2D and 3D settings. More differences are apparent between task completion times of the seven tasks, which are not equally complex (Figure 2). The user's confidence ratings, too, vary more between tasks than between 2D and 3D.

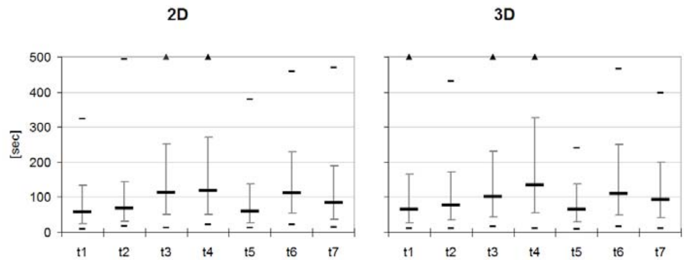


Figure 2. Mean, standard deviation and min/max values in seconds for the log-normally distributed task times for each of the seven tasks in 2D and 3D (▲ = max value outside displayed range)

3.3 Stage III

Stage III moves another step closer to the applied side of the bridge. Having determined that more complex tasks are feasible in 3D and that performance is as good as in the 2D case, data and task complexity is increased by considering a temporal component. The data consist of aggregations of deer sightings for different times of the day. Adding this 'data dimension' requires that an additional information carrying dimension be employed and so the data displays are no longer single bars but rather bar charts accounting for the multi-dimensionality of the data. They are displayed in the virtual equivalent of the environment the data was collected in adding important context. In addition to pre-defined tasks, the participants are asked to work with the visualisation and report the insights into the data gathered (North 2006). We will be interested in seeing whether and when the findings derived through experiments I and II hold.

3.4 Stage IV

Stage IV approaches the research aims in applied settings with data experts as typical in geovisualization. Case studies are valuable for studying single or few cases in depth (Gerring 2004) and learning about the real world applications of the findings of the previous stages. A multiple case design with a selection of diverse cases (Seawright and Gerring 2008) combined with cross case analysis enhances the representativeness of this method and allow us to compare less controlled 'in vivo' experiments involving expert users with those conducted 'in vitro' across our methodological bridge.

4. Conclusions

The selection of the stages as the bridge supports and their related research methodologies is driven by the increasing amounts of context, data and task complexity. We could select other criteria and then the stages might be positioned and methodologically defined differently. But, while exploring a dataset the user's effectiveness and efficiency is influenced by data and task complexity and the application context and that is the combination of aspects we want to analyse and gain insight to.

The meta-framework presented in this paper experimentally explores the visualization of numeric data in 3D desktop based virtual environments by relating methodologically different research stages. The findings from each stage inform those that follow and results from later stages are related back to validate or question earlier ones and thus generating deeper understanding as we develop a 'bridge' of knowledge between 'in vitro' and 'in vivo' research. The results from the first two stages of the bridge sustain our position that this is a valuable research strategy which might also be applicable to other research topics in geovisualization.

6. Acknowledgements

We thank the Swiss National Park for making the deer tracking data available. 'Thank you' also to the many participants in our experiments. We appreciate their efforts in supporting our studies. We also thank the reviewers of this abstract for their helpful comments.

References

Bair, A. and D. House (2007). "A Grid with a View: Optimal Texturing for Perception of Layered Surface Shape." *IEEE Transactions on Visualization and Computer Graphics* **13(6)**: 1656-1663.

Bleisch, S., J. Dykes, et al. (2008). "Evaluating the Effectiveness of Representing Numeric Information Through Abstract Graphics in 3D Desktop Virtual Environments." *The Cartographic Journal* **45(3)**: 216-226.

Brooks, S. and J. L. Whalley (2008). "Multilayer hybrid visualizations to support 3D GIS." *Computers, Environment and Urban Systems* **32(4)**: 278-292.

Cleveland, W. S. and R. McGill (1984). "Graphical Percpetion: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* **79(387)**: 531-554.

Fabrikant, S. I., D. R. Montello, et al. (2006). "The Distance-Similarity Metaphor in Region-Display Spatialization." *IEEE Computer Graphics and Applications* **26(4)**: 34-44.

Gerring, J. (2004). "What Is a Case Study and What Is It Good for?" *American Political Science Review* **98(2)**: 341-354.

Koua, E. L., A. M. MacEachren, et al. (2006). "Evaluating the usability of visualization methods in an exploratory geovisualization environment." *International Journal of Geographical Information Science* **20(4)**: 425-448.

Montello, D. R. (2002). "Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches." *Cartography and Geographic Information Science* **29(3)**: 283-304.

North, C. (2006). "Toward Measuring Visualization Insight." *IEEE Computer Graphics and Applications* **26(3)**: 6-9.

Rester, M., M. Pohl, et al. (2007). Evaluating an InfoVis Technique Using Insight Reports. 11th International Conference Information Visualization, IV2007, Zürich, Switzerland, IEEE Computer Society.

Seawright, J. and J. Gerring (2008). "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* **61(2)**: 294-308.

Wehrend, S. and C. Lewis (1990). A Problem-oriented Classification of Visualization Techniques. IEEE Visualization 90, San Francisco, IEEE.

Yin, R. K. (2003). *Case Study Research: Design and Methods*. Thousand Oaks, Sage Publications.

Zhou, M. X. and S. K. Feiner (1998). Visual Task Characterization for Automated Visual Discourse Synthesis. Computer Human Interaction CHI98, Los Angeles.

Biography

Susanne Bleisch is a PhD student at the giCentre, City University London and a scientific collaborator for the e-learning projects eLML, CartouCHe and GITTA at FHNW. She studied Geomatics at the University of Applied Sciences Northwestern Switzerland FHNW in Muttenz, Switzerland and has research interests in 3D geovisualization, cartography and geoinformatics and e-learning.

Jason Dykes is a Senior Lecturer in Geographic Information at City University London with interests in geovisualization. Co-chair of the ICA Commission on Geovisualization he has developed a number of geovisualization software applications and is lead editor of Exploring Geovisualization (Dykes, MacEachren and Kraak, 2005).

Stephan Nebiker is professor for geoinformatics, photogrammetry and remote sensing at the University of Applied Sciences Northwestern Switzerland FHNW in Muttenz, Switzerland. Among his research interests are virtual globe technologies, 3D geovisualization, mobile geosensors and augmented reality.

A Pilot Study for the Collaborative Development of New Ways of Visualising Seasonal Climate Forecasts

Aidan Slingsby[1], Rachel Lowe[2], Jason Dykes[1], David Stephenson[2], Jo Wood[1], Tim Jupp[2]

[1]giCentre, Department of Information Science, School of Informatics, City University London, Northampton Square, London, EC1V 0HB
Tel. +44 (0)20 7040 8800 Fax +44 (0)20 7040 8845
{a.slingsby | jad7 | jwo}@soi.city.ac.uk
http://www.gicentre.org/

[2]Exeter Climate Systems, School of Engineering, Computing and Mathematics, Harrison Building, University of Exeter, North Park Road, Exeter, EX4 4QF
Tel. +44 (0)1392 264067 Fax. +44 (0)1392 264067
{rl263 | D.B.Stephenson | T.E.Jupp}@exeter.ac.uk
http://www.secam.ex.ac.uk/xcs/

KEYWORDS: visualisation, climate, prototyping, collaboration, uncertainty.

1. Introduction

Seasonal climate forecasts are used by climate scientists, government departments, utility agencies, health agencies and in agriculture. Uncertainties in environmental data and in numerical prediction models make such forecasts inherently probabilistic. Good decisions require that this uncertainty in predictions be communicated to users. Some established methods have limitations, prompting this collaborative work between data visualizers and climate scientists facilitated by the Willis Research Network[1].

We report on the methods developed and prototyped and summarise the informal feedback generated. This informal process is insufficient for forming wider conclusions, but we reflect upon this exploratory work to help inform a planned formal study of collaboration between visualisation and domain-specific experts to develop new and effective modes of visual exploratory data analysis in the domain.

2. Data

Precipitation forecast and observation data expressed as *precipitation anomalies* (deviation from the 25-year mean) for Dec-Jan-Feb periods over 25 years (1981-2005) were provided by ECMWF[2] via the EUROBRISA[3] project. Forecasts from ensemble models (multiple model runs with slightly differing initial conditions), were provided as *(a)* mean forecast anomalies and *(b)* probabilistic forecasts expressed in terms of lower (p_1), middle (p_2) and upper (p_3) terciles (33^{rd} and 66^{th} percentiles) corresponding to *below normal*, *normal* and *above normal* precipitation, respectively.

[1] http://www.willisresearchnetwork.com/

[2] European Centre for Medium-Range Weather Forecasts (http://www.ecmwf.int/)

[3] http://www6.cptec.inpe.br/eurobrisa/

3. Graphical techniques in use and their limitations

Graphical representations of seasonal forecasts are widely used by scientists and agencies. However, the inherent uncertainty that results from modelling dynamic natural systems and therefore the confidence one can place in these forecasts is not always shown effectively by existing approaches (see examples on Met Office, ECMWF, IRI and EUROBRISA websites[4]).

The tercile-based data summarise variation between ensemble members by allocating them into ‘drier than normal’, ‘normal’ and ‘wetter than normal’ tercile bins. These are expressed as probabilities; e.g. if 10% of the ensemble members predict ‘drier than normal’ conditions, 20% lie predict ‘normal’ and 70% predict ‘wetter then normal’, then $p_{wetter}=0.1$, $p_{normal}=0.2$, and $p_{drier}=0.7$. We express these as the lower (p_1), middle (p_2) and upper (p_3) terciles respectively.

Most current approaches either show the probability for only one tercile or they further bin p_1, p_2 and p_3 into nominal categories that may be rather arbitrarily defined, may not cover all scenarios and often use inconsistent colour schemes.

(a) (b)

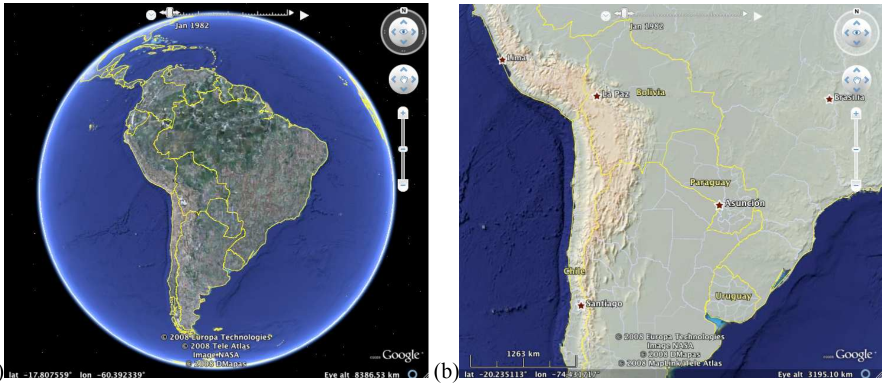


Figure 1. Background datasets in Google Earth.

4. Prototyping ideas and feedback

We addressed some of these limitations through novel, interactive and animated ‘data prototypes’ that support the rapid development of visual ideas. Technologies[5] that enable ‘patchwork prototyping’ (Floyd *et al*, 2007) were employed including: SVG (Figures 2b and 4), KML (Google Earth), Processing (Figure 5) and R (Figure 2c). Google Earth provides built-in aerial imagery and background datasets (Figure 1), which many end-users liked and it was relatively easy to incorporate other datasets such as the elevation model in Figure 1b. The other technologies used are more general-purpose and offer more control over appearance and interaction. Researchers at Exeter University and EUROBRISA project members in Brazil provided informal feedback and suggestions. These users ranged from scientists to decision-makers based in Brazil. For this preliminary study, we did not gather specific requirements from our diverse set of users. Instead, we developed and prototyped ideas for addressing some of the limitations of existing technique identified by the climate scientists and invited informal feedback. This helped us to assess the feasibility carrying out our

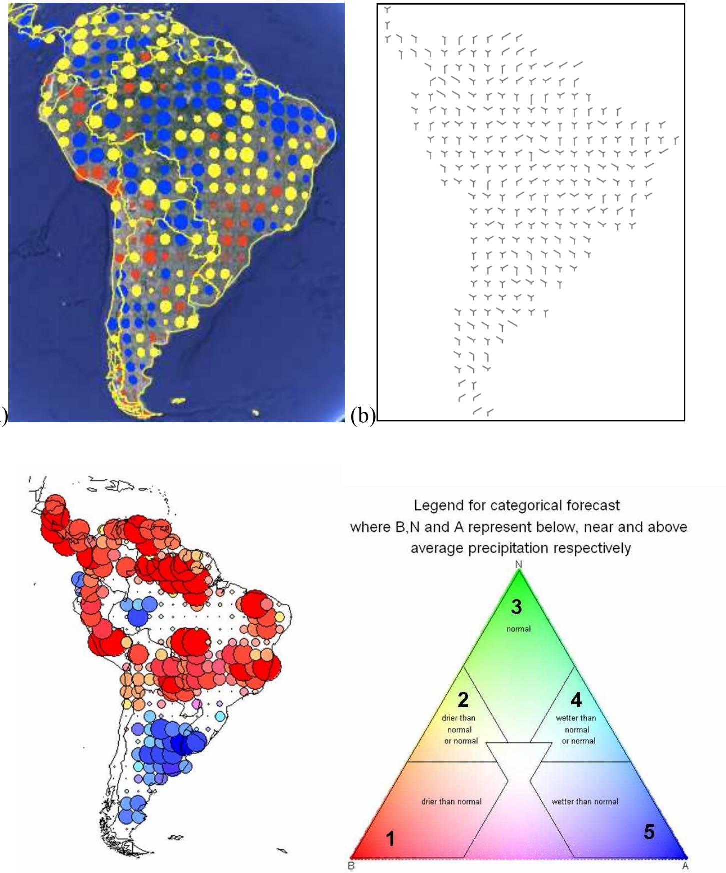
---

[4] http://www.metoffice.gov.uk/, http://www.ecmwf.int/, http://portal.iri. columbia.edu/ and http://www6.cptec.inpe.br/eurobrisa/

[5] http://www.w3.org/Graphics/SVG/, http://code.google.com/apis/kml/documentation/, http://processing.org/ and http://www.r-project.org/

formal study that will deal with specific user requirements from specific user groups of climate scientists at the UK MET office.

(a) (b)



(c)

Figure 2. Tercile-based probabilistic precipitation forecasts. *(a)* Circles coloured by dominant tercile and sized by the sum of the deviations between tercile probabilities; *(b)* Glyphs where arm lengths correspond to p_1, p_2 and p_3 (left, bottom and right arms, respectively). *(c)* The probabilistic forecast (left) is coloured according to the combination of the 3 tercile probabilities given within the colour space shown in the ternary phase diagram (right). Hue varies by angle from the centre and saturation varies by distance from the centre. Colour shows whether drier (p1; red), normal (p2; green) or wetter than normal (p3; blue) conditions are most likely. Corners denote a probability of 1 for each tercile (high saturation, high probability of rainfall in dominant tercile category) and the centre denotes equal probability for all terciles (low saturation, forecast no better than issuing climatology).
Colour is a necessary characteristic of this figure – obtain a colour PDF of this abstract from http://www.gicentre.org/papers/gisruk09/climate.pdf

4.1. Tercile-based probabilistic forecasts

To explore the design space with users we quickly generated initial ideas and prototypes that map different aspects of the data sets using various techniques. In Figure 2 information about all three tercile probabilities is visually encoded at each location concurrently for a single time period. In Figure 2a circles are sized according to the sum of the deviations between the *p*-values; thus the smallest circles occur where $p_1=p_2=p_3=0.33$, where the ensemble forecast tells us little beyond the known climatology. The largest circles occur where $p_1=1$, $p_2=1$ or $p_3=1$, with hue encoding the dominant tercile. User groups in Brazil found these maps easy to interpret. In Figure 2b, glyphs are used where the p_1 (left arm), p_2 (vertical) and p_3 (right) values are encoded through length, with Y-shaped glyphs denoting equal (1/3) probability for all terciles. In Figure 2c colour is derived from plotting the forecast on a ternary phase diagram. Circles are largest for forecasts at the corners and smallest in the centre of the diagram. Initial reactions of users to the latter two design ideas were less positive as they were less immediately interpretable. Sensible concerns were also raised about interpretation of colour by users with colour-impaired vision.

4.2. Comparison of observed and mean forecast precipitation anomalies

Figure 3 compares observed with mean forecast precipitation anomalies. Although there is general agreement between whether observed and modelled precipitation anomalies are positive or negative, observed anomalies were more extreme in magnitude and exhibit stronger spatial patterns.

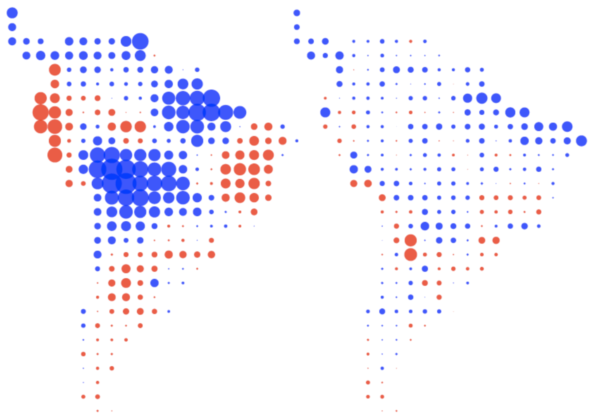


Figure 3. Observed (left) and mean forecast (right) precipitation anomalies (red circles are negative anomalies, blue are positive – size of circle indicates magnitude of anomaly).

4.3. Time series

Time series information is conventionally displayed per location as a graph with time along the *x*-axis. We prototyped some animated maps to show change over time, e.g. the Google Earth based stills in Figure 4 show notable differences between the seasons. We also used SVG to animate Figure 3. SVG supports smooth transitions between time-steps and users liked the aesthetics of this. There is some evidence that smooth transitions can assist in interpretation (Heer and Robertson, 2007), but there is a danger that intermediate stages of the morphing are interpreted as meaningful data points. An interesting side effect of the morphing process for Figure 3's red and blue circles is that the intermediate colour of purple drew attention to changes between positive and negative precipitation anomalies.

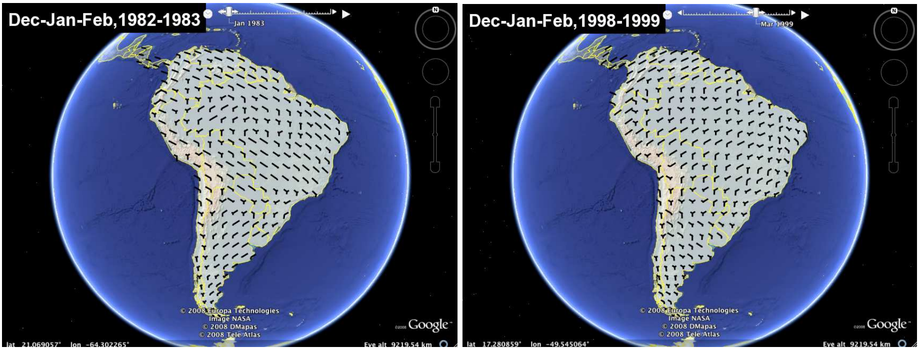


Figure 4. Glyphs showing p1, p2 and p3 values for different seasons.

The use of animation in this way is a common but ineffective technique for trend discovery (Robertson et al., 2008) and interactive techniques may be more appropriate. We rapidly prototyped some ideas in Processing which allows interactive and quite sophisticated prototypes to be created, such as shown in Figure 5 which allows temporal data to be explored spatially and vice versa.

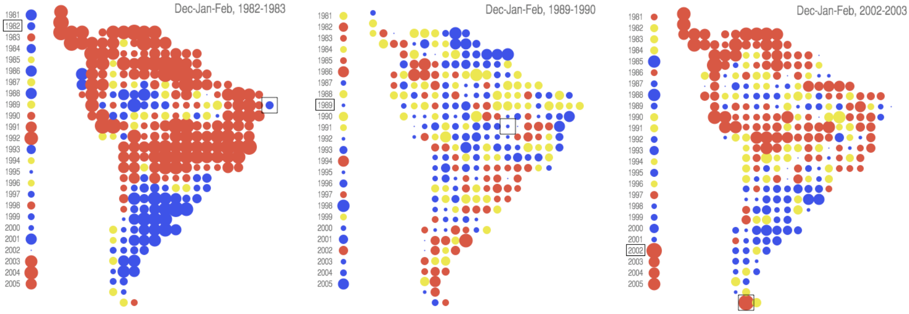


Figure 5. Interactive map that shows the exploration by *grid square* (time-series on the left) or by *season* (map). It also allows seasons to be selected for inclusion in an animation.

4.4. Brier Skill Score

The Brier Skill Score is a measure of the quality of a forecast that applies to each tercile in each grid square over the 25-year period. Transparency is an appropriate means of visually encoding uncertainty (MacEachren, 1992) and is used in Figure 6 to show skill.

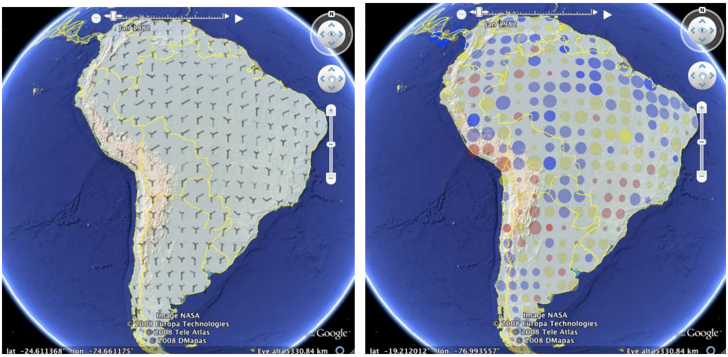


Figure 6. Transparency used to indicate the skill (quality) of the forecast using the Brier Skill Score (BSS). High transparency indicates lower skill.

5. Conclusions and outlook

This pilot demonstrates that cross-disciplinary collaborative work in visualisation is fruitful even where established techniques already exist. All parties involved were exposed to new perspectives and this promoted creative thinking, excitement about the possibilities demonstrated and valuable contributions.

More formal methods are required however. User-centred task-based aims are likely to be core to this activity with long-term workplace-based evaluation. Users initially responded negatively to some of the more complex representations even though they were more information rich. Given time to use them for the tasks for which they were designed may elicit a different response – the Information Visualization community is learning to expect the effective use of sophisticated tools to require training (Shneiderman and Plaisant, 2006). The rapid implementation of ideas as working prototypes facilitated informed feedback and kept all parties engaged. 'Patchwork' data prototypes (Floyd *et al.*, 2007) were successful in our pilot and SVG, KML, Processing and R are suitable technologies that allow ideas to be implemented and discussed quickly.

Our formal study design is based on these conclusions and other methods reported in the literature. Creativity is an important part of the design process and was central to our collaboration. Promoting creativity early the design process leads to better solutions (Jones and Maiden, 2005). Our ongoing design incorporates creativity-promoting activities through the RESCUE methodology, a formal task and workplace-based long-term evaluation (Shneiderman and Plaisant, 2006) and patchwork prototyping (Floyd *et al.*, 2007). This study will be carried out for specific requirements of specific user groups from the UK MET Office.

References

Floyd, I.R., Jones, M.C., Rathi, D. and Twidale, M.B. 2007. "Web Mash-ups and Patchwork Prototyping: User-driven technological innovation with Web 2.0 and Open Source Software". *In System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on. p86.*

Heer, J. and Robertson, G. 2007. "Animated Transitions in Statistical Data Graphics" *IEEE Transactions on Visualization and Computer Graphics 13 (6), pp 1240-1247.*

Jones, S. and Maiden, N., 2005. "RESCUE: An Integrated Method for Specifying Requirements for Complex Socio-Technical Systems". *In J. Mate & A. Silva, eds. Requirements Engineering for Sociotechnical Systems. Ideas Group Inc.*

Robertson, G., Fernandez, R., Fisher D., Lee, B. and Stasko, J. 2008. "Effectiveness of Animation in Trend Visualization". *IEEE Transactions on Visualization and Computer Graphics 14 (6), pp1240-1247.*

Shneiderman, B. and Plaisant, C., 2006. "Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies". *In Proceedings of the 2006 AVI workshop on Beyond time and errors: novel evaluation methods for information visualization. Venice, Italy: ACM, pp. 1-7.*

MacEachren, A.M. 1992. "Visualising uncertain information". *Cartographic perspectives 13, pp10-19.*

Acknowledgements

The authors are grateful to the Willis Research Network and the EUROBRISA network project (F/00 144/AT; kindly funded by the Leverhulme Trust) for funding this work. The dynamical ensemble forecast data were kindly provided by ECMWF as part of the EUROSIP project. The digital elevation model is sourced from USGS and the Google Earth screenshots contain background data from NASA and Europa Technologies (used with permission).

The use of colour was an important part of this study; you can download a colour version of this document from http://www.gicentre.org/papers/gisruk09/climate.pdf

Biographies

Aidan Slingsby is a Willis Research Fellow at City University London; Rachel Lowe is a second-year PhD student at the University of Exeter researching the viability of using seasonal climate forecasts to predict disease risk in Brazil and funded by the Leverhulme network project EUROBRISA; Jason Dykes and Jo Wood are Senior Lecturers in Geographic Information Science at City University London; David Stephenson is Joint Met Office Chair in the Statistical Analysis of Weather and Climate and Director of Exeter Climate Systems at the University of Exeter; and Tim Jupp is a Great Western Research fellow at the University of Exeter, focusing on quantifying uncertainty in the terrestrial carbon cycle (http://www.gwr.ac.uk/).

Polycentric cities and sustainable development: a multi-scaled GIS approach to analysing urban form

Duncan Smith[1]

[1] Centre for Advanced Spatial Analysis,
University College London, 1-19 Torrington Place, London, WC1E 7HB, England
Tel. +44(0)207 7679 1810 Fax +44(0)207 7813 2843
duncan.a.smith@ucl.ac.uk

KEYWORDS: Urban form, sustainable travel, polycentric cities, 3D GIS visualisation, urban planning

1. Introduction

The built environment influences how we live, work and travel in cities. Urban form evolves over time, shaped by changing socio-economic systems and transportation technology (Hall, 1999). The capital-intensive durable nature of the built environment means that this evolution is slow, and structures can persist for centuries, as for example street layouts in many UK cities. It is vital therefore that urban planning decisions are made with a long term perspective, using a strong evidence base.

Currently spatial analysis of the built environment is somewhat limited (Talen, 2003), particularly for city-wide analysis. Direct measures of urban form and function are seldom used due to a lack of widespread access to real estate data, and a lack of techniques to geocode and spatially aggregate this data for the analysis of large urban areas.

This research analyses non-residential property such as business services, retail and public services, which are relevant to a variety of applications such as mapping the structure of urban economic activity, property analysis, and monitoring urban development. Furthermore urban form measures are a key component of sustainable travel indicators (Banister, 2005) as the density of the built environment, mix of uses and public transport accessibility are all significant factors in promoting efficient travel patterns.

A spatial database of commercial property in Greater London is developed, at a series of scales from individual addresses to a 500m raster grid. This multi-scale approach allows both local scale urban form applications and larger scale city-wide analyses. Here the focus is mainly on analysing density and function at the city scale, allowing commercial centres and polycentric structures to be identified. Advocates of polycentric cities argue that greater integration between living and working environments can be achieved through policies such as 'Concentrated Dispersion'- that is the clustering decentralised activities at high densities around public transport nodes (Maguire et al, 2005). This is explored by linking the urban form database to commuting patterns.

2. Creation of the Urban Form Database

The approach utilised here is to geocode real estate data at the fine-scale of addresses, and then aggregate this to larger scales for city-wide analysis. The commercial real estate data comes from the Valuation Office 2005 surveys and includes detailed floorspace and function data. The function data have been classified into general types for this application- including Office, Retail and Local Services.

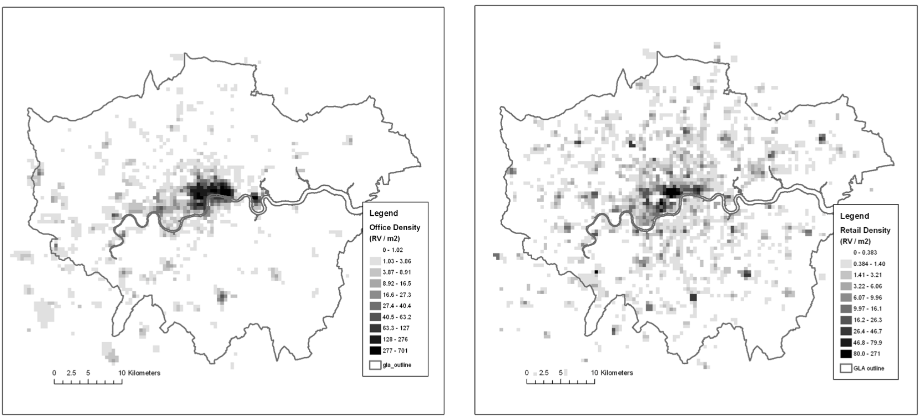
The geocoding process begins with address matching the real estate data to a spatial address database, Ordnance Survey's Address Layer. Aggregation is performed using spatial joins. The purpose of the aggregation methods is to allow viewing across multiple scales. Administrative zone boundaries are avoided in an attempt to reduce MAUP errors (Openshaw, 1984). The intermediate geographies used are street blocks defined by OS road centre line data (useful for density calculations) and street lines (useful for network accessibility measures) as shown in Figure 1. For metropolitan-wide visualisation aggregation to 500m grid is used.

Figure 1. Scales of Analysis and Aggregation. From left to right: Building Level, Block Level, Street Level and 500m Grid.

3. Analysis of Urban Form Data for Greater London

The results reveal varied scales of agglomeration for different urban functions in Greater London (Figures 2 and 3). Office activity is highly monocentric, dominated by central London and spreading into the inner-city. In Outer London the office market is much weaker, with only a few major centres such as Croydon. Outer London office centres have been losing jobs in recent years (Chippendale and Marsh, 2006), potentially increasing commuting distances.

For retail uses the pattern is much more polycentric, with a much greater number of suburban retail centres. Retail functions fit the image of London as a 'City of Villages'.



Figures 2 & 3. Density of office Rateable Value (left) and retail Rateable Value (right) in Greater London 2005

4. 3D Visualisation of Mix of Uses

Three dimensional visualisation techniques can be used to combine density layers and highlight functional diversity. Density values are represented by extruding grid blocks, and different functions are stacked on top of each other as in a 3D graph. Figure 4 uses this technique for office and retail functions. Further data layers (such as residential density) can be included in these stacked 3D visualisations, though as more layers are added, legibility decreases.

Central London is as expected highly diverse, particularly around the West End. In Outer London patterns diverge between the mixed use Metropolitan Centres (such as Croydon and Bromley) and more dispersed developments in the Western Corridor. These dispersed trends are indicative of more car based 'Edge City' style urban forms.

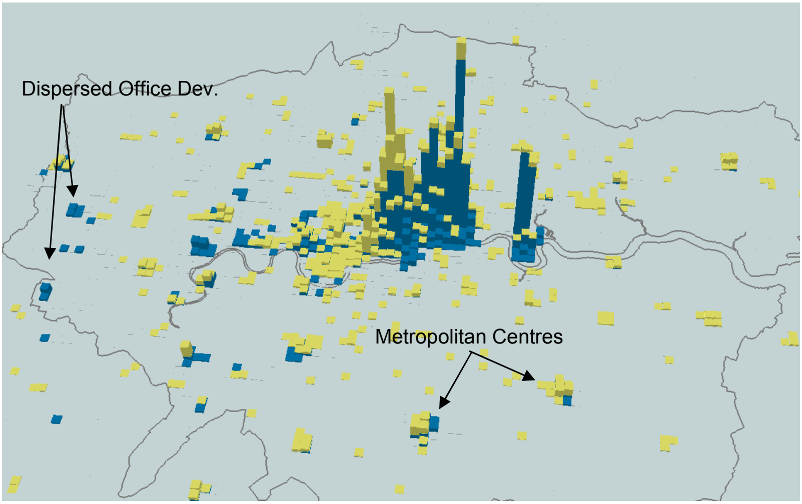


Figure 4. 3D density of Office (dark blue) and Retail (light yellow) Rateable Value in Greater London 2005

5. Integrating Urban Form with Commuting Data

The database can be integrated with travel data to explore relationships between urban form and commuting patterns. Figure 5 maps average commuter carbon emissions (estimated from distance travelled and mode choice) to employment destination zones (Smith, forthcoming). There are distinct travel patterns for particular urban centres. City centre commutes are overwhelmingly by public transport, but due to their long distance nature they are not the most efficient pattern. The Inner City and Outer London town centres achieve more efficient commuting patterns with good live-work integration.

To the West Heathrow airport and surrounding office developments have by far the most carbon intensive commuting. It is probable that the edge city type forms identified in Figure 4 are contributing to these travel patterns. Consideration must be given however to other issues influencing commuting such as employment specialisation, public transport accessibility and the housing market (Smith, forthcoming) so further analysis is needed before robust conclusions can be made.

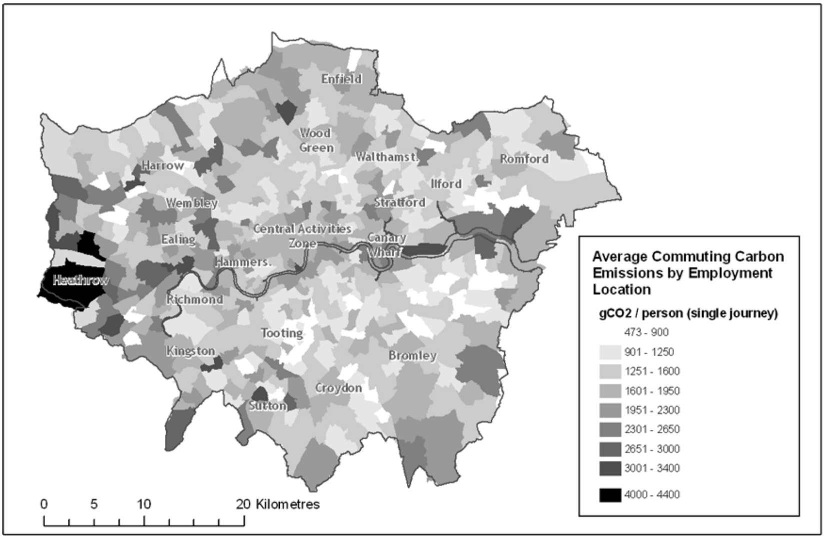


Figure 5. Carbon Emissions per Commuter by Greater London Employment Location (2001 Census)

6. Urban Development Policy Applications

Most urban development occurs incrementally at a local level and planning permissions are decided at a local scale. The advantage of the multi-scale approach of this work is that in addition to city-wide analysis, local scale analysis is possible.

Figure 6 shows the density of office floorspace in Central London at a street block scale. The more detailed block measures identify urban texture, showing parks, squares and areas of high rise development. The blue transparent polygons are areas designated by the planning authority as 'Opportunity Areas'- priority areas for development. The map highlights the early stage of development for the Opportunity Areas in 2005, as they lack office developments (apart from Canary Wharf to the east, which began development in the late 1980s).

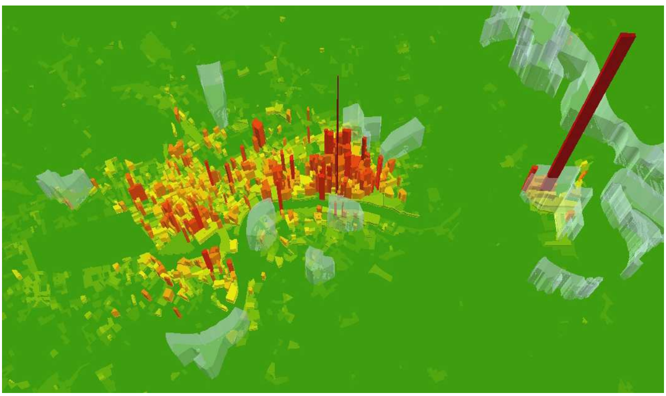


Figure 6. Block scale office density overlaid with GLA Opportunity Areas in Central London

7. Conclusions

This research has developed a spatial database of real estate data for Greater London at multiple scales to improve the empirical grounding of urban form analysis. The results reveal varied scales of agglomeration for different urban functions, with office uses highly monocentric and retail and local services much more polycentric. A more polycentric approach to office based business services could reduce commuting distances with greater live-work integration. There is however the added complication of unsustainable car based office-parks around Heathrow which have the most inefficient travel patterns in Greater London.

Three dimensional visualisations have aided data exploration in this application due to the multidimensional interconnected nature of the data layers. There is also the possibility to use the finer scale data aggregations to link to local planning databases, and integrate such analysis into urban planning practice.

Future improvements of this research could include analysis of temporal change, and a fuller exploration of the factors influencing commuting behaviour.

8. Acknowledgements

This research is funded by ESRC. Data providers include the Greater London Authority Economics Unit, the Valuation Office and the Ordnance Survey.

References

Bannister D (2005). *Unsustainable Transport: City Transport in the New Century*. Routledge Taylor and Francis, London.

Chippendale D and Marsh G (2006). *London Office Policy Review*. Greater London Authority, London.

Hall PG (1999). *Cities in civilization: culture, technology and urban order.* Widenfeld and Nicolson, London.

Maguire S, Edwards M and Twedwr-Jones M (2004). *Intra-Urban Polycentric Development: Suburban Trajectories.* Office of the Deputy Prime Minister, London.

Openshaw S (1984). *The Modifiable Areal Unit Problem.* Geobooks, Norwich.

Smith D (forthcoming). Polycentric employment centres and sustainable travel in London and the South East *Centre for Advanced Spatial Analysis (University College London) Working Paper.*

Talen E (2003). Measuring Urbanism: Issues in Smart Growth Research *Journal of Urban Design* **8(3)** pp195-215.

Biography

Duncan Smith is a 3rd year PhD student at CASA UCL. His research interests are in applying GIS to urban research and planning, in particular to the topics of sustainable development and economic change. Duncan holds an MSc in GIS from the University of Edinburgh.

Feeling the Way: Emotional Mappings

Daniel James Lewis[1], Christian Nold[2] and Muki Haklay[3]

[1]Department of Geography, University College London, Gower Street, London, WC1E 6BT
Daniel.lewis@ucl.ac.uk
[2] http://www.biomapping.net
[3] Department of Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London, WC1E 6BT
m.haklay@ucl.ac.uk

KEYWORDS: Emotion, Mapping, Biomapping, KDE, Critical GIS

1. Introduction

Can emotional responses to the landscape be mapped? This paper introduces an attempt to map the interpersonal, or collective, emotion of a group of volunteers in two places, Greenwich, London, UK and San Francisco, California, USA. Two surface creation methods, Kriging and Kernel Density Estimation, are explored for their applicability to the problem and assessed in terms of the representation they produce.

The work derives from an ongoing community mapping project led by Christian Nold called 'Biomapping'. The project aims to take a measure of change in galvanic skin response (GSR) as an indicator of emotional arousal (this is much like a 'lie detector') and couple the output with a location using a GPS receiver. A subsequent workshop allows for georeferenced comments that give context and insight into the experiences of the participants. Representation has, up until this point, mostly consisted of overlaying individual GPS tracks with GSR indicated by the relative height of the track; this paper aims to push this representation forward to an understanding of the shared, interpersonal emotions connected to places.

Undertaking this work brings to the fore many questions, for instance should 'emotion' be treated as a discrete object or a continuous field? What are appropriate parameters to set in order to derive a surface? What does a representation of change in GSR actually show? Does it really tell us anything about how people feel about a place?

2. The Biomapping framework

Biomapping is using 'active GSR', where a small amount of current is passed through the body and the time taken to fill up a capacitor is recorded. The less time the capacitor take to fill, the less the skin resistance and the greater the level of perspiration present. This process happens once every second, every ten readings are aggregated to create a reading which is given a GPS location.

The measure recorded is change in GSR, rather than absolute GSR, because different people are likely to perspire to different extents depending on biological or environmental factors. Recording the change in GSR allows for a value of variance from a baseline to be established which may account for a change in emotional intensity.

The justification for the use of GSR for affect measurements has a long and somewhat troubled history. Dickson and McGinnies (1966, p584) note that 'it seems clear that GSR is evoked by, among other things, stimuli judged to have affective significance'. However it is very difficult to say for sure that it is measuring anything beyond the physiological activity of sweat glands.

In spite of this much work was conducted in the first half of the 20th century that supported the idea that GSR was linked to emotion or affect, stress and attention or startle response. However, GSR was superseded by Electroencephalography (EEG) and later imaging such as MRI. GSR has been integrated with some curious technologies. Lie-detectors use several physiological responses including GSR in attempting to discern whether a subject is being truthful, use of lie-detectors is generally controversial. Similarly the Church of Scientology uses GSR to conduct ‘stress tests’, this is arguably more sinister as they tend misrepresent the nature of the data gathered and subsume it within a damaging pseudoscience.

The understanding of the data in the context of this work is not that it is ‘scientific measure of emotion’, it is not, or that is taken specifically to represent emotion, it is not proven that it does (but it might). Rather, it is an interesting application of something that lies between art and science and in a way links GIScience back to the cartographic tradition of positioning maps and geographical representation between these two domains.

3. Data

The paper uses two datasets, the key consideration for their choice being that surface creation techniques tend to work better with more data, and the more individual emotional tracks present in an area, the greater confidence we can have in the representation of interpersonal emotion created.

The first dataset was collected in Greenwich, London, between October 2005 and March 2006 and is included in the “Greenwich Emotion Map” (Nold 2006). 80 Individuals contributed to a 7264 point dataset (2731 attributed to men, 4533 to women) covering roughly a $7km^2$ area of Greenwich beside the River Thames. The second dataset came from San Francisco, California and was collected between 30^{th} March and 28^{th} April 2007 and is included in the “San Francisco Emotion Map” (Nold 2007). 98 individuals contributed to a 9808 point dataset (4773 attributed to men, 5035 to women) covering roughly a $10km^2$ area of mid-town San Francisco. It is centred on the Mission district.

