Proceedings of the GIS Research UK 16th Annual conference GISRUK 2008

UNIGIS
Manchester Metropolitan University
2nd - 4th April 2008

Editor: David Lambrick

Proceedings of the GIS Research UK
16th Annual Conference
GISRUK 2008

Manchester Metropolitan University
2nd - 4th April 2008

Published April 2008, Manchester Metropolitan University.

ISBN 978-1-905476-31-2

Welcome

The GISRUK 2008 local organising committee is delighted to welcome you to the sixteenth annual GIS Research UK (GISRUK) conference hosted by UNIGIS UK in association with the Department of Environmental and Geographical Sciences, Manchester Metropolitan University, the School of Environment and Life Science, the University of Salford and the University of Huddersfield.

Established in 1993, the annual GISRUK event continues to attract contributions from across the UK, Europe and beyond. The conference aims are :

- to act as a focus for GIS research in the UK
- to provide a mechanism for the announcement and publication of GIS research
- to act as an interdisciplinary forum for the discussion of research ideas
- to promote active collaboration amongst researchers from diverse parent disciplines
- to provide a framework in which postgraduate students can see their work in a national context.

Manchester and UNIGIS prompted two of this year's themes, GIS and Urban Issues and GIS and Education. We were very pleased to receive a broad and strong response to these themes. Visualisation and technologies for neo-Geographies were the other two themes and as you will see in these proceedings they are also well represented.

We are very grateful for the support of our sponsors to ensure that the cost of attending is affordable by all and provide items such as the conference bags and USB memory sticks. Manchester Metropolitan University has provided accommodation and conference services without which the conference could not have taken place. In particular we would like to acknowledge the large contributions made by Jill Richardson and Tracy McKenna to the smooth running of the conference.

Thanks to the authors and the reviewing committee, standards of papers were once again very high.

We hope you enjoy your stay in Manchester and the conference programme.

Best wishes

David Lambrick
Local Chair
GISRUK 2008

Local Organising Committee:

Richard Armitage	University of Salford
Jianquan Cheng	MMU
Mark Cresswell	MMU
Mark Danson	University of Salford
David Lambrick (Chair)	MMU
Derek Reeve	University of Huddersfield
Jean Reeve	UNIGIS UK
Graham Smith	MMU

National GISRUK Steering Committee:

Katherine Arrell	University of Leeds
Angela Baker	ESRI
Jane Drummond	University of Glasgow
David Fairbairn	University of Newcastle
Bruce Gittings (Chair)	University of Edinburgh
Peter Halls	University of York
Rich Harris	University of Bristol
Glen Hart	Ordnance Survey
David Lambrick	Manchester Metropolitan University
Andrew Lovett	University of East Anglia
Jeremy Morley	University College London
Nick Mount	University of Nottingham
Gary Priestnall	University of Nottingham
Nick Tate	University of Leicester
Adam Winstanley	National University of Ireland, Maynooth
Steve Wise	University of Sheffield
Jo Wood	City University

Contents

UNIGIS Opening Keynote: Professor Henk Scholten, IVM, Vrije Universiteit, Amsterdam

Closing Keynote: Ed Parsons, Google
The Challenges ahead: The role of GI Science in providing geospatial information to the mass-market

Session 2A Visualisation

From Slice and Dice to Hierarchical Cartograms: Spatial Referencing of Treemaps 1
Jo Wood and Jason Dykes, School of Informatics, City University London

Mediating geovisualization to potential users and prototyping a geovisualization application ...9
David Lloyd, Jason Dykes, Robert Radburn, School of Informatics, City University, London and *Leicestershire County Council*

The Effectiveness of Artificial Landmarks for Orientation Tasks within a Large-Scale Virtual Environment 17
Adam Rousell, Claire Jarvis, Jing Li, Department of Geography, University of Leicester, UK

The Visual Exploration of Insurance Data in Google Earth 24
Aidan Slingsby, Jason Dykes, Jo Wood, Matthew Foote[2], Mike Blom[3], Department of Information Science, School of Computing, City University, UK, Willis Research Network and Willis Analytics, London

Using and Evaluating HCI Techniques in Geovisualization:Applying Standard and Adapted Methods in Research and Education 33
Stephanie Larissa Marsh & Jason Dykes, School of Informatics, City University London

Session 2B Land and Energy

Analysing the relationship between landscape composition and preference – the combination of GIS and visualisations 39
Åsa Ode, David R. Miller, *Dept. of Landscape Architecture, Swedish University of Agricultural Sciences,Sweden,Macaulay Institute, Aberdeen*

Estimating land use profiles for diffuse pollution modelling in river catchments 44
Paulette Posen, Andrew Lovett, Michael Hutchins, Amelie Deflandre-Vlandas, Helen Davies, School of Environmental Sciences, University of East Anglia, Norwich and Centre for Ecology and Hydrology Wallingford.

Return of the Stanford Bunny - Definition, Calculation and Application of Visual Topology........50
Sang, Neil, Macaulay Institute,Aberdeen

Assessing bioenergy crop potential in Yorkshire and the Humber using land suitability and multi-objective land allocation 56
Patrizia Tenerelli, Steve Carver Dipartimento PROGESA, Facoltà di Agraria, University of Bari, Italy and School of Geography, University of Leeds, UK

Investigating the Spatial Variation of Natural Energy Potential in the UK 60
Michael Blom, Department of Information Science, School of Computing, City University,UK

Session 3A Spatial Analysis and Methods

Discovering Spatially Multiway Collocations ………………………………………………………...66
Didier Leibovici, Lucy Bastin and Mike Jackson, Centre for Geospatial Sciences, University of Nottingham, UK and School of Engineering & Applied Science, Aston University, UK

Mind the Gap: Reuniting the Tube network with underlying spatial themes …………………...71
Craig Foot, Pragya Agarwal, Deloitte & Touche LLP, London and Department of Civil, Environmental and Geomatic Engineering, University College London, London

Automated schematization using memetic algorithms ………………………………………77
Jerry Swan, Suchith Anand, Mark Ware, Mike Jackson, Centre for Geospatial Science, University of Nottingham, UK and Faculty of Advanced Technology, University of Glamorgan, Wales, UK

Space-time trajectories in a planar and torus-based self-organising map: the importance of eliminating boundary effect ………………………………………………………………………83
Nick Mount and Dan Weaver School of Geography, University of Nottingham, Nottingham

Mapping uncertainty in perceptions of landscape 'wildness' ..…………………………………90
Alexis (Lex) Comber, Steve Carver, Steffen Fritz, Robert McMorran, Justin Washtell, Department of Geography, University of Leicester, School of Geography, University of Leeds, International Institute for Applied Systems Analysis, Austria and Centre for Mountain Studies, Perth College-UHI, Perth.

Session 3B Spatial Literacy

In Search of Spatial Literacy: the Nottingham Sputnik Model ..…………………………………94
Jim Nixon, Robert J. Abrahart, Claire Chambers, Centre for Geospatial Science, University of Nottingham, Nottingham

Developing spatial literacy in secondary education: GIS practicals for key stage 3...... ………102
Alexis (Lex) Comber, Matt Buxton, Claire Jarvis, Jane Wellens, Phil Wood, Department of Geography, University of Leicester, Djanogly City Academy, Nottingham, Staff Development Centre, School of Education, University of Leicester, UK

Noisy School Kids: Using GPS in an Urban Environment …………………………………106
Duncan Whyatt, Gemma Davies, Marion Walker, Colin Pooley, Paul Coulton and Will Bamford, Department of Geography and Department of Communication Systems, Lancaster University

Learning to read the landscape: Interactive Visual Tools for exploring spatial thinking ....……111
Gary Priestnall, Jack March, School of Geography, The University of Nottingham, UK

Session 4A Urban GIS

The Fine Scale Spatial Dynamics of the Greater London Housing Market ...…………………117
Andrew T. Crooks, Duncan A. Smith, Margarethe Theseira, Centre for Advanced Spatial Analysis, University College London and GLA Economics City Hall, London, UK

Visualising London's Suburbs ……………………………………………………………………125
Catherine Emma Jones, Muki Haklay, Sam Griffiths and Laura Vaughan, Department of Civil, Environmental and Geomatic Engineering and The Bartlett School of Graduate Studies, University College London, London, UK

Modelling residential living space for individual properties using digital infrastructure and remote sensing data: a Cardiff case study ……………………………………………………133
Scott Orford, Jonathan Radcliffe, School of City and Regional Planning, University of Cardiff Cardiff

Spatiotemporal interaction modelling and simulation for transport system based on a geosensor network 138
Seong K. Choi[1], Pragya Agarwal Department of Civil, Environmental and Geomatic Engineering, University College London, London

Origin-specific models for analysing commuting flows in Northern Ireland: scale effects and other problems 145
C. D. Lloyd, I. G. Shuttleworth, G. Catney School of Geography, Archaeology and Palaeoecology, Queen's University, Belfast

Session 4B DEMs/DTMs/Morphology

The Accuracy of Gridded Digital Elevation Models Interpolated to Higher Resolution 150
Stephen Wise, Ed Hanna, Tristram Irvine-Fynn, Department of Geography, University of Sheffield, UK

Assessment of DTM quality: a case study using fine spatial resolution data from alternative sources 156
Michal Gallay School of Geography, Archaeology and Palaeoecology, Queen's University Belfast

Developing an automatic process for hydrology generalization: combining a vector GIS topography model with DTM-derived flow direction data 163
Marshall *J. Heap Birkbeck College, University of London*

Using GIS to Reproduce Landscape Evolution in Active Volcanic Environments 177
Amii Darnell, Andrew Lovett and Jennifer Barclay School of Environmental Sciences, University of East Anglia, Norwich, UK

A Statistical and GIS-Based Approach for Morphodynamic Characterisation and Modelling....183
A. M. Bakare, J. G. Morley, R. R. Simons, Department of Civil, Environmental and Geomatic Engineering, University College London

Session 5A Spatial Analysis and methods

Surveying by motion tracking: the application of GIS, camera tracking and animation software to model 3D subterranean landscape from video imagery 191
Anna E Mason, David M Mountain, Jonathan Raper City University, Northampton Square, London

A 3D Scrapbook Approach to Geospatial Analysis 200
S. Wang, S. Brooks, Dalhousie University, Halifax, Nova Scotia, Canada

PDEAR Model Prediction of Rare Protea Species under Climate Change Effects 205
Danni Guo, Renkuan Guo, Guy F. Midgley, and A.G. Rebelo Kirstenbosch Research Center, South African National Biodiversity Institute and Department of Statistical Sciences, University of Cape Town, South Africa