4. Creating Emotional Surfaces

Initially the most important aspect of emotional surface creation is to decide how the GSR data is understood to vary over space – is it continuous, subject to a distance decay function, or is it discrete, tied to the places that the response occurred? The variable being mapped- emotion- is abstract, yet experienced continuously with some rapid changes, therefore a discrete and a continuous surface model have been considered.

The discrete model investigated is Kernel Density Estimation (KDE). A raster image is created from the input GSR point data, each cell in the raster holds a density value based on the values of the input data within a preset window (the ‘kernel’) and subject to a decay function (Gibin et al. 2007). KDE is often used to show ‘hot-spots’, clusters of observations; by weighing the biomapping observations by the change in GSR, the emotional component of the data, areas of elevated interpersonal emotional response can be highlighted.

Kriging was used as the continuous model, this is a geostatistical technique that interpolates the value of raster cells based on a distance weighting of the input data points derived from a semi-variogram that is fitted to their distribution (de Smith et al. 2006). The statistical nature of this technique is to some extent undermined by the uncertainty of the biomapping data, discussed earlier, however as the intent is to produce a useful representation, rather than an analytical output, Kriging provides the best possible result for the data present.

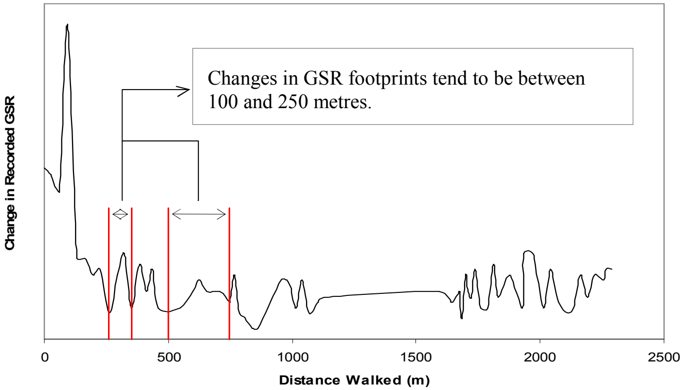


Figure 1: Assessing bandwidth, example of a participant in San Francisco

Both models require the use of a bandwidth, this is standard in KDE, and in Kriging it is applied to mitigate some of the edge effects that arise with automated variable bandwidths. It also allows for a consideration of ‘how far emotion travels’ to be made. Bandwidths were deduced by investigating the distance taken for distinct changes in GSR to become evident (Fig.1) and by considering the environmental context, i.e. how far can a subject justifiably see? Different bandwidths were then used to generate representations which were assessed for their aesthetic quality. Some results of the analysis are shown in Figures 2-4.

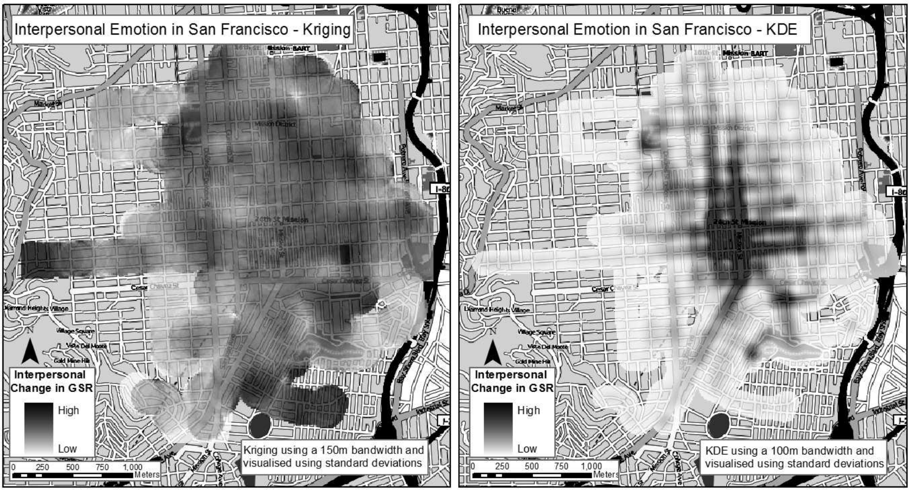


Figure 2: Kriging and KDE surfaces for San Francisco

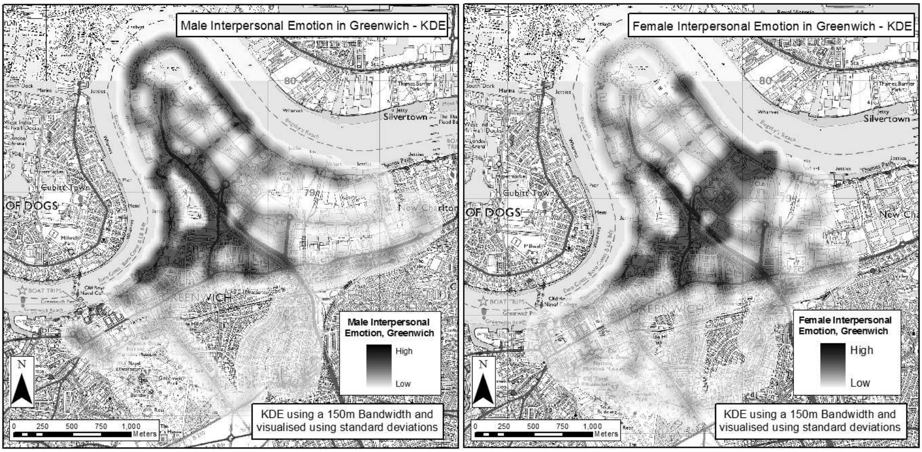


Figure 3: Male and Female KDE surfaces for Greenwich

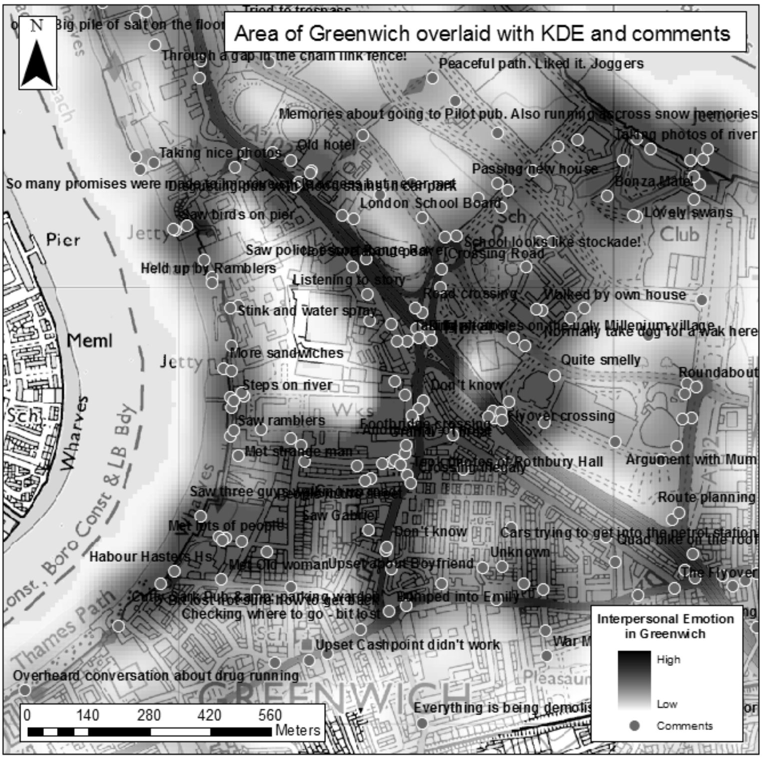


Figure 4: KDE surface for all participants in Greenwich including tagged comments.

5. Results and Discussions

When comparing the surfaces in Figure 2, KDE is a visually more intuitive way of thinking about how interpersonal emotion may vary across space. The hot spotting nature of KDE highlights areas in which a similar change in galvanic skin response was experienced and the viewer is able to understand how they might similarly experience the same area. This is not true of the Kriging surface which seems more inclined to highlight strong individual responses at the borders of the study area rather than collective measurements of emotion elsewhere. The suggestion is therefore that, at least in visualisation-terms, it is useful to treat emotion, or galvanic skin response, as a discrete object.

This reading of the validity of the surfaces is strengthened when georeferenced comments from the participants are added as in Figure 4. Generally the comments pertain to observations and feelings experienced during the biomapping process, aiming to give insight into some of the GSR data. The distribution of these comments closely echo the KDE, but not the kriging, surfaces. The comments also highlight some reasons behind the values of the GSR data and perhaps the KDE surface, people relate areas of high GSR response to: scenic views, green or open spaces, menacing environments or heavy traffic.

In order to extend the study and investigate differences in gender, the surfaces were created for the male and female participants individually, Figure 3 shows male and female experiences in Greenwich. Whilst they are notably different, it is difficult to offer any meaningful conclusions as to why this may be the case. A more incisive attempt may align the intent with that of Kwan's (2002) research into feminist visualisation.

Of course the ethical implications of these representations cannot be ignored, whilst they exclude the possibility of uncovering what an individual feels about an area, the knowledge that communities feel more strongly about some places than others may affect how those places come to be viewed generally. This could create division or rivalry between groups in different places, or be misused.

It is challenging to derive meaning from these representations, due to the nature of the data. Yet, they provide us with an insight into the places mapped. They add to the 'texture' (See Sui 2004) of what we know about these areas and begin to engage with the notion that places are comprised of 'stories so far' (Massey 2005).

6. Acknowledgements

This work was funded through the Mapping change for Sustainable Communities (MCSC) project, one of the 28 Urbanbuzz funded projects aimed at building sustainable communities.
Greenwich Maps are Crown Copyright Ordnance Survey (2008) an Edina Digimap/JISC supplied service. San Francisco maps are from Open Street Map, an OpenGeoData project CC-By-SA 2.0
Emotional Data Copyright Christian Nold, see his website www.biomapping.net for terms.

References

de Smith M, Goodchild M and Longley P (2006) *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Troubador

Dickson H W and McGinnies E (1966) Affectivity in the Arousal of Attitudes as Measured by Galvanic Skin Response. *The American Journal of Psychology* **79**: 584-9

Gibin M, Longley P and Atkinson P (2007) Kernel Density Estimation and Percent Volume Contours in General Practice Catchment Area Analysis in Urban Areas. *GISRUK 2007 Proceedings*

Kwan M-P (2002) Feminist Visualization: Re-envisioning GIS as a Method in Feminist

Geographic Research. *Annals of the Association of American Geographers* **92**: 645-61

Massey D (2005) *For Space*. SAGE, London

Nold C (2006) Greenwich Emotion Map: Project Background (http://emotionmap.net/background.htm) Accessed 13/8/2008

Nold C (2007) San Francisco Emotion Map: Project Background (http://www.sf.biomapping.net/background.htm) Accessed 13/8/2008

Sui D Z (2004) GIS, Cartography, and the "Third Culture": Geographic Imaginations in the Computer Age. *The Professional Geographer* **56**: 62-72

Biography

Daniel James Lewis is a PhD student in the Department of Geography at University College London. He has an MSc in Geographic Information Science also from UCL, and a BA in Geography from the London School of Economics.

Mapping the Geography of Social Networks

Nazanin Khalili, Jo Wood and Jason Dykes

giCentre, School of Informatics, City University London EC1V 0HB
Tel.+44 (0)20 7040 0146 Fax +44 (0)20 7040 8584
{ nazanin.khalili-shavarini.1 | jwo | jad7} @soi.city.ac.uk

KEYWORDS: social networks, volunteered geographic information, locational uncertainty, geovisualization

1. Introduction

Social relations and interaction patterns are visualised in node link graphs in traditional social network maps (Scott 1991, Wasserman and Faust 1994, Wellman and Berkowitz, 1988). The resultant network graphs frequently alter the geometric relations present in the real world in order to emphasize the connectivity and overall view of the networks. Whilst a node's position has considerable potential for carrying information regarding network pattern and structure, no *spatial* information is usually encoded. Location is an important spatial property of social entities (Wellman, 1996) and integrating that information with social network data has the potential of revealing insights into hidden patterns behind communities. The advent of Web 2.0 and popularity of online social networks have resulted in masses of voluntarily submitted locational information being available for study and analysis. Much of this has a geographic component described as Volunteered Geographic Information or VGI by Goodchild (2007). The abundance of VGI provides opportunities for analysing the geography of social networks that has not been used to its full advantage yet and we are interested in visualizing these geographies.

2. Geography in Social Networks

Whilst there is scope for determining multi geographies from large numbers of volunteers the process is by no means straightforward. In reality people are associated with multiple physical locations and VGI provides data through which these might be determined. For example, online geo social networks such as *yelp*[1], *flickr*[2] and *Gypsii*[3] enable us to associate users with home town, point of interest (POI), work place, geo located digital documents, etc. While attempts have been made in the past to relate geography to social entities, (Liben-Nowell, *et al*, 2005, Escher, 2007, Wellman, 1996) they have tended to focus on one specific location for each node, confining members to a bounding box of a city. This naïve geography although useful for variety of purposes (i.e. small world phenomena, analysis of scientific collaboration and friendship network) is of limited or no use for analysis of multiple locations associated with a social entity. This research takes advantage of the wasted geographies in graph presentation of social networks (Viegas and Donath, 2004) visualization and geovisualization techniques (Henry *et al*, 2008, Cui *et al*, 2008 and Henry and Fekete, 2007) for studying and modelling the multi-geographies accessible through VGI.

[1] http;//www.yelp.com
[2] http://www.flickr.com
[3] http://www.gypsii.com

3. Dataset Requirements

A number of appropriate candidate data set exist for this study. Raper (2008) listed 24 recent geo social sites in his blog[4] (17/06/08). Five of these are considered according to the following criteria:

According to the representative sites summarised in table 1 a large number of members (the order of 20,000) are essential requirements in obtaining a representative sample of online socialisers. Secondly, in order to reinforce generalisation of inferences geographically local sites are excluded. The availability of publicly accessible information regarding spatial data (e.g. hometown, POI, workplace, geolocated digital photographs, etc.) and relational data (e.g. public friendship network) are also required. Finally, an Application Programming Interface (API) through which the data may be accessed and manipulated is essential.

The sources listed in Table 1 contain both relational and spatial elements. The table shows that *flickr* can satisfy all the assessed research requirements quite successfully. It is an appropriate and interesting dataset for a number of reasons. Firstly, it contains large number of members and geotagged photos (68 million by Oct 2008) distributed around the world. Secondly the photo collections contain geographic information about members' locations and their uploaded pictures. Thirdly from the social network perspective *flickr* members have a set of publicly available contacts, friends and family networks in their profiles. Finally, it provides an impressive API with potential for mashing up data and retrieving useful information for different needs.

Sites	Large Number	Geo distribution	Hometown	Multiple spatial information	Relational data	API
Yelp	+	-	+	+	+	-
TwitterVision	+	+	+	-	+	+
Flickr	+	+	+/-	+	+	+
Gypsii	-	-	+	+	+	-
POIfriend	+	-	+	+	-	-

Table 1. Assessment results of the selected sites according to research requirements

4. Visualization Requirements

A Java application was developed through *flickrj* API[5] in order to extract the required spatio-social data from *flickr* database. The results are summarised in Table 2.

Spatial Data	Non Spatial Data	Relational Data
(lat, long) for uploaded photos Location of posters (City level)	Number of geo tagged photos Photo Id Photo accuracy Poster's user name Poster's user id	List of public friends of a given poster

Table 2. List of retrieved data from *flickr* database sorted in three categories.

[4] http://isblogs.soi.city.ac.uk/staff/raper
[5] http://flickrj.sourceforge.net/

An extensive evaluation of appropriate visualization packages including those designed for social network analysis was conducted according to the following requirements:

R1. Visualize relational / network data
R2. Visualize spatial data
R3. Calculate social network properties
R4. Calculate statistical properties of data
R5. Visualize large networks
R6. Develop interactive visualizations for enabling exploratory analysis of data
R7. Provide developer with flexibility in design (not being a menu driven package)
R8. Provide books, user manuals, tutorials or discussion boards for users.

Assessment results of the sites according to the criteria mentioned above are summarised in Table 3. As can be inferred two visualization development environments have been found as the most promising tools for visualization of spatio-social data: *Prefuse*[6] and *Processing*[7].

Sites	Relational Data	Spatial Data	SNA Properties	Statistical Properties	Large Networks	Interactivity	Flexibility	Learning Support
Many Eyes	+	-	-	-	-	+/-	-	+
Prefuse	+	+	-	+	+	+	+	+
Processing	+	+	+	+	+	+	+	+
Improvise	+	+	-	-	?	+	-	-
Pajek	+	+/-	+	+	+	-	-	+
Ucinet	+/-	+/-	+	+	+/-	-	-	+

Table 3. Assessment results of the visualization packages.

5. Flickr Friendship Network

In accordance with the requirements identified and considering the size of the *flickr* database our analysis was limited to data sets with the following criteria:

- Users who have geo tagged photos in their collections.
- Photos with the highest available locational accuracy (accuracy level 16).
- Photos uploaded between 9/12/06 to 09/12/08. 15 photos were randomly selected on daily basis during the above two year sampling period.

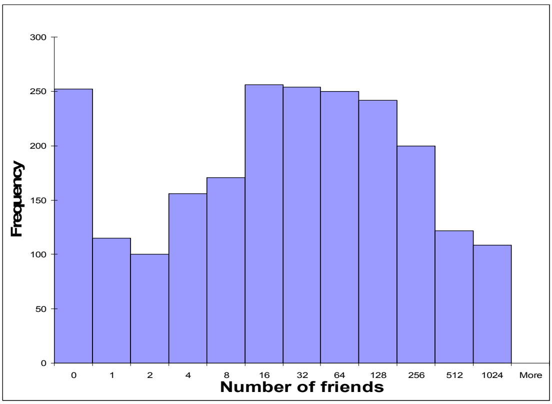


Figure 1. Frequency histogram for number of friends in *flickr* database.

[6] http://prefuse.org
[7] http://processing.org

Figure 1 shows the frequency histogram of the 2231 members who uploaded 12850 geotagged photos during the two year sampling period. The friendship network of *flickr* members follows a *bimodal* distribution with a large population of members with no friends and second population with approximate *log-normal* distribution. The unexpected number of users with 0 and very high numbers of friends indicate the unpredictable behaviour of social entities that can be affected by different factors e.g. exploring the site, limitations on the maximum number of friends etc. On average *flickr* users have 96 friends. The networks of 20 randomly selected members of *flickr* are visualized with *prefuse* and demonstrated in Figure 2 as indicative of the typical patterns of *flickr* friendship networks. Through the experiment became clear that the expectations suggested by the assessment criteria (Table 3) were met successfully in practice and can be concluded that the result of the experiments in practice and with *flickr* sample data are consistent with the assessment and evaluation conducted in section 4. Current work is investigating *processing* and visualizing the same data set with a *processing* application.

As can be seen the network is neither Small World (Watts and Strogatz, 1998) nor Scale-Free (Barabasi, et al, 2003) and the friendship network grows exponentially. Moreover, in the network of only 20 members without any spatial information included and with one of the most effective force directed layout algorithms (Fruchterman and Reingold, 1991) the layout is cluttered. This confirms the fact that, modelling and exploring the geographical layouts of large networks of this type need novel methods to be developed and new design decisions to be made. Therefore, the aim of this ongoing work is to overcome the challenge of developing spatio-social network maps that can be used effectively in identifying social and spatial relations within large data set. The developed maps are also expected to facilitate the evaluation of the role of geographical proximity in online interactions and friendship patterns.

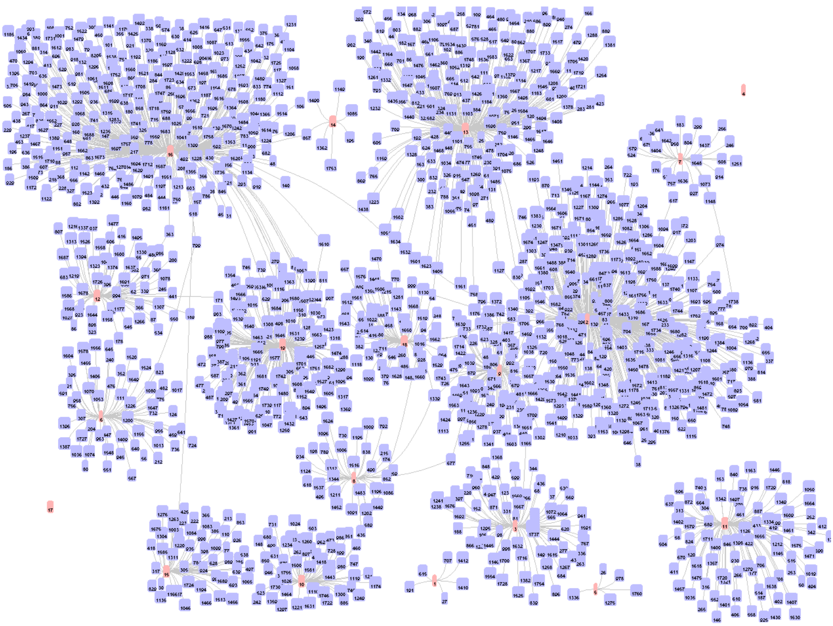


Figure 2. Social network of randomly selected *flickr* members. This visualization excludes the geographical layout. Nodes are laid out according to FruchtermanReingold algorithm (Fruchterman and Reingold, 1991).

6. Conclusion

This ongoing doctoral research aims at determining factors that relate to locational uncertainties of social entities in spatial social networks. This paper argues that VGI in social network context have the challenging properties of being large and multivariate. Consequently, the existing naïve geography of confining social entities in bounding box of a city is not adequate for analysis of multiple locations associated with a social entity. We conclude that the available VGI provide scope for exploring the geographies of social networks through spatial social network maps that support exploration of locational data as well as online relationships and patterns and we have identified data sets and technologies through which this can be achieved. Node duplication technique (Henry, et al, 2008) as a potential solution for visualizing spatio-social relations in a spatially structured social group will be investigated. In addition, privacy issues which might be raised as a consequence of developing new knowledge through the study, analysis and synthesis of VGI will also be considered in due time.

References

Barabasi, A-L, Bonabeau, E. (2003) Scale-free networks. *Scientific American*, 288(5).
Cui, W. et al., 2008. Geometry based clustering for Graph visualization. *IEEE Transaction on Visualization and Computer Graphics*, 14(6), 12771284.
Escher, T. (2007) Researching the geography of social relations: Analysis of the spatial distribution of friendship networks on social network sites. *Third International Conference in e-Social Science*.
Fruchterman T. and Reingold E. (1991) Graph Drawing by Force directed Placement. *Software-Practice and Experience*. 21(11), 1129-1164.
Goodchild M. (2007) Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69, 211-221.
Henry N., Bezerianos, A. and Fekete J.-D. (2008) Improving the readability of clustered social networks using node duplication. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 1317-1324.
Henry, N. and Fekete, J., (2007) MatLink: Enhanced Matrix Visualization for Analyzing Social Networks. Proceedings of the 11th IFIP TC13 International Conference Interact'07, LNCS, 4663, 288-302.
Liben-Nowell D., Novak J., Kumar R., Raghavan P. and Tomkins A. (2005) Geographic routing in social networks. *National Academy of Science*, 102 (33), 11623-11628.
Scott J. (1991) *Social Network Analysis: A Handbook.* London: Sage Publications.
Wasserman S. and Faust K. (1994) *Social Network Analysis*. Cambridge: Cambridge University Press.
Wellman B. and Berkowitz S. (1988) *Structural Analysis in the Social Sciences: A Network Approach*. Cambridge: Cambridge University Press
Wellman B. (1996) Are personal communities local? A dumptarian reconsideration. *Social Networks*, 18(4), 347-354.
Viegas F. and Donath J. (2004) Social network visualization: Can we go beyond the graph? *Workshop on Social Networks for Design and Analysis: Using Network Information in CSCW*.
Watts D. and Strogatz S. (1998) Collective dynamics of 'small-world' networks. *Nature* 393:440-442.

Biography

Nazanin Khalili Shavarini is a second year PhD Candidate at the giCentre, City University London with research interests in visualization and geo social networks. She graduated from Tehran Azad University in 2005 in Computer Hardware Engineering and has an MSc in Information Systems and Technology from City University London.

Jo Wood and Jason Dykes are senior lecturers in geographic information at the giCentre, City University London with interests in geovisualization.

Exploring the links between alcohol policy and the place and time dynamics of vandalism

Ellie Bates[1], William Mackaness[2]

[1]Edinburgh Law School, University of Edinburgh, Old College, South Bridge
Edinburgh EH8 9YL Tel: 07745 406959 Email: e.j.w.bates@sms.ed.ac.uk
[2]Institute of Geography, School of GeoSciences, University of Edinburgh, Drummond Street, Edinburgh, EH8 9XP Tel: 0131 650 8163 Email: william.mackaness@ed.ac.uk

KEYWORDS: vandalism, alcohol, geovisualisation, exploratory spatial data analysis, spatial and temporal scale

1. Introduction

The aim of this paper is to explore the repercussions of a major policy change on alcohol licensing law on patterns of police recorded vandalism in a city in north east England. The study highlights the utility of visual and graphical representations of data in exploring issues relating to crime and place. It demonstrates the value of examining recorded crime data at a variety of spatial and temporal scales when seeking to unravel complex interactions between crime rates, local processes and policy. The study suggests how a single national policy change can have unintended or unanticipated effects at a localised level, both in space and time. It further suggests that when considering why vandalism is conducted in particular places at particular times, the role of alcohol availability and bar density requires further research.

2. Why do this work

Vandalism is a very common crime in England and Wales (Nicholas et al, 2007). However, the place and time dynamics of vandalism remain little understood (Ceccatto & Haining, 2005). It has been suggested there may be links between alcohol availability and vandalism (Bromley & Nelson, 2002) and alcohol and general mayhem (Hayward & Hobbs, 2007). If this is the case then changes in alcohol availability might be expected to affect the distribution of vandalism both in space and time. A recent major change in licensing policy allowing alcohol outlets to stay open for longer, and later, in England and Wales offers an opportunity to assess this. Increased collection of crime data, and improvements in local geo-coding (Chainey & Ratcliffe, 2005) provide an opportunity to examine this change.

3. Methodology

A study area of a city with a high level of vandalism and areas with both high and low densities of alcohol outlets was chosen. Data was examined for 1 year before (prior to 24 November 2005) and two years after the licensing change brought about by the Licensing Act 2003. Police recorded crime data was obtained geocoded to a precision of around 50m along with the ‘time from’ in hours and minutes the offence was recorded as occurring at or after. This comprised data on individual crimes classified according to UK Home Office counting rules as criminal damage; it spanned a period covering 24 November 2004 to 23 November 2007. Details of licences for alcohol outlets were also supplied. Exploratory Data Analysis (EDA) and Exploratory Spatial Data Analysis (ESDA) (Tukey, 1977, Theus, 2005), crime mapping techniques (Chainey & Ratcliffe, 2005) and geovisualisation were used to provide a recursive approach to exploring the change. Free software CrimeStat III (Levine, 2007, Levine, 2006) and Geoda (Anselin et al, 2006, Anselin, 2005) were used for crime mapping and ESDA, proprietary software ArcGIS and ArcSDE were used for visualisation and Microsoft Excel and Access for sorting and grouping data. The use of space time cluster techniques (Assuncao et al, 2007) and Knox and Mantel indexes (Levine, 2007) were considered; these were not used as they appear better suited to detecting clusters of days rather than examining hours of the day. As the aim of this project was to explore the utility of thorough ESDA for exploring phenomena and constructing hypotheses, formal modelling techniques were not used.

4. Key Findings

EDA of the time crime occurs shows higher levels of criminal damage on Friday and Saturday nights, with vandalism most likely to occur between 1700 to 0159, Figure 1a. There is a general trend of lower levels of criminal damage year on year. However, considering weekend vandalism at a given hour in a year as a proportion of all crime in the year, there appears to be a greater concentration of crime occurring between midnight and 6am in the years following the licence law change (labelled N05N06 and N06N07), Figure 1b.

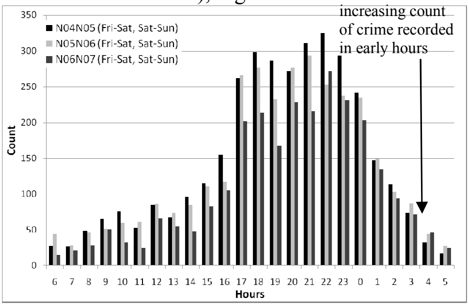


Figure 1a: Count of crimes in a year at given hour on Friday and Saturday 6am - midnight and Saturdays and Sundays midnight to 559am.

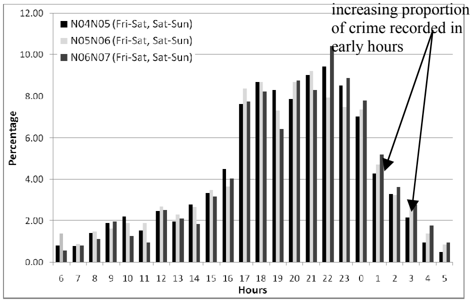


Figure 1b: Percentage of crimes in a year at given hour on Friday and Saturday 6-midnight and Saturdays and Sundays midnight to 559am.

Figure 2 compares vandalism in a year (24 November 2006 – 23 November 2007) with the distribution of outlets licensed to be open for sale of alcohol . A major concentration of alcohol outlets and vandalism can be seen in the city centre (Figure 2a), but this is particularly pronounced between midnight and 4am (Figure 2b).

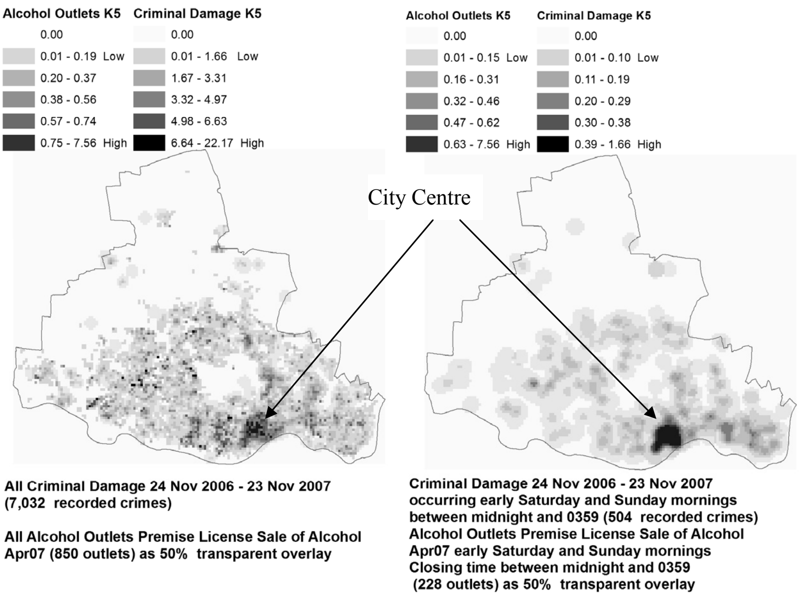
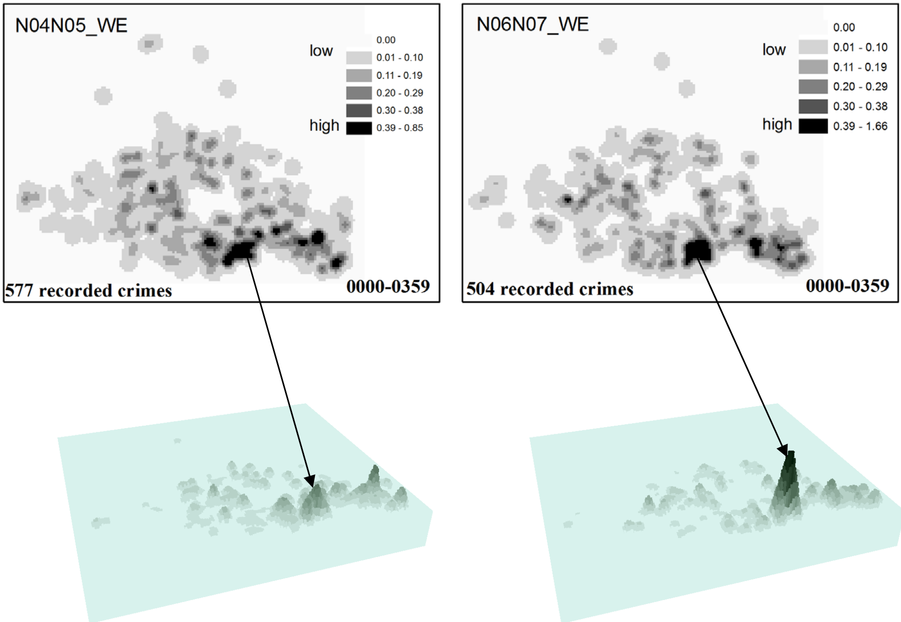


Figure 2a: All Criminal Damage compared with alcohol outlets

Figure 2b:Criminal damage occurring on in the early hours of Saturday and Sunday compared with alcohol outlets

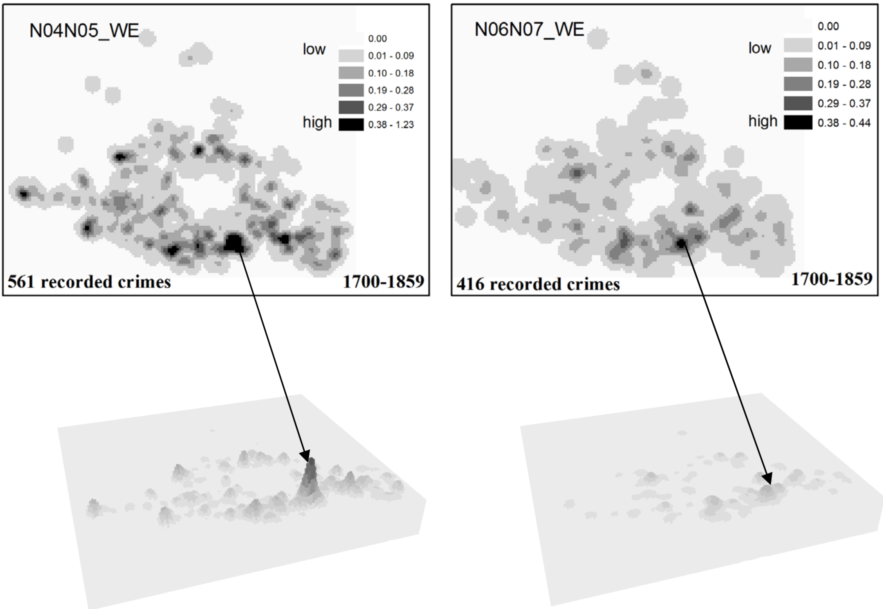
Figure 3 compares criminal damage occurring midnight to 4am on a Saturday or Sunday in the year prior to the licensing change (labelled N04N05) and the second year after the licence change (labelled N06N07).



Criminal damage density maps, mean random nearest neighbour distance to 5 neighbours (K5)
Numbers in legend are multiples of sample mean excluding 0 for based year, 0, 0+mean, 2*mean, 3*mean, 4*mean or higher. Scale 1:200,000 all maps
Each 3D image is a 3D density surface mapped at defined interval of (0.10) N04N05 sample mean excluding 0s

Figure 3: Criminal damage comparing the year prior to the licensing change (N04N05) and the second year after (N06N07) recorded as occurring midnight to 0359 on a Saturday or Sunday, visualised as 2D quartic kernel density map and 3D representation of density surface.

Figure 3 shows a marked increase in concentration of criminal damage in the city centre area following the change to licensing law which allowed alcohol outlets to stay open later. Yet this contrasts with a decrease in levels of criminal damage in the city centre area between 1700 and 1900, Figure 4.



Criminal damage density maps, mean random nearest neighbour distance to 5 neighbours (K5)
Numbers in legend are multiples of sample mean excluding 0 for based year, 0, 0+mean, 2*mean, 3*mean, 4*mean or higher. Scale 1:200,000 all maps
Each 3D image is a 3D density surface mapped at defined interval of (0.09) N04N05 sample mean excluding 0s

Figure 4: Criminal damage comparing the year prior to the licensing change (N04N05) and the second year after (N06N07) recorded as occurring 1700 to 1859 on a Friday or Saturday, visualised as 2D quartic kernel density map and 3D representation of density surface.

These results - suggesting an increase in vandalism in the city centre around 5-6pm which then decreases before rising again from 10pm to 3am – can be accounted for based on routine activity theory (Cohen & Felson, 1979, Scott et al, 2008) temporal constraint theory (Ratcliffe, 2006) and crime pattern theory (Brantingham & Brantingham, 2008) and examination of an individual's movement through space and time in a day. However, wider theories around collective efficacy (Sampson & Raudenbush, 1999, Ceccatto & Haining 2005), situational action theory (Wikström, 2006, Oberwittler & Wikström, 2008), situational crime precipitators (Wortley, 2008) and liminal space (Hayward & Hobbs, 2007) may also have relevance.

Figure 5 suggests how an individual who commits act(s) of vandalism might move between home, work and play through the day. Some places and times provide opportunities to both consume alcohol and commit vandalism but these are constrained by time spent in a particular space, which in turn may be constrained by demands of work or friends or security agencies such as police. The interaction of many individuals, a few of whom commit vandalism, moving through time and space, then leads to observed clusters of vandalism in particular places at particular times.

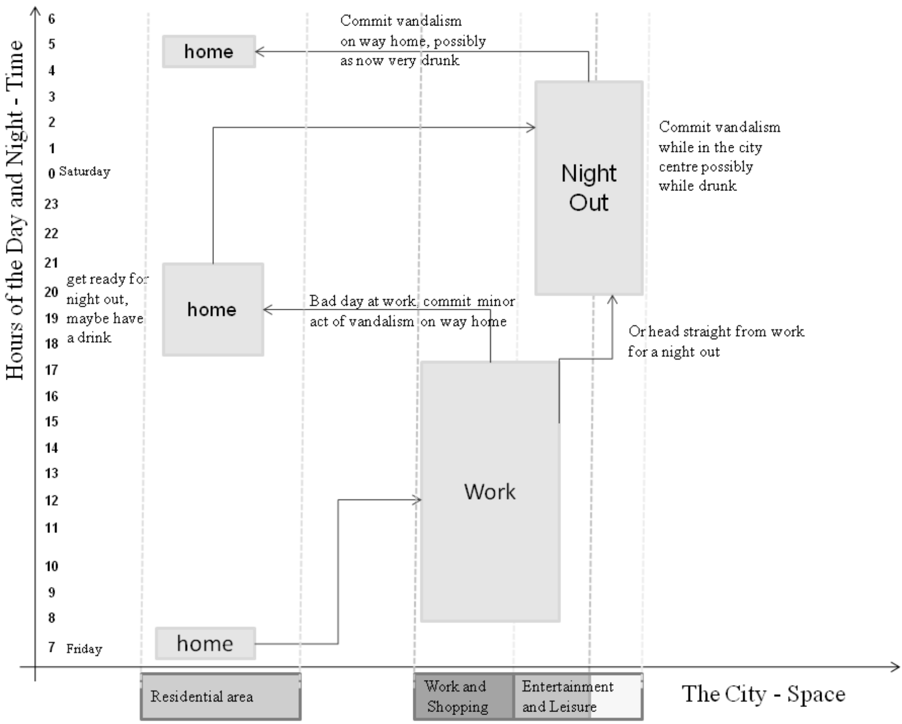


Figure 5. Space time Scenario for Criminal Damage in a City centre

5. Conclusions

A single major policy change has affected the spatial and temporal distribution of recorded criminal damage. This research supports similar findings in other areas of England and Wales (Hough & Hunter, 2008). It seems likely that other local processes and policies not explored here will also have had an effect. However, this research highlights that the city centre appears particularly affected, experiencing increased levels of vandalism into the early hours. Interestingly the city centre appears to act as a ‘vandalism sponge’ – with a shift of activity away from regions within the city to the city centre.

Care needs to be taken in how these results are interpreted; a complex range of factors may influence motivations to undertake vandalism (Cohen, 1984) and levels of police recorded vandalism are likely to substantially under-represent actual levels of vandalism (Kershaw et al, 2008, Bottoms, 2007). Some vandalisms may be more likely to be reported than others – for example those involving insurance claims. Later operating of the night time economy may mean an increase in police presence leading to increased recorded vandalism in the early hours. The study presumes that the ‘time from’ a crime is recorded is an adequate proxy for actual time occurring which may be misleading.

There is a need for much more research in this area, for example why in the presence of apparent guardians and many people does criminal damage still occur? Can and does social cohesion and collective efficacy vary by time of day, and does this affect diurnal levels of vandalism? Are wider cultural explanations and interpretations of the use of public space, particularly relating to liminal space and consumer culture, relevant to understanding links between vandalism and alcohol?

Generally, how can the interaction of crime, place and time be better visualised and understood through quantitative and qualitative methods?
It is important to take account of both spatial and temporal scales because at different scales of observations, different patterns of crime emerge.
There are implications for public crime mapping - specifically: what scale of visualisation is appropriate, whether maps at multiple spatial and temporal scales should be presented, and what levels of interaction with the data should be afforded.

Policing strategies may also need to consider being both time and place specific. For example, city centres or areas with concentrations of bars, may require a specific policy aimed at drunken vandalism which may be quite different to that applied generally to vandalism in other parts of the urban area.

6. Acknowledgements

Research was conducted as part of a MSc by Research in Geographic Information and Society part funded through a School of GeoScience Studentship. Funding for a PhD in Criminology focusing on the place and time dynamics of vandalism has been provided by the Scottish Centre for Crime and Justice Research (SCCJR). We are grateful for comments from Spencer Chainey, Jill Dando Institute of Crime Science; Yolanda Marshall, Newcastle City Council; Lynn Staeheli, University of Edinburgh; Tim Hope, Keele University; Susan McVie, University of Edinburgh, Paul Norris, University of Edinburgh. Special thanks to Safe Newcastle, Newcastle City Council, Northumbria Police in the provision of data.