Testing co-evolutionary hypotheses over geological timescales using GIS 214
Richard J. Butler, Paul M. Barrett, Paul Kenrick, Malcolm G. Penn Department of Palaeontology, The Natural History Museum, Cromwell Road, London

Methods for Predicting Human Behaviour in Emergencies: An Analysis of Scientific Literature 217
Věra Karasová and Glyn Lawson Centre for Geospatial Science, School of Geography, University of Nottingham,UK

Session 5B Ordnance Survey Session

Session 6A Urban GIS

Using Urban Viewsheds for Embedding Geographical Context in Photograph Databases of Urban Areas 223
Thierry Joliveau and Sanjay Rana, Université Jean Monnet - Saint-Etienne, France and CTS, University College London, UK

Visualising Neighbourhood Well Being: GIS in an urban environment ……….........……226
Mike Franks, Oldham Metropolitan Borough Council, UK

Modelling Street Safety in an Urban Environment......………………...……………………230
Claire Ellul, Ben Calnan' Cities Institute, London Metropolitan University, UK

Measuring 'alcohol related' crime and disorder: Examining the impact of extended licensing hours in England ………………..………………………………………………………………236
Andrew Newton,Alex Hirschfield' Applied Criminology Centre, University of Huddersfield, UK

The Importance of Workplace Quality of Life in Determining Office Location ...……………….240
Alice Dalton, 'Faculty of the Built Environment, University of the West of England, Bristol, UK

Session 6B Contributed Geodata and Modelling

OpenRouteService.org is three times "Open": Combining OpenSource, OpenLS and OpenStreetMaps ………………………………………………………………………………..248
Pascal Neis and Alexander Zipf, Department of Geography, University of Bonn, Germany

Exploiting Volunteered Geographic Information to describe Place ..…………...………………252
Ross S. Purves and Alistair J. Edwardes' Department of Geography, University of Zurich, Switzerland

Exploring Volunteered Geographic Information to Describe Place: Visualization of the 'Geograph British Isles' Collection ...……………………………………………………………256
Jason Dykes, Ross Purves, Alistair Edwardes & Jo Wood, giCentre, School of Informatics, City University London and Department of Geography, University of Zurich, Switzerland

A Methodology for Inferring Higher Level Semantic Information from Spatial Databases ..…268
Marie-Kristina Thomson and Roderic Béra' Department of Civil, Environmental & Geomatic Engineering, University College London, UK

UncertML: an XML schema for exchanging uncertainty ……………………...……………275
Matthew Williams, Dan Cornford, Lucy Bastin, Ben Ingram' School of Engineering and Applied Science, Aston University, Birmingham, UK

An Agent-based Approach to Simulating Spatial Patterns of Shifting Cultivation in Vietnam.…280
An The Ngo, Linda Se, Frances Drake' School of Geography, University of Leeds, Woodhouse Lane, Leeds

Session 7A GIS and Education

"I'd rather be watching the telly…": Do rich media approaches offer real teaching and learning benefits for GIS software tuition of digital natives? 285
Nick Mount and Claire Chambers' School of Geography, University of Nottingham, Nottingham

GIS education by another name? Geographical referencing for social scientists ………..…291
David Martin, Samantha Cockings and Samuel Leung, School of Geography, University of Southampton, Southampton, UK

e-Learning in GI Sciences – Approaches towards Sustainable Content …………………......295
Susanne Bleisch, Stephan Nebiker, Joël Fisler, University of Applied Sciences Northwestern Switzerland

Geo-contextualised visualisation for teaching and learning in the field …………………...…300
Claire Jarvis, Gary Priestnall, Gemma Polmaer, Jing Li, Department of Geography, University of LeicesterUK and School of Geography, University of Nottingham, UK

Real-Time Landscape Visualisation: Experiences in a Teaching and Learning Context ……308
Andy Burton, Gary Priestnall, Gemma Polmear, Nick Mount, School of Geography/Institute of Engineering Surveying and Space Geodesy, The University of Nottingham, UK

Session 7B Sociospatial

Social and environmental inequalities in rural England: spatial distributions and the effects of scale ………..316
Steve Cinderby, Annemarieke de Bruin, Meg Huby, Piran White,, Stockholm Environment Institute,Department of Social Policy and Social Work, Environment Department, University of York, York, UK

Socio-economic vulnerability units – modelling meaningful spatial units ……………………..319
Stefan Kienberger, Stefan Lang, Dirk Tiede, Centre for Geoinformatics, Salzburg University, Austria

A method for quantifying equity of access to local services for different demographic groups: combining network analysis with logistic regression using the example of greenspace ……323
Alexis Comber and Chris Brunsdon, Department of Geography, University of Leicester, UK

Session 8A Geodemographics

Despatialising geodemographic propensities ……………………………………………………328
Martin Callingham, School of Geography, Birkbeck College, University of London

Inclusion and exclusion problems in geodemographic targeting ……………………...……335
Jakob Petersen, Pablo Mateos, Paul Longley, Phil Atkinson, Centre for Advanced Spatial Analysis, University College London and Southwark Primary Care Trust, London

Do Local Information Systems Hide the Bigger Picture? An analytical approach to measuring the strength of local boundaries ……………………………………………………………...…342
Sarah Goodwin & Jason Dykes, Manchester Geomatics Ltd, Manchester and giCentre, School of Informatics, City University London, UK

Scenario-based Small Area Population Modelling for Social Infrastructure Planning ………348
Yang Li, Allan J. Brimicombe, Centre for Geo-Information Studies, University of East London, University Way, London, UK

Session 8B Participatory GIS

How to reach the 'hard-to-reach': the development of Participatory GIS for inclusive urban design in UK cities ……………………………………………………………………………..……354
Steve Cinderby, Stockholm Environment Institute, University of York, York, UK

User-centered approach for developing PPGIS …………………………………………………....358
Khitam Shraim, School of Environment & Development, University of Manchester, UK

Participatory Strategic Planning Using a Virtual Reality Environment ……………........……367
David Miller, Paula Horne, Jane Morrice, Jonathon Ball and Pernette Messager Macaulay Institute, Craigiebuckler, Aberdeen

Supporting interactive land use planning of fen meadows in the Netherlands: the Gouwe Wiericke case 372
Gustavo Arciniegas, Ron Janssen, Nancy Omtzigt, and Corinne Cornelisse, Institute for Environmental Studies, Vrije Universiteit, Amsterdam, The Netherlands

From Slice and Dice to Hierarchical Cartograms: Spatial Referencing of Treemaps

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
{jwo | jad7} @soi.city.ac.uk, www.soi.city.ac.uk/~ {jwo | jad7}

KEYWORDS: treemaps, cartograms, tree structures, information visualization, geovisualization

1. Introduction

Treemaps are graphical devices for representing hierarchical information in a compact efficient manner. This paper presents new extensions to the treemap construction process that make them more effective at displaying spatial hierarchical information. Treemaps were first proposed by Schneiderman (1992) as an alternative to the more common directed graph representation of hierarchical information. Figure 1 illustrates how a simple hierarchical arrangement of data can be represented as a directed graph and as a treemap. In the directed graph representation, significant visual emphasis is placed on graph edges (represented as lines), which connect nodes (represented as text) in the network. In the treemap representation, each node in the graph is represented as a rectangle. If a node has child nodes attached to it, the rectangle is subdivided into smaller rectangles representing each of its children.

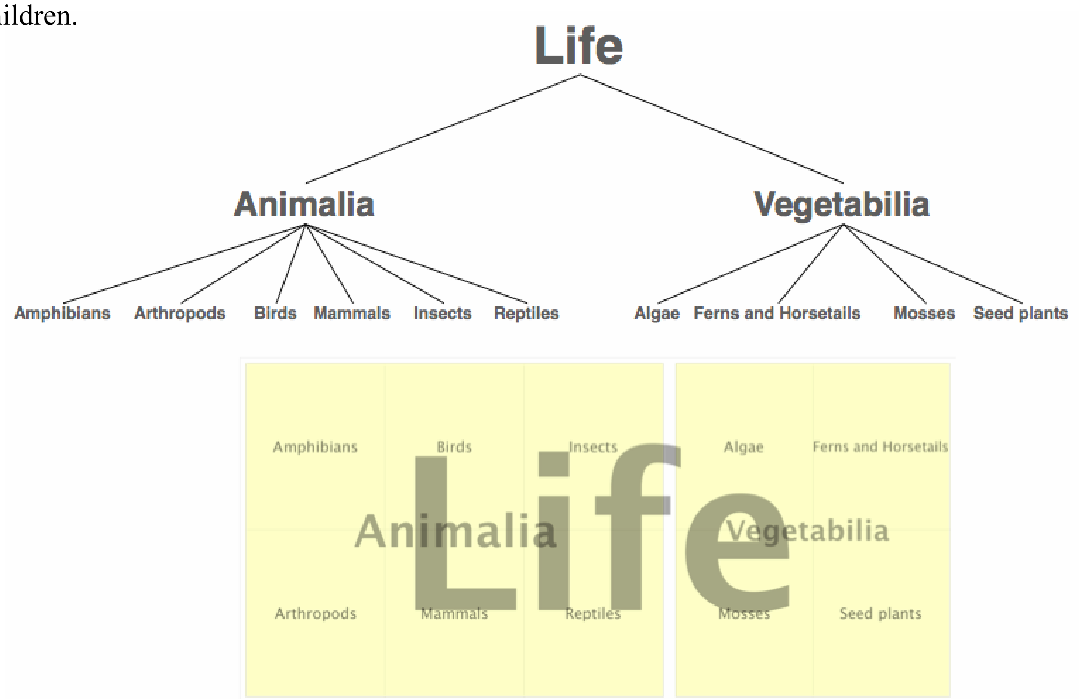


Figure 1. Simple hierarchy represented as a directed graph (upper) and treemap (lower).

Treemaps offer significant advantages over alternative tree representations. Firstly, they are very space efficient. Almost all the graphical space is used to represent data. This is in contrast to the directed graph representation, which is dominantly whitespace. This makes treemaps particularly suitable for large hierarchies and those with high branching ratios. Secondly, space efficiency allows additional visual variables (Bertin, 1983) to be incorporated into the representation. In particular, rectangles may be sized according to some measurement attribute of leaf nodes. Leaf nodes may also be coloured according to some additional attribute. The space efficiency of the representation also allows text labelling of nodes that may be independent of the variables mapped to size and colour.

Treemaps are widely used in information visualization, with applications ranging from the representation of computer file size (Van Wijk, 2002), to gene ontology (Baehrecke *et al*, 2004), to stock market changes (SmartMoney.com, 2007). They appear to be less widely used in the spatial sciences, although the quadtree and other spatial indexing variants (R-tree, *k*d-tree etc.) are analogous mappings of directed trees into 2- and multi- dimensional space. One of the reasons for a lack of uptake in the spatial sciences may the lack of consistency of node location and shape within a treemap, properties that are often important when considering geographic analysis and interpretation. Yet much spatial data is organised in some hierarchy (e.g. postcodes, census units, national and continental boundaries etc.) suggesting a potential for treemap display of geographic information.

2. Problems with TreeMaps

2.1 Poor Aspect Ratios

The original algorithm for treemap construction proposed by Schneiderman (1992) subdivided each node into strips whose width was proportional to the magnitude of the child nodes they represented. At each level of the hierarchy, strips were created in alternating vertical and horizontal directions (see Figure 2, left). This method, subsequently known as ‘slice and dice’, has the advantage of computational simplicity and the preservation of the 1-dimensional order of nodes (nodes A-E and F-I remain in alphabetical order in Figure 2, left), but has the significant problem of producing nodes with very high aspect ratios. This makes visual assessment of relative area difficult. The solution proposed by Bruls *et al* (2000) was the ‘squarified treemap’, which attempted to subdivide any given rectangle to minimise the aspect ratio of child nodes (see Figure 2, right). This makes visual assessment of relative area, as well as node labelling and colouring much easier.

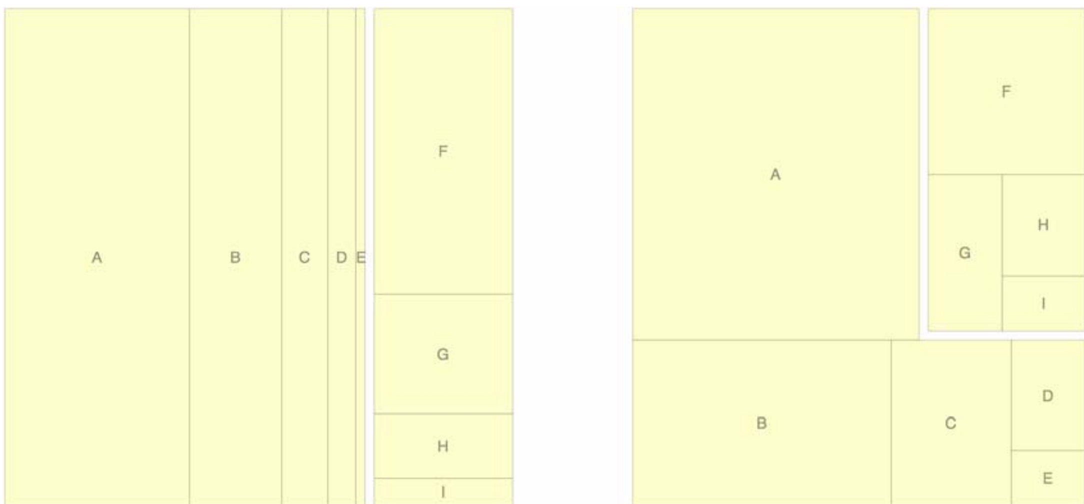


Figure 2. Two-level hierarchy of unequally sized nodes arranged using the *slice and dice* (left) and *squarified* (right) algorithms.

2.2 Visual Indication of hierarchy

Displays of hierarchies with a depth greater than 1 or 2 levels can result in layouts where the hierarchical structure of the data is difficult to interpret. Strategies to emphasise the hierarchy include placing borders of different widths between nodes at each hierarchical level. For deep hierarchies this has the problem introducing too much whitespace between nodes reducing the space available for data representation. Van Wijk and Van de Wetering (1999) proposed 'cushion maps' which produce shaded textures at each level of the hierarchy, emphasising the structure of the graph. While effective at showing the graph topology, the process inhibits the use of colour and text labelling of nodes.

An alternative approach is proposed here whereby colours are allocated to nodes using the following algorithm:

For each node with unallocated colour:
If parent node is undefined and node has no colour allocated,
Count number of sibling nodes, ***n***
Allocate node a random hue from a circular colour space.
*Allocate each sibling node a hue (360/**n**)^o away from last allocated hue.*
Else,
Inherit parent node's colour and assign this colour with random mutation to node.

This approach ensures that top level nodes in the tree have contrasting hues and that sub trees share some common colouring but with sufficient variation to distinguish them. Figure 3 illustrates an example using a subset of London postcodes. Varying the degree of colour mutation allows different levels of the hierarchy to be distinguished.

Figure 3. 15,573 unit postcodes in the London 'E' postcode area. Arranged using the *squarified* layout using the evolutionary colour allocation. *Postcode centroids crown copyright/database right 2008. An Ordnance Survey/EDINA supplied service.*

2. 3 Inconsistency of Positioning

While the squarified treemap layout is widely used, it has one significant drawback. The locations of sibling nodes within a given rectangle are dependent on the order in which those nodes are processed. Sibling nodes are usually ordered one-dimensionally from largest to smallest, but their two- dimensional position will depend on the aspect ratio of the space in which the node must be placed. This can produce a confusing layout of nodes, as illustrated in Figure 4. While it is generally the case that nodes of higher magnitude (darker green in Figure 4) appear towards the top and left of the enclosing space, the relationship with distance from the origin is not consistent. For example, nodes 15 and 16 are of similar magnitude, but at very different distances from the corner. This inconsistency can be quantified by the R^2 correlation coefficient between node order and distance from the origin. The result is that for most squarified treemaps, graphical location is not a reliable indicator of any meaningful characteristics of the node data.

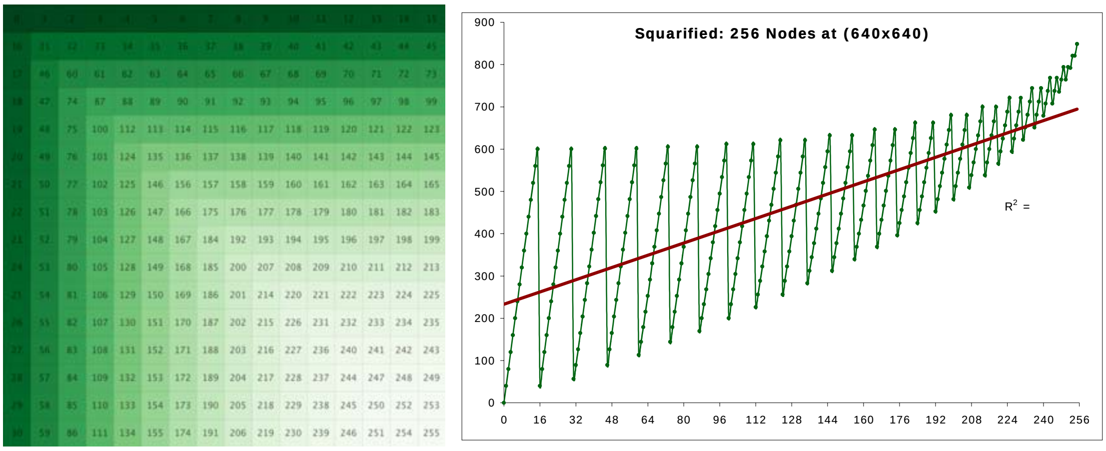


Figure 4. *Squarified* layout of 256 unit sized ordered nodes. Nodes are coloured by order. Graph shows relationship between order (x-axis) and distance from top-left (y-axis).

3. Ordered and Spatial Treemaps

In response to the limitation described in 2.3 above, we propose two new layout algorithms that provide a more consistent mapping from node value to spatial location. These proposed developments enhance the value of location in the representation and improve the layout stability (Bederson *et al*, 2002).

3.1 Ordered Treemaps

In devising an *ordered squarified* layout algorithm, we attempt to provide a consistent relationship between node magnitude and distance from an origin. This is achieved allocating a nominal location to each sibling node such that they are spaced regularly within the containing rectangle:

Function ***OrderedSquarify***:

*Set **r** to be parent rectangle, **s** the set of all **n** sibling nodes and **AllocatePosition(s,r)***
*While **n** > 0,*
*Divide **r** into a grid of **n** cells equally spaced from each other*
*Allocate a row or column of nodes in **s** closest to the current row/column minimising the aspect ratio of allocated nodes.*
*Shrink **r** by the column or row previously processed. Reduce **n** by the number of allocated nodes.*
AllocatePosition(s,r)

Function ***AllocatePosition (s,r)***:
*Order set **s** of **n** nodes by node magnitude.*
*Create a set **d** of **n** locations such that each is equally spaced from its neighbours within the space **r**.*
*Sort items in **d** by distance from origin..*
*Allocate a location to each node in **s** by finding the location in **d** that corresponds to the same position in their sorted sets.*

The result of applying the layout algorithm is a much improved distance consistency. Locations of nodes within a treemap now have more useful meaning (see Figure 5).

Figure 5. *Ordered squarified* layout of 256 unit sized ordered nodes. Nodes are coloured by order. Graph shows relationship between order (x-axis) and distance from top-left (y-axis).

3.2 Spatial Treemaps

A further enhancement for geographically located data is to apply the *ordered squarify* algorithm, but replace the function ***AllocatePosition(s,r)*** with one that simply performs an affine transformation of each node's geographic location such that all the nodes in **s** fit within the rectangle **r**.