References

Anselin, L. (2005) *Exploring Spatial Data with Geoda: A workbook,* Urbana, Univeristy of Illinois.

Anselin, L., Syabri, I. & Kho, Y. (2006) GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* . **36** pp5-22.

ASSUNÇÃO, R., TAVARES, A., CORREA, T. & KULLDORFF, M. (2007) Space-time cluster identification in point process. *The Canadian Journal of Statistics,* 35**,** 9-25

Bottoms, A. E. (2007) Place space, crime and disorder In Maguire, M., Morgan, R. & Reiner, R. (Eds.) *Oxford Handbook of Criminology.* Oxford, Oxford University Press.

Brantingham, P. & Brantingham, P. (2008) Crime Pattern Theory. In Wortley, R. & Mazerole, L. (Eds.) *Environmental Criminology and Crime Analysis.* Culhompton , Devon, Willan Publishing.

Bromley, R. & Nelson, A. (2002) Alcohol-related crime and disorder across urban space and time:evidence from a British city. *GeoForum,* **33** pp239-254.

Ceccato, V. & Haining, R. (2005) Assessing the geography of vandalism: Evidence from a Swedish city. *Urban Studies,* **42** pp1637-1656.

Cohen, S. (1984) Sociological approaches to vandalism. IN LEVY-LEBOYER, C. (Ed.) *Vandalism: Behaviour and motivations.* Amsterdam, Elsevier Science Publications

Chainey, S. & Ratcliffe, J. (2005) *GIS and Crime Mapping,* Chichester, Wiley.

Cohen, L. & Felson, M. (1979) Social Change and crime rate trends: a routine activities approach. *American Sociological Review,* **44** pp588-608.

Hayward, K. & Hobbs, D. (2007) Beyond the binge in 'booze Britain': market led limilization and the spectacle of binge drinking. *The British Journal of Sociology,* **38,** pp437-456.

Hough, M. & Hunter, A. (2008) The 2003 Licensing Act's impact on crime and disorder: An evaluation. *Criminology and Criminal Justice*, **8**, pp239-260

Kershaw, C., Nicholas, S. & Walker, A. (2008) Crime in England and Wales 2007/8. London, Home Office.

Levine, N. (2006) Crime Mapping and the *Crimestat* Program. *Geographical Analysis,* **38**, pp41–56.

Levine, N. (2007) *Crime Stat III A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.1) [Manual]*, Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC.

Nicholas, S., Kershaw, C. & Walker, A. (2007) Crime in England and Wales 2006/7. London, Home Office.

Oberwittler, D. & Wikström, P.-O H. (2008) Why Small is Better In Weisburd, D, Benasco, W. & Bruinsma, G. J. N.(Eds) *Putting Crime in its Place Units of Analysis in Geographic Criminology*, New York, Springer

Ratcliffe, J. H. (2006) A Temporal Constraint Theory to Explain Opportunity-Based Spatial Offending Patterns. *Journal of Research in Crime and Delinquency,* **43** pp261 - 291.

Sampson, R. J. & Raudenbush, S. W. (1999) Systematic Social Observation in Public Spaces: A New Look at Disorder in Urban Neighbourhoods. *American Journal of Sociology,* **105** pp 603-651.

Scott, M., Eck, J., Knutsson, J. & Goldstein, H. (2008) Problem-oriented policing and environmental criminology. In Wortley, R. & Mazerole, L. (Eds.) *Environmental Criminology and Crime Analysis.* Cullompton, Devon, Willan Publishing.

Theus, M. (2005) Statistical Data Exploration and Geographical Information Visualization. In Dykes, J., Maceachren, M. A. & Kraak, M. J. (Eds.) *Exploring Geovisualisation.* Oxford, Elsevier.

Tukey, J. W. (1977) *Exploratory Data Analysis,* United States, ,Addison Wesley.

Wikström, P.-O H. (2006) Individual Settings and acts of crime, situational mechanisms and their explanations of crime, In Wikström, P.-O H.& Sampson, R. J. (Eds) *The explanation of crime: Context mechanisms and development*. Cambridge, Cambridge University Press.

Wortley, R. (2008) Situational precipitators of crime. In Wortley, R. & Mazerole, L. (Eds.) *Environmental Criminology and Crime Analysis.* Cullhompton, Devon, Willan Publishing.

Biography

Ellie Bates is a first year PhD student of criminology in the School of Law at The University of Edinburgh. Her research interests are in the place and time dynamics of vandalism and how this can be explored through a mixed method approach including crime mapping. William Mackaness is a senior lecturer in the School of GeoSciences, at The University of Edinburgh. His research interests lie in the modeling of geographic phenomenon at multiple scales, and visualisation methodologies.

Mapping the sustainability of small business locations

Alice Dalton

Faculty of the Built Environment, University of the West of England,
Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY
Tel. +44(0)117 32 83078 Fax +44(0)117 3283899
Alice.Dalton@uwe.ac.uk

KEYWORDS: small to medium enterprise, sustainability, accessibility, network analysis, office location

1. Introduction

Small to medium enterprises (SMEs, businesses with fewer than 250 employees) account for 63.7% of employment, 99.7% of employers and 53.8% of the economic turnover in the south west region (BERR 2008). Their involvement in long-term sustainable economic development is therefore highly significant. The need for an office premises is a requirement of the majority of SMEs with employees, but the sustainability of these actual office locations in terms of the environment, employees and economic success, is not well documented. Furthermore, empirical research into the needs and decision-making of SMEs is limited. The location of an office building will impact upon carbon dioxide emissions, employee well-being, client/customer contact, general accessibility and overall business success, all of which influence the sustainability of economic development. In order to understand and help identify what is required for sustainable economic development in the future, further research is required.

2. Research aim

This research aims to understand office location decision-making by SMEs at the local and sub-regional scale in the Bristol city-region (Figure 1), in order to analyse the sustainability of office locations, building on current literature and debates in the field. This will provide insight into the economic, environmental and social sustainability of current economic growth for the area and will evaluate the current spatial planning policy framework.

3. Research questions

Three research questions have been established for this research:

1 What determines the locations of office-based SMEs in the West of England?

2 What is a sustainable location for office-based SMEs in the West of England?

3 Are SMEs choosing sustainable locations for their office premises?

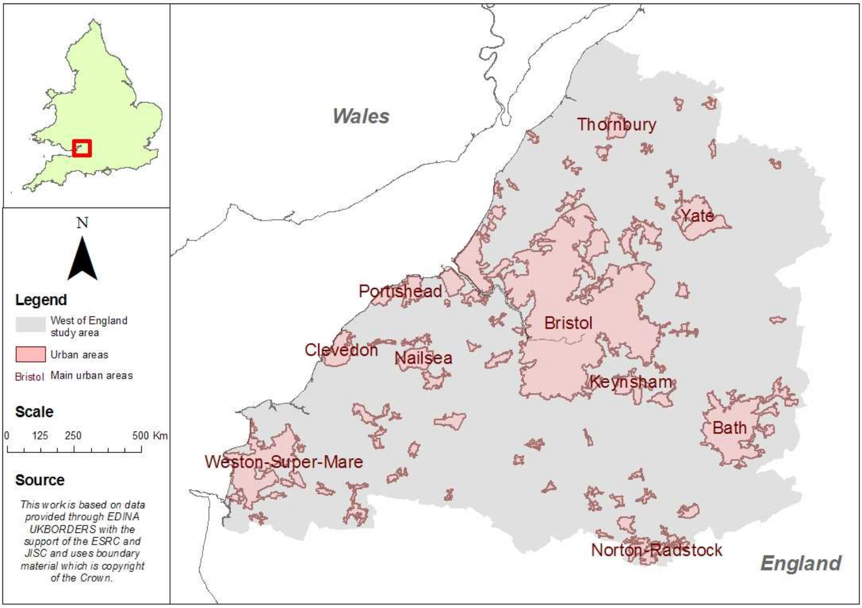


Figure 1. The West of England study area and associated urban areas

4. Research design and methodology

4.1 Definition of terms

Exploratory interview research revealed that staff well-being, proximity to facilities and accessibility are the most important factors for business success. For a location to be environmentally sustainable it can be argued that it must be accessible by public transport, in order that carbon dioxide emissions can be reduced. A sustainable location for an office-based SME is defined as maximising workplace quality of life and accessibility while resulting in minimal environmental impact.

4.2 Sampling procedure

Businesses were selected from the Financial Analysis Made Easy (FAME) (Jordans 2008) database of registered companies. A total of 1025 met the following eligibility criteria: A live SME (EC classification including micro, small and medium enterprises); at least one employee; a non-retail office premises (based on Standard Industry Classification SIC 2003 codes); in the West of England (Bath and North East Somerset, Bristol, North Somerset and South Gloucestershire); contactable by post or email (with addressee and address detail).

4.2 Business survey

An online questionnaire was used to collect responses from the selected businesses. This method was deemed most appropriate due to the sample size, a need to obtain quantitative data for comparison purposes, the assumed Internet accessibility of respondents and the constraints of time, resource and cost. The questionnaire was completed by a total of 215 businesses out of the sample of 1025, a response rate of 21%. The low response rate was expected due to the often quoted difficulty of conducting surveys of small businesses (Dillman 2000; Lewis et al. 2007). Respondents were asked a series of questions including: where their office was located; which factors were important when choosing their office location from a pre-specified list; where their employees were located; modes of travel to work; and the perceived advantages and areas of improvement for their current office location.

4.3 Data analysis

The data analysis stage of this research is ongoing and two main aspects of this have been, or will be, completed. One aspect involves investigating and spatially analysing the responses received from the survey (research question 1). The second aspect concerns mapping the sustainability of locations and office buildings in the study area based on criteria given in the literature and validated by the empirical research (research question 2). The results of these will inform the discussion concerning sustainable decision-making by SMEs (research question 3).

4.3.1 Business survey

Descriptive statistics have been extracted from the responses. The office locations of the business respondents have been georeferenced and mapped in a GIS using Ordnance Survey Code Point® data, enabling spatial analysis to be carried out in a geographic information system (GIS). The distance that employees travel to each office location is being analysed, based on straight distance (using 'Hawths Analysis Tools for ArcGIS', Beyer 2004), road distance and rail distance (using Network Analyst OD Cost Matrix facility in ESRI's® ArcMap™ 9.2). This will reveal specific travel to work areas at the individual business level. By using information provided regarding mode of travel, the carbon dioxide contribution of SME office users in the sub-region is calculated.

4.3.2 Sustainability assessment

Locations were quantified in terms of sustainability, attributing each office with an accessibility index based on the transport accessibility criteria developed by the Building Research Establishment (BRE) Environmental Assessment Method for offices (BRE 2008). For the 'TRA1' criteria (public transport accessibility), National Public Transport Access Nodes (NaPTAN) information from the Department for Transport have been used. For 'TRA2' (proximity to local amenities), Valuation Office Agency data were used to identify the location of cafes, banks, food shops, ATM's and post offices; and Royal Mail data was used to identify Mail Postbox® locations. The Network Analyst OD Cost Matrix facility in ESRI's® ArcMap™ 9.2 was used to find areas within an accessible distance from an office premises. OpenStreetMap data is currently being used to calculate accurate walking distances for over 7000 office locations (Figure 2). Public transport frequency data is being compiled using the National Public Transport Data Repository (NPTDR) (Thales 2009) to assess the quality of services. This gives a specific indication of how sustainable a particular office building could be, which will then be mapped as an accessibility surface across the study area.

Gravity modelling has been used to examine the relationship between office location, sustainability factors and employee location. The Keeble et al (1981) gravity modelling concept is based on the principle of the 'economic potential' of an area being a function of its attractiveness and of its proximity to other urban areas (Copus 1999). The model is used to assess the suitability of areas in the Bristol city-region for office locations, and to assess the influence of accessibility on the suitability of these areas. Locations will be weighted according to the accessibility index calculated above. The model is also used to consider actual employee locations in order to ascertain where the optimum location for an office premises would be.

Survey responses regarding advantages and needed improvements of their office locations, are compared to the indicative sustainability rating of their building. This provides an insight not only into the sustainability of office locations used by SMEs, but also into the appropriateness of sustainability rating systems, suggesting if the criteria are valid for SME office users.

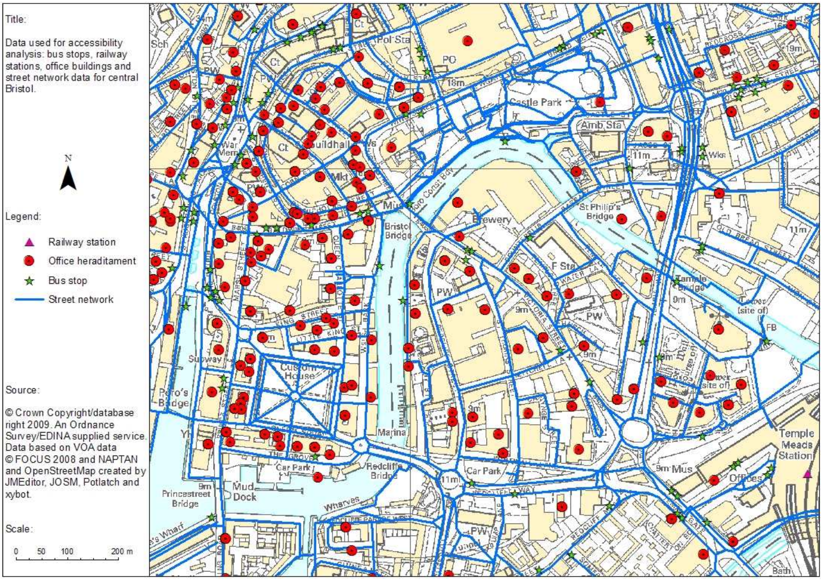


Figure 2. An example of the data used for accessibility calculations

5. Results

Quantitative data collected from the survey reveals that factors of cost, floorspace and broadband availability are most important when choosing an office location (Table 1). On the other hand, analysis of the qualitative comments made by respondents suggests that these factors may not be as important. Factors relating to transport and accessibility accounted for nearly 50% of all responses. Proximity to city-town centre was the most frequently mentioned advantage of respondent's office location. Respondents mentioned amenities, ease of access, atmosphere, centrality and quality of the surroundings. Local environment was the most frequently mentioned factor that the respondents would like to see improved at their office location. Respondents mentioned air quality, street cleanliness, amount of green space, signage, noise, drainage and the 'public realm' generally. In terms of travel mode, 53% of employees access the office by car and 44% by public transport, bicycle or on foot (Figure 3).

Spatial analysis is nearing completion, investigating office locations, accessibility and travel to work distances. Figures 4 to 6 provide an example of the initial analysis of the sustainability of locations in the study area. This suggests that existing premises may not be in sustainable locations according to current guidance. Once the sustainability mapping of the study area is complete, the number and locations of offices in accessible locations will be reported. Less sustainable areas in the Bristol city-region will be revealed. Gravity modelling is in progress and is reinforcing the identification of less sustainable areas. It is also revealing that business decision-making is based on a range of factors other than costs and centrality of premises based on employee location.

Although the analysis phase is still in progress, the research reveals that the heterogeneity of SMEs results in complex approach to decision-making, based partly on 'softer' 'quality of life' criteria. Following further analysis, results from the survey will help to shed light on where businesses want to locate and how important sustainable accessibility is to them. It has been difficult to accurately reflect the criteria adopted by BREAAM on an area-wide scale. In addition, the criteria used may not accurately reflect what is required by smaller businesses to enable their sustainability. This suggests that current methods for assessing sustainability may not be appropriate.

Table 2. Mean scores of location variables from the survey of 215 SME office-users in the Bristol city-region

Factor	Mean score of variable (1 = very important)	Rank of variable
Cost of premises	1.57	1
Floorspace	1.58	2
Broadband Internet availability	1.66	3
Access by car	1.75	4
Proximity to employees	1.90	5
Director's personal preference	2.11	6
Environmental quality of surroundings	2.21	7
Safety/crime levels	2.33	8
Access by public transport	2.44	9
Prestige of location	2.47	10
Access to customer/client	2.63	11
Access to shops and restaurants	2.79	12
Proximity to similar business	3.16	13
Proximity to university	3.50	14

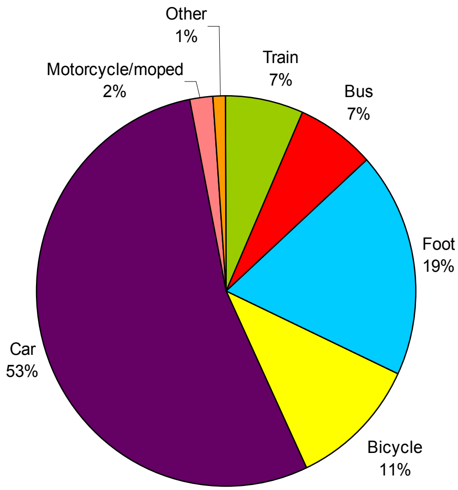


Figure 3. Mode of travel to work of employees working at the offices of the 215 businesses surveyed

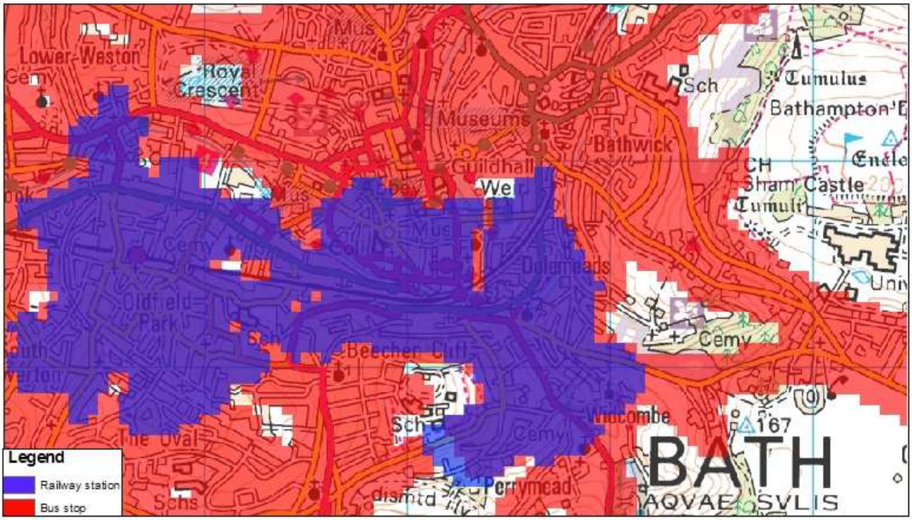


Figure 4. Accessibility to public transport in Bath, Somerset: areas within 1km of railway stations and 650m of bus stops (loosely based on BREEAM criteria)

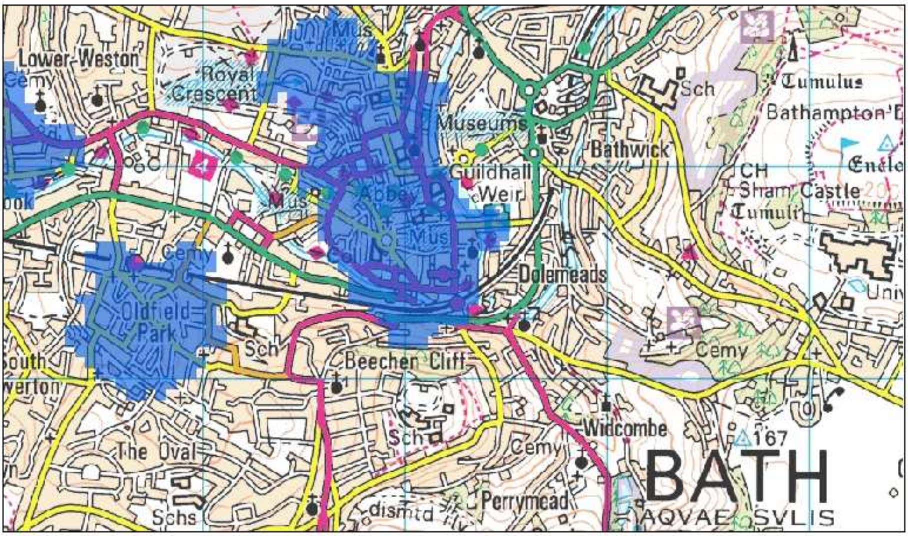


Figure 5. Accessibility to services and facilities in Bath, Somerset: areas within 500m of a food shop, cash point and postbox (loosely based on BREEAM criteria)

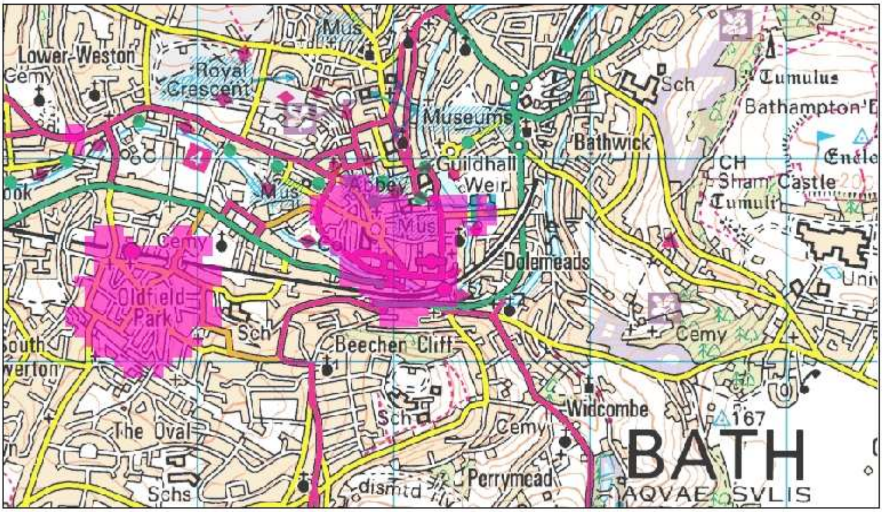


Figure 6. Most sustainable locations for offices in Bath, Somerset: combining the above transport and service accessibility maps (loosely based on BREEAM criteria)

6. Policy implications

The findings of this research have implications for future spatial planning policy, current sustainability assessment methods, encouraging the growth of small businesses, and preparing for sustainable growth in the future.

7. Acknowledgements

This project is part of the Government-funded ‘Great Western Research’ initiative, and is a collaboration between the University of the West of England in Bristol, Bath University and a private geographic information business, Geofutures. Funding is from the South West Regional Development Agency (SWRDA), the University of the West of England and Geofutures.

References

Beyer, HL (2004) Hawth's Analysis Tools for ArcGIS, [Online]. Available: http://www.spatialecology.com/htools [10/12/2008].

Copus, AK (1999) A New Peripherality Index for the NUTS III Regions of the European Union. A Report for the European Commission, Directorate General XVI.A.4 (Regional Policy and Cohesion). ERDF/FEDER Study 98/00/27/130.

Building Research Establishment (BRE) (2008) BREEAM Offices 2008 Assessor Manual. BRE

Environmental & Sustainability Standard. BES 5055: ISSUE 1.0. BRE Global Limited.

Department for Business Enterprise and Regulatory Reform (BERR) (2008) Enterprise Directorate: Small and Medium Enterprise Statistics for the UK and Regions. SME Statistics 2007 for the UK.

Dillman DA (2000). Mail and Internet surveys: the tailored design method. Second edn. New York: John Wiley & Sons, Inc.

European Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises (notified under document number C(2003) 1422) (Text with EEA relevance) (2003/361/EC).

Jordans Limited (2007) Financial Analysis Made Easy (FAME): UK and Irish company information in an instant, Bureau van Dijk Electronic Publishing (BvDEP), [Online]. Available: https://fame.bvdep.com/version-2008215/cgi/template.dll?product=1 [15/01/2008].

Keeble D, Owens PL and Thompson C (1981). The Influence of Peripheral and Central Locations on the Relative Development of Regions. Department of Geography, Cambridge University.

Lewis K, Massey C and Harris C (2007) Learning by doing: six dimensions of complexity in researching SMEs. Qualitative Research in Accounting & Management, 4(2), pp. 151-163.

OpenStreetMap (2009) The Free Wiki World Map, [Online]. Available: www.openstreetmap.org. Accessed 18 February 2009.

Thales (2009) National Public Transport Data Repository (NPTDR), operated on behalf of the Department for Transport, [Online]. Available: www.nptdr.org.uk. Accessed 10 February 2009.

Biography

Alice is in the third year of a PhD at the University of the West of England, Bristol. Alice obtained a BSc (Hons) in Geography (Reading, 2001) then an MSc in Geographical Information Science (Birkbeck, 2005). Alice combined her MSc studies with a position as Geographic Information Specialist at Halcrow.

Real-time Dynamic Simulation of Special Event Crowds using ABM and GIS

Miao Wang and David Fairbairn

School of Civil Engineering and GeoSciences, Cassie Building, Newcastle University,
Newcastle Upon Tyne NE1 7RU, England
Tel. +44(0)191 477654, Fax +44(0)191 2226502
Miao.wang1@ncl.ac.uk

KEYWORDS: Dynamic Simulation, Special Event, Crowd, GIS, ABM

1. Introduction

The most important factor affecting the planning and allocation of resources applied in special events is the crowd participating in the event. Therefore, an understanding of, and a precise prediction capability for, special event crowd dynamics are fundamental for the work of special event planners.

Crowd behaviour is complex, difficult to describe and predict, and although affected by individuals' behaviour, it often shows features that are very different due to the interactive relationships existing amongst individuals themselves and the nature of the collective environment. Experimental investigations have been used for the study of crowds but it has been impossible to capture all such behavioural features, and study is limited by 'practical, ethical, financial, and logical constraints' (Klupfel, Schreckenberg and Meyer-Konig, 2003, p.358). Empirical crowd surveys can be extended by the use of simulation to demonstrate crowd dynamics over a continuous time period and also allow experimenting with different crowd management plans which may not be possible to field-test.

The Agent-Based Model (ABM) has emerged as an effective technique for modelling complex systems, integrating socio-economic spatial dynamics into a single model (Epstein and Axtell 1996) and proving suitable for crowd simulation.

Such ABM-based modelling of crowds has mainly focused on mass evacuation (Egress, GridFlow , CRISP, and EXODUS (Johnson 2005)) and pedestrian traffic in densely peopled areas or in public buildings (STREET, PedGo, SIMWALK, NOMAD, SimPed (Harney 2002). Visitors to special events create different scenarios to these. A special event crowd has the largest positive utility of travel (Ory and Mokhtarian 2005), meaning that individual attendees are willing to spend time on travel and wandering, rather than aiming for minimum navigation time. In addition, the special event crowds' behaviour tends to be located over a large spatial area. Previous work on such crowds (limited to Batty et al (2003)), shows only a final result of peak time crowd spatial features and with all agents released at once at the beginning of the simulation. The major aim of the project described here is to create a model for the real-time simulation and evolution of special event crowd dynamics and its analysis.

2. Methodology

The building of an agent-based simulation model for this research involved the integration of GIS with the ABM software platform CORMAS (agent-based modelling software) (Cormas, 2008). GIS is first used as a tool for integrating data obtained from different sources such as digital maps, paper maps and image files, all preparing for the input data required by a simulation. Secondly, GIS is a good tool for visualizing the simulation result. Though the ABM software chosen has its own visualizing capability, some spatial analysis results such as hotspot maps of congestion areas need to be visualized on the platform of GIS. Thirdly, GIS is used as the spatial analysis tool for prediction of further congested areas in the process of potential risk scenario building. For the first two of these functions loose-coupling integration is adopted, which means by exchanging files the ABM obtains some of its input data from the GIS and produces some of its output in a format that allows import to, further processing with and display in, the GIS. The third function simulates a large number of time steps and it is unrealistic to transfer these as individual layers to a separate GIS file. Thus the functionality required (such as density analysis) is embedded into the ABM system.

The intention of the project is to simulate spatial crowd dynamics at a special event over an extensive site, to determine locations of congestion and risk noting the timing and intensity of such problem hotspots, to determine the impact of mitigating factors (such as improved stewarding or signage) and to assist planners in optimally locating different attractions within the site.

3 Modelling and Simulation

The project used a case study, the 2005 Tall Ships Race (TSR) hosted at the Quayside, Newcastle upon Tyne and Gateshead over a four day period in July 2005. The event occurred within an irregularly shaped site of approximately 2 by 0.25 km along the north and south banks of the River Tyne. Contemporary Tall Ships Races annually attract more than a hundred competing ships to a hosting seaport (usually in Europe), among them some of the largest sailing ships in existence. These sailing ships provide a stunning attraction for visitors who can look at them and participate in an associated festival (with musical performances, stalls, funfairs, exhibitions, promotional events, and entertainment on site, in addition to visits to the ships themselves). More than one million visitors participated in TSR 2005.

In order to model the crowd, the ABM software required the identification of entities and the construction of an initial class diagram showing attributes, methods and structural relationships. While the attributes can characterize the entity, methods are the tasks which the entity can undertake in the model (Figure 1). The generic ABM model entities to the left of Figure 1 have been supplemented by the specific entities created for this project, all portrayed in UML. The ‘Tourist’ is the agent object in TSR model and the ‘Cell’ is the spatial entity object. ‘LargeAttraction’ (i.e. a facility which the crowd might be interested in visiting) and ‘Congestion’ (areas defined as having a certain density) are aggregated spatial entities objects in the TSR model. Tourist behaviour is encoded in rules and programmed in the frame of the Tourist object by setting parameters such as ‘targets’ and ‘records’. The behaviour of Tourists can be classified into five linked categories: entering the site; self-positioning; visiting attractions (which itself may involve identifying target, walking towards target); lingering; and exiting. The way these agents behave and interact among themselves will influence the dynamics of the model.

Agents representing visitors were introduced dynamically through a series of entrances to the site throughout the day in a manner which replicated the predicted numbers. It was decided to set one modelling time step equivalent to 4 seconds in reality and modelling a 16 hours day therefore meant 14400 time steps. Every time step, the Tourist agents undertake a range of possible actions such as ‘enter’, ‘findCoodinate’ and ‘addTargets’. Further setting of Tourist parameters such as ‘records’ and ‘lifeTime’ enables the agents to have memory and ensure structured, focussed and intelligent perambulation. This allows for the running of dynamic simulations which can ensure that the correct number of visitors (Tourist agents) are introduced to the area over a one-day period, and that each agent (representing the visitors) can react to the attractions available, and to congestion patterns which emerge.

The embedding of spatial analysis functions into the ABM, which has obviated the need for tedious export of results from the ABM platform to GIS for each time step in the simulation, has been accomplished by setting dynamic evolution of the Cell in the TSR model. The parameter ‘riskValue’ and the method ‘countRiskValue’ help to evaluate the potential congested area and the parameter ‘conFre’ and the method ‘countConFre’ evaluate the potential high frequency actual congestion hotspots. Thus, spatial analysis such as hotspot detection can be done on the platform of the ABM.

The Cell entities are raster grid cells created from scanned OS Master Map data, supplemented by landuse information, event site information from the local council’s site plan and the locations of the Tall Ships themselves from the Berthing plan (see Figure 3).

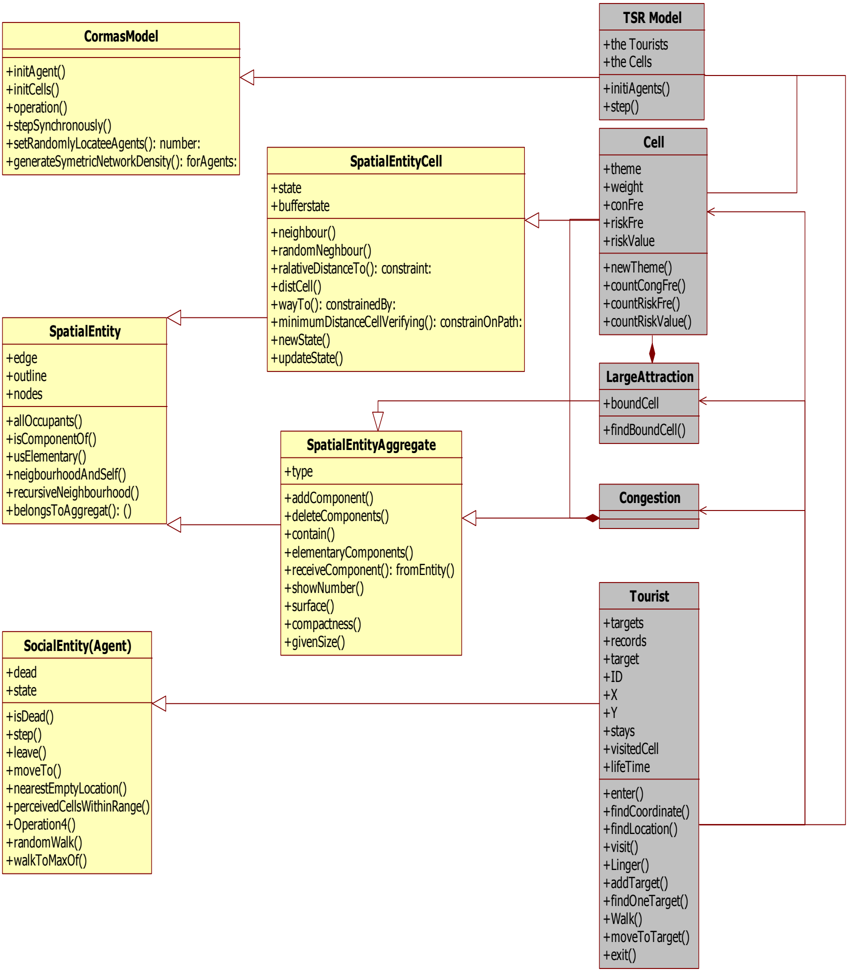


Figure 1 Hierarchy of CORMAS abstract classes (left and middle) **and the TSR model for crowd simulation** (to the right)

4 Results

During the simulation, results showed realistic modelling of crowd size and distribution. Furthermore, the model automatically and continually demonstrated the potential ‘risk’ areas of highest crowd density: potentially close to pedestrian standstill, but with the possibility of agents moving to adjacent cells from their over-crowded locations. Figure 2 shows, using the visualisation capabilities of the CORMAS platform, a specific example of the areas at ‘risk’ (in this case, a moment during the peak visitor period in the day). 14400 such visualisations were prepared to simulate a complete day with 100,000 visitors in total. These cells at risk could be aggregated to highlight areas of highest potential frequency of risk. In addition, the model also highlighted the actual areas of ‘congestion’ (standstill, when agents could not move to adjacent cells) at each time point.

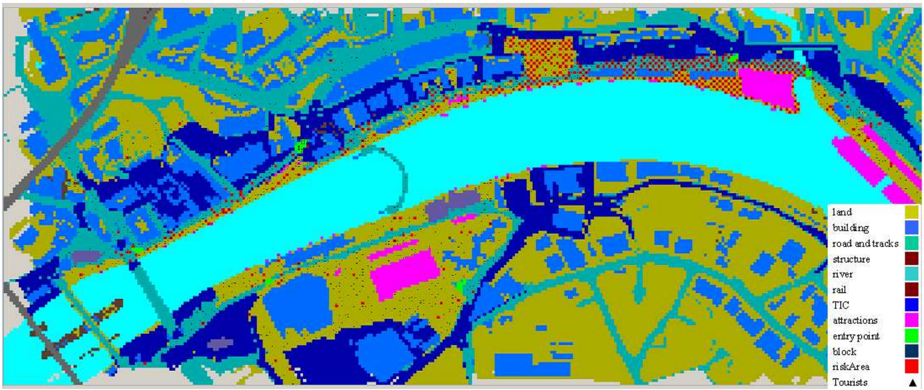


Figure 2 Risk areas detected on the base map (dots in cells indicate risk, at one time step in the day)

The model was then modified to demonstrate the effects of changing parameters, specifically those related to tourist numbers, tourist behaviour and special event assistance that can be provided, such as signposting and volunteer guides. The impact of improved signage was improvised by widening the agents' radius of perception, in effect bringing more attractions onto the list of targets. Figure 3 shows that this has an effect on the number of cells labelled as being at 'risk' of congestion.

Figure 4 has increased the number of visitors to a more realistic 200,000 per day and the resulting actual congestion (not just risk of congestion) pattern is visualised in ArcGIS. When the number of Tourists is doubled, the number of congestions tripled or even quadrupled. The pattern of actual congestion areas appears to be close to some entry points for the smaller number of Tourists, spreading to a more universal picture of congestion with the higher number.

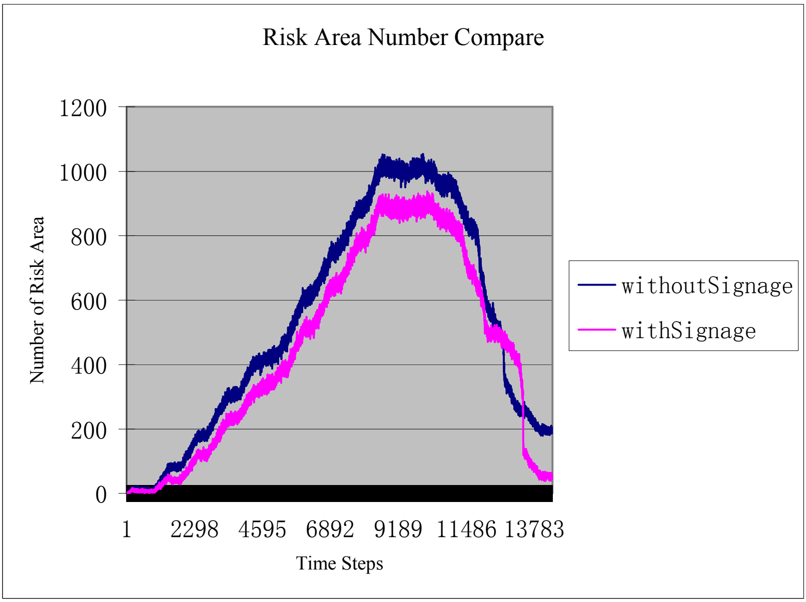


Figure 3 Comparison of numbers of risk areas (cumulative number of cells at risk without signage=7571567, cumulative number of cells at risk with signage=6474820)

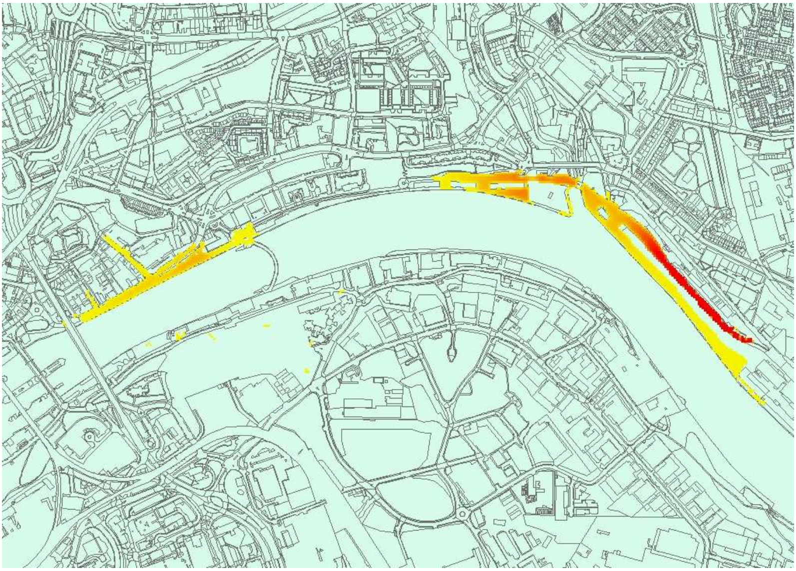


Figure 4 Summary high frequency congestion area map for one day (the shaded areas have been congested at least once during the day, the darker areas showing highest frequency congestion areas)

5 Conclusions

The real-time TSR simulation model has provided a means of better understanding the spatial features of crowds and provided ways to visualize the spatial and temporal performance of special event crowds. It has shown that GIS technology can be integrated with ABM to establish a spatial-temporal process model aiming to capture the spatial features of the evolution of a complex system. Although the Tourist behaviour rules designed in this research were very simple, they imply a great potential for simulating intelligent agents by using ABM. It is possible to modify the visitors' behaviour rules to let the simulation results become much closer to reality.

6 References

Batty, M., J.Desyllas, and E.Duxbury (2003) "The discrete dynamics of small-scale spatial events: agent-based models of mobility in carnivals and street parades." International Journal of Geographical Information Science **17**(7): 673-697

Cormas (2008) Software available at http://cormas.cirad.fr/en/outil/outil.htm (accessed December 2008)

Epstein, J. and R.Axtell (1996) Growing Artificial Societies. Brookings Institution Press

Harney, D. (2002) "Pedestrian modelling: Current methods and future directions." Road & Transport Research **11**(4): 2-12

Johnson, C. (2005) "Lessons from the evacuation of the World Trade Center for the development of computer-based simulations" Cognition, Technology & Work **7(4)**: 214-240

Klupfel, H., M.Schreckenberg and T. Meyer-Konig, (2003) "Models for Crowd Movement and Egress Simulation" Traffic and Granular Flow '03, Berlin, Springer**:** 357–372

Ory, D. and P.Mokhtarian (2005). "When is getting there half the fun? Modeling the liking for travel" Transportation Research Part A **39** 97-123

Miao Wang is a PhD student in the School of Civil Engineering and Geosciences, Newcastle University, having previously lectured in tourist geography in China. The research described here was successfully presented as an MPhil thesis in 2008.