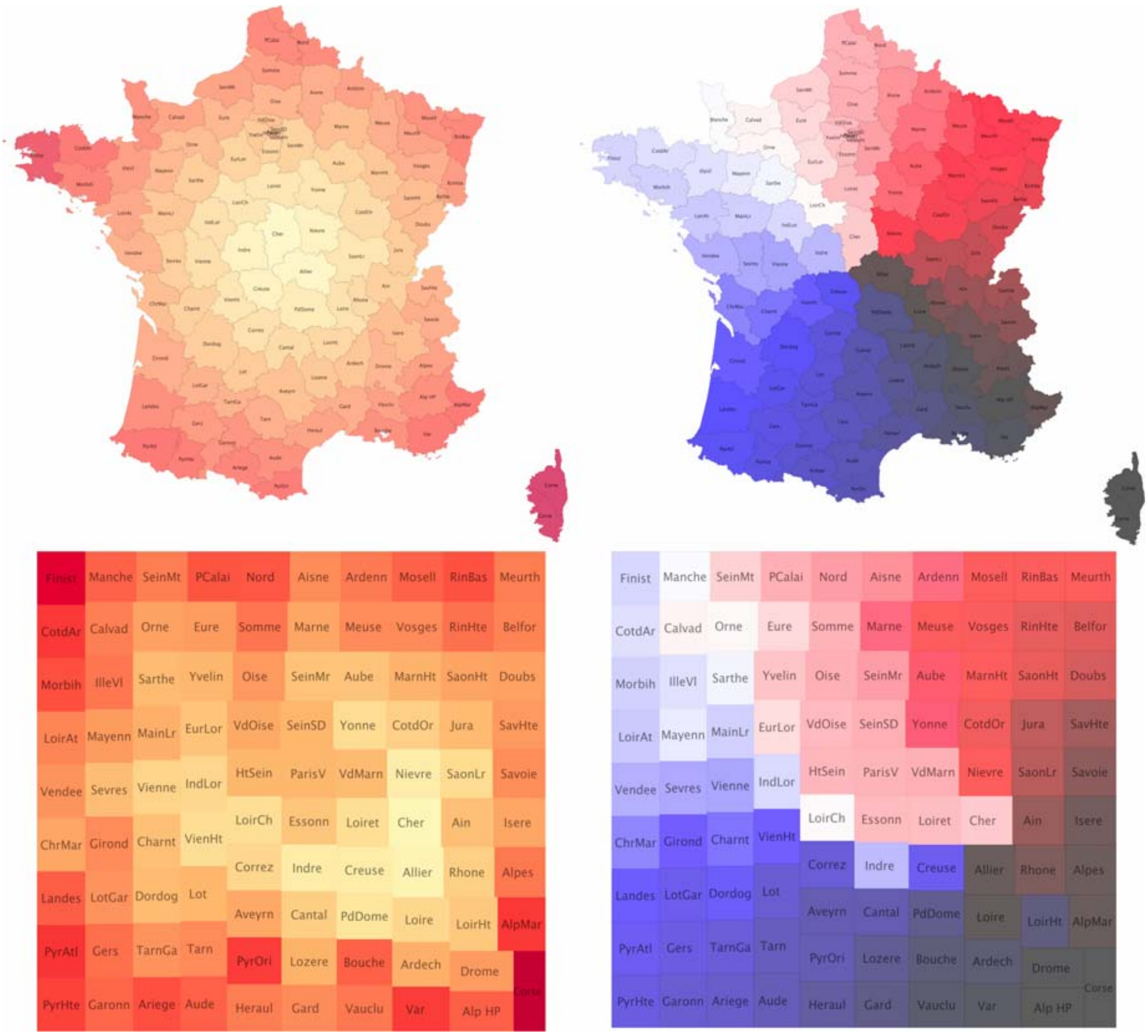


Figure 6. *Spatial treemaps* of the departments of France. Left images show departments coloured by distance from the French centroid, right shows direction from the centroid.

The result is a layout that approximately preserves the original location of items represented as nodes (see Figure 6 where colouring is used in this instance to demonstrate the spatial distortion that results from the transformation).

Effectively, the spatial treemap becomes a cartogram where node size can be used to represent any measurement scale variable. Figure 7 shows how the hierarchical structure of postcodes could be used as a basis for cartogram mapping of settled areas.

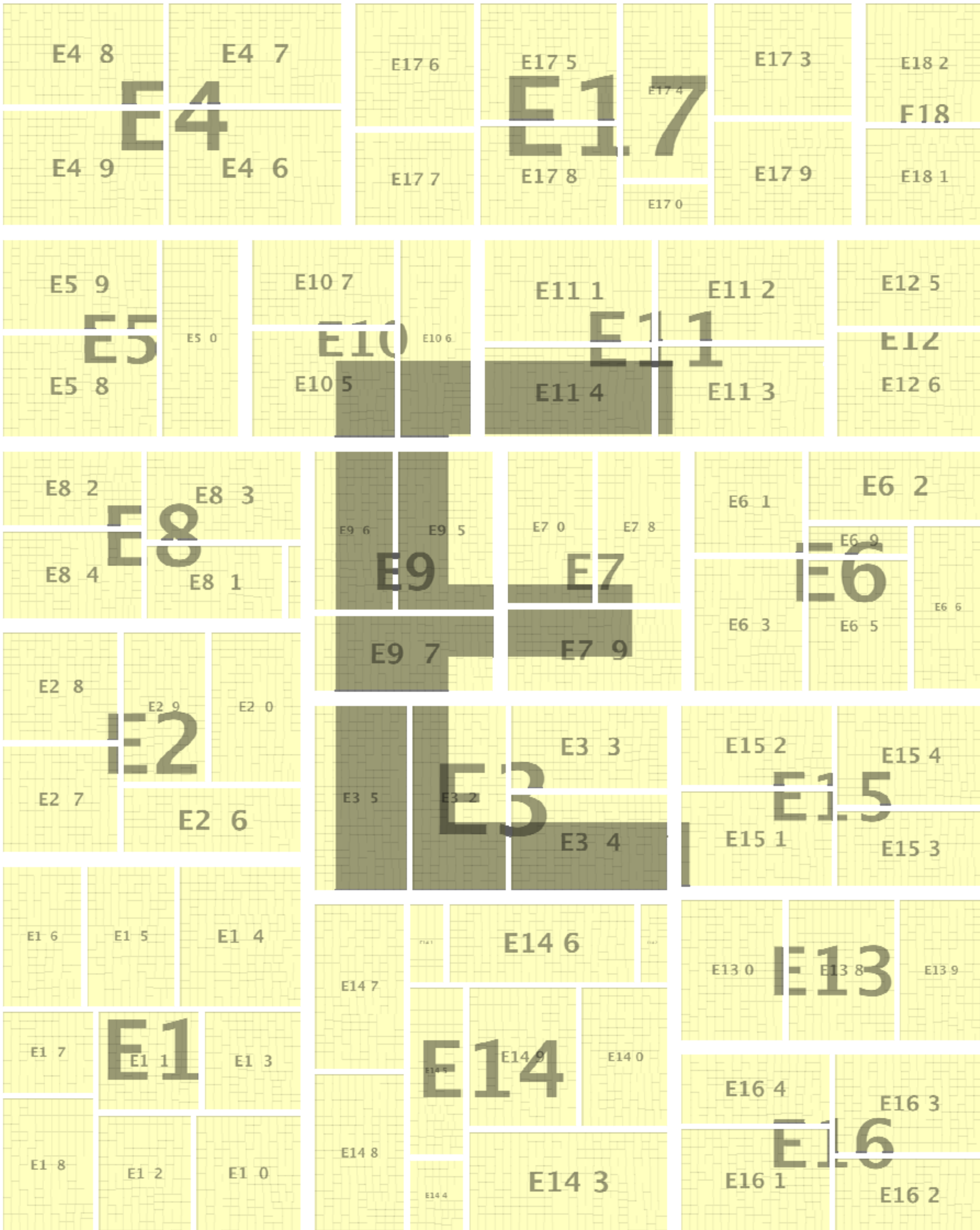


Figure 7. East London postcodes mapped using the s*patial* layout. Colour and/or size could be mapped to each node representing two variables associated with each unit postcode. *Postcode centroids crown copyright/database right 2008. An Ordnance Survey/EDINA supplied service*

4. Conclusions and Further Work

We have proposed three new enhancements of the treemap generation process. Evolutionary colour allocation allows hierarchical structure to be emphasised. Ordered squarified layouts provide a more consistent mapping of node magnitude to layout, thus making the interpretation of treemaps easier. Spatial treemaps allow geographic data to be mapped efficiently, retaining hierarchical structure as well as approximate geographic location and mapping node size to some other measurement variable.

Further work is being carried out on improved colour spaces, on metrics to quantify space and topology distortion, as well as various applications (insurance risk, Slingsby *et al*, submitted; photographic information retrieval, Dykes *et al* submitted; and mountain feature mapping).

References

Bederson, B., Shneiderman, B., and Wattenberg, M. (2002). Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies. ACM Transactions on Graphics, **21(4)**, pp. 833-854.

Baehrecke, E., Dang, N., Babaria, K., and Shneiderman, B. (2004). Visualization and analysis of microarray and gene ontology data with treemaps. *BMC Bioinformatics*, **5(84)**

Bertin, J. (1983). *Semiology of graphics*. University of Wisconsin Press.

Bruls, M., Huizing, K. and Wijk, J. V. (2000). Squarified Treemaps. www.win.tue.nl/~vanwijk/stm.pdf (accessed 20th December, 2007).

Dykes, J., Purves, R., Edwardes, A. and Wood, J. (submitted) Exploring volunteered geographic information to describe place: Visualization of the 'Geograph British Isles' collection. Submitted to GISRUK, 2008

Shneiderman, B. (1992). Tree visualization with tree-maps: 2-d space-filling approach. ACM Transactions on Graphics, **11(1)**, pp.92-99.

Slingsby, A., Wood, J., Dykes, J., Foote, M. and Blom, M. (submitted) The visual exploration of insurance data in Google Earth. Submitted to GISRUK 2008.

SmartMoney.com (2007) A map of the Market, *www.smartmoney.com/marketmap/* (accessed 21st December, 2007)

Van Wijk, J. (2002) SequoiaView, w3.win.tue.nl/nl/onderzoek/onderzoek_informatica/visualization/sequoiaview (accessed 21st December, 2007)

Van Wijk, J., and Van de Wetering, H. (1999). Cushion treemaps: visualization of hierarchical information. *Information Visualization*, (Proceedings. 1999 IEEE Symposium Information Visualization), pp.73-78.

Biography

Jo Wood and Jason Dykes are senior lecturers in geographic information at the giCentre, City University London with interests in developing and using interactive graphics to explore complex structured data sets. Jo is particularly interested in analysing surfaces and hierarchies in applications ranging from geomorphometry to insurance risk. He is the author of the LandSerf GIS and other software packages in Java. Jason has particular interest in the development and evaluation of geovisualization applications and the statistical analysis of spatial pattern.

Mediating geovisualization to potential users and prototyping a geovisualization application

David Lloyd[1], Jason Dykes[1], Robert Radburn[2]
[1]School of Informatics, Room A520, City University, Northampton Square, London, EC1V 0HB
Telephone: +44 (0)20 7040 0212, Fax: +44 (0)20 7040 8584
Email: d.lloyd@soi.city.ac.uk, Web: www.soi.city.ac.uk/~at775

[2]Policy, Research & Information, Chief Executive's Department,
Leicestershire County Council, County Hall, Glenfield, Leicestershire, LE3 8RA

KEYWORDS: geovisualization, card sort, user-sketching, prototyping, crime and disorder reduction

1. Introduction

After establishing user requirements (Lloyd, Dykes, Radburn, 2007) human centred techniques can be used to progressively prototype geovisualization applications. User evaluation has been limited in visualization (Ellis and Dix, 2006) and a handful of studies focus on geovisualization users at the design stages (e.g. Ahonen-Rainio and Kraak, 2005). This paper outlines work to mediate geovisualization tools/interactions to potential users and shows how HCI prototyping techniques can aid the development of a geovisualization application. Our case study involving the crime reduction researchers in Leicestershire County Council (LCC) gives insight into which prototyping techniques work and which need modifying because of the unique characteristics of using interactive computer maps as interfaces to information exploration.

2. Approach

2.1 Mediating geovisualization to potential users

Card sorting and *user sketching* are techniques that can be deployed for establishing user insight into design ideas (Faiks and Hyland, 2000; Tohidi el, 2006). We tested these HCI techniques in a geovisualization context by providing potential users ("users") with a tutorial on candidate geovisualization tools and interactions. The material was selected from the literature given the divergence in suggestions made during interviews with geovisualization experts (Lloyd, Dykes, Radburn, 2007).

2.1.1 Card sorting

Card sorting is frequently used at the beginning of the prototyping process to "discover users' mental model of an information space" (Nielsen and Sano, 1995). The technique was evaluated to investigate three factors. *Firstly*, to identify plausible crime research tasks the users might perform with geovisualization, grouping these according to the users' notions of similarity to identify those of greatest importance. Results were consolidated and used as the basis for specific crime tasks in the sketching exercise (see 2.1.2). *Secondly*, to identify how well potentially useful geovisualization techniques could be elicited. Users grouped the tutorial techniques according to their helpfulness in exploratory research. *Thirdly*, to understand how well card sorting could establish the extent of user engagement and understanding of geovisualization (Stein, Baxter and Leinhardt, 1990). The users were asked to sort the geovisualization techniques freely (Figure 1). The results were analysed through hierarchical clustering (Dong, Martin & Waldo, 2001).

Figure 1. Free card-sorting of geovisualization tools and interactions

2.1.2 Sketching

Enabling users to sketch their ideas can facilitate reflection and encourage deeper interpretation and analysis in human-centred design (Tohidi et al., 2006). To test this HCI approach in the context of geovisualization, users were asked to sketch ideas for specific crime tasks identified from the first card sort (Figure 2). Seven sketches were produced by three people in 30 minutes. The number of elements in each design corresponding to the tutorial was counted. A count was also made of elements not included in the tutorial, and the results analysed by crime task. The sketching counts results and hierarchical card sort analysis were synthesized to quantify the results derived from these methods.

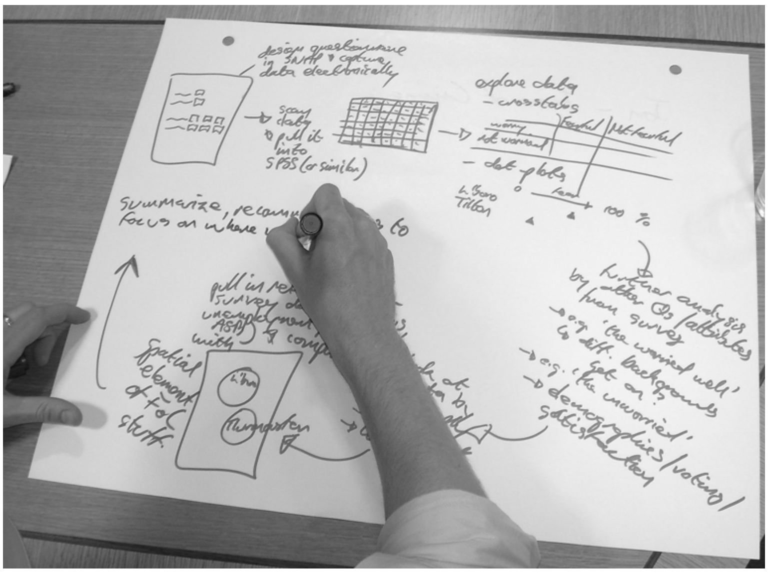


Figure 2. User in course of sketching a "Fear of crime" application

2.2 Low-level prototyping

Low-level prototypes are useful in human-centred design as they are simple, cheap and quick to produce (Preece, Rogers, Sharp, 2002) and help with mediation between user and developer. Various techniques exist and were selected according to audience, stage of development, speed, longevity, expression, style, medium and fidelity (Arnowitz, Argent, Berger, 2007). Wireframe prototyping (Tullis, 1998), paper prototyping (Rettig, 1994) and digital interactive prototyping (Arnowitz, Argent, Berger, 2007) are appropriate low-level prototyping techniques to test in a geovisualization context in the early stages of development.

2.2.1 Wireframe prototyping

Two outline designs for a geovisualization application were developed from the mediation exercises and from our user understanding (Lloyd, Dykes, Radburn, 2007). These were shown to users in the form of *wireframe* prototypes that explored subjective reactions (Tullis, 1998). Wireframe prototypes are “intended to provide an early approximation to a software idea” (Arnowitz, Argent, Berger, 2007). The wireframes were created quickly and with minimum resources using office software and dummy data (Figure 3).

Individual interviews were conducted where the wireframes’ scope, structure, layout and interactions were explained. A “think aloud” protocol (van Someren et al., 1994) was employed with pre-generated questions to prompt user responses about the designs, their relevance and potential. Sessions were audio recorded for later transcription and analysis. Each user was asked which of the two wireframes was preferred, and that shown in Figure 3 was favoured by the majority.

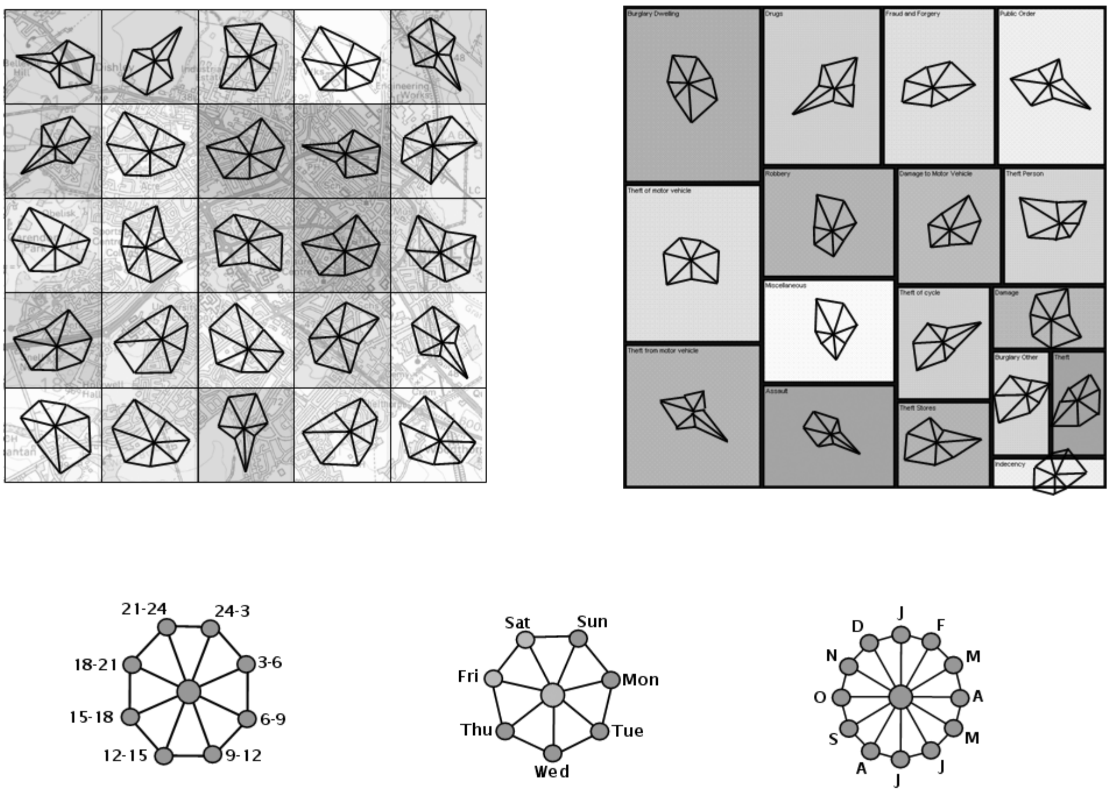


Figure 3. The wireframe prototype of the chosen application design combining thematic map, crime treemap and temporal glyphs

2.2.2 Paper and Digital Interactive prototyping

Typically, a more realistic prototype is developed from a wireframe before committing resource and emotion to writing code (Cohen et al, 2004). We used two prototyping approaches to compare their performance in the geovisualization context. Both used paper to emphasize the prototypical nature of the applications, one consisting of pre-printed maps, graphics and input sheet – the "paper prototype" (Figure 4). The other used a paper input sheet to control a screen with a far wider selection of digital maps – the "digital interactive prototype" (Figure 5). Both methods involve users performing real tasks with a rudimentary interface. A paper prototype (Rettig, 1994) is manipulated by someone 'playing computer' (Snyder, 2003). A digital interactive prototype is "similar to a paper prototype ... [that] can range from a more narrative style…to a fully interactive high fidelity coded prototype" (Arnowitz, Arent, Berger, 2007).

With paper prototyping, "complex or subtle interaction usually can't be simulated perfectly" (Snyder, 2003), but the technique does have the potential to show a range of functionality – for example, our paper prototype included a treemap that could not be included in the digital interactive prototype within the same development time.

Our experience with the wireframes indicated that real data must be used in these low fidelity/low cost prototypes and indicated the importance of short but realistic tasks. Both prototypes used a six year data series from an LCC database containing 42,000 crimes for a 5km square urban area, but employed different high-volume crime types to avoid a learning effect ("Assault" for paper; "All Other Theft" for the digital interactive prototype). Approximately the same amount of time was taken to develop both prototypes.