A Spatial Structuration Heuristic for Integrated Automated Map Generalisation with Attribute and Geometry

Didier G. Leibovici, Jerry Swan, Suchith Anand and Mike Jackson

Centre for Geospatial Science, University of Nottingham, NG7 2RD, UK
Tel. +44(0)115 846 8408 Fax +44(0)115 951 5249
Didier.leibovici@nottingham.ac.uk, www.nottingham.ac.uk/cgs

KEYWORDS: map generalization, classification, competing algorithm, entropy, thematic maps

1. Introduction

Map Generalization is the process by which coarse scale maps are to be derived from fine scale maps, balancing the amount of real-world information with visual confusion, Robinson et al. (1995). This requires the use of operations such as simplification, selection, displacement and amalgamation of features that are performed subsequent to scale change. Recently, focusing on the attribute values of the geometrical objects, some research has been on thematic maps, Liu et al. (2003), Steiniger and Weibel (2005) and Revell (2007), with application in demographic maps, soil maps, land cover and land use maps. In ecological, public health or epidemiological monitoring, presenting maps to decision makers needs often to be "simplified" for better communication or privacy concerns, Waller and Gotway (2004), Curtis et al. (2006).
For these situations, algorithms need to consider the ontology associated with the theme and/or statistical clustering methods as well as geometrical transformations. In Leibovici et al. (2008) two approaches were considered: performing sequentially geometrical and attribute generalisation algorithms, or integrating the two types of transformations in a combined optimisation approach. This paper extends the research on the second approach.
A general generic framework algorithm is described to allow competing optimization at each iteration step. A similar approach is done in Neun (2008) using web services and multi-criteria selection and collaboration filtering for orders of transformations, Foerster et al. (2008). The herein approach is nonetheless seen as operating at a micro-level or atomic level, therefore proposing a real integration between the two types of generalization. The paper focuses particularly on the spatial structuration goal based on a spatial entropy approach. Emphasising on some potential extensions, this generic approach allows also direct comparisons of the sequential and combined approaches, and the possibility of testing families of operators in competition with more traditional ones.

2. Integrated Optimisation approach

This map generalisation approach aims to iteratively transform locally the objects of the map, in order to give a "simpler" visual perception, while "preserving" most of the information. At each iteration, an *admissible transformation* is selected among the collections of geometric transformations and attribute transformations, increasing the most the impression of "spatial structuration" in the map. Within this *minimax* strategy, only the maximisation is really part of the optimisation; the minimisation part is left to the term of *admissible*: initial settings of transformations. A full *minimax* could be done either by looking for the transformation with a maximum heuristic among all families of transformations but minimum among its own family, or by adding a second heuristic.

2.1 Map characteristics

A map ***M*** is described by characteristics related to the objects composing the map: the geometry characteristic ***geo***, and the attribute characteristic ***att***. Each characteristic is referring to some object properties and is measured by one or more characteristic descriptions or variables.

- For ***geo*** the characteristic descriptions, c_{geo}, can be ***(#vertices***, ***circularity***, ***orientation ....)*** as here, the objects are seen as polygons, c_{geo} can refer to one or to the whole vector of variables.
- For ***att*** the variables are the attribute variables ***(v1, v2, ...)***, but functions of these can also be

used.

The ***circularity*** ratio, defined as ***4π Area /Perimeter***^{2}, measures the compactness of the shape, with a maximum of one for a circle, and a minimum limit of zero for a narrow and infinitely long shape. The ***orientation*** of a polygon is not clearly defined here, but could be for example the collection of the angles of the segments (making the polygon) and any functional derived from it. This characteristic description could be useful when considering schematisation where one would like to "align" as much as possible the polygons.

2.2 Characteristics classes

Either just by their different values occurring within the map or by a grouping associated to them, the characteristic descriptions define for each characteristic a system of equivalent classes. c_{geo} has kc_{geo},classes, and c_{att} has kc_{att} classes.

These classes illustrate the distributions of the characteristic descriptions (variables).For each characteristic, each of the ***n*** objects in the map belongs to a particular class. During the generalisation process, we have the option of keeping records of the changes of its membership, and/or transforming the object for its characteristic variable values to match the one representing the class it belongs to, *e.g.* for a numerical variable using the mean or median of the object values within the class, for the ***#vertices*** variable, transforming "as much as possible" the objects so their number of vertices (***#vertices)*** gets closer the min or the median for the entire class of polygons, *i.e.* reducing the variation within the initial classes.

Notice it would be possible to split the list of variables attached to a characteristic to define either a multidimensional membership function, or sub-characteristics class memberships, by then enhancing or refining the descriptions and differentiations of the objects of the map.

2.3 Micro Transformations

For attribute and geometric characteristics a set of transformations operating directly on the characteristic classes and/or on the map objects allow the map to evolve to a generalised display. Geometric transformations come from usual approach of operators in generalisation, Jones and Ware (2005), whilst attribute transformations are derived from the general statistical clustering literature, with or without spatial constraint, Leibovici et al. (2008).

2.4 Heuristic for Spatial Structuration

The Shannon entropy ***H()***, of a distribution describing the spread and organisation of the information, has already been investigated in GIS, *e.g.* Bjorke (1996). As the entropy increases the organisation or structuration diminishes and become more uniform. This is illustrated by the fact that for ***k*** classes a uniform distribution will have maximum entropy and a more spiky distribution will have lower entropy, becoming zero for one class getting the entire sample. Using the same principle, the objective function (***Of***) or heuristic value to maximise is, for a characteristic ***i***, for a transformation T_i operating according to the characteristic description c_i on the map ***M***:

$$Of_i(T_i,(c_i,M)) = -H(p_k) = \sum_{k=1}^{n_{c_i}} p_k \log(p_k) \tag{1}$$

where p_k are proportions of the classes, ($p_k = 0$,the contribution to the entropy is ***0***). To take into account the spatial domain, Li and Claramunt (2006) introduced some spatial weights in this formula, in our approach the focus is on the distribution itself. As we are looking at spatial organisation, the distribution of cooccurrences of classes is used instead, and p_{mk} refers then with a multi-index ***mk*** to the proportion of collocations within a given distance ***d*** of the classes expressed in the multi-index. Figure 1 gives an explicit formulation of the collocation entropy, using the table of spatial cooccurrences, here of order ***3***. Distance of collocation and order of collocation are then parameters of this spatial entropy calculation. In order to normalise the entropy, regarding to the number of classes, one can use the ratio to the uniform case, that is, $p_k = 1/nc_i$, or $p_{mk} = 1/nc_i^3$ for collocation entropy of order ***3***, which gives an entropy of ***log(***nc_i***)*** or collocation entropy of ***log(***nc_i^3***).*** With normalised spatial entropy using collocation distribution, the formula (1) becomes:

$$Of_i(T_i,(c_i,M)) = -H_N^{C,d}(p_k) = log(n_{c_i}^3)^{-1} \sum_{mk} p_{mk} \log(p_{mk}) \tag{2}$$

For the attribute characteristic looking for more spiky distribution of cooccurrences of the classes will clearly transform the map towards a "simpler" perception. For the geometric characteristic this is not always obvious and depends on the characteristic description used. For example with the number of vertices (***#vertices***) as characteristic description, having collocations of polygons with the same number of vertices will not necessarily provide a simple perception, but if some polygons are already simplified (a small number of vertices) the collocations with them will force the algorithm to simplify the neighbourhood polygons as well.

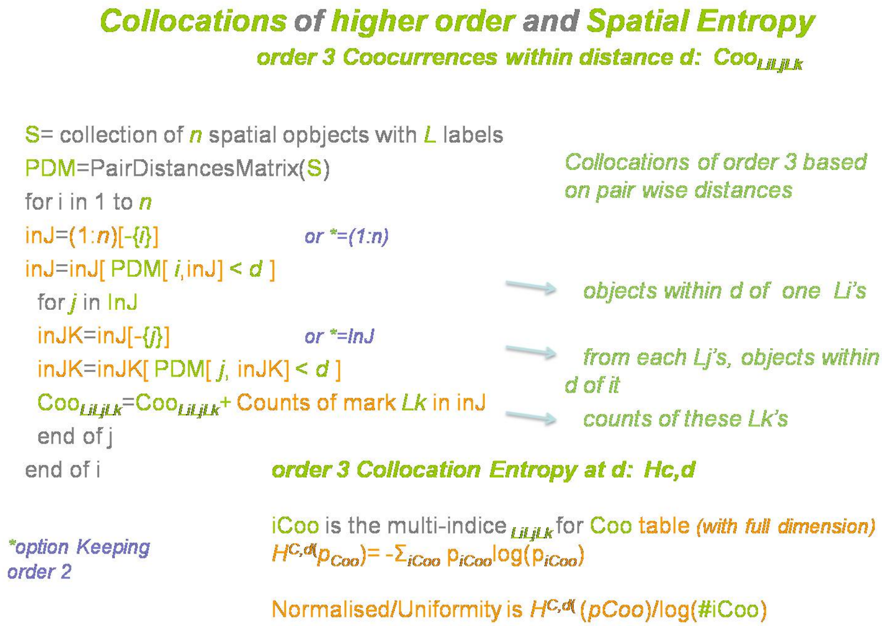


Figure 1. Spatial Collocation Entropy and Algorithm to compute cooccurences table for collocations of order ***3***

2.5 Combined Generalisation Algorithm

For each characteristic seen through their characteristic description, a list of transformation operators has to be defined before running of the algorithm. These transformations are said admissible in term of minimal or micro-transformations simplifying the map with underlined minimum loss of information. Applying a particular transformation may have impacts on the other characteristic descriptions; therefore the optimisation nonetheless combines the characteristics by looking for the best transformation among the whole list of operators according to the heuristic given in equation (2) but also needs to combine these heuristics. This is done either with a multidimensional optimisation:

$$\tilde{T} = arg \max_{\substack{i \in \{att,geo\} \\ T_{att} \in \mathfrak{F}_{att} \\ T_{geo} \in \mathfrak{F}_{geo}}} (Of_{geo}(T_i,(c_i,M)), Of_{att}(T_i,(c_i,M))) \quad (3)$$

where $\mathfrak{F}_{att}$ and $\mathfrak{F}_{geo}$ are families or lists of operators associated to the mentioned characteristic. With a simple summation of the heuristics as they are always positive and they have been normalised, one gets:

$$\widetilde{T} \; = \; arg \max_{\substack{i\in\{att,geo\} \\ T_{att}\in\widehat{\Im}_{att} \\ T_{geo}\in\widehat{\Im}_{geo}}} \left(Of_{geo}(T_i,(c_i,M)) + Of_{att}(T_i,(c_i,M))\right) \qquad (4)$$

Notice the multidimensional optimisation implies the use of a pseudo-order, here in dimension 2; for more than 2 dimensions this may be preferred as the sum.

The use of one micro local transformation may lead to a non significant change in the heuristics, so a strategy could be to increase the number of transformations (composition of operators) until a change of the objective function occurs. The objective function will then look like:

$$\widetilde{T} \; = \; arg \max_{\substack{i_1,\ldots,i_{ns} \\ n_s minimal \\ \in\{att,geo\} \\ T_{att}\in\widehat{\Im}_{att} \\ T_{geo}\in\widehat{\Im}_{geo}}} \sum_{m_c\in\{att,geo\}} Of_{m_c}(T_{i_{n_s}} \circ \ldots T_{i_2} \circ T_{i_1},(c_i,M)) \qquad (5)$$

where the minimality of $\boldsymbol{n_s}$ means it is the smallest number such as the increase of the whole objective function is "significant" (increase or threshold).. In other words, if for $\boldsymbol{n_s = 1}$ there is no best $\boldsymbol{T}$ one looks for the best $T_{i_2} \circ T_{i_1}$ so $\boldsymbol{n_s}$ is set to **2**, and so forth. The composition order of the $\boldsymbol{n_s}$ operators may play an important role in the optimisation

3. Further Discussion

The generic framework, developed with java using geotools libraries, will be available as a web processing service. The flexibility of this framework allows exploring computationally and algorithmically some properties of the integrated combined approach:

- direct comparison of, and with, sequential approaches,
- multi-competing optimisation
- constraints setting
- schematisation operators

Concrete examples will be demonstrating the potentials expressed in this abstract, in the domain of demographic and public health mapping.

References

Bjorke, J.T (1996) Framework for Entropy-based Map Evaluation. Cartography and Geographic Information Science, 23(2), p.78-95(18).

Curtis, A Mills, JW, Leitner, M (2006) Keeping an eye on privacy issues with geospatial data. Nature 2006, 441:150.

Foerster, T Burghardt, D Neun, M Regnauld, N Swan, J and Weibel, R (2008): Towards and Interoperable Web Generalization Services Framework – Current work in progress. Proceedings of the 11th ICA workshop on Automated generalization and Multiple representation.

Jones, C.B and Ware, J.M (2005) International Journal of Geographical Information Science: Map Generalization Special Issue (Taylor and Francis), Volume 19, Number13 8-9, pages 859-997.

Leibovici, D.G Swan, J Anand, S and Jackson, M (2008) Combining Attribute with Geometry for Automated Generalization and Schematisation. AutoCARTO, USA, 6-10 Septembre 2008.

Li, X and Claramunt, C (2006) A spatial-based decision tree for classification of geographical information , Transactions in GIS, 10(3), Blackwell Publishing, pp. 451-467.

Liu, Y Molenaar, M Ai, T and Liu, Y (2003) Categorical Database generalization aided by Data Model. Proceedings of the 21st International Cartographic Conference 201, 10-16.

Neun, M (2008) Automated processing for map generalization using web services. GeospatialWeb Services workshop University of Nottingham 2008, http://www.nottingham.ac.uk/geography/geowebservices/

Revell, P (2007) Generic Tools For Generalising Ordnance Survey Base Scale Landcover Data International Cartographic Association 10th conference on Generalization and Multiple Representation, 2-3 August 2007, Moscow.

Robinson, A.H Morrison, J.L Muehrcke, P.C Kimerling, A.J and Guptill, S.C (1995) Elements of Cartography, 6th Edition, New York: John Wiley & Sons.

Steiniger, S and Weibel, R (2005) A Conceptual Framework For Automated Generalization and its Application to Geologic and Soil Maps. Proceedings 22nd International Cartographic Conference, La Coru-a (Spain), 11-16 July 2005.

Waller, L. A., and Gotway, C. A. (2004). Applied spatial statistics for public health data. Wiley series in probability and statistics. Hoboken, N.J., John Wiley & Sons.

Biography

Dr Didier G. Leibovici, is a Research Fellow in geospatial modelling and analysis, with previous posts as statistician in epidemiological/medical imaging research and as geomatician for landscape changes in agro- ecology. Interests refer to interoperability and conflation models with cross-scales for integrated modelling applications within an interoperable framework with chaining web services.

Dr Jerry Swan, is a Research Fellow in Geocomputation. His approach includes optimization, graph theory, symbolic computation, knowledge representation, machine learning, semantics and ontologies. Jerry is particularly involved in the Persitent Test Bed project commissioned by OGC, AGILE and EUROSDR.

Dr Suchith Anand, is an Ordnance Survey Research Fellow at the Centre for Geospatial Science, University of Nottingham. His research interests are in open source GIS, automated map generalization, geohydroinformatics, mobileGIS, location based services, optimization techniques and asset management.

Pr. Mike Jackson is Director of the Centre for Geospatial Science. Prior to this he worked in industry at QinetiQ, Hutchison 3G, Laser Scan in various geospatial specialist and executive roles and in research for NERC. Mike is non-executive director of the Open Geospatial Consortium, and has research interests in combining new technologies such as positioning, pervasive computing and location based services for geo-informatics applications

Characterizing maps to improve on-demand cartography - the example of European topographic maps

Laurence Jolivet

IGN/COGIT, 2 avenue Pasteur, 94 160 Saint-Mandé
Tel. +33 1 43 98 80 00, Fax +33 1 43 98 81 71
Email: laurence.jolivet@ign.fr

KEYWORDS: on-demand mapping, European maps, Web interface, data customisation

1. Context and approach

An on-demand map must meet the map maker goal: to convey a specific message to readers. It implies that the map design process integrates somehow the map maker goal. Existing mapping applications already offer tools for customisation. The help concerns the access to data, the possibility to choose among predefined styles and the display of a map in a specified format. It is one of the aims of the Web Map Context OGC standard (2006) that supports the specifications of a map from a WMS. A user is given the opportunity to realise a personalised mapping of available data, but he is bound to know which data are to be selected and which styles are adapted to these data and to the general map specifications. Our objective is to provide an automatic map making process on the Web that would allow a good translation of initial specifications into a map design so that the map convey the right message to readers.

The difficulty in the automation of mapping processes comes from the various manners of realising a map and from the potential alternatives in interpreting the specifications concerning the map goals and constraints. No generic method exists for map making, even though it follows well-defined logical steps. Our proposal is to follow the steps of traditional mapping process thanks to implemented cartographic knowledge within a Web environment. Our technical response is a chain of modular Web Services that helps in defining a formal map description and in providing an adapted data selection and symbolisation to the map description. The map description model includes legend themes, legend lines, data relations - association, order, difference - and reading levels as listed by Bertin (1967). This general proposal is presented in Bucher (2007) and in Jolivet (2008).

To accurately take into account the map objectives, we have extracted from theory cartographic principles; and studied how users characterize existing maps. Indeed, one challenge is to be able to determine which map specifications are important to obtain an on-demand map and how to translate them in the map display. For example, Regnauld (2007) presents different types of ontologies to formalise knowledge in order to deploy them in an automatic generalisation system. Especially in a Web context, specifications have to be sufficiently accurate and formal for automated interpretation of user objectives.

In the next parts, we first identify what characteristics of a map could be included in a formal map description. Then, we focus on how to translate abstract legend definition into displayed symbolisation. Hence, we study legends perception in a more sensitive way. As an example, legends of European topographic maps are analysed thanks to comparative works and user tests. A prototype of an automated on-demand map process has been implemented: the map maker may specify different map characteristics and the system interprets them into changes in data selection and in display styles of the working map.

2. Identifying relevant items to specify a map

2.1 Characterization from existing maps

Cartographic theory suggests expressing map specifications as a scale, a format, a geographical area and so on. We focus on how specific map specifications can be integrated in a map description and can be translated on legend elements - data content and symbolisation - so that to correspond to a particular map objective. A map description includes indeed the symbolisation, with colours perception, that is linked to the content of the map within geometric representations and legibility considerations as mentioned in Brewer (2007).

Our approach to identify relevant elements for map specification is to study how we describe existing maps. It means by extracting what can be seen on the whole maps, what thematic data emerge, what terms correspond to the map. A first experiment is described in the work of Dominguès (2006), where characterization is carried by qualifiers. Some of these words tend to be impartial like “net”, “contrasted”, and others are more ambiguous to use: “beautiful”, “original”. These elements are part of the map formal description. They can be given by analysing from the map display, how its initial specifications have been translated. Decisions during the process can be deduced from the map display. For instance, choices in the legend could have been made from the type of maps: risks maps, topographic maps; from the objective of display: communication, analyse; from the user profile; from the importance given to specific data or from symbolisation preferences.

2.2 Examples of legend characterisation from the European topographic maps

Following the experiment mentioned above, we have studied how people tend to describe various topographic maps. The goal is to identify relevant properties they could use to specify the map they need. Topographic maps at the scale 1:50 000 were considered from 15 European states. A set of maps was constructed with the legends applied to three extractions of IGN data. These three extractions correspond to area types - urban, rural and mountainous. To adapt a European legend to French data model, we tried to keep the consistency with the thematic elements: hydrology, roads, buildings, land use, topography.

Figure 1. IGN data displayed with symbolisations of topographical maps from Austria, the Netherlands and Finland

A user test was performed on the constructed maps. Firstly, couples of opposite-meaning qualifiers were defined after Dominguès (2006). Marks were asked to evaluate how these qualifiers correspond to the map from the legend. The qualifiers deal with colour perception: “drab vs. luminous”, “pastel vs. garish”, “non esthetical vs. beautiful”, “sober vs. rich”. The qualifiers also deal with the whole symbolisation on the map: “realistic vs. artistic”, “balanced vs. emphatic”, “untidy vs. accurate”, “heavy vs. lighten”. Second, a question focused on the themes rendering: the goal is to underline which symbolisation choices put forward the specific cartographic thematic elements listed earlier. Last, users indicated how the legend appeared as adapted to the different area types.

The test highlights sets of legends associated by the same qualifying terms. It also highlights correlations between the qualifiers. Legends considered as 'heavy' will be most likely qualified as also 'untidy' and 'non esthetical'. In the same way, 'sober' is associated with 'pastel' and 'realistic'. The thematic elements are emphasized by the legend. It appears logically that the visual importance of a theme is due to its own symbolisation and also to interactions on the map with the other symbolisations. Then, some legends were defined as adapted to certain types of objects repartition corresponding to the area types. For example, a legend would be defined as more adapted for urban areas if the buildings are in darker colours, whereas mountainous areas would support better visualisation of the topography.
Elements of legend characterisation emerged from the test results. They give indications on how adapting a map legend to the map formal description. Associations of the coloured symbolisation are in particular important to translate map specifications on the data to display.

3. Presentation of a prototype and conclusion

In our prototype for on-demand mapping on the Web, the collection of user specifications is supervised from a graphical interface. The specifications are integrated to the map description: geographical area, scale, map type and objective. Importance on data and preferences on the symbolisation can also be fixed. Requests according to user specification are made on characterised European legends. The characterisation allows the selection of an adapted legend. If it is necessary pre-defined legend can be adapted thanks to cartographic principles. In figure 2 is illustrated a map display automatically provided from user specifications.

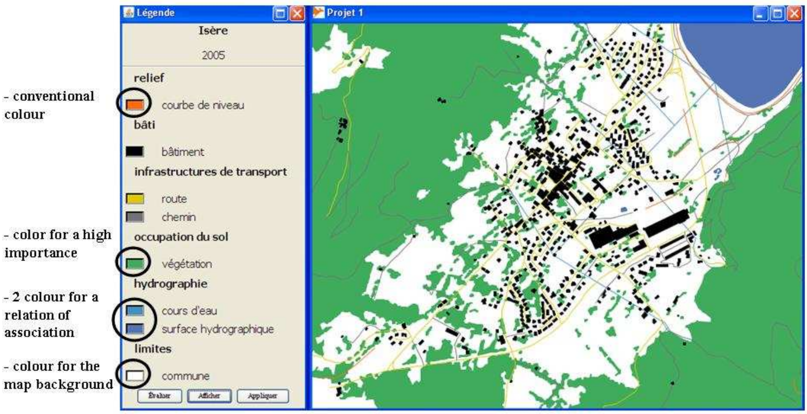


Figure 2. Data display integrating user specifications, here a 1:50 000 pre-defined topographic legend with importance given to the vegetation theme

The presented work aimed at integrating implicit knowledge in a formal map description for an automatic on-demand maps design process. The approach concentrates on the characterisation of existing maps. It allows the request of pre-defined legends for adapting the map display to specific user needs. It would be interested to enlarge map customisation and to be able to propose more adapted legends.

References

Bertin J. (1967). *Sémiologie Graphique*. Édition Mouton, Paris.

Bucher B. Jolivet L. Buard E., Ruas A. (2007). *The need for Web services.* In J.M.Ware and G.E.Taylor (Eds), W2GIS, Springer Lecture Note in Computer Science, Cardiff, UK.

Brewer C. Hatchard G. and Harrower M. (2003). *ColorBrewer in print: a catalog of color schemes for maps*. Cartography and Geographic Information Science, vol. 30, n°1, 2003, pp. 5-32.

Christophe S. (2007). *Legend design by a novice user on the Web.* GIDAYS'07, Münster, Germany, 10-12 September 2007.

Dominguès C. Bucher B. (2006). *Legend design based on map samples.* In proceedings of the 4th GIS Science conference, extended abstracts, Raubal, Miller, Frank and Goodchild (Eds), IFGI Prints, Münster, September 2006.

Jolivet L. (2008). *On-demand map design based on user-oriented specifications*. AutoCarto conference, Shepherdstown (WV), USA, 8-11 September 2008

Open Geospatial Consortium. (2006). *OpenGIS® Web Map Service Implementation Specification. OpenGIS® Implementation Specification. Beaujardiere (de la),* J. (Ed), 2006. http://www.opengeospatial.org/standards/wms∞∞.

Regnauld N. (2007). *A distributed system architecture to provide on-demand mapping.* In proceedings of the ICC Conference 2007.

Biography

Laurence Jolivet is a member of the COGIT laboratory of the IGN France. Her works concerns the access to mapping functionalities by the Web. It focuses on automatic processes for on-demand mapping from user specifications to maps display.

GIS, Reassurance Policing and 'Signal' Crimes

Paul Richards, Paul Longley, Alex Singleton

University College London,
Department of Geography and CASA, Pearson Building, Gower Street London WC1E 6BT
Tel. +44 (0)20 7679 0500
Email, paul.richards2@ucl.ac.uk, plongley@geog.ucl.ac.uk, a.singleton@ucl.ac.uk

KEYWORDS: reassurance policing, police databases, crime mapping, police activity mapping, London

1. Introduction

The present deployment of GIS within policing is most obviously apparent in the widespread use of crime mapping (see Chainey and Ratcliffe 2005). The established practices of crime mapping are undergoing change with the advent of new requirements for police to produce crime maps for the public under Recommendation 4 of a Statistics Commission (2006) report and recent Home Office circulars (e.g. Home Office 2008). This requirement needs to be balanced against the concerns of the Information Commissioner about privacy (Guardian 2008a), adverse public perceptions (e.g. BBC 2008) and issues of copyright (Guardian 2008b). Taken together, these concerns have potential to widen the 'reassurance gap' in UK policing.

Millie and Herrington (2004) discuss the 'reassurance gap' in the following terms: "Recent years have seen falls in recorded crime in England and Wales, although the perception for many has remained that crime is rising. This has been dubbed the *reassurance gap* and is closely related to concerns over fear of crime and public confidence in the police. *Reassurance policing* (RP) is seen as a means to address this 'gap'." The perceived link between this gap and confidence in the police means that the Metropolitan Police Service (MPS) has come to see its management as a key strategic outcome. The context to reassurance policing is that:

- Communities should feel engaged with, confident in and satisfied with the police service
- Security should be improved and the public should feel safe
- Crime, disorder, vulnerability and harm should be prevented and reduced
- More offenders should be brought to justice

(MPA & MPS 2007)

Since April 2004, a key element of reassurance policing in London has been the introduction of Safer Neighbourhood Teams (SNT), each consisting of a uniformed police sergeant, a number of constables and dedicated Police Community Support Officers (PCSO). These have been created for each of the 624 council ward areas in Greater London (MPS 2008). In addition there are non-ward based SNTs in areas requiring additional police resources. By doing this, the MPS has adopted the 'reassurance' policing style that has been viewed favourably by the Home Office (Tuffin et al 2006). As the name suggests, reassurance policing seeks to increase the community's sense of security and confidence in the police.

2. The Problem

Reassurance policing involves communities and their police sharing their perceptions of crime problems in their areas (Innes 2005). Key to this shared view is the creation of crime maps which contain detail down to ward and sub-ward levels. Yet to date, the crime maps that have been produced on the Web by police and the government have not matched people's perceptions of crime in their communities. Such mapping is based up on selected police recorded crimes, and there are widely understood concerns about public lack of confidence in officially produced statistics (Office for National Statistics 2005), shared even by the MPS Commissioner (Daily Express 2008). Academics have suggested (e.g. Innes 2005, Sparks et al 2001) that communities base their perceptions of crime upon diverse sources of information that bear little relationship to the crimes that are recorded by police. Innes (2005) concludes that the most significant influence upon perception are "signal" crimes or incidents. These signal events often arise from disorder and anti-social behaviour, yet they are not usually recorded as crimes by the police. Signal crimes are important because by definition they change people's behaviour or attitudes (Innes 2004, Innes and Fielding 2002, Innes 2003, Innes et al. 2004). Crime maps based on police recorded data are always likely to provide an incomplete picture of crime in an area for the obvious reason that police need to discover the crime or have it reported to them. Even when this is done, legal definitions, Home Office counting rules and performance related pressures from the local police and the Crown Prosecution Service (CPS) all influence what is actually recorded. However, even with these limitations there is still vast potential for improving the quality of crime maps using the police crime recording database. For instance, there is an obvious potential benefit in highlighting signal crimes and showing how police are dealing or have dealt with them.

3. Towards more effective use of police response data

This paper adopts a new and innovative perspective upon crime mapping from the stance that public understanding of "crime" is influenced by the process of reporting rather than the technicalities of recording it. What members of the public report may not be considered by the police to be recordable crimes, but they are nevertheless key to understanding the perceptions of reporting individuals. If the report involves a visible emergency response by the police, the knowledge of the response but not necessarily the nature of the report itself is likely to spread within the community. Police responses that involve a multiple number of police units are likely to meet some of Innes' (2005) criteria for a 'signal crime'.

This paper reports on the outcome of analysis of a new dataset that records police responses to incidents, based on calls from the public and patrolling officers' own discoveries. The challenge is to extract data that provide an accurate indication of levels of crime and disorder in areas that can be used to interpret and complement other datasets. The other datasets discussed in this paper are the police crime and GPS datasets.

An important advantage of the way in which police responses are logged in this database is that it provides a good indication of where concentrations of high visibility police activity occur. Reassurance policing is based on the premise that the public find it reassuring to have police patrolling on foot in their neighbourhoods (Povey 2001). It may also be the case that police responding to incidents in fast cars can also bring reassurance but ironically this may also be a factor in the perception of an increasing crime rate in an area. These are interesting nuances that parallel the debate around the value of publishing crime maps, and are also being investigated in this research.

Over the last four years, the MPS has conducted intensive research into community perceptions of policing, crime, disorder and anti-social behaviour in seven of the 624 council ward areas in Greater London. This paper reports on the analysis of police response data that has been carried out on these wards. Analysis of four of the seven wards is used here to show the how the response data compares with the crime data. Figure 1 shows total number of crimes recorded by the police in each ward in the financial year 2007/8. It should be noted that Roehampton experienced the lowest number of police recorded crimes of the four wards, about half that of New Cross.

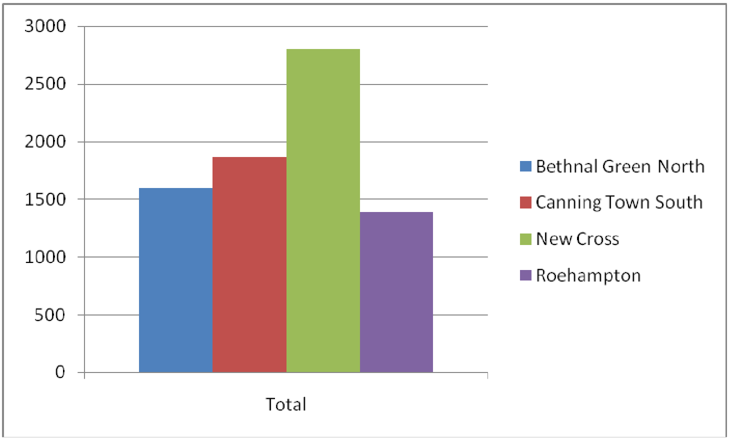


Figure 1: Total number of police recorded crimes in 2007/8 in the 4 wards

Figure 2 shows the total number of calls in each ward made by members of the public in 2007/8 that have resulted in police responding in "emergency mode" where five or more units have been assigned. These units are most likely to have been police vehicles that cover a whole borough of approximately 20 wards. Assignment of five or more units to the same call represent a significant proportion of any borough's response resource, illustrating the importance of the incident to police and the likely visual and audio impact upon the public. It is a hypothesis of this ongoing research that a significant proportion of these incidents are "signal".

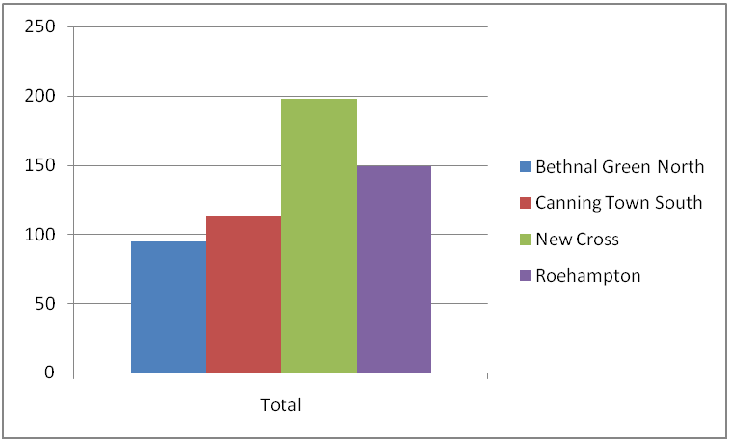


Figure 2: Total number of signal incidents in 2007/8 for the 4 wards

Figure 2 shows that even though Roehampton had fewest police recorded crimes it had the second highest number of signal incidents. Further analysis of the response data shows that in Roehampton a disproportionate number of signal incidents (when compared to the other three wards) occur in the evening, night and at the weekend meaning that a person working office hours outside the Borough could gain a perception that the crime levels in Roehampton are similar to New Cross, contrary to the police perception that the crime level is about half that of New Cross. For example, Figure 3 shows that on Saturdays there are more signal incidents during the year in Roehampton than in New Cross.

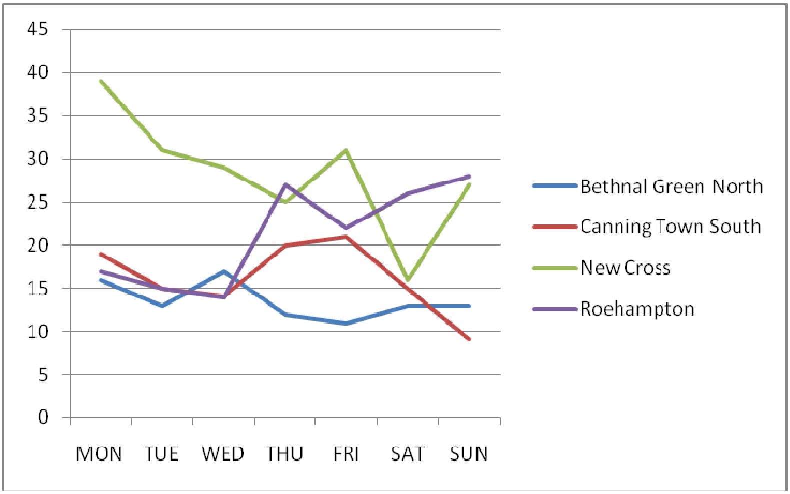


Figure 3: Number of signal incidents generated by calls from the public distributed over the week

The police response dataset allows incidents that have been initiated by police to be differentiated from those that have been initiated by calls from the public. Analysis of these data suggests that the level of police response teams' proactivity and visibility in Roehampton was more consistent with the level of police recorded crime than the level of signal incidents.

As discussed above this mismatch of perception is likely to lead to a 'reassurance gap' and therefore dissatisfaction with police performance. There are indications of this in the MPS tracking wards survey report for Roehampton ward.

4. Additional data for analysing police response and visibility

A police vehicle responding to an emergency call will usually use flashing blue lights and, if impeded, will use traffic sirens and when safe may break the speed limit. The response dataset does not include the location of response vehicles when assigned, but just the location of the incident and the location of the caller if given. So another dataset is required to assess the impact of police vehicles in "emergency mode" and travelling to incidents through other wards to get to the emergency call. All MPS response vehicles now have mobile data terminals that transmit a GPS signal every 15 seconds when switched on. Trevor Adams' group at the Metropolitan Police has carried out some exploratory work plotting these signals in the Borough of Lewisham.

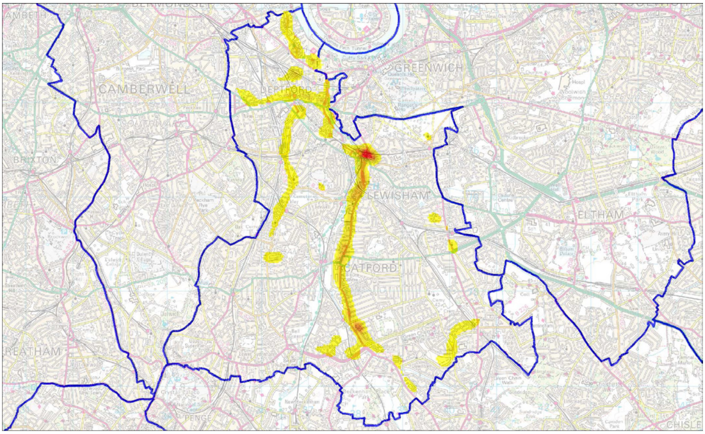


Figure 4: Map of the London Borough of Lewisham showing police response car GPS locations between 2200hrs and 0000hrs

Figure 4 is a hotspot map of police vehicle activity. This map changes over the day but there is constant activity up and down the north/south Bromley Road through Catford. The GPS signal provides vehicle speed so it should be possible to deduce when it is acting in response mode. Once this analysis is carried out it will be able to indicate which areas are regularly affected by police vehicles responding to emergencies. Residents in these areas perception of the crime rate may be influenced by this.

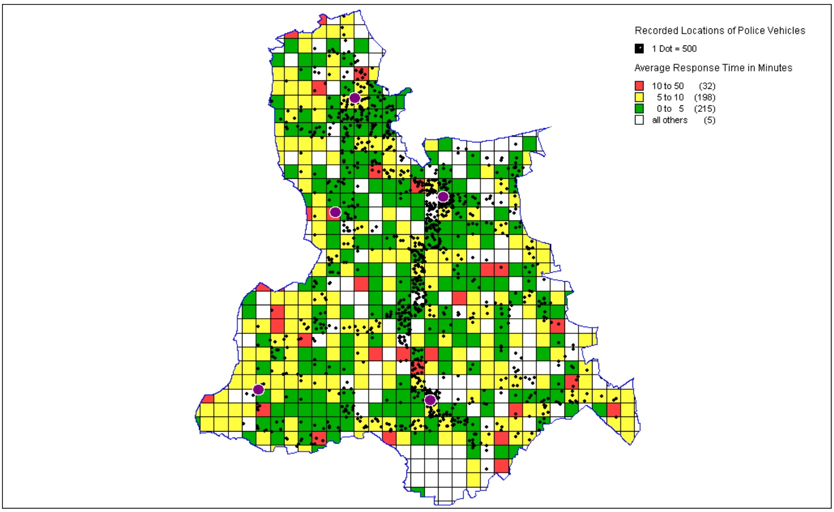


Figure 5: Map of the London Borough of Lewisham showing police response car locations in one calendar month (each dot represents 500 signals at that location) and average response times in minutes

Figure 5 presents a different way of representing response car locations by showing locations that have had 500 GPS signals transmitted from them in a calendar month. This is overlaid by grid data of police response times from the police response database. This allows further analysis of the dynamics of police visibility and the efficiency of police responding quickly to calls. A conclusion that could be drawn from the analysis discussed in this paper is that police should be more selective in which calls they respond to in emergency mode. The danger of so doing is that reassurance is reduced by police not responding quickly enough to calls that the community think are important.

In the next 18 months it will be possible to analyse the locations of all patrolling officers in the MPS from the GPS signal transmitted from their police radio (McClimont 2008). It will then be possible to assess the amount of police visibility there is at different locations and the impact on reassurance.

5. Conclusion

Reassurance policing is about visibility, accessibility and familiarity (Povey 2005):
"*It would be naïve to focus simply on visibility – reassurance flows as much from the style of policing as the visible presence. A police car speeding by with lights flashing and sirens blaring signals trouble. The 'feel-good factor' comes instead from officers who are known and accessible – preferably on foot patrol – and who are skilled at engaging with local communities and their problems.*"

Creating a feel-good factor about reassurance policing requires officers who are skilled at engaging with local communities and their problems rather that the inaccessible speeding car. Tactical and strategic deployment of these resources can be considerably enhanced using previously unused datasets to add relevance, direction and a sense of joint achievement. Three datasets have been discussed in this paper; crime, police response and GPS. These and other police datasets will continue to be researched along with other GIS data with the aim reducing the fear of crime and increasing communities' confidence in the police.

6. Acknowledgements

The author wishes to acknowledge the data supplied by the MPS and particularly the use of analysed data provided by Neil McClimont.

References

BBC (2008) *Will Crime Maps work? 26th August 2008.* Accessed at http://news.bbc.co.uk/1/hi/magazine/7577659.stm
Chainey, Spencer and Ratcliffe, Jerry (2005) *GIS and Crime Mapping*. Chichester: Wiley
Daily Express (2008*) Little faith in crime stats - Blair 18th June 2008*. London: Daily Express
Guardian (2008a) *Is the Information Commissioner against crime maps? 5th June 2008*. Manchester: Guardian
Guardian (2008b) *OS move raises the political stakes. 19th November 2008*. Manchester: Guardian
Home Office (2008) *New interactive local crime maps in every area*. Press release 28th July 2008. London: Home Office. Accessed at http://press.homeoffice.gov.uk/press-releases/Crime-Maps
Innes, Martin (2005) 'Reinventing Tradition?: Reassurance, Neighbourhood Security and Policing. C*riminal Justice* 2004; 4; 151
Innes, Martin (2003) *Understanding Social Control: Deviance, Crime and Social Order*. Buckingham: Open University Press.
Innes, Martin and Fielding, Nigel (2002) 'From Community to Communicative Policing: Signal Crimes and the Problem of Public Reassurance', *Sociological Research Online* 7(2). Accessed at http://www.socresonline.org.uk/7/2/innes
Innes, Martin, Trudy Lowe, Helen Mackenzie, Philip Murray, Colin Roberts and Lisa Twyman (2004) *The Signal Crimes Perspective: Interim Findings*. Guildford: University of Surrey.

McClimont, Neil (2008) Conversation September 2008
Metropolitan Police Authority and Metropolitan Police Service (2007) *Policing London Strategy for 2007- 10.* London: MPA & MPS
MPS (2008) http://www.met.police.uk/saferneighbourhoods/about.htm ; London: MPS
Millie, Andrew and Herrington, Victoria (2004) Reassurance Policing in Practice: Views from the Shop Floor *British Society of Criminology* volume 7/002
Povey, Keith (2001) *Open All Hours*. London: HMIC.
Sparks, Richard, Girling, Evi and Loader, Ian (2001) 'Fear and Everyday Urban Lives', *Urban Studies* 38(5–6): 885–98.
Statistics Commission (2006) *Report 36 Crime Statistics: User Perspectives*. London: Statistics Commission
Tuffin, Rachel, Morris, Julia and Poole, Alexis (2006) Home Office Research Study 296: An evaluation of the impact of the National Reassurance Policing Programme. London: Home Office RDSD

Biography

Paul Richards is a first year PhD student at UCL who holds at ESRC CASE award sponsored by the Metropolitan Police Service for researching "Real time geodemographics for reassurance policing and crime prevention" He is a retired police inspector having worked 30 years policing London. He has been awarded an M.Phil for Criminology and Information Systems.

Professor Paul Longley, Department of Geography, UCL. Paul Longley's research interests concerns the use of GIS and quantitative methods in urban analysis. He is Deputy Director of Centre for Advanced Spatial Analyses at UCL and UCL Site Manager of the 'Spatial Literacy in Teaching' (Splint) Centre of Excellence in Teaching and Learning.

Dr Alex Singleton completed his PhD in November 2007 and is now a Research Fellow in the Department of Geography and Centre for Advanced Spatial Analysis at University College London.

Constructing a timely GIS dataset for the City of Sydney's Census geographies to study the impacts of infill developments

Piquer-Rodriguez, M*. and Ghosh, S†.

* Urban Research Centre, University of Western Sydney
Level 6, 34 Charles Street, Parramatta, NSW 2150, Australia
www.uws.edu.au/urban Phone: +61 28833 5910; Fax: +61 29891 5899
Email: m.piquerrodriguez@uws.edu.au
† Email: s.ghosh@uws.edu.au; Phone: +61 28833 5903; Fax: +61 29891 5899

KEYWORDS: boundary, census geographies, urban form, spatio-temporal comparability.

1. Introduction

Infill urban development – the redevelopment and intensified usage of land within the existing urban landscape – has emerged as a normative goal of urban policy across the globe; with Sydney no exception. However, there has been very little critical consideration of infill urban development's wider consequences. In 2005 more than half of the houses for new inhabitants in Sydney were to be accommodated through infill development. Sydney's future well-being has therefore been staked not on continued suburban expansion, but rather on a re-configured urban form with much greater population densities.

Urban form changes over time; and so do the geographical boundaries used to study population variability (Census boundaries) to better represent dynamic urban systems. However, these changes to Census boundaries often mean different census are not comparable (i.e. over time). As Martin et al. 2002 state, the linkage of census geographies is one of the four major obstacles to analyze intercensal social change. The case of the Australian Census does not differ from others. Spatio-temporal comparability at its basic spatial unit (Manley et al. 2006) is impeded by several factors; described later in the paper. And Census counts have a data gap for 1996 CDs that forces us to use a different count that may show larger enumeration counts.

This paper reports on an ongoing research project that is examining the urban and social implications that infill development has had throughout the City of Sydney in Australia. The paper specifically focuses upon the methodological problems faced in constructing a timely GIS data set to examine population changes between 1996 and 2006 where new developments have occurred.

The paper begins by introducing the main data set used to construct the GIS. Following this, a particular set of problems associated with using Australian census data to provide temporal comparisons in urban areas are examined. A set of approaches used to resolve these issues are discussed; those particularly focused upon are concerned with issues relating to the spatio-temporal comparability of Censuses. This is followed by the method used to approach such issues. The paper concludes by discussing the generic and Australian-based methodological problems faced.

2. Dataset

Population changes in Australia over time can be studied by looking at the differences in the latest Census variables from 1996 to 2006. The Australian Bureau of Statistics (ABS) produces Census data every 5 years. It also generates and standardizes the geographical classification of spatial units for the Census. The basic spatial unit of the Census is the Collection District (CD) followed by the Statistical Local Area (SLA)/Local Governmental

Area (LGA)[1] (ASGC, 2008). Although ABS has released for 2006 experimental mesh block boundaries for Australia (ABS, 2008) they do not contain the same amount of extensive information than the current CDs. When working at local spatial scales, the LGA Census data is also found too general to depict geographical differences within the area of interest. Therefore, we decided to work with the lowest level of spatial aggregation available, the CD. By doing this we are trying to minimize the aggregation problem of the Modifiable Areal Unit Problem (MAUP) (Openshaw 1977; Openshaw and Taylor 1984) that is present in the Australian Standard Geographic Classification (ASGC) of the Census data (Blake et al. 2000). Choosing the CDs as the basic spatial unit of this study also provides maximum flexibility in further aggregations of the data in case it would be necessary.

Even then, ABS Censuses' spatial and temporal comparability at its basic spatial unit (CD) is not possible for the place of usual residence count, although it is available for SLAs and for geographies above this level. The Place of Usual Residence count represents a more real count of the usual residents of an area and is less likely to be influenced by seasonal factors (ABS, 2006a). This count is only available from the 2001 Census CDs onwards. In order to compare CDs Census data from 1996 to 2006 we had to use Place of Enumeration count that is available for 1996, 2001 and 2006 CDs. Place of Enumeration data is based on the count of people's location on the Census Night. The differences between the two counts are specified in the 2006 census population measures (ABS, 2006a). Table 1 shows the differences in selected CDs' total population for both the counts.

Table 1: Total population in selected CDs in Sydney, 2006

CD code	Place of Usual Residence	Place of Enumeration
1400123	216	976
1400103	536	1690
1402018	811	1030
1400705	466	485
1400509	777	755
1440202	968	994
1401504	717	706

(Source: ABS 2006b, ABS, 2006c)

The differences in the two counts make the comparability of realistic social-spatial CD variables more difficult. Moreover, the spatio-temporal comparability of Census CDs is hindered by poor look up tables that potentially could allow the best fit among boundaries changes of censual years (see ABS, 2006d for a review on the major problems with the current ASGC). The ABS provides comparability tables (look up tables) for changes in the CDs' boundaries through time. However these changes correspond to quantitative measures such as percentage of dwelling change more than spatial changes in the Census boundaries. Even more, these comparability tables do not have consistent change codes among different census years. Table 2 exemplifies the inconsistencies the Census look up tables for CDs have that make them complicated to use. Figure 1 shows the differences in boundaries for some of the CDs in the 1996, 2001 and 2006 Census.

Therefore, due to the difficulties faced when using spatial comparisons of Census CDs' through time, this study required the construction of temporally comparable census geographies. To our knowledge, this has never been done before for an Australian urban area.

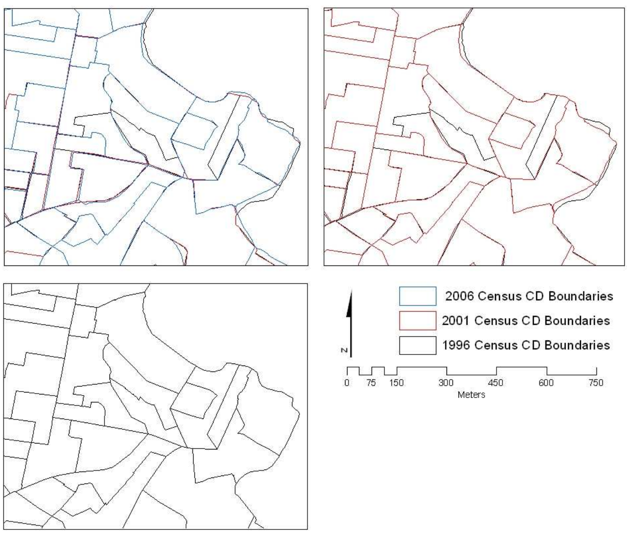
[1] For an extensive explanation on the ABS Census' Zone design please consult Blake el al. 2000.

Table 2: Censuses look up tables for facilitating the comparability among the 1996-2001, 2001-2006 and 1996-2006 Census.

Change code 01-06 and 96-01	Explanation
0	2006 CD is perfectly comparable to the 2001 CD.
1	Although a change to the CD boundary has occurred, there is no change to population and dwellings.
2	CD is comparable within a 2% dwelling limit. That is a boundary change has occurred, however, no more than 2% of the 2001 dwellings in the CD have been subtracted from or added to the CD.
3	CD is comparable within a 10% dwelling limit. That is a boundary change has occurred, however, no more than 10% of the 2001 dwellings in the CD have been subtracted from or added to the CD.
4	CD is not comparable due to a split into two parts (the 2001 boundary being retained around the two new 2006 CDs).
5	As for 4 above except that the split has involved the creation of three or more new CDs.
6	CD is not comparable due to the extent of boundary variation.
7	CD is not comparable because it is an amalgamation of two 2001 CDs to give a new 2006 CD (the 2001 outer boundaries are retained).
8	CD is not comparable due to an amalgamation of three or more 2001 CDs, retaining the same outer boundaries.
9	The 2006 CD has changed slightly when compared with the 2001 CD due to the supply of new basemap. The CD boundary follows the same features, however, due to the new basemap the boundary has been re-aligned accordingly.

Change code 96-06	Explanation
0	CD is comparable
1	CD is not comparable

(Source: ABS, 2006e)



Source: ABS, 2006

Figure 1: Examples of differences in CDs Census boundaries for 1996, 2001 and 2006.

3. Methodology

We first integrated Census CD's variables into a GIS for three different years, 1996, 2001 and 2006. Due to the substantial amount of Census variables used, some GIS stability problems were experienced. These were overcome by importing Census data tables into a File Geodatabase. Editing bugs within the GIS software were detected and resolved in cooperation with the GIS software provider. Our approach to creating consistent geographies was based on the manual adjustment of the three Census years' boundaries. The adjustment of the Census boundaries was done by adopting a freeze history approach (Norman et al. 2003 and Blake et al 2000). The 1996 Census boundaries were picked as the reference zone system and subsequent edits of zone boundaries were made to fit this year zone boundary. Norman et al. 2003 criticize the use of the freeze history approach by saying that it is not a solution that can be imposed post hoc because the frozen zones become less appropriate with the passage of time. In our case, the 1996 boundaries were found to require less labour to be adapted to and that it is why we chose them as original boundaries. A set of rules on spatial adjustments was developed facilitating and standardizing the future replication of the exercise on different geographies. This set of rules as well as a more detailed methodology will be further detailed in subsequent papers.

The labour intensive exercise of adjusting manually the Census boundaries to early geographies can be justified by the fact that "using traditional techniques for adjusting one geography to another one, such as spatial areal interpolation, are less reliable for individual population characteristics like income or education which vary systematically across space in unrelated ways to the overall distribution of population" (Blake et al. 2000 pp. 158, 172). Blake et al. 2000 (p. 173) also supported the approach of the freeze history as a solution to allow comparativeness of the ABS Census data maintaining temporal consistency by amalgamating older geographies.
An accuracy analysis of the efficiency of such an approach in relation with conventional techniques for comparing census boundaries in time is designed and yet to be implemented.

4. Conclusions

Unlike other countries, such as the UK, the integration of time-series local-level census data remains an unresolved issue in Australia (Boyle and Feng, 2002; Gregory, 2002). The lack of standardized comparability on geographical units hampers the generation of urban research, particularly at local scales. In fact, due to the difficulties in comparing Census data spatio-temporally researches tend to use Census data either studying a single year with a high degree of spatial detail or studying a long term change in coarse geographical units (Gregory and Ell 2005). The time consuming task of manually modifying Australian geographic boundaries, making them comparable over time, remains costly and potentially introduces data errors. New ABS geographies, set of release in 2011 (ABS, 2006e), are set to again neglect the pivotal issue of comparability over time. The task of comparing different censuses will not be possible at important spatial scales until a national standard approach is taken.

Acknowledgements

This research was supported by an International Research Initiatives Grant from the University of Western Sydney, Sydney, Australia. The authors would like to thank Dr. Mark Davidson for his additions to the paper as well as the anonymous referees for their valuable comments.

References

Australian Standard Geographical Classification (ASGC). 2008. Available at: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/D4B482AE72A3CCF9CA2574CF00136AB8?opendocument (accessed on 11th November 2008)
Australian Bureau of Statistics (ABS) (2006a). *Population Measures*, Cat No. 2914.0, Available at:

http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/2914.0Main%20Features255002006?opendocument&tabname=Summary&prodno=2914.0&issue=2006&num=&view= (accessed on 17th February, 2009)
Australian Bureau of Statistics (ABS) (2006b). *ABS 2006 Census Data Packs – Place of Usual Residence – Basic Community Profile*. Release 2, Cat no.2069.0.30.001, Canberra, Australia.
Australian Bureau of Statistics (ABS) (2006c). *ABS 2006 Census Data Packs – Place of Enumeration Profile Release 2*, Cat no.2069.0.30.004, Canberra, Australia.
Australian Bureau of Statistics (ABS) (2006d). *Review of the Australian Standard Geographical Classification*: Available at: http://www.abs.gov.au/ausstats/abs@.nsf/mf/1216.0.55.001(accessed on 17th February, 2009).
Australian Bureau of Statistics (ABS) (2006e). *Changes to geographic areas between censuses,* Available at: *http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/2914.0Main%20Features215002006?opendocument&tabname=Summary&prodno=2914.0&issue=2006&num=&view*= (accessed on 17th February, 2009)
Australian Bureau of Statistics (ABS) (2008). *Mesh Blocks Digital Boundaries, Australia, 2006* , Cat No. 1209.0.55.002, Canberra, Australia
Blake, M., Bell, M. and Rees P. (2000). Creating a temporally consistent Spatial Framework for the Analysis of Inter-regional migration in Australia. *International Journal of Population Geography*, **6**,155-174
Boyle,P. and Feng, Z. (2002). A method for integrating the 1981 and 1991 British census interaction data. *Computers, Environment and Urban Systems*. **26**, 241-256
Gregory,I.N. (2002). Time-variant GIS databases of changing historical administrative boundaries: A European Comparison. *Transactions in GIS*, **6(2)**:161-178
Gregory,I.N. and Ell,P.S. (2005). Breaking the boundaries: geographical approaches to integrating 200 years of the census. *J.R. Statist. Soc. A*. **168**, part 2, 419-437
Norman,P., Rees,P. and Boyle,p. 2003. Achieving data comparability over space and time:creating consistent geographicall zones. *International Journal of Population Geography*. **9**, 365-386
Manley,D., Flowerdew,R. and Steel,D.(2006). Scales, levels and processes: studying spatial patterns of British census variables. *Computers, Environment and Urban Systems*. **30**, 143-160
Martin,D., Dorling,D. and Mitchell,R. (2002). Linking censuses through time: problems and solutions. *Area* **34.1**, 82-91.
Openshaw, S. (1977). A geographical solution to scale and aggregation problems in region building, partitioning, and spatial modelling. *Transactions of the Institute of British Geographers*.**2**, 459-72
Openshaw, S. and Taylor, P.J. (1984). The modifiable areal unit problem. *Concepts and techniques in modern geography*. Vol 38. Norwich: Geo Books.

Biography

Maria Piquer-Rodriguez is an Environmental Scientist by training with a masters degree in Geographical Information Systems from the University of Wageningen (Holland). She joined the Urban Research Centre in January 2008, before she worked for the University of Almeria, Spain. Her main interests are land-use change detection and the interactions of the urban fringe with its surroundings.

Sumita Ghosh is an architect and planner specialising in sustainable urban forms, environmental performance indicators and tools. She completed her Ph.D. from the University of Auckland, New Zealand. She has applied Geographic Information Systems (GIS) in local scale urban sustainability research projects. She has working experience in teaching, research and professional fields in Australia, New Zealand and India.

Spatial variation in personal exposure of parking attendants in Leeds to carbon monoxide and ultrafine particles

Anil Namdeo, Ana Pareira

Transport Operations Research Group, School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, NE1 7RU
Tel. (+44 191 2228486) Fax (+44 191 2226502)
Email anil.namdeo@newcastle.ac.uk

KEYWORDS: parking attendants, personal exposure, ultrafine particles, carbon monoxide, Leeds

1. Introduction

Recent studies show that thousands of premature deaths every year in Europe may be associated with traffic air pollution (WHO, 2000). Emissions from vehicles affect the quality of the air we breathe and have the most impact close to where traffic is greatest, e.g., in the centres of towns and cities and near busy roads (DFT, 2004). The real exposure of the urban population to many airborne pollutants could be different from (and often greater than) that measured by fixed monitoring stations, and it is the outdoor urban workers such as traffic police, parking wardens, street sweepers or newspaper vendor that are likely to undergo higher than average exposure to traffic pollutants (Violante *et al.*, 2006; Romieu, 1999). Studies have shown that the exposure of individuals to an air pollutant is dependent on their daily activity and the location and time allocated to these activities (Vellopolou and Ashmore, 1998).

2. Leeds Study

A study was carried out investigate the temporal and spatial variation of personal exposure of Leeds City Council parking attendants to two pollutants related to traffic and with known health effects, carbon monoxide (CO) and ultrafine particulates (UFP). Three areas (Figure 1) were selected according to the parking attendants' usual location of work; two in the Leeds city centre (Park Row and Market Area) and one in a suburban area (Headingley). This choice was based on Gulliver and Briggs (2004) suggestion that a selection of different routes used for personal monitoring provides contrasting traffic conditions and therefore provides a wide range of exposure levels. Park Row area is characterised by several storey buildings on either side, predominantly businesses such as banks and restaurants. Market Area is mainly commercial in nature and is characterised by a busy bus corridor with several storey buildings on both sides. Leeds Market open car park is located at the centre of this area. Headingley area is approximately 3km North West of City Centre. A main radial route and bus-corridor into the city runs to the east of this area. The area has a varied environment, with a residential area in its centre and the presence of the Leeds Met University Campus next to it, generating high traffic throughout the terms.

High resolution real time measurements (10s average) were taken in the three locations during the winter and the summer in 2007. A pilot survey was carried out in November 2006 to test the practicality of the data collection method. The CO and UFP measurements were taken by using the portable monitors carried out by parking attendants during their beats. Portable GPS devices were carried by the parking attendants to locate their position and later match with corresponding CO and UFP measurements. The first set of CO measurements carried out by the parking attendants was undertaken for three days including a Saturday during winter in February 2007. Summer surveys were carried out in July 2007 also for three days including a Saturday. Data was collected from the start of the parking attendants work shift between 0700h and 0800h and ended at the end between 1600h and

1800. UFP data collection was also carried out during winter in February 2007 and during summer in July and August 2007. For the accuracy of the measurements, zero and span tests were carried out before each set of surveys with all CO and UFP monitors used for the study.

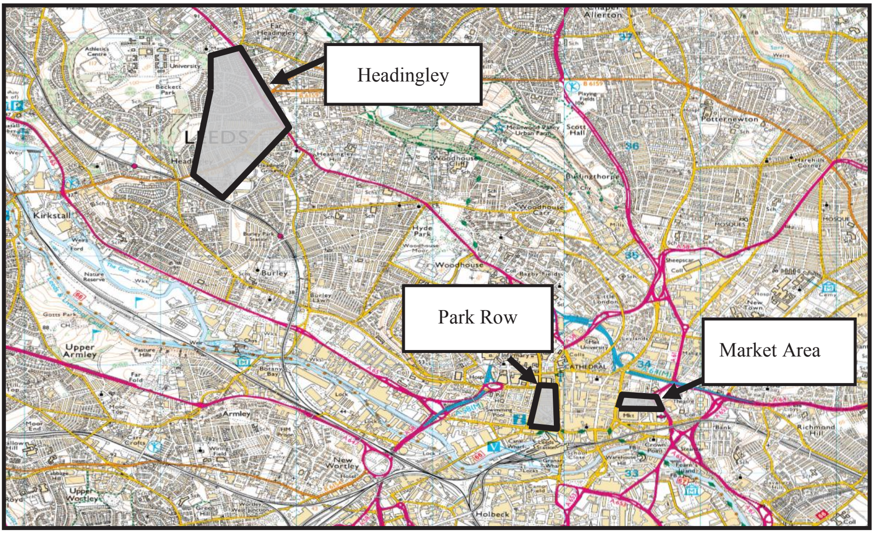


Figure 1: Study Area Location Plan

2. CO Exposure Levels

A total of 24 usable sample sets were available for analysis. These consisted of 17 CO sets (eight in winter and nine in summer) and seven UFP sets (three in winter and four summer). Average CO exposure levels in three areas are shown in Table 1. The CO daily mean concentrations were mostly below the UK National Air Quality Objective of 8ppm. However there is a wide variation in exposure levels between the seasons and weekdays. Summer levels are higher than winter levels in Market Area and Headingley whereas this variation is small in Park Row. As expected weekday levels are higher than weekend levels indicating the influence of traffic activity. Figures 2-4 show spatial variation in personal exposure of parking attendants to CO. These figures clearly show that there is a large variation in exposure levels depending on the location of parking attendants with respect to traffic activity. If the attendants were near busy roads (e.g. Otley Road in Headingley) or in street canyons (Park Row) then they are exposed to higher CO levels compared to the locations which are not near busy traffic (e.g. side streets). Market Area presents an interesting story; here parking attendants' beats are near shops and in an open car park area where due to frequent stopping, stopping and idling of cars in search of parking spaces has created some hot-spots as shown in Figure 4.

Table 1: CO exposure levels in three areas

Sampling Area and Day		CO (ppm)					
		Summer 2007			Winter 2007		
		Mean	Maximum	Standard Deviation	Mean	Maximum	Standard Deviation
Park Row	Weekday	1.388	5.330	0.820	1.178	4.155	0.378
	Saturday	0.888	3.421	0.433	0.428	4.406	0.482
Market Area	Weekday	1.204	5.860	0.708	0.270	3.008	0.332
	Saturday	0.767	1.902	0.241	-	-	-
Headingley	Weekday	1.782	9.255	1.483	0.380	4.091	0.543
	Saturday	2.196	4.922	0.414	0.314	4.532	0.436

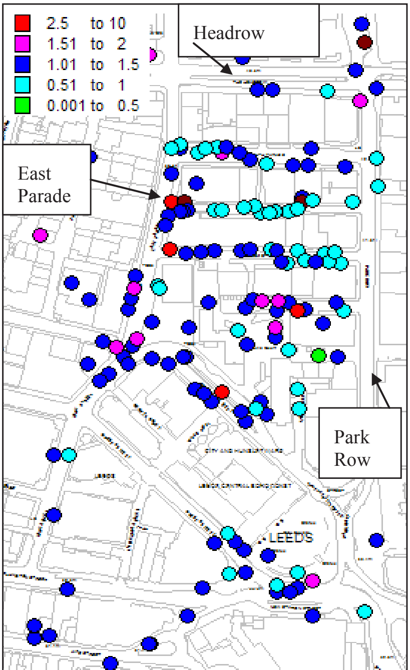


Figure 2. Spatial Variation in CO Exposure Levels in Park Row

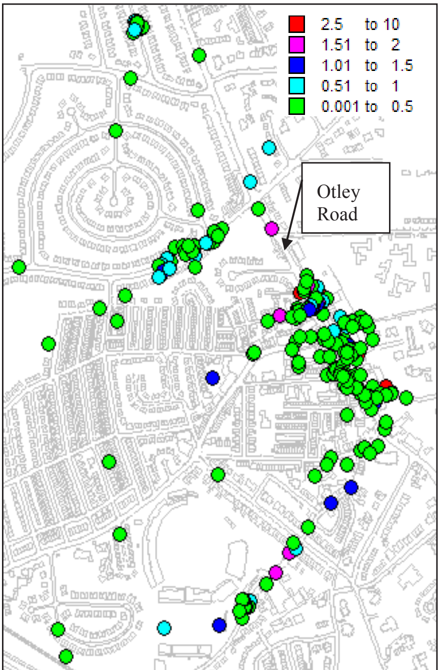


Figure 3. Spatial Variation in CO Exposure Levels in Headingley

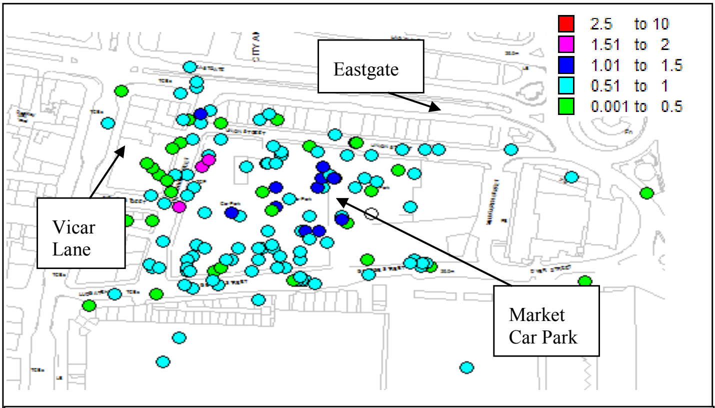


Figure 4. Spatial Variation in CO Exposure Levels in Market Area
(c) Crown Copyright Ordnance Survey Map. An Edina/JISC supplied service.

3. Ultrafine Particles Exposure Levels

UFP are measured as number of particles per cm^3 volume of air. There are no standards set for UFP; though fine and ultrafine particles have been linked with adverse effect on human health especially respiratory diseases (WHO, 2005). Contrasting to CO exposure levels, UFP levels are higher in winter than in summer by a factor of 2 to 3 (Table 2). UFP exposure levels are generally higher during AM peak hours compared to PM peak hours. Parking attendants in Park Row and Market Area are exposed to higher UFP levels than in suburban area of Headingley. Figures 5-7 show spatial variation in personal exposure of parking attendants to UFP.

Table 2: Ultrafine particles exposure levels

Sampling Area and Time		UFP (particles/cm^3)					
		Summer 2007			Winter 2007		
		Mean	Maximum	Standard Deviation	Mean	Maximum	Standard Deviation
Park Row	AM	42203	371084	29724	104413	308414	40993
	PM	32955	192924	20619	82033	361234	48155
Market Area	AM	31431	128484	13276	109884	302632	53884
	PM	39842	126777	18208	67861	226189	27936
Headingley	AM	32127	206011	26748	74615	234871	41450
	PM	31885	221301	23594	26193	201140	25639

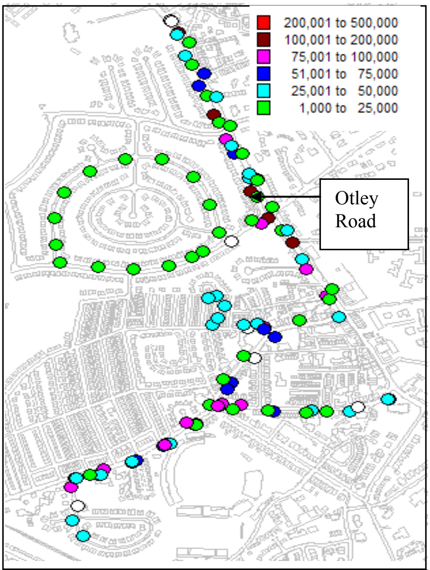


Figure 5. Spatial Variation in UFP Exposure Levels in Headingley

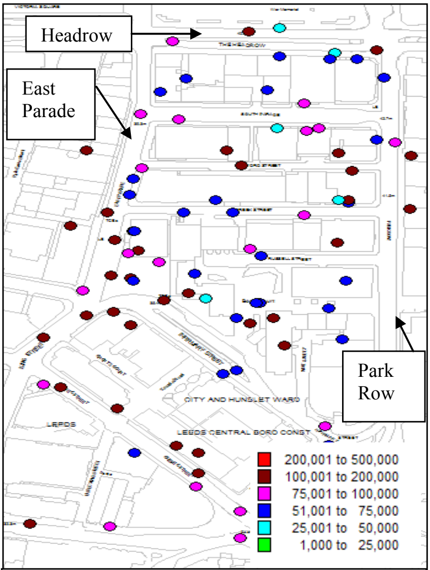


Figure 6. Spatial Variation in UFP Exposure Levels in Park Row

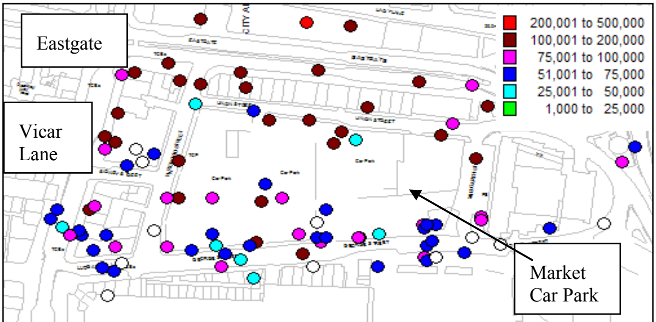


Figure 7. Spatial Variation in UFP Exposure Levels in Market Area.

UFP concentrations were highest in Park Row followed by Market Area and Headingley. The effect of traffic activity is clearly evident from Figure 5 where parking attendants are exposed to an order of magnitude higher UFP levels near busy roads compared to the residential area.

4. Conclusions

This study concludes that parking attendants are exposed to varying levels of pollutants in their beat. The variation in CO and UFP concentrations is directly related to the traffic activity and the topography and geometry of the area their work routes go through. The results presented here are preliminary and further analysis is planned.

References

Gulliver J and Briggs DJ (2004) Personal exposure to particulate air pollution in transport microenvironments *Atmospheric Environment* **38 (1)** 1-8

Romieu I (1999). Epidemiological study of health effects arising from motor vehicle air pollution. In Schwela D and Zali O (1999) *Urban Traffic Pollution* E & F Spon (London)

Vellopoulou AV and Ashmore MR (1998) Personal exposures to carbon monoxide in the city of Athens: I. Commuters' exposures. *Environment International* **24 (7)** pp 713-720

Violante FS, Barbieri A, Curti S, Sanguinetti G, Graziosi F and Mattioli S (2006) Urban atmospheric pollution: Personal Exposure versus fixed monitoring station measurements. *Chemosphere* **64**, pp 1722 - 1729

WHO (2000). *Air quality guidelines for Europe*. WHO Regional Publications, European Series No. 91. WHO (Copenhagen, Denmark)

WHO (2005). *Particulate Matter air pollution: how it harms health*. Fact sheet EURO/04/05, WHO (Berlin, Copenhagen, Rome)

Range Queries over Trajectory Data with Recursive Lists of Clusters: a case study with Hurricanes data

Fernanda Barbosa, Armanda Rodrigues

CITI / Departamento de Informática
Faculdade de Ciências e Tecnologia da UNL
2829-516 Caparica, Portugal
(fb,arodrigues)@di.fct.unl.pt

KEYWORDS: Information Retrieval, Geographic Information Systems, Trajectory Data, Range Query, Metric Data Structures

1. Introduction

With the rapid increase in the use of location-acquisition technologies (GPS, GSM networks, etc.), large amounts of spatio-temporal datasets will soon be accumulated. In different application domains, we need to represent moving entities, i.e. to collect the successive location positions of a given entity, which form the trajectory for that entity. For example, in the analysis of geographic events (migrating birds, hurricanes or people on the move) in Geographic Information Systems (GIS), information associated with a geographic event is described as a collection of positions, stored over time.

The key idea is that moving entities are described by a sequence of positions in k-dimensional space. Each position in the sequence represents the entity's location at a given time. Thus, a trajectory for a moving entity in a k-dimensional space is viewed as a line in a k + 1-dimensional space, where time is an additional dimension.

In all these applications, there is a need to identify similarities on the temporal evolution of geographically positioned entities. For example, find all animals whose migration flow is similar to the swallow's migration or to find all hurricanes that have a route similar to that of Katrina hurricane. So, we need to compare the evolution of movements and decide if they are, to a degree and according to some measure, close (similar).

In most cases, the database in question are very large, so if the similar search leads to an exhaustive search in the database, the response time will be very long and the search will became ineffective, especially because data objects stored in modern databases involve high dimensionality. For this reason, it is necessary to introduce new techniques which can deal with this problem effectively.

In recent years, the problem of similar searching in spatial-temporal data received the attention of several researchers. Most of the actual work in objects trajectories is focused on three aspects: data representation of trajectories, distance function to measure similarity between any two trajectories and an efficient similar searching technique.

In order to deal with the high dimensionality of data objects, dimensionality reduction techniques are proposed. These techniques enable the use of a suitable indexing structure on the reduced objects, in order to provide an efficient similar search. The Symbolic Aggregate approximation (SAX) technique [1], which is the first symbolic representation for time series, enables the use of existing text-retrieval techniques and algorithms developed for this context.

The choice of a distance function which best represents the degree of similarity for trajectories depends on the application domain in question. Some of the functions used for moving objects are the Euclidean Distance (ED), the Manhattan Distance (MD), the Dynamic Time Warping (DTW), the Longest Common Subsequences (LCSS), the Edit Distance on Real sequence (EDR), the Edit distance with Real Penalty (ERP) [2] and the Extended Edit Distance (EED) [3]. Some of these distances are metric (ED, ERP and EED). When this distance is metric, the set of elements defines a metric space.

In order to provide efficient similar searching, several metric data structures have been proposed [4, 5]. In these data structures, the database is divided into partitions which are based on distances calculated between a set of selected objects and the remaining objects. Space partitions seek to minimize the need for exhaustive search. This means that, at search time, some subsets are discarded and others are exhaustively searched. The distance-based indexing method may be pivot based or cluster based [4]. Some of the data structures using the pivot-based method are the Vp-Tree [6] and the Mvp-Tree [7]. There are variants of the pivot-based method, used in LAESA [8]. Some of the data structures using the cluster-based method are Gh-Tree [4], GNAT [9], Sa-Tree [10], KNN graph [4], LC [11, 12] and RLC [13, 14].

Our main goal is to evaluate the use and the efficiency of range queries in metric data structures in different application domains, such as geographic events, image and music processing. In this work, we address the problem of querying a hurricanes database using the RLC data structure with the Edit distance with Real Penalty (ERP). Work with similar progress is under way using the same hurricanes database, in which the data dimension was reduced using the SAX technique and the metric distance is the Extended Edit Distance (EED).

Similar evaluations were realized in natural language dictionaries (English, French, German, Italian, Portuguese and Spanish) using the Levenshtein distance [14] and the Extended Edit Distance [15].

The rest of the paper is structured as follows. In Section 2, we recall some basic notions on range queries in metric spaces. Section 3 summarizes the essentials of the RLC data structure, by presenting its definition and briefly describing how some algorithms work. Section 4 is devoted to the characterization of the metric space over trajectories data, while Section 5 reports on experimental results. Conclusions and future work are drawn in Section 6.

2. Range Queries in Metric Spaces

A metric space is a pair (U,d), where U is a set of objects, called the universe, and d: $U \times U \Rightarrow \Re^{+}_{o}$ is a function, called distance, that satisfies the three following properties:

1. Strict positiveness: $d(x,y) \geq 0$ and $d(x,y) = 0 \Leftrightarrow x = y$;
2. Symmetry : $d(x,y) = d(y,x)$
3. Triangle inequality: $d(x,y) \leq d(x,z) + d(z,y)$.

A database over a metric space (U,d) is a finite set $B \subseteq U$. Examples of metric spaces include the sets $\Re^n$ (for any n>1) with the Euclidean distance or the set of all strings with the Levenshtein distance [14].

The problem raised by range queries is to yield the set of all database objects whose distance to a given object does not exceed a certain amount. Formally, given a database B over a metric space (U,d), a query point $q \in U$, and a query radius $r \in \Re^{+}$, the answer to the range query (q,r) is the set $\{x \in B \mid d(x,q) \leq r\}$.

Metric data structures seek to minimize the number of distance computations performed in range queries.

During the computation of a range query (q,r) in a database over a metric space (U,d), triangle inequality and symmetry are used to discard elements of the database without computing the associated distance to the query object. Given a query element q and a query radius r, an element x may be left out without the need for evaluating d(q,x). This will arise if: (1) there is a pivot p where $|d(q,p) - d(x,p)| > r$, in the pivot-based method, or (2) x belongs to the cluster with center c and $|d(q,c) - d(c,x)| > r$, in the cluster-based method. In these cases, it is not necessary to compute d(q,x) since we know that $d(q,x) > r$, based on the triangle inequality.

3. Recursive Lists of Clusters

A Recursive List of Clusters (RLC) is a dynamic metric data structure, which was built on the List of Clusters (LC) data structure [12].

3.1. Definition of RLC

The definition of RLC relies on the notions of cluster and list of clusters. Let B be a database over a metric space (U, d).

A cluster of B is a triple (c, r, I), where:

- The centre c of the cluster: $c \in B$;
- The radius r of the cluster: $r \in R^{+}_{o}$;
- The interior I of the cluster is a subset of the elements of the database, excluding c, whose distance to c does not exceed r: $I \subseteq \{x \in B \mid 0 < d(x, c) \leq r\}$.

We denote by $O_{c,r,I}$ the set $I \cup \{c\}$ of all objects that occur in a cluster (c, r, I).

The main property of a list of clusters, besides being a sequence of clusters with the elements of the database, is that any object belongs to the first cluster that may contain it. A list of clusters L for B is a sequence of clusters in B, $L = < (c_1, r_1, I_1), \ldots , (c_n, r_n, I_n) >$, for some $n \geq 0$, that satisfies the three following properties:

- The set of objects in L is B: $B = \cup_{i=1}^{n} O_{ci,ri,Ii}$.
- Clusters are pairwise disjoint, that is, for every $i, j = 1, 2, \ldots, n$: $i \neq j \Rightarrow O_{ci,ri,Ii} \cap O_{cj,rj,Ij} = \varnothing$.
- Every element occurs in the first cluster in which it fits, i.e., for every $i = 2, \ldots, n$, every $x \in O_{ci,ri,Ii}$, and every $j = 1, \ldots, i-1$: $d(x, c_j) > r_j$.

In the RLC data structure, every cluster has the same radius. As a consequence, it is unlikely that all clusters' sizes are similar. To accommodate the cases in which clusters have just two or three elements and those in which they have thousands of objects, clusters may be implemented in two different ways.

Let ρ be a positive real number and α be a positive integer. A recursive list of clusters for B with radius ρ and array capacity α is a list of clusters for B of fixed radius ρ, $L = < (c_1, \rho, I_1), \ldots , (c_n, \rho, I_n) >$, for some $n \geq 0$, that satisfies the following property, for every $j = 1, \ldots, n$:

- If the size of the interior I_j does not exceed α, I_j is implemented with an array;
- Otherwise, I_j is a recursive list of clusters (with radius ρ and array capacity α).

Remark that an object may belong to several clusters, hierarchically organised. For the sake of the efficiency of the range query algorithm (to be explained later), every object x is associated with a sequence, which includes the distances from x to the centres of the clusters to which x belongs.

3.2. Construction and Range Query Algorithms on RLC

Now, let us summarize the main steps of the RLC construction and range query algorithms. For further details, the reader is referred to [13].

The data structure construction is done by a succession of insertion operations. To perform the insertion of an object x in a RLC, the list is iterated until a cluster C = (c, ρ, I) where x fits is found, i.e. $d(x,c) \leq \rho$. If no such cluster exists, a new cluster (with centre x and empty interior) is created at the tail of the list. Otherwise, x will be inserted in I. Two distinct situations can happen:

- When I is an array, either there is space for an additional element, and the insertion into the array is performed, or the array is full. In this case, I becomes a new list of clusters where x and every object in the array are inserted.
- When I is a list of clusters, the insertion of x in I is (recursively) performed.

In the range search algorithm, the list is iterated and the relationship between each cluster C = (c, ρ, I) and the query (q, r) is established, based on the distance d(q, c) from the query point to the cluster's centre, and on the two radii (ρ and r). Obviously, c belongs to the answer set iff $d(q, c) \leq r$. Four distinct situations can happen:

- If the query ball is inside the cluster, the search continues in the cluster's interior and the iteration stops.
- If the query ball contains the cluster, all objects in the cluster's interior are automatically collected into the answer set and the iteration continues.
- If the query ball intersects the cluster (without containment in either direction), the search proceeds in the cluster's interior and the iteration continues.
- If the query ball is outside the cluster, the cluster's interior is ignored and the iteration continues.

Apart from collecting or leaving out entire clusters' interiors, without computing the distances of their objects to the query point, objects are also immediately added into the answer set or dropped when the search algorithm reaches an array. To understand how this can be done, let x denote an arbitrary object and $m=d(q,c)-r$. It can be proved that:

- If $d(x, c) \leq m$ then $d(x, q) > r$ (i.e., x does not belong to the answer set); and
- If $d(x, c) \leq -m$ then $d(x, q) \leq r$ (i.e., x belongs to the answer set).

This is why every object x in the database is associated with the sequence of distances from x to the centres of the clusters to which x belongs.

4. Hurricanes Metric Space

In this section, we present the hurricanes trajectories database, which were used in our experiments. All of them came from an Atlantic hurricanes database available in http://weather.unisys.com/hurricane/atlantic/. The objects are hurricane trajectories and the metric function used is Edit distance with Real Penalty (ERP) [2]. Each hurricane is described by name, year, identification (id) number and temporal series, which represents the trajectory data associated. In our database, a hurricane is identified by the year and the id number. The trajectory data is defined for one specific moment (day, month and hour) with the associated position originally in latitude and longitude. These coordinates were converted to the x-y plane. So, our trajectory data is a time series $S = \langle(t_1,x_1,y_1), \ldots, (t_n,x_n,y_n)\rangle$. The hurricanes database is composed of 1,331 trajectories. From the original database, we eliminated duplicate hurricanes and those hurricanes with only one observation in the data set.