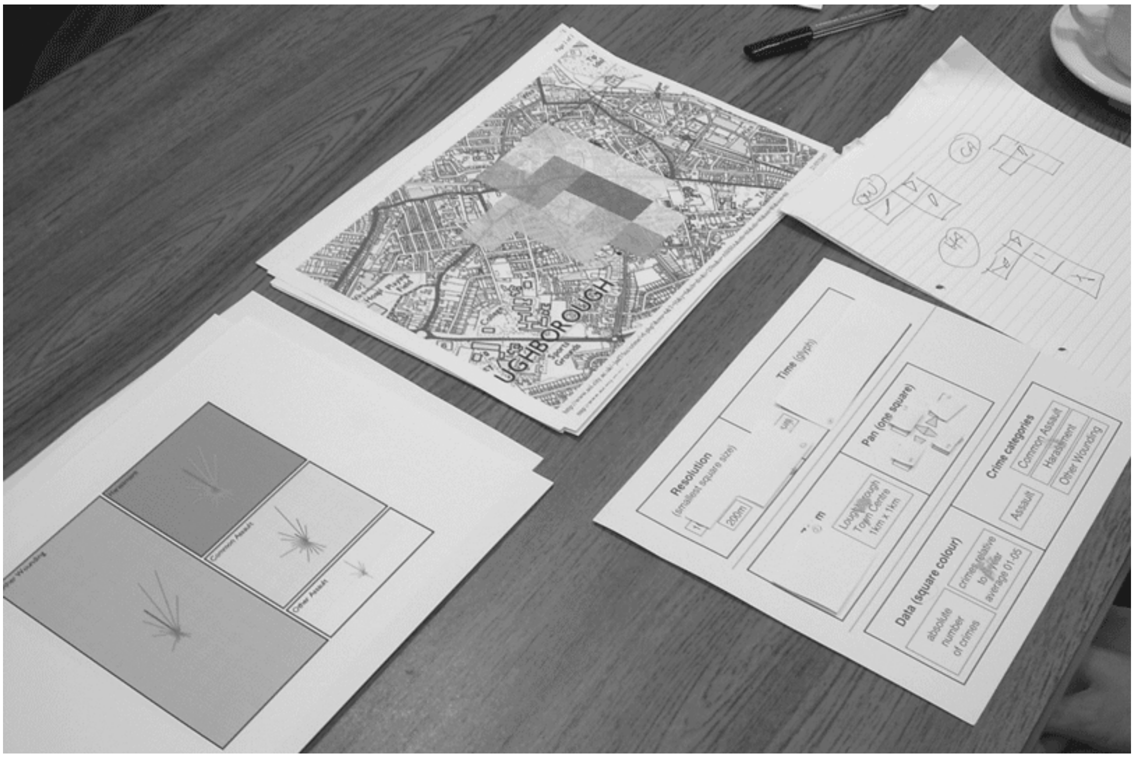


Figure 4. Paper prototype during user testing showing (left to right) treemap with temporal glyphs, thematic map with temporal glyphs and the input sheet for user interface

Both prototypes were tested on individual LCC researchers to explore the functionality/flexibility trade-off when comparing paper and digital prototyping. Sessions lasted 1½-2 hours to avoid overtiring users, with prototype order randomised. Task order was consistent with successive tasks increasing in complexity; tasks followed the taxonomy of Koua, MacEachren and Kraak (2006). Users were asked to "think aloud" during tasks (van Someren, Barnard, and Sandberg, 1994; Marsh, Dykes, and Attilakou, 2006) and were questioned after each task to determine both the extent of their ideation and their experiences of working with the prototype. Tasks were audio recorded and transcribed, the text coded and conceptualised using qualitative data analysis software (Muhr, 2004) to identify usability issues, evidence of how the prototypes performed in a geovisualization context and draw comparisons between them.

Figure 5. Digital interactive prototype showing thematic map and temporal glyphs during potential user testing. The paper input sheet is not shown. Note the absence of treemaps.

3. Findings

3.1 Card sorting

Card sorting successfully identified crime task clusters that formed the basis of user sketching and for advancing the application design. Where users sorted geovisualization techniques by usefulness in crime data exploration, the three users found 23, 27 and 29 (out of 29) "most helpful". This – plus user comment during the card sort - suggests the tutorial did not give users sufficient understanding to apply geovisualization techniques effectively to their exploratory work. However the card sort successfully identified this issue. The "free" card sort of geovisualization techniques produced a subjectively logical, high-level classification using hierarchical cluster analysis indicating that users had grasped a wider geovisualization tool/interaction taxonomy. Clusters corresponded to "mapping tools" (11 tools/interactions in cluster), "advanced/new data analysis" (9), "established data analysis" (5) and "interactions" (4).

3.2 Sketching

Seven user sketches of three crime tasks produced 78 elements of which 30 were from the 29 tools/interactions of the geovisualization tutorial. Of these, an average 6.3 elements were found on "night-time economy" sketches compared to 3.5 and 2.0 on sketches relating to other crimes indicating that "night-time economy" represented a good candidate task for a geovisualization application.

Ten of the 29 geovisualization techniques were absent from user sketches, eleven were used once and six were mentioned twice. "Star plot/glyph" appeared three times; "line graphs" four times. Sketching appears to be a useful and quick way to determine potentially acceptable techniques with users.

Synthesising the results of the high level geovisualization tool/interaction clustering with the element count in user sketching yielded an average element count of 1.50 (interactions); 1.4 (established data analysis); 1.22 (advanced/new data analysis) and 0.55 (mapping tools). The relatively low number for "mapping tools" indicates the importance of non-spatial tools to these users currently.

3.3 Wireframes

Design was strongly led by data considerations and during the prototyping, users expected to see patterns that they knew existed in the data. They were puzzled by their absence in the wireframe prototypes and generally did not engage with them until the "discrepancies" had been resolved. A key finding therefore is that the use of real user data is important to attract user engagement in geovisualization prototyping. The complex graphically dense nature of geovisualization interfaces meant that even the data-free wireframes took a significant amount of time to construct in office software. The wireframes contained many individual elements and envisaged interactions and coordinated multiple views. It was time-consuming to explain these possibilities to users. It is possible that along with the "finished" look of the wireframe, this might have inhibited user criticism.

3.4 Paper and Digital Interactive prototyping

In both cases users formed hypotheses based on the information in the underlying map image. Users had strong expectations about the patterns they observed in the prototypes and the importance of real data was reinforced many times. Because of the use of real data in the paper prototype, it did not have the flexibility associated with the medium – certainly new, realistic "screens" could not be generated "on the fly" and this limited exploration. The intense engagement of users with both prototypes, their willingness to hypothesise about cause and effect, and the ideation that occurred throughout the prototyping sessions were striking. This contrasted significantly with the data-free wireframing.

Constraints on the ability to show interactivity across multiple views with the paper and digital interactive prototypes limit the applicability of results to date to geovisualization. This key aspect of the prototyping process will be addressed when a fully coded prototype is specified, built and tested on users and results compared. Work to date is based upon a case study of a small numbers of users in one domain. The work will be repeated in another domain to draw comparisons between the two cases and to determine whether findings can be generalized.

4. Conclusions

This paper details the use of human-centred prototyping techniques to progress the development of a geovisualization application and will be of use to those in the geovisualization community who seek ways in which to create applications that meet the needs of real users.

We found card sorting to be a useful technique for task specification and the verification of task credibility, but less effective for establishing user understanding in the context of a multifaceted subject such as geovisualization. The task of mediation between geovisualization expert and domain expert is relatively difficult (Lloyd, Dykes, Radburn, 2007) and so card sorting and sketching may be beneficial.

The positive results from employing a user-centred perspective include high levels of user engagement and evidence of ideation. Our experience suggests that prototyping is a route that geovisualizers could employ successfully to build applications fitted to user needs. It is important to use real data with which expert users are familiar in geovisualization prototyping, both to gain engagement with users, and to help them learn about the nature of geovisualization (exploration and ideation) and geovisualization techniques. The use of paper and digital interactive prototypes with real data seems to help overcome the knowledge gap identified above and to develop and establish user understanding of geovisualization techniques. However, the development of sophisticated prototypes contradicts many of the objectives of low-fidelity prototyping such as speed of deployment, low cost, short longevity, and low fidelity (Arnowitz, Arent, Berger, 2007). Digital interactive prototypes that use high level scripting techniques to provide access to real data may offer a solution in the context of geovisualization and play an important role in the process of human-centred geovisualization design. We may infer that human-centred geovisualization design should expect less of users and developers at the outset than is the case in other forms of human-centred design and rely upon low-level digital prototyping to develop applications and mutual understanding. Equally, we should not underestimate the importance of developing relationships between geovisualization experts and domain experts. In our experience, a series of human-centered activities has developed knowledge and trust in all parties and resulted in much reflection that will have a positive impact upon practice.

5. Acknowledgements

This work is supported by EPRSC and Leicestershire County Council through industrial CASE award EP/C547438/1. The help and cooperation of the Crime and Disorder Reduction Team at LCC is acknowledged with gratitude.

References

Ahonen-Rainio P and Kraak M-J (2005) Deciding on Fitness for Use: Evaluating the Utility of Sample Maps as an Element of Geospatial Metadata *Cartography and Geographic Information Science* **32(2)** pp101-12.

Arnowitz J, Arent M and Berger N (2006) *Effective Prototyping for Software Makers.* Morgan Kaufmann, San Francisco.

Cohen CF, Birkin SJ, Garfield MJ and Webb HW (2004), Managing conflict in software testing *Communications of the ACM*, **47(1)** pp76-81.

Dong J, Martin S & Waldo P (2001) *A user input and analysis tool for information architecture* paper presented to Conference on Human Factors in Computing Systems (CHI 2001) Seattle, WA, USA.

Ellis G and Dix A (2006) *An explorative analysis of user evaluation studies in information visualisation* paper presented to Beyond Time and Errors: Novel Evaluation Methods For Information Visualization (BELIV '06), Venice, Italy.

Faiks A and Hyland N (2000)'Gaining user insight: a case study illustrating the card sort technique *College & Research Libraries* **61** pp349-57.

Koua EL, MacEachren A and Kraak M-J (2006) Evaluating the usability of visualization methods in an exploratory geovisualization environment *International Journal of Geographical Information Science* **2(4)** pp425-48.

Lloyd D, Dykes J and Radburn R (2007) *Understanding geovisualization users and their requirements – a user-centred approach* paper presented to GIS Research UK 15th Annual Conference, National Center for Geocomputation (NCG), Maynooth, Ireland.

Marsh SL, Dykes J and Attilakou F (2006) *Evaluating a Geovisualization Prototype with Two approaches: Remote Instructional vs. Face-to-Face Exploratory*, paper presented to Information Visualization 2006, London, England.

Muhr T (2004) *User's Manual for ATLAS.ti 5.0* ATLAS.ti Scientific Software Development GmbH, Berlin.

Nielsen J and Sano D (1995) SunWeb: user interface design for Sun Microsystem's internal Web *Computer Networks and ISDN Systems* **28(1-2)** pp179-88.

Preece J, Rogers Y and Sharp H (2002) *Interaction Design: beyond human-computer interaction*, Wiley, Chichester.

Rettig M (1994) Prototyping for tiny fingers *Communications of the ACM* **37(4)** pp21-7.

Snyder C (2003) *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces.* Morgan Kaufmann, San Francisco.