In our experiment, the similarity between two hurricanes is based on the similarity between the associated trajectories data, which is computed through Edit distance with Real Penalty (ERP). Let S = <(t_{1s},x_{1s},y_{1s}), …, (t_{ns},x_{ns},y_{ns})> and T = <(t_{1t},x_{1t},y_{1t}), …, (t_{mt},x_{mt},y_{mt})> be trajectories data, let p = (x_p,y_p) and q = (x_q,y_q) points in x-y plane and let d be the distance between two points p and q, denoted by d(p,q), defined in Equation 1.

$$d(p,q) = \sqrt{} (x_p-x_q)^2 + (y_p-y_q)^2 \quad (1)$$

The ERP between S and T, denoted by ERP(S,T), is defined in Equation 2. This is a metric function, as demonstrated in [2].

$$\begin{aligned}
&ERP(S,T) = \Sigma_{1..m}\, d((x_{it},y_{it}),(0,0)), \ \text{if } n = 0;\\
&ERP(S,T) = \Sigma_{1..n}\, d((x_{is},y_{is}),(0,0)), \ \text{if } m = 0; \qquad (2)\\
&ERP(S,T) = \min \{ERP(<(t_{2s},x_{2s},y_{2s}), \ldots, (t_{ns},x_{ns},y_{ns})>, <(t_{2t},x_{2t},y_{2t}), \ldots, (t_{mt},x_{mt},y_{mt})>)\\
&\quad + d((x_{1s},y_{1s}), (x_{1t},y_{1t})), ERP(<(t_{1s},x_{1s},y_{1s}), \ldots, (t_{ns},x_{ns},y_{ns})>, <(t_{2t},x_{2t},y_{2t}), \ldots, (t_{mt},x_{mt},y_{mt})>)\\
&\quad + d((0,0), (x_{1t},y_{1t})), ERP(<(t_{2s},x_{2s},y_{2s}), \ldots, (t_{ns},x_{ns},y_{ns})>, <(t_{1t},x_{1t},y_{1t}), \ldots, (t_{mt},x_{mt},y_{mt})>)\\
&\quad + d((x_{1s},y_{st}),(0,0))\}, \text{ otherwise}
\end{aligned}$$

5. Experimental Results

The goal of this section is to understand how RLC behaves with hurricanes databases. For each database, four files were generated. The smallest has 250 hurricanes randomly chosen from the database, and was used as the set of query hurricanes. The other three are random permutations of the hurricanes databases. The justification for making use of three equal sets (the databases) lies in the fact that the final shape of the data structure depends on the order in which the objects occur in the input of the construction algorithm.

For each database, we submitted the set of 250 query trajectories with query radii 1, 2, 3 and 4. The average number of objects (and the corresponding percentage of the database size) retrieved by each query is indicated in Table 1.

Table 1 - Average number of objects retrieved and corresponding percentage of the database.

Query Radius	Number	Percent
1	5	0,38%
2	16	1,20%
3	28	2,10%
4	41	3,08%

The results obtained by the 250 queries in the hurricanes database are presented in Figure 1. The presented results are the mean of the results achieved by submitting the set of queries to the three databases, with query radii 1, 2, 3 and 4.

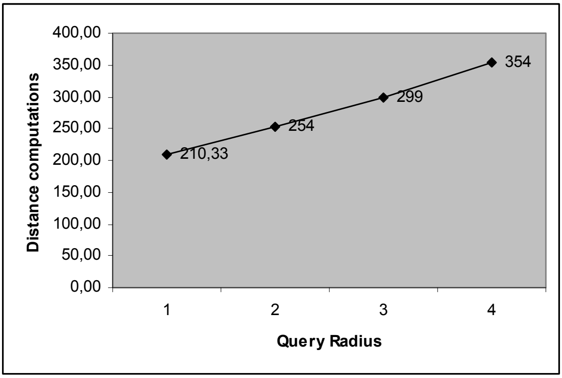


Figure 1 - Average number of distances computed to query the hurricanes database.

In an exhaustive search, for each query we need to compare the query trajectory with all the trajectories in the database. So, for each query we need to compute 1,331 distances. The average number of distance computations (and the corresponding percentage of the database size) retrieved by 250 queries is indicated in Table 2. The average number of the distance computations for each query is small, the percentages vary between 15,8% and 26,6%.

Table 2 – Average number of distance computations performed and corresponding percentage of the database.

Query Radius	Distance Computations	Percent
1	210,33	15,80%
2	254	19,08%
3	299	22,46%
4	354	26,60%

These experimental results confirm that the RLC data structure minimizes the number of distance computations performed in range queries. During the computation of one range query in the hurricanes database, many of the hurricanes are discarded, without computing the associated distance to the hurricane query.

6. Conclusions and Future Work

The ubiquity of real world metric spaces with distinct features led us to evaluate the performance of range queries with the RLC data structure, in different application domains. The experiment in [14, 15], evaluates range queries with the RLC over natural language dictionaries.

In this work, we proposed a technique for speeding up range queries in a hurricanes database. With respect to the trajectories data representation and the Edit distance with Real Penalty (our metric space), we need to compare these results with our similar work over the hurricanes database, where the trajectories data are described by symbolic representation, using the SAX dimensionality reduction technique, and the metric distance is the Extended Edit Distance used in [15]. Once we have these results, we will select the metric space, which should leads to the best results in our range queries. Further on, we will need to evaluate the performance of range queries with the RLC data structure, i.e. to compare RLC performance with the performance of other metric data structures.

References

[1] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, “Locally adaptive dimensionality reduction for indexing large time series databases”, *SIGMOD*, ACM, vol. 30 (2), June 2001, pp. 151-162.

[2] L. Chen, *Similarity-based Search Over Time Series and Trajectory Data*. Ph.D. thesis, University of Waterloo, Canada, 2005.

[3] M. Fuad, and P. Marteau, “The Extended Edit Distance Metric”, *Proc. Of Content-Based Multimedia Indexing*, IEEE Xplore, doi 10.1109/CBMI.2008.4564953, June 2008, pp. 242-248.

[4] H. Samet, *Foundations of Multidimensional and Metric Data Structures*. Morgan Kaufmann Publishers, San Francisco, USA, 2006.

[5] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín, “Searching in metric spaces”, *Computing Surveys*, ACM, vol. 33 (3), September 2001, pp 273-321.

[6] P. Yianilos, “Data structures and algorithms for nearest neighbor search in general metric spaces”, *Proc. of the 4th Annual SIAM Symposium on Discrete Algorithms*, ACM, Philadelphia, USA, 1993, pp. 311-321.

[7] T. Bozkaya, and M. Ozsoyoglu, “Distance-based indexing for high-dimensional metric spaces”, *Proc. of SIGMOD International Conference on Management of Data*, ACM, New York, UK, 1997, pp. 357-368.

[8] M. Micó, J. Oncina, and E. Vidal, “A new version of the nearest-neighbour approximating and eliminating search algorithm”, *Pattern Recognition Letters*, vol. 15 (1), 1994, pp 9-17.

[9] S. Brin, “Near neighbor search in large metric spaces”, *Proc. of 21st International Conference on Very Large Data Bases*, San Francisco, USA, 1995, pp. 574-584.

[10] G. Navarro, “Searching in metric spaces by spatial approximation”, *Very Large Data Bases Journal*, Vol. 11 (1), 2002, pp 28-46.

[11] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín, “Searching in metric spaces”, *Computing Surveys*, ACM, Vol. 33 (3), 2001, pp 273-321.

[12] E. Chavez,. and G. Navarro, “A compact space decomposition for effective metric indexing”, *Pattern Recognition Letters*, Vol. 26 (9), 2005, pp 1363-1376.

[13] M. Mamede, “Recursive Lists of Clusters: a dynamic data structure for range queries in metric spaces”, *Proc. of 22th International Symposium on Computer and Information Sciences*, Istanbul, Turkey, 2005, pp. 843-853.

[14] M. Mamede, and F. Barbosa, "Range Queries in Natural Language Dictionaries with Recursive Lists of Clusters", *Proc. of 22th International Symposium on Computer and Information Sciences*. Ankara, Turkey, IEEE Xplore, 2007, doi: 10.1109/ISCIS.2007.4456857.

[15] F. Barbosa, "Similarity-based retrieval in high dimensional data with Recursive Lists of Clusters: a study case with Natural Language Dictionaries", submitted to International Conference on Information management and engineering (ICIME 2009), Kuala Lumpur, Malaysia.

Residential mobility during pregnancy in the north of England

Susan Hodgson[1], Mark Shirley[2], Mary Bythell[3], Judith Rankin[1,3]

[1] Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK
Tel. +44(0) 191 222 3823 Fax +44(0) 191 222 8211 E mail: susan.hodgson@ncl.ac.uk
[2] Institute of Research on the Environment & Sustainability, Newcastle University, Newcastle, UK
[3] Regional Maternity Survey Office, Newcastle upon Tyne, UK

KEYWORDS: Migration bias, mobility, exposure error, exposure misclassification

1. Introduction

Many epidemiological studies assign exposure to an individual's residence at a single time point, such as birth or death. This approach makes no allowance for migration and may result in exposure error, leading to reduced study power and biased risk estimates (Armstrong, 2003). Pregnancy outcomes, with short exposure windows and lag periods, are less susceptible to this bias, however data from North American populations indicate that pregnant women are a highly mobile group (Canfield et al., 2006, Fell et al., 2004, Khoury et al., 1988, Shaw and Malcoe, 1992, Zender et al., 2001). The only study to look at mobility during pregnancy in the UK found that 23.1% of mothers moved house during pregnancy (Dolk, 1997). While these findings are interesting they may not be generalisable to the population in the north of England which, anecdotally, is considered very stable.

We assessed mobility in pregnant women in the north of England using data from the Northern Congenital Abnormality Survey (NorCAS), a prospective, population-based register of congenital anomalies.

2. Methods

Data on pregnancies delivered between 1st Jan. 1985 and 31st Dec. 2003 were extracted from NorCAS, a registry of congenital abnormalities covering the former northern health region (Boyd et al., 2005, Richmond and Atkins, 2005) (figure 1).

Figure 1: Geographic coverage of NorCAS (shaded area).

Eligible cases had a gestational age at delivery of ≥24 weeks (a viable delivery (Louis et al., 2004)) to make the findings more easily generalised to the majority of pregnancies in the region that result in a healthy baby.

The NorCAS data includes address at booking, typically at a gestational age of ~13 weeks, and at delivery, so we were able to assess mobility between these time points. Postcodes at booking and delivery were geocoded using grid references from the ONS All Fields Postcode Directory (Office for National Statistics) assigned to each postcode centroid.

As individual level data on socio-economic status were not collected, the area-level index of multiple deprivation scores 2004 (IMD) (Office of the Deputy Prime Minister, 2004) at the lower layer super output area (SOA) level were assigned to each postcode.

3. Results

Out of 14,885 cases in the NorCAS dataset over the period 1985-2003, 11,559 (77.7%) referred to unique pregnancies with a gestational age at delivery of ≥24 weeks, and were therefore eligible for analysis.

Of the 7,919 (68.5%) eligible women for whom the address at booking and delivery were known, 705 (8.9% (95% CI 8.3 - 9.5)) moved between booking and delivery. Of the women who moved the mean and median moving distance was 9.7 and 1.4km respectively (table 1 and figure 2).

Table 1: Residential migration in the NorCAS dataset 1985-2003.

Residential migration	Frequency	%
No move	7214	91.1
Moved	705	8.9
Of the women who moved:		
Moved within postcode	29	4.1
Moved up to 500m	154	21.8
Moved 500-<1000m	99	14.0
Moved 1-<2km	107	15.2
Moved 2-<5km	129	18.3
Moved 5-<10km	75	10.6
Moved ≥10km	79	11.2
Unable to geocode	33	4.7

The mean maternal age at delivery was significantly lower in women who moved compared to women who did not move (mean age 25.4 versus 27.3 ($p<0.01$) in movers versus non movers respectively). The percentage of women moving house decreased with increasing age group ($p<0.01$).

Women who moved house between booking and delivery had a higher average IMD score (i.e. were more deprived) than women who did not move (mean IMD score at booking 38.3 versus 33.7 ($p<0.01$) in movers and non movers respectively). The percentage of women moving house increased linearly with increasing quintile of IMD score at booking ($p<0.01$).

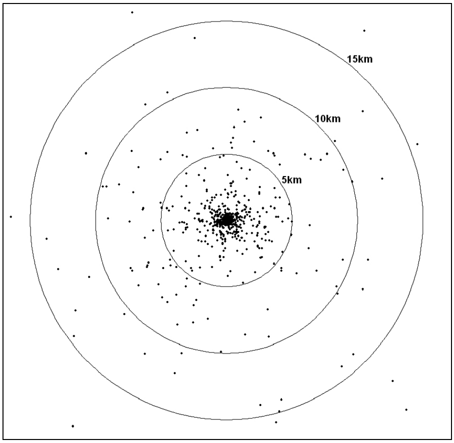


Figure 2: Scatter diagram of residential moves made between booking and delivery by women in the NorCAS dataset (93% of movers represented, the remaining 7% moved further afield). Each point represents one woman making a residential move away from her address at booking (addresses at booking have been set to the centre point of the plot).

4. Discussion

Residential mobility of pregnant women in the north of England is considerably lower than that reported in North America and lower than the previous figure of 23.1% for the UK (Dolk, 1997). Consistent with these studies we found that mobility was higher in younger women and in women living in more deprived areas, and that the majority of moves were over a relatively short distance (Canfield et al., 2006, Dolk, 1997, Fell et al., 2004, Khoury et al., 1988, Shaw and Malcoe, 1992, Zender et al., 2001).

We used prospectively collected data from a long-standing, high quality register of congenital anomalies, however, there are several limitations which may restrict the generalisability of our results. Firstly, the NorCAS data represent a specific subset of pregnancies affected by a congenital anomaly, and as such may not be wholly representative of all pregnancies occurring in the region, most of which result in a healthy infant. Nonetheless, several studies have shown that the patterns of migration during pregnancy are similar in mothers of infants with and without congenital anomalies (Canfield et al., 2006, Shaw and Malcoe, 1992). Secondly, our findings relate to mobility between booking and delivery, which reflects a period of 24 weeks on average, or about 60% of the pregnancy. The percentage of women moving house in this period is likely to be lower than the percentage of moves taking place throughout the whole pregnancy. Dolk et al (1997) found that 19.2% of women moved in the last six months of pregnancy, a similar period to that captured by NorCAS, supporting our view that even if the figure of ~9% is an underestimate, the mobility of pregnant women in the north of England is still relatively low compared to women elsewhere in the UK. Thirdly, almost a third of cases were missing an address at delivery, meaning our assessment could be biased if these missing data are related to migration. We carried out sensitivity analyses to assess the impact of this missing data on our estimate of migration, and conclude that a lower realistic figure of mobility in this dataset is 6.1%, and a higher realistic figure is 15.3%; the best estimate of mobility in this dataset is likely to be somewhere between these two figures, most likely between 8.9 to 10.9%.

These data support anecdotal evidence that this population is comparatively stable, however the mobility we have observed may still introduce misclassification or error into an exposure assessment relying solely on postcode at delivery, and migration should still therefore be considered a potential source of bias in future studies.

5. Acknowledgements
NorCAS is funded by the DoH PPR. JR is funded by a personal award from the DoH NIHR. We are grateful to all the Link Clinicians in the Northern region for their continued collaboration and support of the NorCAS.

6. References

Armstrong B (2003) Exposure Measurement Error: Consequences and Design Issues. In Nieuwenhuijsen MJ (Ed.) *Exposure Assessment in Occupational and Environmental Medicine.* Oxford University Press.

Boyd PA, Armstrong B, Dolk H, et al. (2005) Congenital Anomaly Surveillance in England--Ascertainment Deficiencies in the National System. *BMJ,* 330**,** 27.

Canfield MA, Ramadhani TA, Langlois PH, et al. (2006) Residential Mobility Patterns and Exposure Misclassification in Epidemiologic Studies of Birth Defects. *J Expo Sci Environ Epidemiol,* 16**,** 538-43.

Dolk H (1997) The Influence of Migration in Small Area Studies of Environment and Health - Migration During Pregnancy. *ONS Longitudinal Study Update***,** 6-8.

Fell DB, Dodds L & King WD (2004) Residential Mobility During Pregnancy. *Paediatr Perinat Epidemiol,* 18**,** 408-14.

Khoury MJ, Stewart W, Weinstein A, et al. (1988) Residential Mobility During Pregnancy: Implications for Environmental Teratogenesis. *J Clin Epidemiol,* 41**,** 15-20.

Louis JM, Ehrenberg HM, Collin MF, et al. (2004) Perinatal Intervention and Neonatal Outcomes near the Limit of Viability. *Am J Obstet Gynecol,* 191**,** 1398-402.

Office for National Statistics (2008) Postcode Directories [Computer File]. ESRC Census Programme, Census Dissemination Unit, Census Geography Data Unit (UKBORDERS).

Office of the Deputy Prime Minister (2004) The English Indices of Deprivation 2004 (Revised). ODPM Publications.

Richmond S & Atkins J (2005) A Population-Based Study of the Prenatal Diagnosis of Congenital Malformation over 16 Years. *Br J Obstet Gynaecol,* 112**,** 1349-57.

Shaw GM & Malcoe LH (1992) Residential Mobility During Pregnancy for Mothers of Infants with or without Congenital Cardiac Anomalies: A Reprint. *Arch Environ Health,* 47**,** 236-8.

Zender R, Bachand AM & Reif JS (2001) Exposure to Tap Water During Pregnancy. *J Expo Anal Environ Epidemiol,* 11**,** 224-30.

Biography
I am a lecturer in environmental epidemiology. My research focus is exposure assessment, spatial epidemiology and routine health data analyses. I have applied these techniques to explore health impacts of exposure to heavy metals, organic pollutants and air pollution, and have undertaken routine health data investigations around numerous pollution sources.

vizLib: Developing Capacity for Exploratory Data Analysis in Local Government – Visualization of Library Customer Behaviour

Robert Radburn, Jason Dykes, Jo Wood

giCentre, School of Informatics, City University London, EC1V 0HB
Telephone: +44 (0)20 7040 0212, Fax: +44 (0)20 7040 8584
E : {sbbd476 | jad7 | jwo} @soi.city.ac.uk, W : http://www.soi.city.ac.uk/~ {sbbd476 | jad7 | jwo}

KEYWORDS : libraries, local authority, exploratory data analysis, visualization Processing

1. Introduction

A growing public expectation for better local services and a new local government performance framework (The Audit Commission, 2008) means that local authorities are increasingly needing to use their data holdings to understand citizens. 'Strong and Prosperous Communities' (DLCG, 2006) requires all local authorities to demonstrate that they understand their communities by deploying their local administrative data. The systematic analysis of large structured spatio-temporal data sets is difficult, but Visualization and Visual Analytics (Thomas and Cook 2005) may provide solutions - the NVAC centres are supporting the visual exploration of geographic data in areas such as security and health (e.g. Bhowmick *et al*., 2008).

2. Context and Rationale

Leicestershire County Council provides services to more than 600,000 people. Its Research and Information Team uses rich innovative maps and graphics to inform evidence-based policy (e.g. Radburn, 2008). To extend this work researchers need access to visual techniques for integrating, synthesising and exploring large structured spatial data sets. Each has unique characteristics, and so generic packaged solutions can be inadequate (Fry, 2007). Alternatively, flexible visualization environments through which informative dynamic interfaces can be rapidly developed to suit the data and task in hand may empower local authority researchers in their analysis and improve the evidence base upon which decisions are made (Lloyd *et al*., 2008).

The skills required to use these technologies are rare in local authorities and data sources are typically unexplored at present. We address this skills gap through the ESRC funded UPTAP programme which supports researchers in developing data analysis skills with the aim of Understanding Population Trends and Processes. An UPTAP User Fellowship has enabled an LCC researcher (Radburn) to work with City University for five months to develop the visualization skills required in an important service delivery area – Libraries. Emerging technologies and developing techniques are being used to exploit previously unexplored data stores and gain knowledge of population trends and processes in Leicestershire

3. Leicestershire Library Records

Libraries must be sensitive to the changing needs, aspirations and requirements of their users - a challenge in the face of budgetary pressures and changes in modern lifestyles. Leicestershire Library Service (LLS) routinely collect data about the 400,000 residents who are registered with their local library in the TALIS database - one of the most detailed data sets that the local authority possesses containing detailed customer behaviour and preferences (including lending records on books, films, and music) at 54 libraries across the county on a weekly basis.

4. Approach

Within the national libraries service readily available socio-economic data sources are used to profile communities (EMMLAC, 2005). But knowing your customer is different from understanding how they engage with the service. Just as supermarkets use data driven marketing techniques (Humby *et al*., 2003) the TALIS database could be used to understand which customers deliver what value and in what proportion through analysis of the spatio-temporal behaviour of library users. LCC has already identified the utility of applying marketing techniques to a small sub-set of the TALIS database (Radburn *et al*., 2007) concluding that:

i. there was no such thing as an 'average' customer
ii. the classic Pareto 80/20 rule held (Novos, 2004)
iii. those who visited the library most recently are more likely to revisit

Analysing the data from all 54 libraries to see whether these characteristics are spatially variable could lead to information overload – hence the desire to use visualization to explore the following kinds of questions:

i. Who are the best customers using the library service? Which areas are they living in?
ii. How do priority groups use the library service?
iii. Can the library database provide an up-to-date snapshot of the changing demographics of the local area?

To address these types of research questions through visualization the fellowship consists of three stages: skills development, the production of bespoke visualization prototypes and the communication of the results to relevant communities.

5. Preliminary Results

Our initial manipulation and visualization of the TALIS database has employed the high level open design tool *Processing* (Fry & Reas, 2007; Fry 2008). *Processing* encourages an exploratory approach to data visualization as code linking graphical methods can be quickly configured and deployed. Feedback is rapid and graphical so data can be visually queried very efficiently (Fry, 2008). For example, the total book lending for Leicestershire's 15,000 postcodes can be quickly visualized with basic interactions to display useful information about the data in a short 'sketch' consisting of a few lines of *Processing* code (Figure 1).

The *Processing* environment is designed for small visualization projects consisting of a few pages of high level code that are rapidly modified as the process of visual enquiry progresses (Fry, 2008). Such sketches can be rapidly amended to show various characteristics of the data (Figure 2) and combine data from diverse sources (Figure 3). Useful snippets of code can be retained and incorporated to further simplify the process and ensure efficiency. For example a zoom function was added to the TALIS visualization from an emerging library of *Processing* functionality being developed at the giCentre. The learning curve involved in *Processing* is not a steep one and even these initial interactive graphics reveal complex spatial structure that has not been considered previously by LCC.

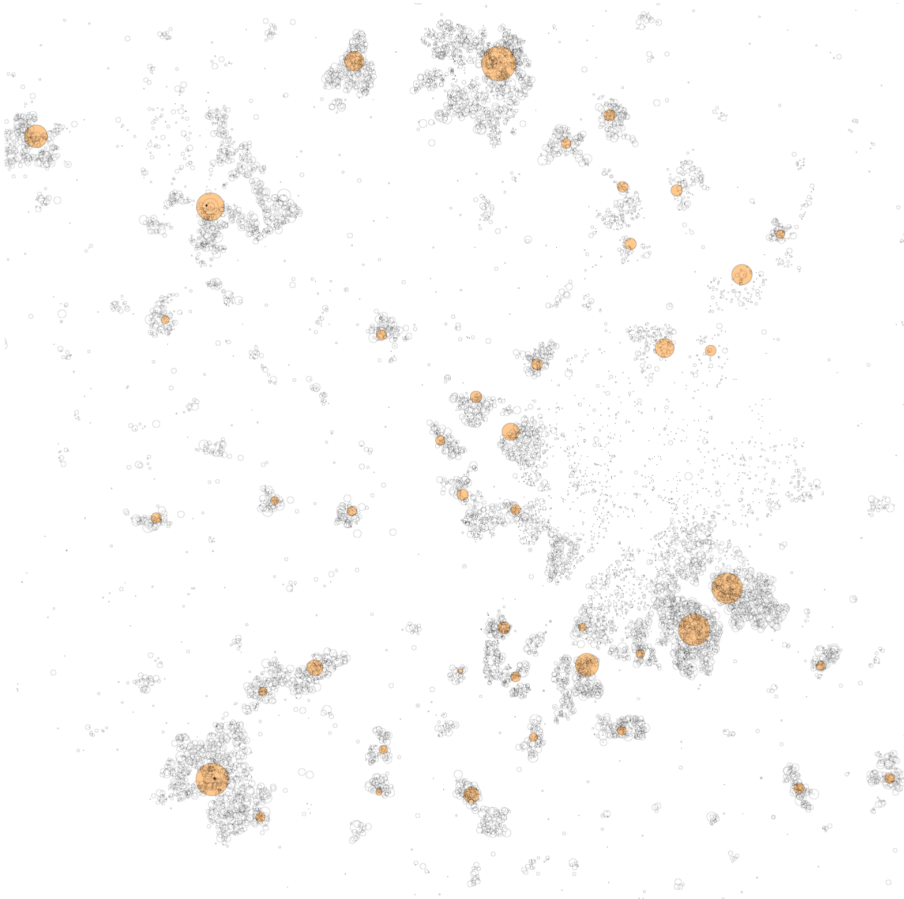


Figure 1. The structure of library usage in part of Leicestershire. Libraries are represented by shaded symbols sized according to annual loans.15,000 unit postcodes are sized according to annual loans in an interactive *Processing* application that zooms, pans and provides instant graphical details on demand.

Figure 2. Detail for part of Leicestershire selected through interactive functionality. Left - Symbol size and density reflects numbers of loans; Right – colour composites show proportions of loans in adult fiction (red), adult non-fiction (green) and other (blue) categories. An item of interest is interactively selected and detail provided through a rapidly responding and configurable graphical interface.

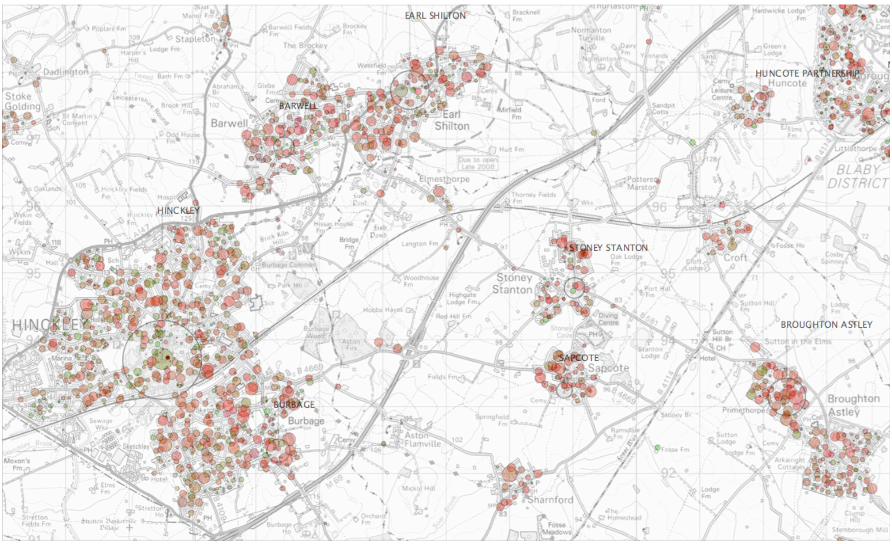


Figure 3. Further detail around Hinckley showing RGB colour composites and an OS LandRanger backdrop. © Crown Copyright / Database right 2008. An Ordnance Survey / EDINA supplied service.

When compared to alternative development environments such as Flash or SVG, *Processing* provides high levels of flexibility in an environment that has the advantages of being based upon (and sometimes drawing directly upon) a fully functional and robust programming language (Java). This means that it is extremely quick and enables us to calculate the positions of symbols and draw them to screen rapidly enabling large datasets to be visualized and animated very effectively. Consequently we can explore TALIS by considering libraries and customer types through combinations of highly interactive implementations of bespoke and novel graphics.

These include: recency / frequency plots (Radburn et al., 2007) that segment behaviour and allow ‘best customers’ to be selected and library profiles to be compared; spider plots showing the spatial relationships between libraries and customer locations; spatial treemaps (Wood and Dykes, 2008) in which space-filling layouts are utilised to generate data-dense graphics. These can be conditioned by time, type of loan and other characteristics in *Processing* in real time. Evidence that morphing is beneficial for transitioning between graphical alternatives (Heer and Robertson, 2007) has led us to use develop novel interactions that vary the visual parameters used in graphics and rapidly morph between combinations of these techniques. Figures 4 and 5 show examples.

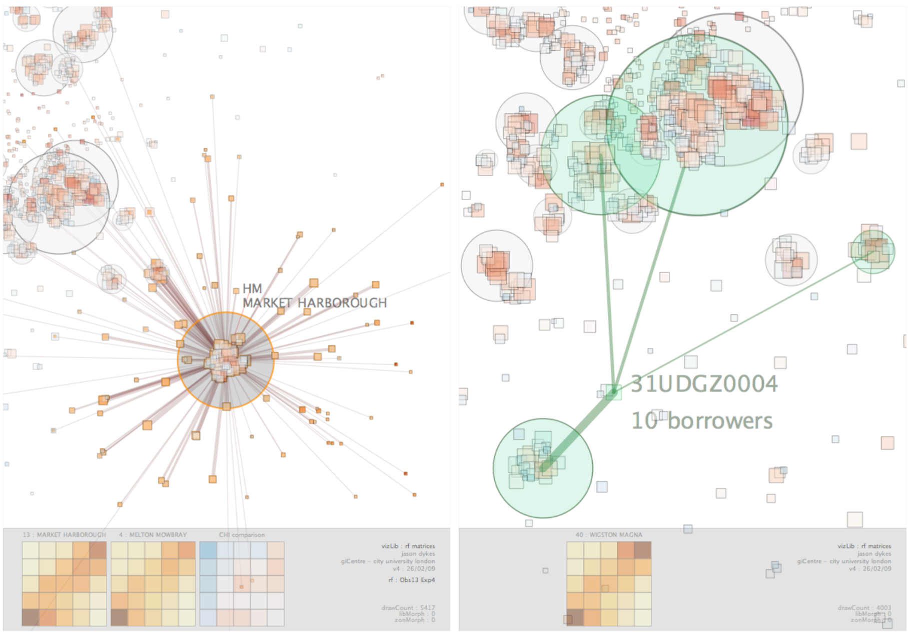


Figure 4. – Interactions showing recency / frequency (RF) and spider plots. The former show numbers of users in recency (columns) / frequency (rows) quintiles, with most recent and frequent users at the top right. The ‘CHI comparison’ allows RF plots for interactively selected libraries to be compared, here revealing that Market Harborough has consistently fewer users in the low recency quintile and more recent / frequent customers. The latter link libraries with users (left) and vice versa (right).

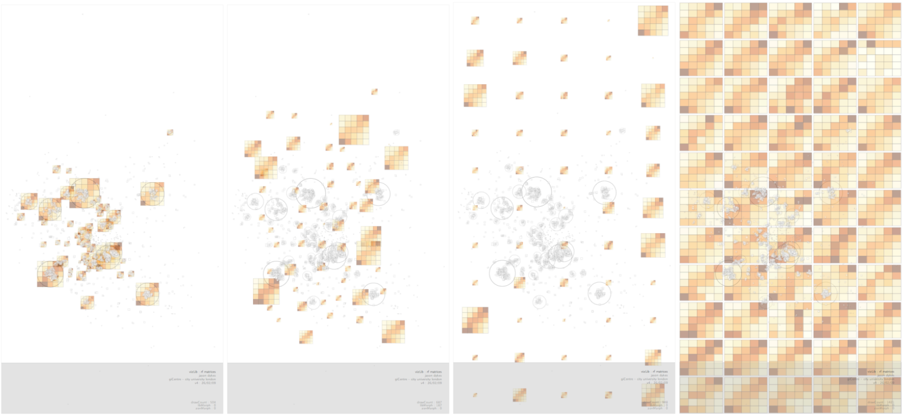


Figure 5. – Fast animated transitions change spatial ordering from geographic coordinates in which RF plots overlap to a spatially ordered treemap in which they are distinguishable but retain some geography in their layout. The transition can be considered as 'ordering' and 'visualization change' in Heer and Robertson's taxonomy.

6. Summary and Conclusion

The government is demanding that local authorities use local data more effectively and base decisions around customer behaviour. For library services a spatial approach to understanding users can help address the three questions that organisations need to answer when entering into a dialogue with customers (Novos, 2007):

- What will you say?
- Who will you say it to?
- When will you say it?

Visualization can help us answer these specific questions but the skill set required to develop and use flexible interactive graphical applications that suit particular combinations of data set and task (Andrienko *et al.*, 2005) is rare in local authorities. The *vizLib* project is embedding geovisualization into an organisation through an investment in people and their skills, rather than the generation of software 'solutions' expected to fit all situations. Our initial work with *Processing* suggests that it is a promising approach for applied geovisualization that makes a number of novel and potentially useful visualization methods possible. Interesting trends have been noted and insights gained from our initial work as we compare libraries, customers and places and consider their spatial variation.

Acknowledgments:

This work is supported by an ESRC UPTAP User Fellowship RES-163-27-0017 and Leicestershire County Council. Data kindly provided by Leicestershire Library Service.

References:

Andrienko, G., Andrienko, N., Dykes, J., Mountain, D., Noy, P., Gahegan, M., et al. (2005). Creating Instruments for Ideation: Software Approaches to Geovisualization. In J. Dykes, A. MacEachren & M.-J. Kraak (Eds.), Exploring Geovisualization (pp. 103-125). Amsterdam: Elsevier.

Bhowmick, T., Robinson, A. C., Gruver, A., MacEachren, A. M., & Lengerich, E. (2008). Distributed Usability Evaluation of the Pennsylvania Cancer Atlas. International Journal of Health Geographics, 7(6).

Department for Communities and Local Government. (2006). Strong and prosperous communities: the local government white paper. Retrieved 12/12/08, from http://www.official-documents.gov.uk/document/cm69/6939/6939.asp

East Midlands Museums Libraries and Archives Council. (2005). How to… Develop a Community Profile: A guide for museums, libraries and archives. Retrieved 12/12/08, from http://tinyurl.com/EMMLAC2005

Fry, B. (2008). Visualizing Data. Sebastopol, CA: O'Reily, 382 pp.

Fry, B., & Reas, C. (2007). Processing: A Programming Handbook for Visual Designers and Artists. Cambridge, MA: The MIT Press, 736 pp.

Heer, J., and Robertson, G., 2007, Animated transitions in statistical data graphics: IEEE Transactions on Visualization and Computer Graphics, v. 13, p. 1240-1247.

Humby, C., Hunt, T., & Phillips, P. (2003). Scoring Points: How Tesco is winning customer loyalty: Kogan Page, 276 pp.

Novos, J. (2004). Drilling down: Turning Customer data into profits with a spreadsheet: booklocker.com, 196 pp.

Radburn, R. (2008). Understanding the Loughborough Foodscape. Leicester: Leicestershire County Council.

Radburn, R., Thomas, N., Forster, P., & Pye, S. (2007). Beyond the comfort zone. Public Library Journal, 22(3), 20-23.

The Audit Commission. (2008). Comprehensive Area Assessment Joint Inspectorate Proposals for Consultation. Retrieved 12/12/08, from http://www.audit-commission.gov.uk/caa/downloads/CAAConsultation08.pdf

Thomas, J. J., & Cook, K. A. (2005). Illuminating the Path: The Research and Development Agenda for Visual Analytics: Pacific Northwest National Laboratory and Department of Homeland Security.

Wood, J., and Dykes, J., 2008, Spatially Ordered Treemaps: IEEE Transactions on Visualization and Computer Graphics, v. 14, p. 1348-1355.

Biographies:

Robert Radburn is an ESRC UPTAP research fellow at City University London developing capacity for visual exploratory analysis in local government, and a Senior Research Officer at Leicestershire County Council working to support evidence-based policy.

Jason Dykes and Jo Wood are senior lecturers in geographic information at the giCentre, City University London with interests in geovisualization.

Developing a New Approach to Geovisualization for Health Professionals

Jessica Wardlaw[1], Mordechai (Muki) Haklay[1] & Catherine Emma Jones[1]

[1]Department of Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London, WC1E 6BT
Tel. +44 (0)20 7679 4188 Fax + 44 (0)207 9161887
Email. J.wardlaw@ucl.ac.uk

KEYWORDS: Geovisualization, usability, health, Google Maps, social marketing

1. Introduction

Geovisualization draws on cartographic and geographic tradition, integrating the representation and analysis of geospatial information with more recent developments in scientific and information visualization, and exploratory data analysis (EDA). There is a vast range of digital data resources that include geospatial referencing (MacEachren *et al.* 2004). For public health, geovisualization aids understanding of the geographic distribution of diseases (MacEachren *et al.* 2004). These datasets are multivariate, and the complex relationships among variables are often unknown. Combining cartographic methods with those from information visualization and EDA can provide researchers and analysts with a range of tools for visually and statistically exploring these relationships.

For GIS professionals, concepts such as layers, zooming in and out of a map, panning and clicking on objects for information form part of every-day language. For health professionals, who are more familiar with Microsoft Office software, such concepts are challenging and require a steep learning curve, therefore the skills and knowledge required to operate such geovisualization tools are often complex (Jones *et al.* 2009, in press). Furthermore, current tools are presented to the end user with predefined datasets, when users in the healthcare sector often wish to incorporate their own much more detailed local datasets into their analysis. For this reason, amongst others, there is still a lack of users of geovisualization, particularly in the health sector (Higgs *et al.* 2005).

This paper describes an application that has been developed to pilot an alternative approach to geovisualization in health. The application borrows usability conventions from outside GIS, providing a more accessible geovisualization tool for health care professionals. The application was developed at Dr Foster Intelligence[1] for the purposes of identifying areas for social marketing campaigns. Social marketing uses data about people's characteristics and lifestyles to promote healthy behaviour. The target users of the tool are employees in Dr Foster Intelligence who work with Primary Care Trusts (PCTs) to decide where health interventions should be targeted and where health services should be commissioned.

2. Dr Foster Social Marketing Tool

The Dr Foster Social Marketing Tool is a Google Maps mashup, which integrates data from multiple sources onto the map interface. Google Maps is a free web mapping service application, which is behind many online map-based services including the Google Maps website. Since the release of its Application Programming Interface (API) in 2006, the Google Maps interface has been embedded on third-party websites. By June 2007 an estimated 50,000 mashups had been created (Haklay *et al.* 2008).

Google Maps was developed to provide street maps, a route planner, and a business locator for several countries. Most of the mashups contain the equivalent of push pins that have been located on a map, with information attached to the pin. However, the GMap Creator tool, developed by the

[1] Dr Foster Intelligence is a company that specialises in the dissemination of healthcare information

Centre for Advanced Spatial Analysis, inspired more sophisticated mashups such as the London Profiler (www.londonprofiler.org). It creates tile layers from ESRI shapefiles for overlaying onto a Google Maps interface. This was a major factor in the choice of Google Maps for the Social Marketing tool.

2.1 Innovative Functionality of the Social Marketing Tool

Zooming/panning - Desktop GIS lack a standard functionality, with respect to zooming and panning so by using the Google interface (Harrower and Sheesley 2005), it is likely that users will already be familiar with how to navigate around the map. This is because Mashups have made Google Maps very familiar to general internet users. Consequently users of the Social Marketing tool will be more accustomed to Google's zooming and panning functionality than to other GIS equivalents. Moreover, internet users have become so familiar with the Google approach that other web mapping providers such as Multimap are now implementing the same conventions. GeoWise InstantAtlas™[2] applications require the user to click on arrows at the edge of the map window to pan, a paradigm that is significantly different.

Adding datasets to the map - Contrary to standard GIS functionality of adding layers by clicking on check boxes, datasets can be added by dragging them from the list on the left hand side of the window and dropping them onto the map interface. This is aligned with Microsoft Windows tools such as Windows Explorer; general computer users may find this more intuitive, so whilst this represents a departure from conventional GIS functionality it is more inline with the functionality health care professionals are familiar with. The datasets in the system have been organised according to the public health issue for which they inform (e.g. smoking, obesity, diabetes, alcohol). This assists the user to immediately locate appropriate information which can be used strategically to inform social marketing campaigns. This functionally assists in reducing the cognitive load resulting from searching for datasets in a long list and makes the application more useful as the data are grouped appropriately to match the questions the users want to ask of it. To aid the conceptual design of this new functionality we have included segmentation data such as the Output Area Classification (Figure 1, bottom image) and GP Practice locations. This type of data adds to the contextual understanding of where service facilities are located and provides information about the social landscape of the different neighbourhoods where social marketing campaigns may be implemented. By including this type of data the tool can aid local knowledge building.

Multi-scale base mapping - Social marketers require this functionality due to the different scales of health data; additionally different background mappings are more appropriate for different tasks. For example, if the user wishes to identify green space for the promotion of physical activity it might be easier for them to use satellite imagery.

Uploading new datasets - This process is required to, for example, incorporate the results of offline analysis such as clustering of multiple datasets into the tool. New datasets can be added as these become available, and tile layers from other mashups such as the London Profiler or MapTube can also be incorporated. The uploading of new datasets to the tool has been simplified for non-expert web developers by means of a Data tab at the top of the window.

Comparing multiple datasets - This is required for evidence-based decision-making; it is often important to look at more than one dataset in making decisions about where health promotions should be targeted. A slider bar, which can easily be identified, allows the user to view the data and the background mapping simultaneously. Layers can also be reordered, so that GP Practice points will be overlaid above a tile layer for example (Figure 1). By comparison in desktop GIS, this functionality is

[2] GeoWise InstantAtlas™ is a software package that assists users to create their own online atlas, to publish data as tables, charts and maps on the same interface.

hidden within the cartographic options, and the ability to change the opacity of data is often restricted to raster data.

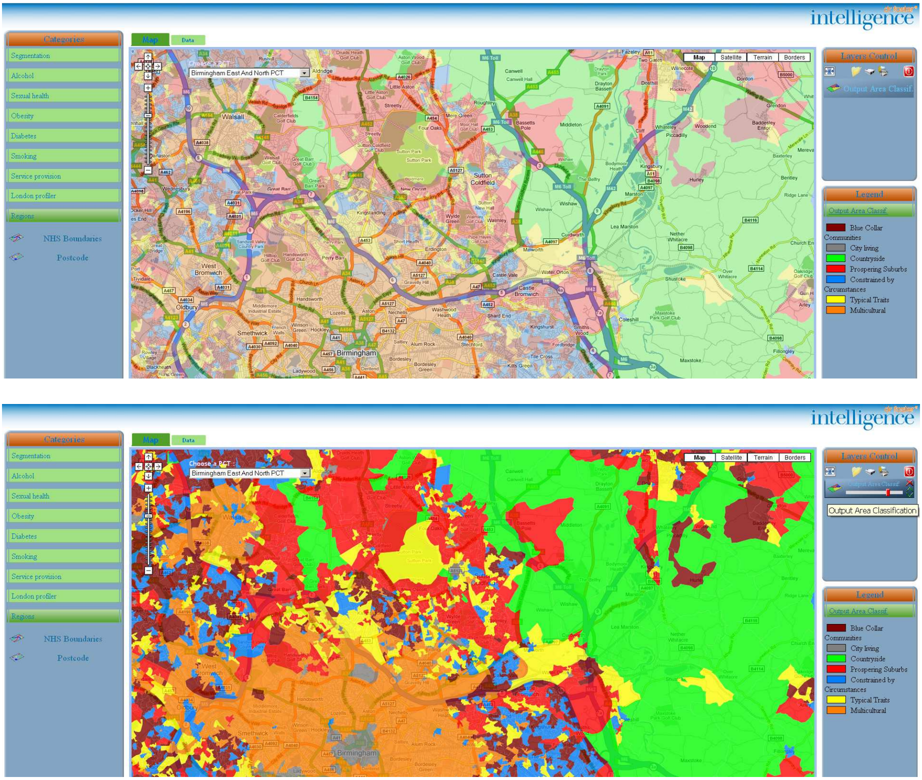


Figure 1. Varying Opacity in the Social Marketing Tool

No plug-ins required - Users in the health sector are often restricted in terms of installing new software on their computers, because of local user access restrictions set by computer administrators, The Google Maps interface does not require any plug-ins, which is a distinct advantage over web mapping tools that require Flash or Scalar Vector Graphics (SVG) plug-ins, and so provides universal usability for health care professionals with access to the internet..