Stein MK, Baxter JA and Leinhardt G (1990) Subject-Matter Knowledge and Elementary Instruction: A Case from Functions and Graphing *American Educational Research Journal* **27(4)** pp639-63.

Tohidi M, Buxton W, Baecker R and Sellen A (2006) *User sketches: a quick, inexpensive, and effective way to elicit more reflective user feedback*, paper presented to 4th Nordic conference on Human-computer interaction: changing roles, Oslo, Norway.

Tullis T (1998) *A method for evaluating Web page design concepts*, paper presented to Human Factors in Computing Systems CHI 98, Los Angeles, CA, USA.

van Someren MW, Barnard YF and Sandberg JAC (1994), *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes.* Academic Press, London.

Biography

David Lloyd is a third year PhD student researching the use of human-centred techniques in geovisualization at City University London in collaboration with Leicestershire County Council.

Jason Dykes is a Senior Lecturer in Geographic Information at City University London with interests in geovisualization. He is co-chair of the ICA Commission on Geovisualization, has developed a number of geovisualization software applications and is co-editor of Exploring Geovisualization Dykes, MacEachren and Kraak (2005).

Robert Radburn is a Senior Research Officer at Leicestershire County Council working with a variety of partners and agencies to support evidence-based policy. He is a champion of the use of information visualization and geovisualization techniques to inform service delivery within regional and local authorities.

The Effectiveness of Artificial Landmarks for Orientation Tasks within a Large-Scale Virtual Environment

Adam Rousell, Claire Jarvis, Jing Li

Department of Geography, University of Leicester UK
TEL: +44 (0)116 252 3808 FAX: +44 (0)116 252 3458
Email: chj2@le.ac.uk
KEYWORDS: Landmarks, Visualisation, Virtual Reality, Virtual Environment, Orientation

1 Introduction

A primary use of virtual environments is to allow people to explore and interact with an environment that they either do not have access to, or as a training facility for tasks that they will perform in the real world. In many of these uses, it is essential that the user can successfully navigate this environment and to do this, they need to know where they are (both location & orientation). Further to this problem, it is also important that users can successfully perceive distances and scales to allow them to perform these tasks accurately. In Figure 1, we outline a number of scenarios where difficulties in navigating and spatial perception within a virtual environment could compromise their effectiveness in a variety of applications, including those related to teaching and learning and public engagement more generally.

Teaching vignette A

You are aware that students have a tendency to 'switch off' on coach journeys between stops on orientation days in the field. However, it is really important that they gain a sense of direction relative to the coast and mountain ranges in order to understand particular landforms and processes at different locations in your field area. As one measure to try and overcome this difficulty, you build an immersive virtual environment to enhance the students' sense of scale and location before they go into the field. However, it soon becomes clear that students find it exceedingly difficult to locate themselves in this environment relative to a map and further that difficulties estimating distances in the VE are contributing to this problem.

Teaching vignette B

You wish to demonstrate a real-time virtual environment in a public participatory context, looking at the potential impact of wind turbines both in terms of visual and sound impact in a natural landscape. You would like groups of the "local public" to be able to place potential wind farms in the region and debate their impact in relation to places of particular emotional significance for them in the area. However, for this process to work, the participants must be able to find both locations of personal significance and to place the wind turbines in the landscape. Both prove much more difficult than you expected, as participants keep getting "lost" in the VE.

Teaching vignette C

You wish to teach students to sample particle sizes in a manner more lively than the traditional statistics lecture, while reserving precious time in the field for undertaking research as opposed to training in sampling. You decide instead to develop an immersive environment that can represent terrain based on underlying GIS data layers in a realistic way, in which students can point to features and include them in a virtual collection bag for later analysis. To your horror, you find that while students enjoy "collecting" virtual data samples in this way, their sense of scale as regards distance and height in the VE means that they find collecting a representative sample genuinely difficult.

Figure 1. Examples of complex space-time communication issues in a geomorphological teaching context

There has been extensive research into the cognitive processes of navigation and orientation, and how different features aid in the construction of cognitive maps. However, one of the main problems discovered is that people do not often perceive many features in the Virtual Environment (VE) as they would in the real world. They tend to get disorientated easily (Wiener 2007) and lack the spatial judgement that is easily perceived in real environments when calculating distances (Interrante et al. 2007) and avoiding obstacles (Suma *et al.* 2007). For navigating large scale virtual environments it is essential that the use can orientate themselves accurately (Juan-Espinosa *et al*. 2000).

From a geographer's perspective, the use of a top-down overview map as an inset is a common orientation device and a potential means of assisting the user to navigate in a broader landscape context. Such an overview map is implemented in the Google Earth interface, and indeed we use a very traditional example below in Figure 1. However, Sjölinder *et al.* (2005) suggest that the use of an overview map is of little use when learning an environment as it aids in the construction of route knowledge rather than survey. As the information is directly presented to the user they tend not to learn that information, and so using an overview map in an application aimed at aiding in the learning of an environment ready for visiting that location seems of little benefit. It has also been found that in other applications that require navigation, such as alternative route determination, survey knowledge presented in the form of an overview map is generally not used and route knowledge is used instead (Janzen *et al.* 2001).

The research presented here assesses whether introducing artificial landmarks into a large scale environment with little in the means of natural or manmade landmarks aids in orientation tasks. Further research is carried out in the form of dynamically colouring the landmarks depending on the distance from the user's location to ascertain whether colour can be used as an adequate distance scale.

2 Background and Methodology

2.1 Study Region

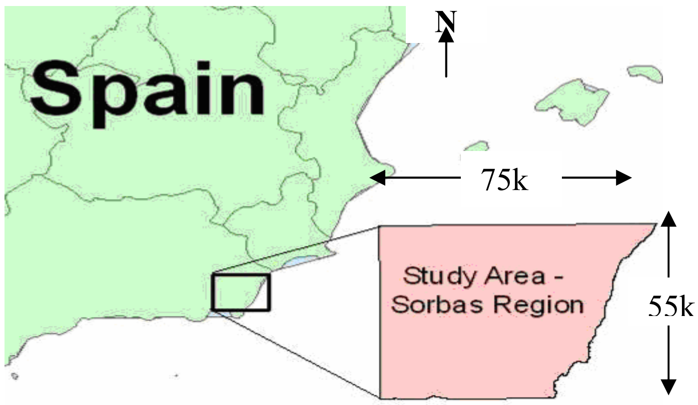


Figure 2 - Study area

The study uses a virtual representation of the Sorbas region in South-Eastern Spain, a relatively unpopulated inland area with a single coast line in the confines of this study area (Figure 2). The location (lower left corner) of the study area was at 542718.3 Easting and 4076991.2 Northing (using a WGS 1984 Complex UTM Zone 30N projection), and was 75 km wide by 55 km tall. This region was selected so that the VE could have a future use as a training environment for geography students at the

University of Leicester; each year, a number of students undertake field studies in the Tabernas and Sorbas regions.

2.2 Construction of the Virtual Model

The software used to construct the environment for the trials to take place in was the Bionatics Blueberry 3D geometry engine in conjunction with Paradis VegaPrime. Blueberry 3D geometry was used to create the terrain and VegaPrime to add the functionality required for the interactive analysis and stereo projection to take place.

2.2.1 Interactivity and Landmarks

The environment was created in the Bionatics Blueberry software using various datasets to create a realistic representation of the environment. Further functionality was then added to this environment using VegaPrime to allow the user to navigate and to include the artificial landmarks. Code was also implemented to change the colour of the landmarks dynamically, as well as to allow the use of multiple viewports so that the images could be projected in stereo using specialist equipment.

2.3 User Trials

The user trials consisted of a brief fly around the environment by the user, before they were placed in nine different virtual locations without the ability to move forward and backwards within the scene. Once placed at these sites, the triallists were required to plot the location they believed themselves to be in at within the virtual environment on a paper map. This map comprised of the basic features within the environment such as buildings, roads, contour lines and the coastline. Urban features were also marked on the map which corresponded to the landmarks shown in the environment.
In this experiment, 15 trials were undertaken consisting of test subjects between the age of 21 and 42, and comprising of 11 males and 4 females.

2.3.1 Artificial Landmark Representations

To ascertain the effectiveness of using artificial landmarks to represent the towns, each test subject undertook nine trials at three different locations. One third of these trials had no artificial landmarks, another third had the artificial landmarks displayed but in one colour, and the final third had the same landmarks shown but they were dynamically coloured to represent distance. The order of locations in which users began the trials was altered to avoid geographical bias in the results.

2.3.1.1 No Markers

At the first three locations the user was placed in, no additional information was provided about the environment other than the visual terrain (Figure 3).

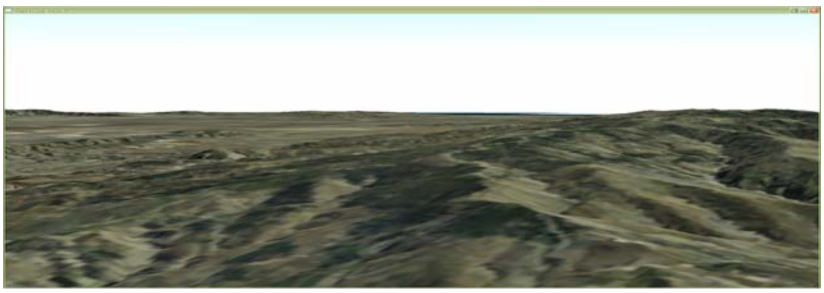


Figure 3: No markers

2.3.1.2 Plain Markers

At the next three locations, 10 markers coloured yellow were displayed in the form of a cone pointing down to an urbanisation. These markers were labelled to indicate the name of the particular town they were pointing towards, which also corresponded to towns marked on the paper map (Figure 4).