2.2 Standard Health Geovisualization Functionality

Alongside the functions already listed, the Social Marketing tool incorporates many standard health specific functions.

Zoom to PCT - The function to zoom directly to a PCT was implemented because social marketing campaigns are commissioned by PCTs. The first thing that a social marketer would need to be able to do with the tool is to zoom to the PCT for which they are devising a social marketing campaign (Figure 2). This reduces the need to navigate around a map and immediate sets the zoom to cover an extent that is useful to the user.

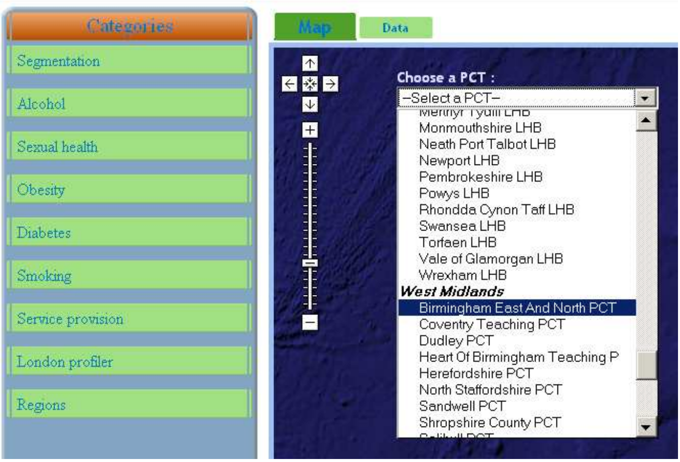


Figure 2. Zoom to PCT functionality in the Social Marketing Tool

GP Practice Charts - The Google Maps interface can be integrated with Google Charting. This was used in the Social Marketing tool to provide charts of GP Practice data on the prevalence of the major public health issues, such as diabetes, coronary heart disease and obesity (Figure 3).

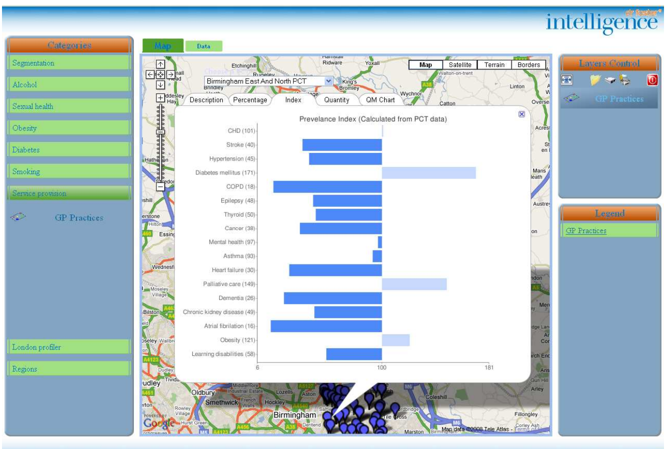


Figure 3. GP Practice data charted in the Dr Foster Social Marketing Tool

3. Discussion and Conclusion

While technological advances in the field of geovisualization are undeniable, many challenges remain. To support real-world knowledge construction and decision making, some of the most important challenges involve dissemination of geovisualization tools (MacEachren *et al.* 2004). There is still a lack of users of such sophisticated geovisualization tools, which are both useful and usable in the "real world". One such area where there is a lack of users is the health sector.

In the Social Marketing Tool that this paper reviews, the Google Maps functionality will be familiar to many users. However, the customised functions may not be as intuitive. The Social Marketing Tool has attempted to move away from conventional GIS functionality and implement more familiar functionality to improve the user experience. For this reason, analytical functionality which is standard in GIS software has not been developed for the Social Marketing tool. The tool has been designed for the presentation and overlay of datasets. Customised functionality in the Social Marketing tool includes adding layers by dragging and dropping the dataset onto the interface and changing the opacity of the layer by using a slider bar. The uploading of new data, despite being greatly simplified, still requires the technical expertise of a web developer. GIS terminology has also been used in some parts of the tool, for example tile layer, marker layer, legend and layer.

The Social Marketing Tool has attempted to provide proof of concept for improving the user experience by moving away from GIS conventions, to bridge the gap between health geovisualization tools and their users. Task analysis and usability tests are now needed to determine whether this goal has been achieved.

4. References

Haklay M, Singleton A and Parker C (2008) Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass* **3 (October 2008)** pp2011–2039

Harrower M and Sheesley B (2005) Designing better map interfaces: A framework for panning and zooming *Transactions in GIS* **9(2)** pp77-89

Higgs G, Smith D and Gould MI (2005) Findings from a survey on GIS use in the UK National Health Service: organisational challenges and opportunities *Health Policy* **72(1)** pp105-117

Jones CE, Haklay ME, Griffiths S and Vaughan L. (2009, in press) A less-is-more approach to geovisualization - enhancing knowledge construction across multidisciplinary teams *International Journal of Geographical Information Systems.*

MacEachren AM, Gahegan M, Pike W, Brewer I, Cai G, Lengerich E and Hardisty F (2004) Geovisualization for Knowledge Construction and Decision Support *IEEE Computer Graphics and Applications* **24(1)** pp13-17

5. Acknowledgements

This work was carried out through a partnership between Dr Foster Intelligence and University College London. This Partnership received financial support from the Knowledge Transfer Partnerships programme (KTP). KTP aims to help businesses to improve their competitiveness and productivity through the better use of knowledge, technology and skills that reside within the UK Knowledge Base. KTP is funded by the Technology Strategy Board along with the other government funding organisations.

The authors would like to acknowledge the hard work of Sylvain Dauriach, at the National School of Geographical Science in Paris, who developed the Social Marketing Tool. Guidance for writing this paper was also given by Claire Ellul from the Department of Civil, Environmental and Geomatic Engineering, University College London.

The authors would also like to acknowledge the Google Maps technology that can be seen in the figures and state that there is no intention of making the website a commercial product; It is for the use of Dr Foster Intelligence employees only. Output Area boundaries are reproduced with the kind permission of the Ordnance Survey.

6. Biography

Jessica Wardlaw is a Research Associate at University College London and is working at Dr Foster Intelligence through a Knowledge Transfer Partnership. At Dr Foster Intelligence, Jessica is introducing GIS mapping techniques into the analysis and visualisation tools that are used to process healthcare statistics. She has a BA in Geography from the University of Oxford and is currently focused on the mapping of public health issues and health inequalities.

Initialising and Terminating Active Contours for Vague Field Crisping

Mark M. Hall, Christopher B. Jones

School of Computer Science, Cardiff University
5 The Parade, Roath, Cardiff CF24 3AA, UK
+44 (0)29 2087 4812, +44 (0)29 2087 4598
{M.M.Hall, C.B.Jones}@cs.cardiff.ac.uk
http://gis.cs.cardiff.ac.uk

KEYWORDS: Spatial language, vagueness, active contours, crisping, algorithms

1. Introduction

Many aspects of spatial language concerned with real-world features and the relationships between them are essentially vague. While this vagueness is managed quite effectively in natural language communication between people, there are currently only very limited facilities for interpreting such language when used to communicate with computers. To overcome this shortcoming a field-based model has been developed to provide a continuous representation of the spatial aspect of this vagueness.

Using fields to represent spatial information is not a new idea. Couclelis (1992) described how people alternate between a field and object views of the world, something a GIS also needs to be able to do. Yamada et al. (1992) describe a field-based system for locating the area described by a spatial expression. Liu et al. (2008) present a general field model and Guo et al. (2008), use fields to model the uncertainty and error in expressions such as "10km N of Kuala Lumpur".

2. Vague Fields for Representing Spatial Information

The field as defined in this work is a matrix of field intensity values that vary between 0 and 1. The semantics of these field intensity values is that they represent how confident the field is that people would describe the cell as part of the area described by the expression. For an expression such as "Photo of a tree near Stackpole", the value of each field cell specifies how "near" it is to "Stackpole".

For some purposes, like ranking results in an IR task, the vague field can be used directly, but for many other methods it is necessary to transform the continuous field into the object representation that is used in most GI methods and systems. Such a transformation is called a crisping and various methods exist for creating such a crisp representation, with one frequently used method being thresholding, where the crisp boundary is drawn so that all field cells within the boundary have a field value higher than the threshold and all cells outside a lower value. The problem of this approach is that it can create very jagged outlines, even for seemingly smooth fields and that the integration of further constraints on the outline can be difficult. To overcome these shortcomings a more flexible crisping method based on active contours has been developed.

3. Active contour based crisping

Active contours are energy minimising splines, that were initially developed for delineating boundaries in medical image analysis Kass et al. (1988), but have also been used in a GIS context by Guilbert and Saux (2008). In most cases two energies are defined that shape the active contour. An internal energy maintains the cohesion of the active contour, and an external energy defined by the domain warps the active contour into the desired shape. Hall and Jones (2008) gives a detailed description of how these energies are iteratively applied to the active contour to create the boundary (see example in Figure 1). This paper expands upon that work and provides approaches on the problems of active contour initialisation and termination.

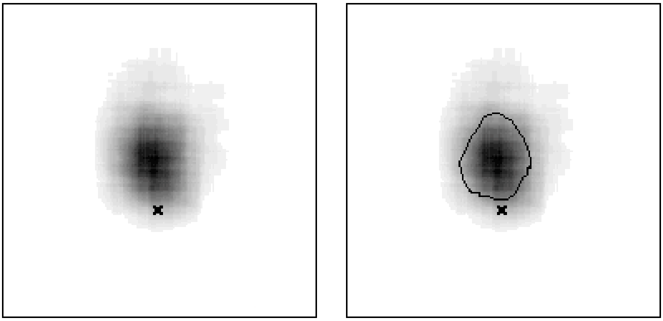


Figure 1. Vague field for the spatial preposition "north of" (left) and a crisping of the field (right, black line). The black x marks the position relative to which the field models"north-of-ness".

3.1. Active Contour Initialisation

Initialising the active contour is an important question, as the initial shape of the active contour directly influences the result quality and processing time. A simplistic approach would be to initialise the active contour as a rectangle around the whole field. While this is guaranteed to produce a consistent result, it is computationally inefficient. A more precise initial shape is required, without sacrificing result quality.

The basic concept for this more optimal initialisation is to use an arbitrary thresholding of the field as the initial shape. More complex spatial expressions such as "Old tree near Stackpole, north of the lily ponds in Pembrokeshire, UK" create multiple fields ("near Stackpole", "north of the lily ponds", "Pembrokeshire", "UK") and to avoid having to threshold the combination of these fields the active contour is initialised using only the field created by the first spatial sub-expression ("near Stackpole"), based on the assumption that the first sub-expression is the focus of the whole expression. The threshold is calculated on this field and then to simplify the initial shape the convex hull of the threshold points is determined, the vertices of the convex hull becoming the control points of the active contour.

3.2. Active Contour Termination

The second problem is how to determine when to stop the active contouring process. The simplest approach is to place a hard limit on the number of iterations that the active contour runs through. This has one major problem and that is that it is impossible to find one number of iterations that always produces good quality results and is computationally efficient. If the number of iterations is too small, then the result might not be correct, as there were still major changes happening in the active contour's shape. On the other hand choosing too high a number leads to wasted iterations, after the active contour has already achieved its final shape. To overcome this a dynamic termination based on the total energy (the sum of energies acting on all control points) in the active contour and the number of modified control points is presented. Clearly evaluating the number of control points that move between iterations is the most precise method, but also computationally very intensive, so the total energy changes are used as filters to determine when to check control point modification.

In order to be able to evaluate the active contour changes, a history of the control point locations and total energies for the previous ten iterations is maintained. On the total energies two thresholds α and β are used to determine when to check control point locations, where α is the maximum total change between the oldest and newest item in the history and β the maximum change between consecutive steps. In the shape comparisons the newest shape is compared to all previous shapes by calculating the number of control points that have moved. The comparison is based on the set of control point coordinates and not control point identity, in order to not falsely detect control points switching location as active contour changes. The active contour terminates if the number of modified control locations is zero or a hard limit of 2000 iterations is reached to guarantee termination.

4. Conclusions

Vagueness in natural language expressions needs to be modelled to enable their use in GIS, and a field-based model has been developed for this task. To integrate this continuous vague field with existing, crisp GIS methods and systems an active contour based crisping algorithm has been developed. This paper focuses on how to initialise and terminate this algorithm.

For the initialisation problem an algorithm was presented that determines the central field in the spatial expression and then uses a combination of thresholding and convex hull calculation to provide an initial shape for the active contour. Termination is based on a two-step approach that combines the total energy acting on the active contour and the movement of the active contour's control points to determine when the final active contour state has been reached. The combination of these two methods improves the efficiency of the crisping algorithm, while maintaining the result quality.

Future work will focus on evaluation. A series of human-subject tests is planned to determine how people rate the crisp shapes. Further spatial prepositions and handling of more complex spatial expressions is also being investigated.

5. Acknowledgements

This material is based upon work supported by the European Community in the TRIPOD (FP6 cr n°045335) project.

References

Couclelis H (1992) People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space* **639/1992** pp65-67

Guilbert E and Saux E (2008) Cartographic generalisation of lines based on a B-spline snake model *International Journal of Geographical Information Science* **22(8)** pp847-870

Guo Q et al. (2008) Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach *International Journal of Geographical Information Science* **22(10)** pp1067-1090

Hall M M, Jones C B (2008) A Field Based Representation for Vague Areas Defined by Spatial Prepositions *Methodologies and Resources for Processing Spatial Language, Workshop at LREC'2008*

Kass M et al. (1988) Snakes: Active Contour Models *International Journal of Computer Vision* **1(4)** pp321-331

Liu Y et al. (2008) Towards a General Field model and its order in GIS *International Journal of Geographical Information Science* **22(6)** pp623-643

Yamada A et al. (1992) Reconstructing Spatial Image from Natural Language Texts *Proc. COLING-92* pp1279-1283

Biography

Mark M. Hall

Mark Hall is a PhD student focusing on spatial computational linguistics. His research interests lie within the areas of geographic information systems, semantics, linguistics, cognitive sciences and ontologies. Currently his focus is on spatial language, how it is used and what people understand when they hear or read spatial expressions.

Christopher B. Jones

Chris Jones has been Professor of Geographical Information Systems (GIS) in the School of Computer Science at Cardiff University since 2000, having held previous academic positions at the University of Glamorgan, and the University of Cambridge. Research interests include geographical knowledge representation and indexing methods for spatially-aware web search engines.

Terrain Identification using mobile handsets

John Tullis, Gary Priestnall, Nick Mount

School of Geography, The University of Nottingham, Nottingham, NG7 2RD, UK
Tel. +44 (0) 115 9515443 Fax. +44 (0) 115 95429
LgxJT2@nottingham.ac.uk

KEYWORDS: Visualisation, Terrain Panoramas, Location Based Services,

1. Introduction

Recent advances in mobile computing together with widespread uptake of positioning technologies such as GPS have spawned numerous developments in the field of location based services (LBS) and mobile GIS. Much research effort into LBS has been devoted to automobile navigation systems and pedestrian wayfinding in urban environments (Raper, 2007). But relatively little attention has been focused on the potential of LBS to aid users in rural or mountainous environments (Wood *et al*., 2006). This paper will report on ongoing efforts to redress that balance by focussing on a task commonly undertaken by outdoor users such as hillwalkers: namely, the identification of hill names on mountain panoramas. The project's overall aim is to optimise the portrayal of terrain information on mobile devices to assist users in identifying key features in mountain environments. This abstract will first justify the rationale for the research, before outlining a novel approach to gathering empirical data on the effectiveness of mobile phone based panorama rendering techniques.

2. Research rationale

The ability to identify named landscape features (hill summits in particular, but also lakes and other points of interest) is useful for many reasons. First, it satisfies hillwalkers' unending curiosity as to which peak is which. Second, it fosters greater knowledge of unfamiliar upland environments. And third, with improved environmental awareness, users may be less prone to navigation errors or disorientation (Langmuir, 2004). Traditionally, walkers seeking to find out the names of visible hills have had three options to help them: topographic map and compass, printed diagrams such as Wainwright's annotated panoramas (Wainwright, 1960), or tourist panoramas mounted on display panels or trig points at prominent viewpoints. Unfortunately, none of these methods is perfect. Feature matching with map and compass is viable only for simple terrain (Figure 1); in areas of more complex terrain, even the most experienced map readers may struggle to reliably identify distant features (Figure 2) (Pick *et al*., 1995). Printed panoramas and tourist display panels are useful in situ, but only exist for a limited number of fixed locations, so are rather inflexible for users on the move.

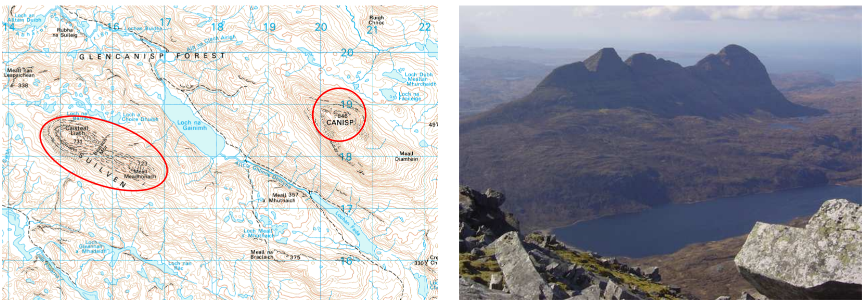


Figure 1. Simple terrain identification: Identifying Suilven from Canisp (NW Scotland) is straightforward, even for novice map readers.

Figure 2. Reliably identifying hill names in regions of complex terrain can be difficult to impossible using map and compass alone, even for experts.

3. Digital approach - ViewRanger

One innovative alternative to the terrain identification problem, which has the potential to circumvent some of the limitations discussed earlier, harnesses the power of mobile GIS to calculate which peaks can be seen from any particular location. This approach takes the form of a commercially available mobile phone mapping application called ViewRanger (see www.viewranger.com), whose 'Panorama View' (shown in Figure 3 below) annotates hill names onto terrain sketches. Although ViewRanger gives users the option of 3D textured panoramas, which convey depth well, this project only exploits ViewRanger's 'no shading' mode, because black outlines on a white mobile phone screen yield excellent contrast, even in bright sunlight.

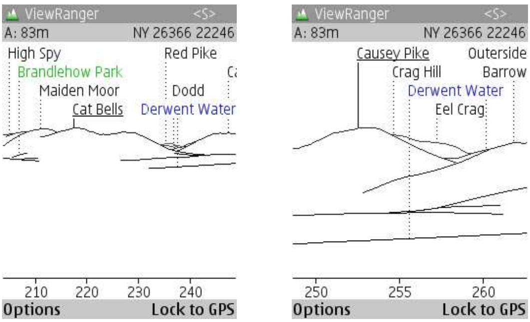


Figure 3. ViewRanger's Panorama View output, examples from England's Lake District.

The abstraction of terrain sketches or *caricatures* from DEMs is not new, and has been pursued by a numbers of researchers for many years prior to ViewRanger's release. The following disciplines share an interest in calculating critical breaks of slope from elevation data: Non-Photorealistic Rendering (DeCarlo *et al*., 2003; Visvalingam and Dowson, 1998); Machine Vision / Autonomous Navigation (Gienko *et al*., 2003); Cartography (Imhof, 1982); and Map Generalisation (Peucker and Douglas, 1975). Central to all these research efforts is an appreciation of terrain sketches' tremendous power to convey a great deal of information on surface form with extreme economy of lines. Yet little research effort has been devoted towards investigating how such an efficient system of terrain portrayal might be optimized for identifying hill names, let alone on mobile devices with inherently limited screen sizes and processing power. This project capitalises on ViewRanger's panorama engine as a platform for conducting an empirical study on the effectiveness of various panorama renderings, which will in turn feed into a revised methodology for generating terrain panoramas that are easier and quicker to use.

4. Experimental setup

A range of experimental methods—lab based and field based—are currently in use in a bid to firstly identify and isolate the factors that impede ViewRanger panorama recognition, and secondly to test whether improvements in digital panorama design lead to statistically significant improvements in users' accuracy rates and response times. The setup can also be used to compare the effectiveness of traditional approaches (e.g. printed Wainwright panoramas) to ViewRanger's digital approach.

The laboratory setup employs a series of Nokia smartphones running a customised 'Python for S60' application which logs (via Bluetooth) participants' reaction times and accuracy rates when matching terrain sketches on the phone screen to photographs of the real world displayed on a 24" monitor. For example, Figure 4 illustrates a terrain sketch (left) with its corresponding real world scene (right). By displaying a number of such image pairs to participants it is possible to reliably discern which types of terrain are easier to match than others. Early results indicate that scenes with one or two prominent features are 'easier' to match, although determining the precise nature of (and importance of) visually salient features is an ongoing challenge.

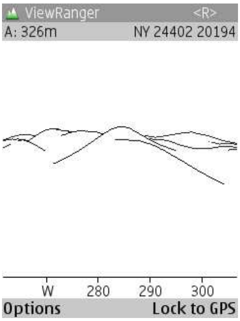


Figure 4. ViewRanger panorama alongside real world photograph of the same scene

Recently, terrain rendering software has been used to produce more scientifically robust lab experiments. By substituting artificial landscape renderings (created in Terragen, see http://www.planetside.co.uk/terragen) for photographs of the real world, it is possible to manipulate factors such as DEM resolution, terrain coarseness, number of peaks, prominence of peaks, or vertical exaggeration with much greater control. This technique facilitates fairer comparison between experimental conditions, because almost all factors other than the independent variable in question can be held constant.

Quantitative and qualitative field based experiments, which aim to complement the laboratory work, feature as part of two University of Nottingham Geography fieldtrips to England's Lake District.

5. Future Directions

Once an improved set of rules has been formulated for creating terrain panoramas that are more conducive to manual alignment with the real world, the obvious progression is to fuse these with an augmented reality (Liarokapis *et al*., 2006) style interface that overlays hill names (or other information) onto live video captured from the mobile phone's digital camera. A prototype system has been developed for the Nokia N95, although it is still some way from being fully integrated with the orientation data (pitch, roll and yaw) provided by a 3D inertial sensor. This hugely compelling concept is already taking hold on other platforms, with Google's new G1 phone (which contains a digital compass) featuring an application called Wikitude (http://www.mobilizy.com/wikitude.php), the developers of which claim that "users may hold the phone's camera against a spectacular mountain range and see the names and heights displayed as overlay mapped with the mountains in the camera" (Mobilizy, 2008). Without wanting to promote reckless over-reliance on electronic navigation aids in the outdoors, it is clear that the future holds enormous potential for combining digital terrain identification, mobile GIS, LBS and augmented reality techniques.

6. Conclusions

This paper has reported on efforts to improve the success with which digital panorama identification tools such as ViewRanger can allow hillwalkers to identify terrain features such as hill names. A novel phone-based experimental system has been described, which is at present being implemented to gather empirical data on the effectiveness (as evidenced by accuracy rates and reaction times) of different digital panorama techniques. The results of this research will offer the opportunity to design better, more effective terrain sketches, which could be applied to a wide range of applications, including mobile phone based augmented reality.

7. Acknowledgements

I am grateful to Craig Wareham and Mike Brocklehurst, of ViewRanger, for their support and technical assistance. The funding for my project is provided by a University of Nottingham School of Geography studentship.

References

DeCarlo, D., A. Finkelstein, et al. (2003). Suggestive contours for conveying shape. SIGGRAPH 2003.

Gienko, G., E. Levin, et al. (2003). Geographic Information to Support Vision-Based Approaches for GPS-Independent Autonomous Navigation. Proc. of ASPRS 2003 Annual Conference, Anchorage, USA.

Imhof, E. (1982). Cartographic Relief Presentation, De Gruyter.

Langmuir, E. (2004). Mountaincraft and Leadership. Manchester, MLTE.

Liarokapis, F., V. Brujic-Okretic, et al. (2006). "Exploring Urban Environments using Virtual and Augmented Reality." Journal of Virtual Reality and Broadcasting 3(5): 1-13.

Mobilizy Smartphone Solutions (2008) Wikitude Travel Guide. Available online: http://www.mobilizy.com/wikitude.php Accessed 10 December 2008.

Peucker, T. K. and E. G. Johnston (1972). "Detection of surface-specific points by local parallel processing of discrete terrain elevation data." Computer Graphics and Image Processing 1975(4): 375-387.

Pick, H. L., M. R. Heinrichs, et al. (1995). Topographic Map Reading. Local Applications of the Ecological Approach to Human-Machine Systems. P. A. Hancock, J. Flach, J. K. Caird and K. Vicente. New York, Erlbaum: 255-284.

Raper, J. (2007). "A critical evaluation of location based services and their potential." Journal of Location Based Services 1(1): 5-45.

Sousa, M. C. and P. Prusinkiewicz (2003). "A few good lines: Suggestive drawing of 3D models." Computer Graphics Forum 22(3): 381-390.

Visvalingam, M. and K. Dowson (1998). "Algorithms for sketching surfaces." Computers and Graphics 22(2-3): 269-180.

Wainwright, A. (1960). A Pictorial Guide to the Lakeland Fells: North Western Fells: 6, Frances Lincoln.

Wood, M., D. G. Pearson, et al. (2006). Comparing the effects of different 3D representations on human wayfinding Location Based Services and TeleCartography G. Gartner, W. Cartwright and M. P. Peterson. Berlin, Springer: 345-356.

8. Biography

John is studying for a PhD at the University of Nottingham, School of Geography. He presented at GISRUK 2006 in Maynooth, and was a postgrad helper at 2005’s GISRUK in Nottingham.

An Agent-based Model of Shifting Cultivation: Issues of Dynamic Land cover Validation

An The Ngo[1], Linda See[2], Frances Drake[3]

[1]School of Geography, University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, UK
Tel. (+44 343 3300) Fax (+44 343 3308)
Email: geotan@leeds.ac.uk

KEYWORDS: agent-based modelling, swidden agriculture, Vietnam

1. Introduction

Shifting cultivation, which is a rudimentary cultivation method (FAO, 2003; Gilruth *et al.*, 1995), is practiced widely in Vietnam. Shifting cultivation has continued to increase despite attempts by the Vietnamese government to reduce encroachment upon the forests, which are indispensable to shifting cultivators (Vien *et al.*, 2006). In some areas they have implemented systems of forest protection but large latent conflicts can occur whenever management measures are not sufficient.

In this context, a decision support tool in the form of an Agent-based Shifting Cultivation Model (ASCM) (Ngo *et al.*, 2008) has been developed. This tool can help decision makers analyze alternative solutions in order to assess a range of management policies, in particular whether or not forest protection policies could bring sustainable outcomes to the shifting cultivators. However, the ASCM can only produce reliable results within its experimental frame if the behaviours captured within the system are properly validated. Since the model contains many interacting and complex routines, validation is therefore the most important process in the development of the model.

This paper considers the main aspect of the validation framework, which consists of the calibration and verification process of the ASCM. Many validation procedures have been implemented in this research such as population development, household income and land ownership. However, only vegetation transition is selected for discussion in this paper because the vegetation dynamics are the final and the most important outcome of the model, resulting from complex interactions between human agents and land resources. Reliable results of vegetation distribution could help decision makers identify optimal policy solutions for management practice.

2. Calibration of vegetation transition within the ASCM

The trend of vegetation transition, especially in the fallow fields, is quite consistent within the northern uplands of Vietnam (Tran, 2007) and is illustrated Figure 1. To formulate the transition rules for the ASCM, the Markov approach (Barker, 1989; Vanclay, 1994) was applied in combination with a fuzzy-logic-controlled cellular automata model. The dynamic rule governing the change of vegetation status over time in the model is generalised to include both spatial and historical aspects as follows:

$$S_{t+1} = f\left(S_t, P_t, T_t\right) \qquad (1)$$

where S_{t+1} and S_t is the vegetation state of patch S at time t and $t+1$; P_t is the collective gross development at time t, depending on both identified internal and external forces of patch S; and T_t are the transition rules. Detail of these parameters will be described below.

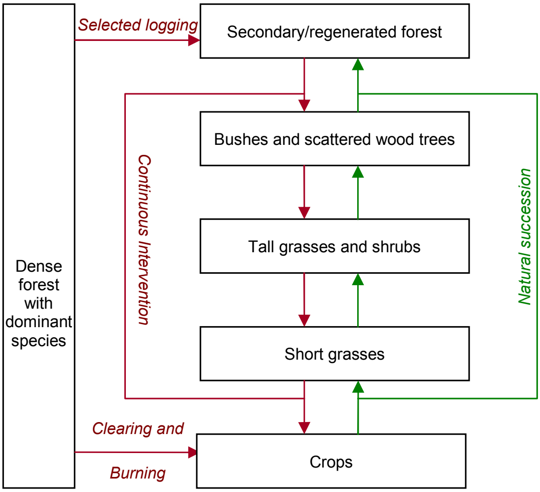


Figure 1. Vegetation transition processes in swidden fields

2.1. The vegetation development P_t

The collective gross development (P_t) is itself a function of natural succession (i.e. natural cumulative growth), neighbourhood effects, human intervention, and historical issues such as fallow age. Land-cover that is predominantly trees will be parameterised using data from VN-LUDAS (Le *et al.*, 2008), which is well validated for upland Vietnam. Shrub cover will be empirically parameterised from previous research (Do, 1994; Tran, 2007) and data collected during fieldwork at the study site in 2008.

Trees: The development probability of forest tree cover can be estimated using a set of numerical equations related to stand basal area (${}^{t+1}P_{Gr}$) at time $t+1$, which is calculated from ${}^{t}P_{Gr}$, natural cumulative growth ${}^{t}Z_g$ and basal area removed due to human activity (i.e. logging, firewood collection) as follows (Le *et al.*, 2008):

$$^{t+1}P_{Gr} = {}^{t}P_{Gr} + {}^{t}Z_G - G_{removal} \quad (2)$$

$$G_{removal} = G_{Logged} + G_{damage} + G_{mortality} / T \quad (3)$$

where G_{logged} is the basal area logged by human agents, G_{damage} is standing basal area damaged immediately by the logging operation, and $G_{mortality}$ is the basal area lost as tree mortality occurs over some years (T) after the logging event, expressed as follows:

$$G_{damage} = P_{Gr} * (0.0052 * G_{Logged} / g_{Logged} + 0.0536) \quad (4)$$

$$G_{mortality} = P_{Gr} * (0.0058 * G_{Logged} / g_{Logged} + 0.0412) \quad (5)$$

$$Z_G = a * (P_G)^{\varepsilon} - b * (P_G) \quad (6)$$

The parameters in equations 4 to 6 are: $a = 1.290014$; $b=0.033948$; $\xi = 10^{-6}$ (See Le (2005) for how to empirically calculate a and b); $G_{logged} = 0.0.38$; and $T = 3$ years.

Shrub and grass: This type of cover on natural land is not directly time dependent, especially in the areas that have been left for fallow for more than 10 years, and are quite stable for two reasons. First, these areas have unsuitable biophysical conditions for transition from shrub to secondary forest. Second, the shrub and grass community has reached an equilibrium state between endogenous species. This fact suggests that the development probability of shrub cover correlates positively with the time factor at the yearly development stage (<5 years). At longer times (> 8 years), if the cover is still a shrub for some reason, the chance of upward transition is negatively associated with time.

2.2. The transition rule T_t

The upward transition rule T_t is represented by a fuzzy-set of transition probability μ and user-identified transition threshold μ_0. The transaction probability μ is estimated from P_t and the identified range of the lower and upper development bounds ($[\alpha\ \beta]$ and $[\gamma\ \delta]$). T_t is also formulated separately for different types of vegetation cover.

Tree cover: Due to the stand basal area P_{Gr} is estimated independently with neighbourhood and soil conditions; the transition possibility is thus calculated by a linear monotonically increasing membership function (Cox, 2005):

$$\mu(P_t,\alpha,\beta) = \begin{cases} 0 & P_t < \alpha \\ (P_t-\alpha)/(\beta-\alpha) & \alpha \le P_t \le \beta \\ 1 & P_t > \beta \end{cases} \tag{7}$$

where α = 18.28 and β = 32.94 (m^2ha^{-1}) are the means of the basal area of the open and dense forests measured in Vietnam upland forests (Le *et al.*, 2008).

Shrub and grass cover: are captured through a trapezoidal fuzzy set (Cox, 2005):

$$\mu(P_t,\alpha,\beta) = \begin{cases} 0 & P_t < \alpha \\ (P_t-\alpha)/(\beta-\alpha) & \alpha \le P_t \le \beta \\ 1 & \beta \le P_t \le \gamma \\ (\delta-P_t)/(\delta-\gamma) & \gamma \le P_t \le \delta \\ 0 & \delta > P_t \end{cases} \tag{8}$$

The above parameters are adopted from field surveys separately for shrub and grass as provided in Table 1.

Table 1. The transition parameters of shrub and grass land-cover

Cover	Parameters (years)				Conditions
	α	β	γ	δ	
Shrub	2	8	9	30	3 or more neighbours patches are the forest; good soil
	3	9	10	25	3 or more neighbours patches are the forest
	3	10	11	25	Otherwise
Grass	1	3	5	30	3 or more neighbours patches are the forest; good soil
	1	4	6	25	3 or more neighbours patches are the forest
	1	5	7	25	Otherwise

The transition rule can be established by comparing μ and μ_0; if $\mu > \mu_0$ the cover will change from a lower to higher development order; otherwise it will remain the same.

The downward transition of the forest tree cover is predicted based on the lower bounds of each land-cover type. The lowest basal areas in Vietnam upland forests are measured by Le (2005) as $15.54m^2ha^{-1}$ and $28.56m^2ha^{-1}$ for open forest and dense forest, respectively. These parameters are adopted in the ASCM to identify whether or not a forest cover is converted to the lower order. A forest patch with P_t less than 15.54 will be re-identified as shrub cover in the simulation program.

The downward transition of the shrub and grass lands is mainly predicted from grazing. The transition rules are formulated based on field observations as follows:

$$C = \begin{cases} 0 & visit < 144 \\ (visit - 144)/(1095 - 144) & 144 \le visit \le 1095 \\ 1 & visit > 1095 \end{cases} \tag{9}$$

where C is the probability of downward vegetation conversion due to grazing activities. The relationship between T and C is formally expressed as:

$$T_t = \mu - C \tag{10}$$

If $T_t < 0$, shrub cover will be converted to bare land. The μ_0 parameter is then optimised by tuning its value to best match the proportions of land-cover type derived from satellite images and model outputs. $\mu_{0\text{ - }tree}$ and $\mu_{0\text{ - }shrub}$ are 0.5 and 0.8 respectively.

3. Verification of the simulation

The verification of the ASCM was undertaken by comparing the model outputs (run from 2000 onwards) with ground truth points collected during surveys in 2007 and 2008. Additionally, a pixel-by-pixel comparison of the land-cover predicted by the model and the land-cover map derived from remotely sensed data in 2006[1] was also conducted. In order to reconcile the spatial resolution from all the data sources, the smallest pixel size of $28.5m^2$ was chosen, which matches the resolution of the TM images. Overall accuracies of the tests are shown in Table 2.

Table 2: Results of accuracy test between model outputs against observation data

	2006 From satellite image	2007 Survey data (N=125)	2008 Survey data (N=98)
Overall Accuracy (%)	46	27	29
Standard Error	0.50	0.99	1.12
Standard Deviation	2.73	5.42	6.12
Number of repetition (Model runs)	30	30	30

The results above are quite significant compared to previous research, most of which has simply looked at the test of tendency but not a pixel-by-pixel comparison (Haggith *et al.*, 2003; Jepsen *et al.*, 2006; Leisz *et al.*, 2005; Sulistyawati *et al.*, 2005). Moreover, the results of Wada *et al.* (2007) for a 500m resolution indicated only a 5% accuracy.

[1] The unsupervised classification method was firstly applied to the TM satellite images. The land cover map was then modified by stacking and comparing the classified map with an aerial photo (1x1m) of the same area.

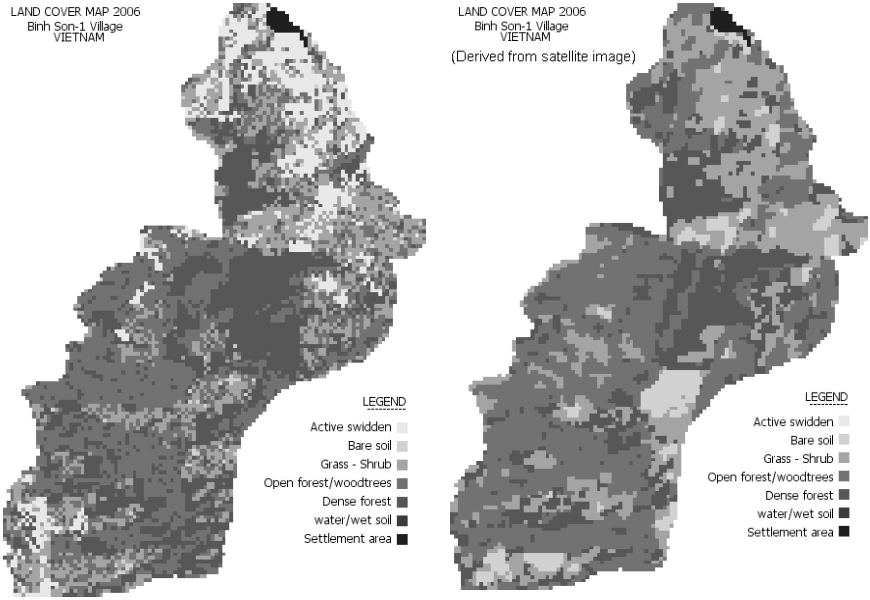


Figure 2. The model output land cover map (left) and land cover map derived from the satellite image (right). There is no active swidden class in the map derived from the satellite image. All bare soil and short grass could potentially be active swidden.

4. Conclusions

Although there are other validation processes not described here, the results to date indicate that the model appears to provide a realistic representation of the farmers' decision-making processes. Once the validation is complete, the ASCM will be used for scenario analysis to test out various alternative policy options.

5. Acknowledgements

We would like to acknowledge the support of the Vietnamese government and the School of Geography, University of Leeds in funding this research project.

6. References

Barker, W.L., 1989. A review of models of landscape change. *Landscape Ecology*. pp.111-133.

Cox, E., 2005. Fundamental Concepts of Fuzzy Logic. *Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration.* San Francisco, Morgan Kaufmann.

Do, D.S., 1994. Shifting Cultivation: Its social, economic and environmental values relative to alternative land-use. *IIED Forestry and Land use.* London, International Institute for Environment and Development.

FAO, 2003. Shifting cultivation. [online]. [Accessed 15 Jan 2007]. Available from World Wide Web: [http://www.fao.org/docrep/x5382e/x5382e03.htm].

Gilruth, P.T., Marsh, S.E. and Itami, R., 1995. A Dynamic Spatial Model of Shifting Cultivation in the Highlands of Guinea, West-Africa. *Ecological Modelling*. 79, pp.179-197.

Haggith, M., Muetzelfeldt, R. and Taylor, J., 2003. Modelling decision-making in rural communities at the forest margin. *Small-Scale Forestry*. 2**,** pp.241-258.

Jepsen, M.R., Leisz, S., Rasmussen, K., Jakobsen, J., Moller-Jensen, L. and Christiansen, L., 2006. Agent-based modelling of shifting cultivation field patterns, Vietnam. *International Journal of Geographical Information Science*. 20**,** pp.1067-1085.

Le, Q.B., 2005. *Multi-agent system for simulation of land-use and land cover change: A theoretical framework and its first implementation for an upland watershed in the Central Coast of Vietnam*, Gottingen, ZEF.

Le, Q.B., Park, S.J., Vlek, P.L.G. and Cremers, A.B., 2008. Land-Use Dynamic Simulator (LUDAS): A multi-agent system model for simulating spatio-temporal dynamics of coupled human-landscape system. I. Structure and theoretical specification. *Ecological Informatics*. 3**,** pp.135-153.

Leisz, S.J., Nguyen, T.T.H., Nguyen, T.B.Y., Nguyen, T.L. and Tran, D.V., 2005. Developing a methodology for identifying, mapping and potentially monitoring the distribution of general farming system types in Vietnam's northern mountain region. *Agricultural Systems*. 85**,** pp.340-363.

Ngo, T.A., See, L. and Drake, F., 2008. An Agent-based Approach to Simulating Spatial Patterns of Shifting Cultivation in Vietnam. *GISRUK 2008.* Manchester, Unigis.

Sulistyawati, E., Noble, I.R. and Roderick, M.L., 2005. A simulation model to study land use strategies in swidden agriculture systems. *Agricultural Systems*. 85**,** pp.271-288.

Tran, D.V., 2007. Solutions in fallow management: swidden agriculture experience in Vietnam's uplands.

Vanclay, J.K., 1994. *Modelling Forest Growth and Yield: Applications to Mixed Tropical Forests*, Wallingford, CAB International.

Vien, T.D., Leisz, S.J., Lam, N.T. and Rambo, A.T., 2006. Using traditional swidden agriculture to enhance rural livelihoods in Vietnam's uplands. *Mountain Research and Development*. 26**,** pp.192-196.

Wada, Y., Rajan, K.S. and Shibasaki, R., 2007. Modelling the spatial distribution of shifting cultivation in Luangprabang, Lao PDR. *Environment and Planning B-Planning & Design*. 34**,** pp.261-278.

Biography

An The Ngo is a third year PhD student in the School of Geography, University of Leeds. Back in Vietnam, Mr. Ngo is a lecturer in GIS application for natural resource management at Hanoi Agricultural University.

Dr Linda See is a Senior Lecturer in GIS and Computational Geography. Her research interests include soft computing techniques (fuzzy logic, neural networks and genetic algorithms) and agent-based modelling as applied to a range of spatial problems: flood forecasting, geodemographics, crime reduction, etc.

Dr Frances Drake is a Senior Lecturer in meteorology and climatology. Her research interests are in environmental governance and scientific citizenship. She is particularly interested in how climate change is being made amenable to government and technology choice in environmental issues.

Geographical Information for Archaeology: The Vanishing Landscapes of Syria

Robert Dunford*, Jennie Bradbury, Graham Philip and Nikolaos Galiatsatos
[1]Department of Archaeology, University of Durham Science Site, South Road, Durham, DH1 3LE
Tel. (44(0)1913341167) Fax (44(0)191 334-1101)
*Corresponding author: R.W.Dunford@durham.ac.uk
KEYWORDS: Archaeology, Syria, Corona, Ikonos, GIS

1. Landscape Archaeology

The focus of archaeological research in the Middle East has long been dominated by the excavation of large sites, with far less attention being paid to an investigation of the geographical context within which these sites are situated. Without this contextual information it is difficult to fully understand the changing nature and scale of human activity across the landscape. To address this, projects have begun to focus on *landscape archaeology*, where the integration of information regarding track ways and trading routes, water courses, field boundaries, agricultural and forest cover and underlying geology is invaluable, providing a context for human activity in the landscape. GIS, with its ability to consolidate such landscape features as digital data layers and then analyse their spatial patterns, provides a tool by which archaeological datasets can be investigated with reference to their geographical context.

2. Vanishing Landscapes of Syria

Figure 1. The Vanishing Landscape of Syria: a cairn field a) 2002 b) 2005; during preparation for agriculture.

In Syria, such analyses are particularly pertinent as the archaeological landscape is under significant threat of erasure as increasing demands for resources have lead to widespread agricultural development. This, in turn, has lead to sites of archaeological significance such as settlements, enclosures and burial structures being lost under the bulldozer (Figure 1). To address this projects such as the Settlement and Landscape Development in the Homs Region (1998-2006; Philip *et al.*, 2002; Bridgeland et al., 2003), Vanishing Landscapes of Syria (2007- ongoing; UoD, 2008a) and Fragile Crescent (2008-ongoing; UoD, 2008b) have been working at the landscape scale for many years, collating data. These projects have collated data drawn from archaeological field prospection, interpretation of existing maps and reviews of the published literature in order to preserve valuable information before it is lost.

Remotely sensed data has played a key role in enhancing these analyses, particularly following the declassification of US reconnaissance satellite imagery such as CORONA and IKONOS (Donoghue et al., 2002; Galiatsatos, 2004). With spatial resolutions less than 10m, and imagery from the late

1960s, and therefore prior to the commencement of much of the agricultural development, these systems allow for the identification of landscape features that have subsequently been lost. This allows both for the mapping of changing patterns of archaeological evidence, as well as targeting of further field surveys in an attempt to determine whether any archaeology has survived the site removal. Figure 2 highlights some of the datasets integrated within a GIS framework within the Vanishing Landscapes project.

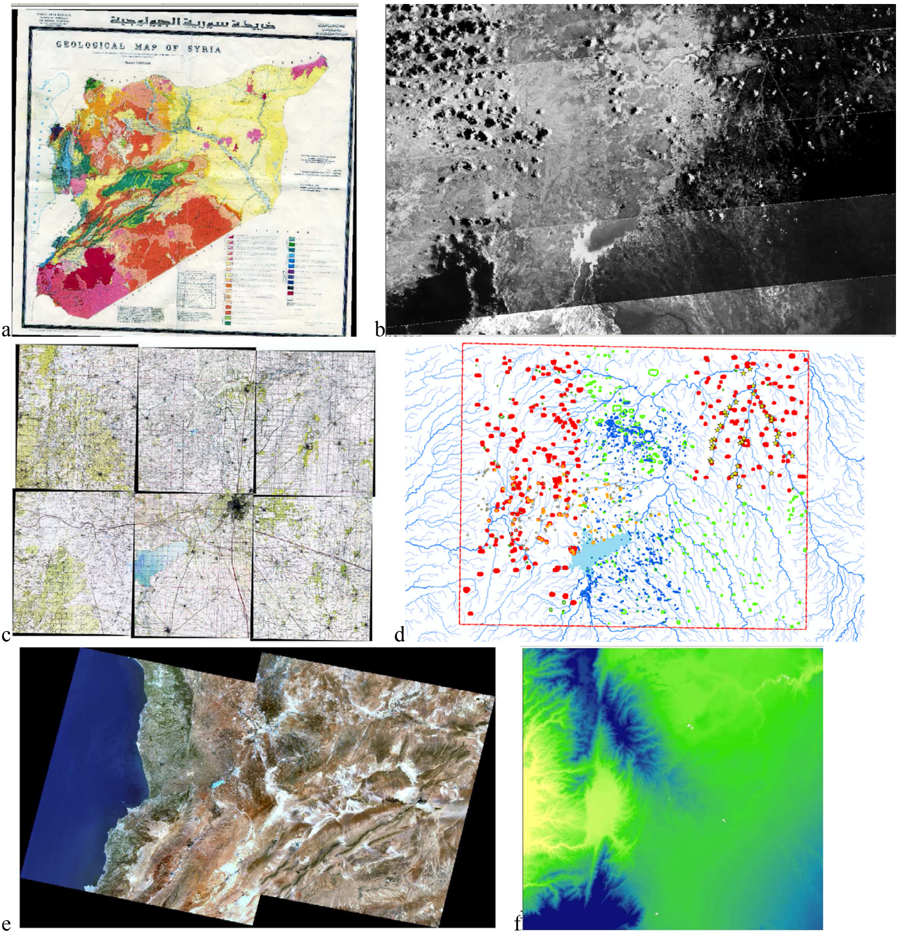


Figure 2. Consolidation of datasets within the Vanishing Landscapes Project: a) Geology Maps b) Corona 10m Satellite imagery (c. 1969) c) Map Data (Russian, Syrian and French survey maps) d) Vector data layers (including archaeological survey data, rivers and map interpretation data) e) Landsat 30m multispectral imagery f) SRTM digital elevation model.

3. Conclusion

This paper reviews the role of Geographical Information Science within the context of the Vanishing Archaeological Landscapes of Syria project run by the Archaeology Department at the University of Durham. It presents and evaluates the impression of the archaeological landscape that is possible when diverse data layers are drawn together to provide an archaeological context. We argue that the integration of data in this manner is not only an important analytical tool, but also provides a useful theoretical framework by which the changing archaeological palimpsest can be recognised and considered. Furthermore the paper evaluates these conclusions in light of the reliability of the input data and the GIS analyses applied to them. Considerations of the future of GIS as a tool for sharing archaeological data with the country of origin will also be discussed.

4. Acknowledgements

The work presented in this paper is funded by the Leverhulme trust and has been undertaken in cooperation with the Syrian Directorate of Antiquities and Museums. Thanks are extended to colleagues within the Settlement and Landscape development in Homs Region and Vanishing Landscapes projects.

5. References

Bridgland, D.R., Philip,G., Westaway, R. & White, M. (2003). A long Quaternary terrace sequence in the Orontes River Valley, Syria: a record of uplift and of human occupation.*Current Science* **84**(8): 1080-1089.

Donoghue, D.N.M., Galiatsatos, N., Philip, G. & Beck, A.R. 2002. Satellite imagery for archaeological applications: case study from Orontes valley, Syria. In *Aerial archaeology:developing future practice*. Robert W. Bewley & W. Rączkowski Amsterdam: IOS Press.

Galiatsatos, N. (2004). Assessment of the CORONA series of satellite imagery in landscape archaeology: a case study from the Orontes valley, Syria. Department of Geography. Durham, Durham University. PhD.

Philip, G., Jabour, F., Beck, A., Bshesh, M., Grove, J., Kirk, A. & Millard, A. (2002), Settlement and Landscape Development in the Homs Region, Syria: research questions, preliminary results 1999-2000 and future potential. *Levant* **34**: 1-23.

UoD (2008a), Vanishing landscape of Syria: ground and space mapping of a diverse world http://www.dur.ac.uk/archaeology/research/projects/?mode=project&id=362, *Electronic Resource*: accessed 12/12/08.

UoD (2008b), The Fragile Crescent Project (Middle East): Settlement Change during the Urban Transition, http://www.dur.ac.uk/archaeology/research/projects/?mode=project&id=391, *Electronic Resource*: accessed 12/12/08.

Biography

Robert Dunford is a Postdoctoral Research Associate in the Department of Archaeology whose research interests are the application of geographical information systems, remote sensing and the world wide web for applications in Environmental Management and Archaeology.

Jennie Bradbury is a Doctoral Student in the Department of Archaeology whose doctoral research focuses on the Cairn Landscapes of the Northern Homs region in Syria.

Graham Philip is a Professor in the Department of Archaeology; his research interests fall into three main areas: landscape archaeology, artefact studies, and the nature of early complex societies.

Nikolaos Galiatsatos is a Postdoctoral Research Associates in the Department of Archaeology whose research interests are in the Remote Sensing of Archaeological Landscapes with a special focus on the use of CORONA.

Creating and Maintaining Street Orienteering Maps using OpenStreetMap

Oliver O'Brien

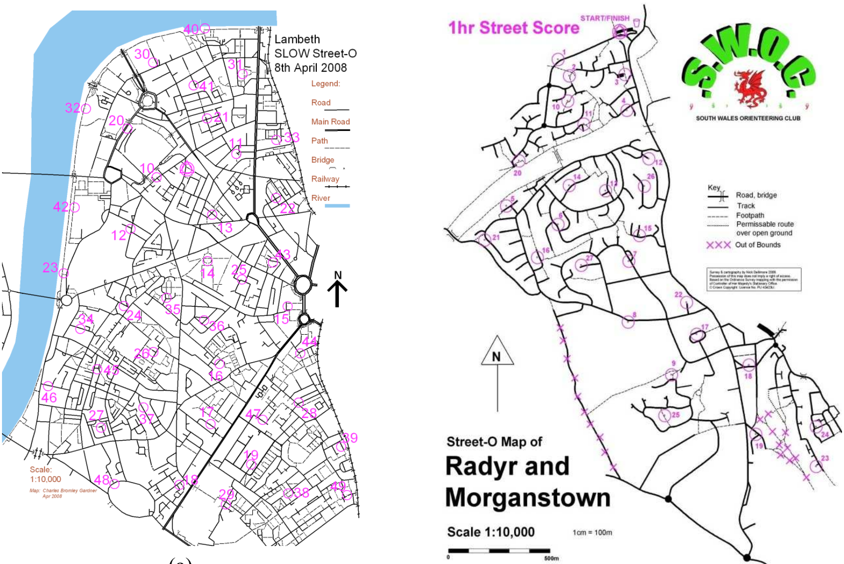
Department of Geography, University College London
Pearson Building, Gower Street, London WC1E 6BT
Tel. +44 (0) 20 7679 0510
o.obrien@ucl.ac.uk

KEYWORDS: orienteering, running, route choice, OpenStreetMap, Quantum GIS

1 Introduction

Street orienteering (street-O) is an informal form of the competitive navigational sport of orienteering. Meets are typically held in urban or suburban areas on weekday evenings, and require the competitor to visit as many checkpoints (controls) as possible within a certain time, typically one hour. At each control, a question on the competitor's "clue sheet" accompanying their map can be answered. Success at a street-O race requires the ability to plan and run an efficient route between as many higher scoring controls as possible, while avoiding being penalised for a late return to the start.

The map used is very minimalistic, typically only showing roads and paths. Street names, point features and contours are not normally included. Areal features are only shown if very significant, such as major rivers. Unlike regular orienteering maps, there is no de jure standard for street-O mapping and designs vary from club to club. A typical scale is around 1:10000 and the map is normally printed on an A4 sheet of paper.



(a) (b)

Figure 1. Sample street-O maps. (a) Lambeth, central London. Cartography: Charles Bromley-Gardner. (b) Radyr and Morganstown in Cardiff. Cartography: Nick Dallimore.

2 Street-O Map Production

Street-O maps have traditionally been produced using an application called OCAD (OCAD AG, 2005), in a three stage process. Firstly a base-map is scanned from a street-atlas or taken from screenshots of Ordnance Survey or Google Maps data. The street network is then digitised. The manually constructed representations are then evaluated by a field-survey, which identifies extra geographic features (e.g. alleyways) which are missing from the map and assesses suitable locations for controls. Once the ground truthing is complete, a final course layout can be created and the event maps produced. Manual street-O map production is therefore a laborious and time consuming process. The research presented in this paper aims to demonstrate, through the implementing of a GIS, an efficient and automated way in which maps suitable for street-O can be created.

Using a GIS to process and create street-O maps allows the direct use of vectored and spatially referenced base data, thus avoiding manual digitisation from a base raster. Additionally, use of a GIS enables spatial metadata not intended for the final printed map to be stored. Examples could include possible future control sites, previously used start venues with comments as to their suitability, and comments about possible road traffic conditions. Additionally, at the course design stages, spatial calculations could be conducted in a GIS to examine optimum route possibilities. Post-race, this information might also be used by competitors to analyse their actual route run and improve their route planning skills for future races.

3 The Street-O GIS Architecture and Implementation

A number of open-source desktop GIS applications are in active development, and are very promising for the display and production of simple maps from spatially-referenced data. Quantum GIS (QGIS) is used in this paper as it has a straightforward user interface and is compatible with multiple operating systems. Additionally, it also has a map composer, allowing for creation of cartographically simple maps directly within the application.

A PostGIS database is used to store the orienteering-specific data, such as controls and course annotations. It has good integration with QGIS and is free to use. Although in this paper a local database is used, if it were accessed across a network it could allow such data to be easily shared between mappers within a club or wider community.

OpenStreetMap (OSM) is used as the data source for the base-map. The quality and completeness of its coverage is variable (Haklay, 2008) but rapidly improving, and is already sufficient, for the application outlined in this paper, in many urban areas.

The data is obtained using the XML export function on the OSM website. A custom Python script filters out non-relevant features, transforms the XML to the GML specification (OGC, 2008a), and converts significant areal features into polygons (OSM stores areas as closed directional polylines). The script simplifies and enhances the way that overlapping features are indicated, to ensure that such features are stacked appropriately on the orienteering map. For example, water overlapping with a road is assumed to normally be below it, except where it is specifically marked as on a bridge passing above the road.

Any errors or omissions in the OSM data that are discovered during the ground truthing of the initial maps can be corrected using the “Potlatch” Flash-based editor on the OSM website, and the data re-exported. This benefits the wider OSM community and minimises non-orienteering data (i.e. corrections to base maps) stored on the orienteering system.

GML layer files are created to record each feature with topological ordering. These data are then imported into a map in QGIS as layers, projected and styled appropriately. QGIS's styling is relatively unsophisticated, not yet making use of the OGC's Style Layer Descriptor (OGC, 2008b) specification, but is sufficient for street-O maps. A duplicate layer for bridged roads is used and styled to emulate the normal cartographic "cased" representation of bridges.

Orienteering-specific data can be added by connecting to and editing various tables set up in a PostGIS installation – one each for controls, start/finish locations, out-of-bounds areas and annotations. The last table's data would not appear on the final printed map – for the others, an appropriate field (e.g. event date) could be used in the filter to pull the appropriate subset of data for the event concerned.

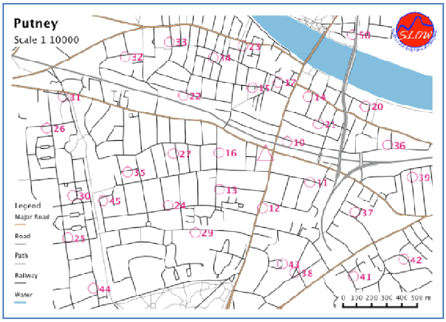
4 Case Study: Putney

The method has been tested with the QGIS application and a PostGIS database locally installed on a computer running Mac OS X, as a prototype. The Putney area in south-west London has an existing street-O map, allowing a comparison to be made (Figure 2.) The data is obtained from OSM via the web interface described above and filtered. The resulting layer files are added to a new QGIS project, re-projected and styled.

A number of cartographic problems arise which can be managed, relating to "end-capping" of the strokes, lack of generalisation and unconnected polylines. Comparison with the existing street-O map and online mapping reveal some data discrepancies, although, after ground truthing, the OSM-derived data was found to be more correct in the majority of cases.

Next, the relevant tables in PostGIS are included in the project, and the start and controls added. Control questions and answers can also be added, which could simplify production of the clue sheet.

(a)



(b)

Figure 2. Street-O maps of Putney, south-west London. (a) The original map. Cartography: Dorte Torpe Hansen. (b) Map produced from OpenStreetMap data using Quantum GIS, coloured according to the standards for conventional orienteering maps that some existing street-O maps, but not the original Putney map, follow.

Finally, QGIS's Print Composer is used to select the appropriate portion of the map and add adornments, including an automatically generated legend based on the styling present, a title, scale, scale bar and club logo. The map can be saved as an SVG, PDF, or printed directly.

5 Future Work

The application as presented is sufficient for creating street-O maps, however in the future it may be beneficial to evaluate a range of other data including:

5.1 Contours

Managing the orienteering and base-map data in a GIS allows spatial data from other sources to be added. Contours can be added quite easily – public domain digital elevation model data can be converted to contours and the resulting shapefiles added to the map (Allan, 2008). The contours produced using this method however are not "smoothed" and tend to contain anomalous features, and therefore would only be useful for areas that include significant hills.

5.2 Blue Plaques

One common source of clue questions for the London-based street-O maps is the inscriptions on the "blue plaques" managed by English Heritage (EH). To provide a quick way to identify potential control sites without first visiting the area, the EH website's plaque list can be obtained, the locations geocoded using the Google Local web service, and a GPX file created containing their locations and inscriptions. These can be seen as a layer of blue circles in Figure 3.

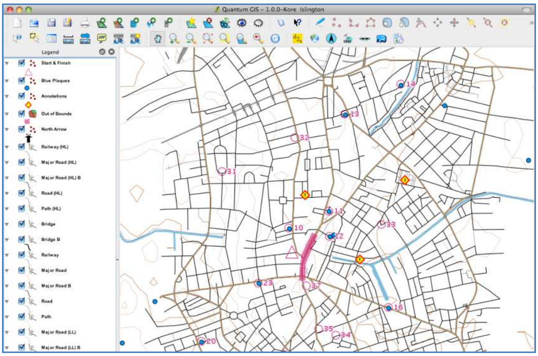


Figure 3. Screenshot of QGIS showing contours, blue plaque locations and annotations, added to a street-O map of Islington, north London.

6 Routing and the Orienteering Problem

Optimum routing calculations can be performed on the data – this is of interest to competitors during post-race analysis. The "pyroutelib" OSM library developed by White (2007) can be used to find theoretically optimum routes, once the control locations are added as appropriate nodes on the adjacent polyline segments on the network – this is achieved with a Python script.

Computing a best route for the whole race, assuming a fixed running speed and limit, is known as the Orienteering Problem (Tsiligrides, 1984 p797), a variant of the Travelling Sales Problem (Gutin, 2002) and is hard to do computationally (Chekuri, 2008) – heuristics are needed for large numbers of controls. The research field studying combinatorial optimisation problems like these is very active.

7 Conclusions

The process described above provides a fast and accurate way to create street-O maps of areas where OSM coverage is sufficiently complete and accurate. As OSM's contributions continue, it will become feasible to produce street-O maps for more and more areas. Both the orienteering and OSM communities would benefit from field-checking observations, while a separate but centralised store of orienteering data would prove particularly useful for organisers in subsequent years. The process is still not as easy as it needs to be for widespread adoption in the community – however the automation capabilities of QGIS and further development of open-source GISes and source data can simplify and enhance the process.

Acknowledgements

The author would like to thank Dr Alex Singleton of UCL for his feedback on this paper, Dr Jason Dykes of City University for supervising the MSc dissertation that this paper was derived from, and the various street-O mappers who answered a questionnaire and provided map data files. The figures above show maps derived from OpenStreetMap data (© OpenStreetMap contributors), and maps derived from Ordnance Survey mapping (© Crown Copyright. All rights reserved.)

References

Allan A et al (2008) Contours, *OpenStreetMap Wiki*. Available from: *http://wiki.openstreetmap.org/index.php/Contours* [Accessed 4 December 2008]

Chekuri C et al (2008) Improved Algorithms for Orienteering and Related Problems. *In: ACM-SIAM Symposium on Discrete Algorithms, 20-22 January 2008 San-Francisco, California.*

Gutin G (2002) *The Traveling Salesman Problem and Its Variations,* Dordrecht: Kluwer Academic Publishers.

Haklay M (2008) How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England. Under review in *Environment & Planning B* and available at *http://tinyurl.com/osmquality* [Accessed 4 December 2008]

OCAD AG (2005) OCAD 9: Smart for Cartography, *OCAD Website*. Available from: *http://www.ocad.com/en/Ocad9Flyer.pdf* [Accessed 4 December 2008]

OGC (2008a) *Geography Markup Language (GML) Encoding Standard*. Available from: *http://www.opengeospatial.org/standards/gml* [Accessed 4 December 2008]

OGC (2008b) *OpenGIS Styled Layer Descriptor Profile of the Web Map Service Implementation Specification*. Available from: *http://www.opengeospatial.org/standards/sld* [Accessed 4 December 2008]

Tsiligrides T (1984) Heuristic Methods Applied to Orienteering, *Journal of Operational Research Society* **35** pp797-809

White O (2007) Pyroutelib, *OpenStreetMap Wiki*. Available from: *http://wiki.openstreetmap.org/index.php/PyrouteLib* [Accessed 4 December 2008]

Author Biography

After a degree in Physics at the University of Oxford and four years working in the City of London as a financial software developer, Oliver took his MSc in GIS at City University, and now works as a research assistant and programmer at UCL, as part of an ESRC project led by Dr Alex Singleton.

Improving the London GreenMap - Comparing Approaches to Displaying Large Numbers of Points in GoogleMaps

Claire Ellul[1], Fabrice Marteau[2], Muki Haklay[1]

[1]Dept. of Civil, Environmental and Geomatic Engineering, University College London,
Gower Street, London, WC1E 6BT
Tel. +44 (0) 20 7679 4118 Fax +44 (0) 20 7380 0453
c.ellul, m.haklay@ucl.ac.uk

[2]Ecole National des Science Geographique, 6-8 avenue Blaise Pascal, Champs-sur-Marne
77455 Marne La Vallee, Cedex 2, France

KEYWORDS: Google Maps, Performance, Web GIS, Community Mapping

1. The London GreenMap

The London GreenMap (London21 2008a) was the first web-based Green Map in the world and was launched in early 2004. It is a core component of the activities undertaken by the London21 charity (London21 2008b), which is a network of community groups, individuals and representatives who work in all parts of Greater London to help create a greener, healthier and more sustainable city. Members of the network add information to the map about green organisations or events. The general public can then browse the map and search the events (e.g. an eco-design fair, a lecture on corporate social responsibility) and organisations (e.g. green campaign groups, cultural and leisure activity centres, community groups) to find out what is happening in their locality. They can also receive text messages about events taking place within a given radius of their home location and having themes of particular interest.

There are around 1500 organisations listed in the London21 database and the GreenMap forms an important component of fund-raising exercises. The ability to demonstrate, at a glance, the depth and breadth of organisations forming part of the network is fundamental. However, the GreenMap is implemented on the Google Maps Application Programming Interface (API, Google 2008). This provides the ability to overlay vector data (points, lines and polygons) onto the base mapping images supplied by Google, but cannot display large numbers of interactive points (see Purvis *et al.*, Chapter 7). Currently, an unsatisfactory work-around has been implemented – the GreenMap displays a random collection of organisations (***Figure 1***).

This paper presents the results of an investigation into improvements to the Green Map to meet the following requirements:

- The ability to display all organisations simultaneously on the map, in under 30 seconds
- The ability to click on any of these organisations and retrieve further information

Given the wide range of technical skills and computational power within the user base of the GreenMap, any selected solution should be compatible with both the Internet Explorer and the Firefox web browsers, and perform well on a low-specification computer.

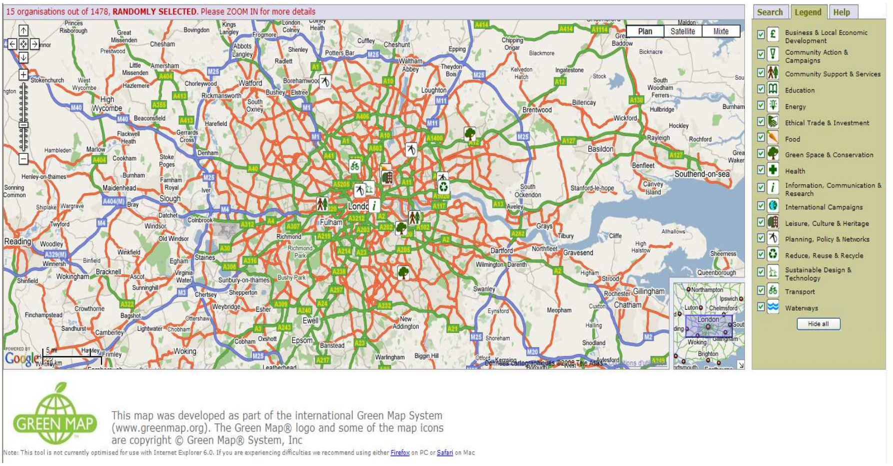


Figure 1 - Current London GreenMap showing 15 randomly selected Organisations

2. Potential Solutions

A number of potential solutions were identified, as follows:

- Option 1 - use vector drawing functionality (embedded in the web browser) to overlay the points on top of the Google Map. For the Firefox browser, this is provided in the form of Scalable Vector Graphics (SVG, Mozilla 2008). Internet Explorer uses Vector Mark-up Language (VML, Microsoft 2008). Using a vector-based approach has the advantage that all points are created with interactivity. However, additional code is required to redraw the points when the user zooms in or out (with the Google Maps API this redraw functionality is inbuilt)[1].
- Option 2 - Generate a series of raster images of the points at various different scales, and overlay these on the Google Maps. This can has the advantage that performance is optimised. However, should a new point be added, the rasters must be regenerated. No interactivity is possible using this approach.
- Option 3- Clustering the points, where an evenly distributed sub-set of the points is selected for display (rather than the random selection used in the current solution). Whilst providing interactivity, this solution does not meet the requirement to show the depth and breadth of the organisations data when zoomed out and was discarded.

3. Performance Tests

Performance tests were carried out using Options 1 and 2 above, and baseline tests were also carried out using Google Maps API point display functionality. The tests were carried out on two computers, one having a 1.9 GHz AMD Athlon Processor and 1.5GB RAM and the other having a Pentium D 3 GHz processor with 3.5GB of RAM. These were selected as being fairly representative of the range of machines used by the end-users of the GreenMap.

Performance measurement tools were used for the time measurements – Firebug (an add in which allows developers to edit, debug, and monitor live web pages) was used for FireFox and HTTP Watch Basic (which provides similar monitoring and debugging tools) for Internet Explorer.

[1] Note that while SVG plug-ins are available for Internet Explorer, the use of plug-ins is felt to be inappropriate for the non-technical user-base of the GreenMap.

Tests were carried out for 0 to 1500 points determine performance changes as data volume increases. To exclude database query time from the test execution time, the organisation data was held as KML files. Each test was executed ten times, and it was assumed that the loading time for the Google base mapping is consistent over all the tests.

4. Results

Table 1 below summarizes the results obtained for the 1500 point test for the varying combinations of Firefox/SVG and Internet Explorer/VML on both slow and fast computers.

Table 1 - Performance Obtained for 1500 points for the various Tests

Number of Points	Test	Average Loading Time (s) – FireFox/SVG – Slow Computer	Average Loading Time (s) – Internet Explorer/VML – Slow Computer	Average Loading Time (s) – Firefox/SVG – Fast Computer	Average Loading Time (s) – Internet Explorer/VML – Fast Computer
1500	Google Maps API	100.00	128.00	35.78	107.00
1500	VML and SVG without Icons	20.93	16.04	12.53	8.52
1500	VML and SVG with Icons	24.58	108.00	12.15	8.77
1500	Raster – (GMap Creator)	6.63	6.15	4.03	4.05

4.1 – Google Maps API

In all cases 1500 points took over 30 seconds to display. However, the Firefox browser performs notably better than Internet Explorer, giving almost acceptable performance on the higher-specification 'Fast' computer.

4.2 – VML and SVG

The VML and SVG tests were divided into two parts – firstly the tests were run to display standard simple circles without images or icons. The second round of tests associated existing GreenMap icons with each point to determine the impact these had on performance.

For both VML and SVG, results for simple points were more than adequate for the GreenMap requirements. With icons, the Firefox browser gave adequate performance. However, issues were encountered with Internet Explorer on the lower-specification computer, with an overall time of 108 seconds to display 1500 points.

4.3 GMap Creator/Raster Approach

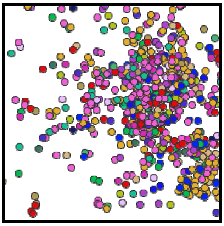


Figure 2 – A Transparent PNGs created by the GMap showing all Organisations

The CASA *GMap Creator* (Centre for Advanced Spatial Analysis, CASA 2008) takes ESRI shape (.shp) files as input and generates transparent Portable Network Graphics (PNG) images at varying scales (Figure 2). These can then be overlaid onto the Google Map base, as shown in Figure 3. Points can be coloured according to a theme (in this case, organisation type).

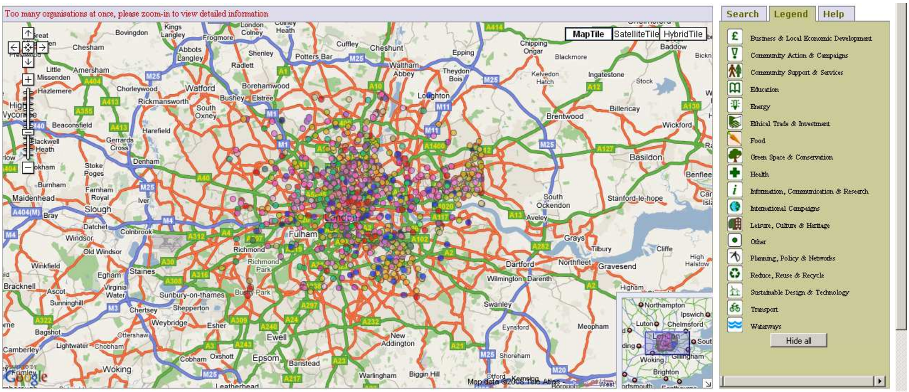


Figure 3 – Results obtained from the GMap Creator – Organisations Coloured by Theme

Results obtained using this approach were nearly identical for both browsers, and in all cases significantly lower than those obtained using the vector drawing approaches.

5. Comparisons

Figure 4 shows a graph of the results obtained on the slower computer for all tests.

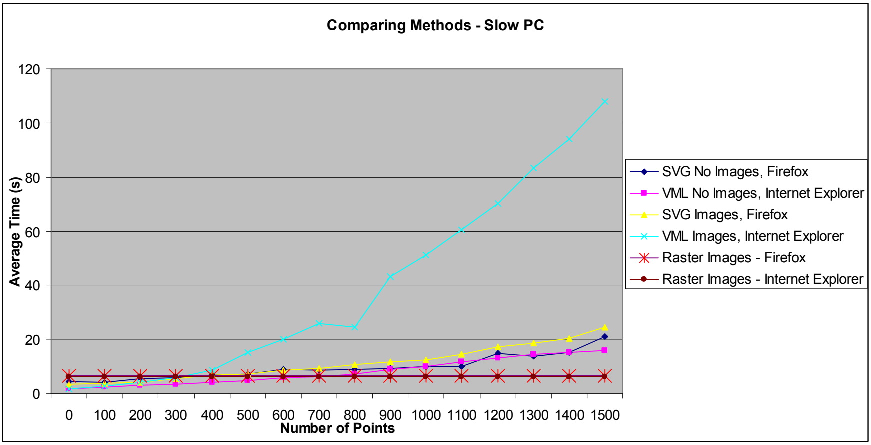


Figure 4 - Comparing Methods for increasing numbers of points - Slow Computer

For the Firefox browser the use of SVG gives the required results for all datasets up to 1500 points, with a highest performance time of 24.58s. However, an issue arose in Internet Explorer using VML in combination with images, with an average performance of 108s for 1500 points on the slower computer. Given this, and the fact that good performance was obtained with the Raster Images in all cases, a compromise solution has been implemented. When the user is zoomed out (over 300 points to be displayed), the raster images are displayed (see Figure 3). When zoomed further in, the vector Google Maps points are displayed (see Figure 5). The cut-off of 300 organisations was selected by examining the performance of the Google Maps API under varying point numbers (average display time for 300 points in Internet Explorer is 17.13 seconds, and for 400 points is 23.25 seconds).

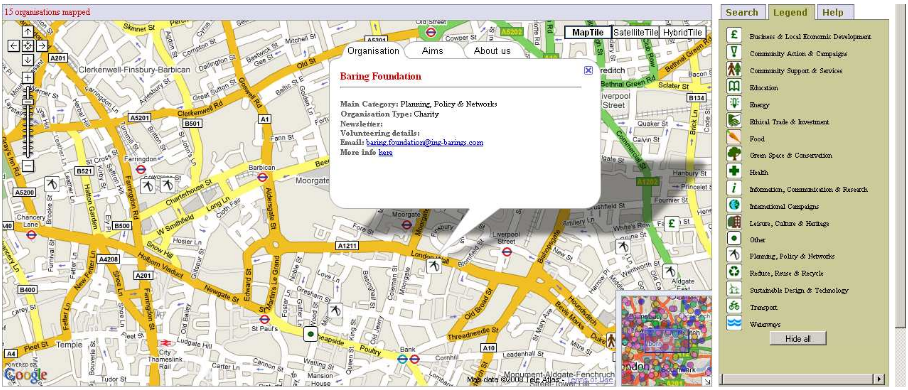


Figure 5 – Final Solution – Using Raster Images when zoomed out (see Figure 3) and Interactive GoogleMaps Points when Zoomed In

This solution does provide the ability to view the entire dataset of organisations, whilst maintaining some level of interactivity. Additionally, the use of in-built Google Maps functionality to display the points avoids the requirement to write separate, browser-specific, code for the VML and SVG display. However, as it is not currently possible to incorporate custom icons into the image, further work is required to determine an appropriate method to replace the standard point symbology provided by the GMap creator. A method to automatically re-generate the raster images is also required.

6. Further Work

An investigation into the slow performance given by Internet Explorer for VML with images may lead to a resolution of the issue described above, which may be caused by lack of image caching. Alternative combinations of vector and raster data (for example VML/SVG with Raster maps) could also be examined. Further work could also take into account user preferences – users may prefer slower performance but with clearer icons and more interactivity.

7. Acknowledgements

This project was supported the UCL-led UrbanBuzz Programme, within which UEL is a prime partner.

8. References

CASA (2008) *The GMap Creator* [online], Available from: http://www.casa.ucl.ac.uk/software/gmapcreator.asp [Accessed 12th December 2008]

Google (2008) *Google Maps Application Programmers Interface [online] Available from: http://code.google.com/apis/maps/* [Accessed 12th December 2008]

London21 (2008a) *About the Map* [online], Available from: http://greenmap.london21.org/about_map.php [Accessed 12th December 2008]

London21 (2008b) - *About Us – Networking for a Green and Sustainable City* [online], Available from: http://www.london21.org/page/16 [Accessed 12th December 2008]

Microsoft (2008) *VML Reference* [online], Available from: http://msdn.microsoft.com/en-us/library/bb263898(VS.85).aspx [Accessed 12th December 2008]

Mozilla (2008) *Mozilla SVG Project* [online], Available from: http://www.mozilla.org/projects/svg/ [Accessed 12th December 2008]

Purvis M, Sambells J and Turner C (2006). *Beginning Google Maps Applications with PHP and Ajax,* Springer-Verlag, New York, USA.

7. Biography

Claire Ellul completed a PhD on Topological Relationships between 3D Objects, and is currently working as a Post-Doctoral researcher at University College London and London Metropolitan University. Prior to commencing her PhD, she spent 10 years working as a GIS consultant and software engineer.

Surface roughness as a landscape index

Katherine Arrell, Steve Carver
School of Geography, University of Leeds, LS2 9JT
Tel. (+44) 0113 3433343 Fax (+44) 0113 3433308
Email k.arrell@leeds.ac.uk

KEYWORDS: DEM, terrain analysis, landscape, topographic characterisation

1. Introduction

Greater accuracy and higher resolution terrain data from direct measurements (for example SAR/LiDAR/TLS) have created a wide range of opportunities for detailed landscape analyses previously hampered by a lack of suitable data. Further, the increasing size and volume of these datasets necessitate quantitative data generalisations and metadata that can inform process studies, for example drainage density and relative relief. A number of studies have attempted to extract geomorphically significant measures from digital elevation data (Pellegrini, 1995; Wood, 1996; Burrough, et al., 2000; Arrell et al., 2007), these have largely attempted to characterise landscape elements and thus infer geomorphic process. Attempts to characterise or classify landscapes holistically still remain under developed and would provide useful metrics for digital elevation data analysis. This paper looks at the development of measures of surface roughness as a multi-scale index for characterising landscape types.

We propose that the methods outlined here can provide landscape characterisations that reflect surface geomorphology, differentiating between surface types e.g. fluvial vs. glacial, erosional vs. depositional, soft vs. hard geology, when these landscape types exhibit different surface roughness scaling trends. We propose that scaling roughness trends will provide more meaningful measures than semi-variograms alone where local variability in surface properties governs the convergence and divergence of mass and energy which form critical controls on surface processes.

2. Study Area

LiDAR data for the upper Wharfe Yorkshire, the Odenwinkelkees glacial valley Austria and Cley-next-to-the-Sea were used as test datasets. These varied landscapes were selected to assess the robustness of the outlined technique to differentiate between landscapes. SRTM data were used for the same areas.

3. Methods

A number of different measures of surface roughness were used to assess their ability to differiate between landscapes. Roughness was measured as the standard deviation of each elevation, slope, aspect, curvature and convergence/divergence, comprising five different measures in total.

The standard deviation of surface roughness was measured within increasing kernel windows from 3x3 cells upward and stored per pixel for each different kernel size. These data were exported as an ASCII file and plotted against kernel window size to look at local, focal and global trends in surface roughness. This process was repeated with each of the five quantifications of surface roughness.

Resultant graphs bear resemblance to semi-variograms showing the increasing dissimilarity of locales with distance.

Data for each DEM were plotted and used to characterise trends for landscape types using a line of best fit through the data (Figure 1). The confidence with which we can account for the variability in the data with a trend line through roughness data will diminish as the scatter of points at each kernel size increases. The landscape types defined by their trend surfaces will also allow regions within the data to be identified. These may reflect key landforms or constituent landscape elements present at specific scales. Optimal landscape differentiation and hence classification will occur where surface roughness is different for different landscape types at the same kernel size and where surface roughness changes for the same landscape type at different kernel sizes.

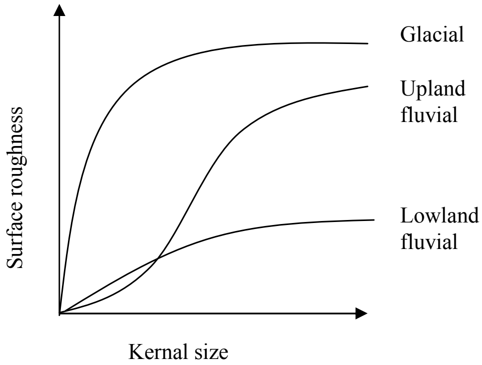


Figure 1. Schematic of plotted landscape types

4. Results and Discussion

The proposed technique will identify indicative roughness trends for different landscapes through analysis of absolute magnitudes of, and scaling trends in, surface roughness. The method proposes the identification of key or indicative kernel sizes for different landscape types defined by the scale of key landforms (for example valleys, slopes, cliffs). A measure of the robustness of this technique to classify landscape types and the ability of different roughness measures to differentiate between different landscape types will be assessed by testing the resulting classifications on a range of "unseen" terrain models.

References

Arrell KE, Fisher, P, Tate N, Bastin L (2007) A fuzzy c-means classification of elevation derivatives to extract the morphometric classification of landforms in Snowdonia, Wales, *Computers & Geosciences* **33** pp1366–1381.
Burrough PA, van Gaans P, MacMillan R, (2000) High resolution landform classification using fuzzy k-means. *Fuzzy Sets and Systems* **113** pp37–52.
Pellegrini G (1995) *Terrain shape classification of digital elevation models using eigenvectors and Fourier transforms* Ph.D. Dissertation, State University of New York College of Environmental Science and Forestry, Syracuse, NY.
Wood J (1996) *The geomorphological characterisation of digital elevation models*. PhD Dissertation, University of Leicester, /http://www.soi.city.ac.uk/_jwo/phd/S (accessed 13 April 2007).

Biography

Katherine Arrell is a lecturer in physical geography at the School of Geography, University of Leeds. Her research interests are largely focused on terrain analysis and characterisation and regional modelling of glacier mass balance.