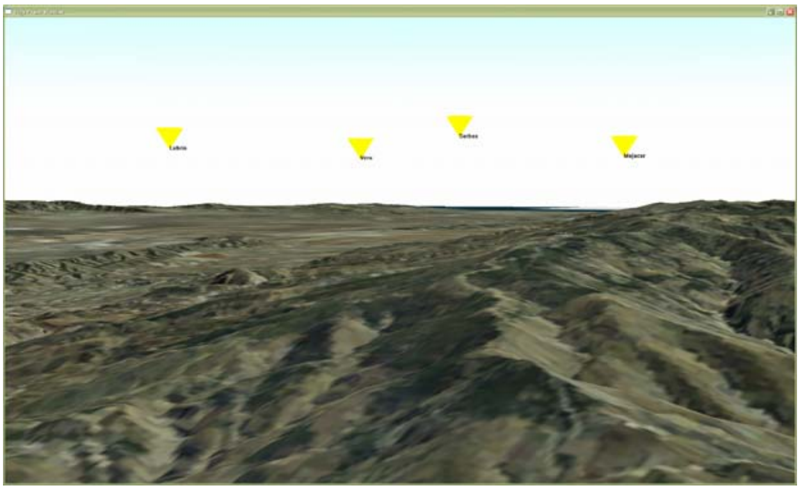


Figure 4: Plain markers

2.3.1.3 Dynamically Coloured Markers

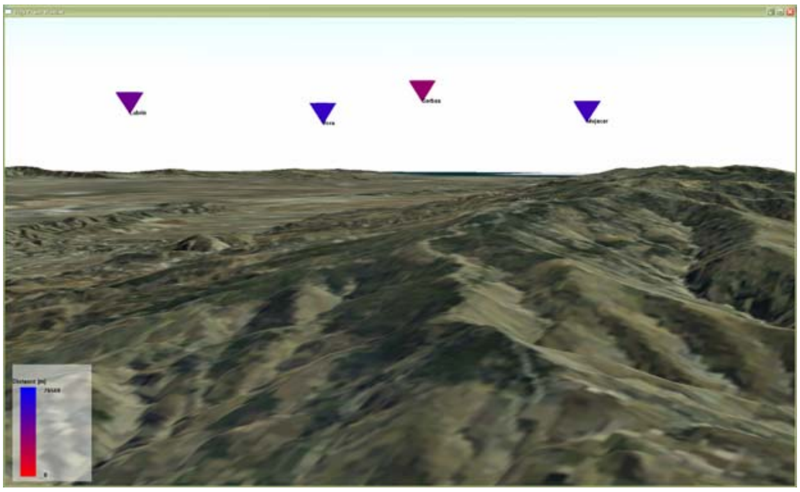


Figure 5: Dynamically coloured markers

Finally, the remaining three locations also had the same markers displayed, but here the markers were dynamically coloured depending on their distance from the user's location. If the marker was close to the user then it appeared red, and if it was far away it appeared blue. In between these extremes a graduation of colour was used going from the red through to magenta and then to blue. A legend was provided to the user on the display showing that this was the case (Figure 5).

2.3.2 Post Trial Feedback

Once all locations had been plotted on the paper map, the user was asked a series of questions about their 3D stereo landscape experience and to ascertain what methods they used to determine their orientation.

3. Results

The results presented here were gathered by measuring distance values on the paper maps between perceived locations as marked by the participants and their actual geographical locations. Additionally, aural feedback was collected from the participants both to provide additional qualitative indications as to the orientation techniques users employed in the study, as well as to provide support when analysing the quantitative findings.

Numerical results, in the form of distance error from the actual point, were measured (1 cm = 3.3 km) and analysed using the R statistical package. A t-test was performed between each of the groups of data obtained (no markers, plain markers and colour markers). The results from these trials are indicated in Table 1. The P- value from this test represents the probability of the two distributions in fact being the same distribution.

Table 1: t-test results

t-test	P value
No markers & plain markers	1.548e-05
No markers & coloured markers	7.443e-06
Plain & coloured markers	0.8167

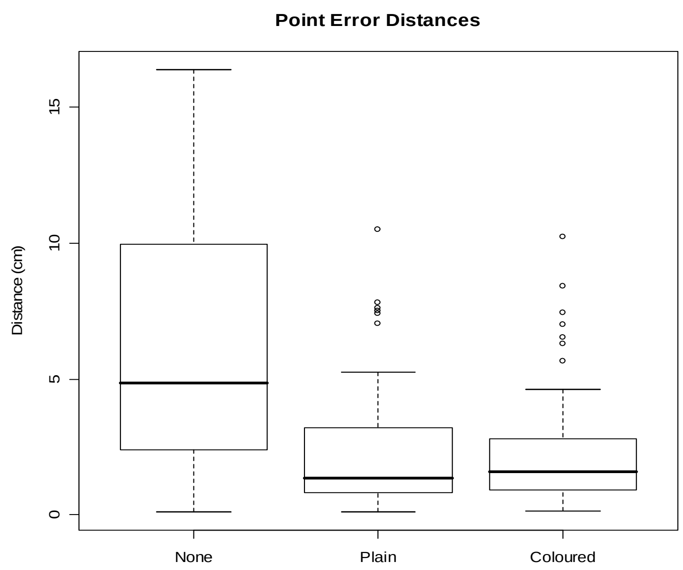


Figure 6: Box plot of error distances

From these results, it is evident that whilst both the plain and coloured markers show a distinctly statistically different distribution from the no-marker results, there is no statistical significance in the distribution of the results between the coloured and plain markers. This can be seen when viewing the box plots of each distribution, as shown in Figure 6

This suggests that in fact, knowing the distance from the landmark was not needed by the test subjects when determining their orientation within the environment.

Qualitative feedback suggested that the features used by participants for the orientation tasks were associated with the same representations as used in the Urban Image Theory, in that for larger scales, boundaries and zones (coastlines and terrain covers) are used for general locality, and then paths and nodes (roads) for more local determination. The artificial landmarks were used mainly as global features, owing to their method of implementation.

4 Discussion

From the statistical analysis of the distributions of the error, it has been found that whilst the addition of artificial landmarks into the environment greatly increased the accuracy of the orientation tasks, the further addition of distance information in the form of a colour scale had negligible impact.

A further feature found from the results was that the accuracy of orientation tasks greatly varies depending on the location. For example, trial locations that did not have any distinguishable visible features such as buildings or roads were much more inaccurate in terms of orientation tasks.

Owing to the similarity of results between the trials using plain markers and coloured markers, and post trial feedback given by the users, it was concluded that in fact knowing ones distance from the landmarks makes no difference to accuracy. Topological relationships between the landmarks are determined and this information then used to derive a user location.

However, the main item determined from this research was that in large scale virtual environments where there is little in the way of visible landmarks, it is essential that the user is provided with navigational assistance. Perhaps to a geographer, this may seem self evident, yet in the virtual reality literature much emphasis have been placed on the problem of spatial perception and less on potential solutions. This is particularly the case in regard to open and, to less an extent urban, landscapes; extent work largely focuses on internal environments.

The usefulness or otherwise of style of artificial landmarks presented in this paper, in particular their location in the sky, clearly has a scaled component related to longer distance navigation and a fly-through perspective. At shorter distances, selected man-made landmarks of a more familiar type such as signposts at road junctions are likely to prove of generic value while the highlighting of distinctive buildings may especially benefit those with prior local knowledge of an area. Equally, a mapped inset showing a small scale top-down overview of the area may be beneficial. While as we reported above, initial experiments using map insets by others have been less than successful, it may be that a spatial literacy component to their effectiveness will prove important and that familiarity with topographic mapping is a necessary pre-condition if such insets are to improve spatial navigation and wayfinding. Further research is underway to investigate these matters in the context of geographical, regional landscapes in particular.

Acknowledgements

Data was obtained from Instituto de Cartografía de Andalucía, which was gained access to from the University of Leicester. Acknowledgement is also deserved for NERC who funded my Postgraduate University studies.

5 References

Interrante V., Ries B., Lindquist J., and Anderson. L. 2007; Elucidating Factors that can Facilitate Veridical Spatial Perception in Immersive Virtual Environments; *IEEE Virtual Reality Conference 2007*; Charlotte, North Carolina, USA

Janzen G., Schade M., Katz S., and Heermann T. 2001; Strategies for detour finding in a virtual maze: the role of the visual perspective; *Journal of Environmental Psychology;* 21: 149-163.

Juan-Espinosa M., Abad F. J., Colom R., and Fernádez-Truchaud M. 2000; Individual differences in large-spaces orientation: g and beyond; *Personality and Individual Differences;* 29: 85-98.

Sjölinder M., Hook K., Nilsson L. G., and Andersson G. 2005; Age differences and the acquisition of spatial knowledge in a three-dimensional environment: Evaluating the use of an overview map as a navigation aid; *Int. J. Human-Computer Studies*; 63: 537-564.

Suma E. A., Babu S., and Hodges L. F. 2007; Comparison of Travel Techniques in a Complex, Multi-Level 3D Environment; *IEEE Virtual Reality Conference 2007*; Charlotte, North Carolina, USA

Wiener J. M. 2007; Can People Not Tell Left From Right in VR? Point-to-origin Studies Revealed Qualititive Errors in Visual Path Integration; *IEEE Virtual Reality Conference 2007*; Charlotte, North Carolina, USA

Biographies

Adam Rousell: *Having just completed my studies at postgraduate level of GIS (MSc), I am currently working as a web developer. My research interests include virtual reality and geographical visualisation, with the intention of doing a PhD in these fields.*

The Visual Exploration of Insurance Data in Google Earth

Aidan Slingsby[1], Jason Dykes[1], Jo Wood[1], Matthew Foote[2], Mike Blom[3]

[1]The giCentre, Department of Information Science, School of Computing, City University, Northampton Square, London, EC1V 0HB, UK
Tel. +44 (0)20 7040 8800
a.slingsby@soi.city.ac.uk; jwo@soi.city.ac.uk; jad7@soi.city.ac.uk

[2]Willis Research Network, Ten Trinity Square, London, EC3P 3AX, UK
footem@willis.com

[3]Willis Analytics, Ten Trinity Square, London, EC3P 3AX, UK
Tel. +44 (0)20 7488 8111
blomm@willis.com

KEYWORDS: insurance, geovisualization, interactive data exploration, Google Earth, KML

Introduction

The global insurance industry is concerned with the management of financial risk. The distribution and level of potential risk to an individual insurance company is dependent on its portfolio, comprising individual policies, each of which has its own specific terms, conditions and exceptions. In many cases, insurers will look to transfer some or all of the aggregate risk of these policies through a second level of insurance, termed reinsurance. In order to manage and spread their risk, individual insurance and reinsurance companies must each fully understand the composition and spatial distribution of their exposure (the specific objects that they insure) and the potential losses that they might incur. Insurance cover for losses incurred from catastrophic events is a particularly important part of the insurance industry. Catastrophic events and exposure are inherently spatial, yet the availability and use of spatial data and methods for assessing risk in the insurance industry has been inconsistent (BNSC/Infoterra, 2001). The potential for the use of such information is significant, particularly in the areas of underwriting where the location and characterisation of risk (resulting from events such as floods, earthquakes and windstorms) affects the pricing and rating of insurance and catastrophe risk assessment where models are used to estimate the frequency and severity of potential loss for particular portfolios.

Visualization and geovisualization techniques can both complement and help communicate the results of GIS and other analyses in the exploration of multivariate datasets (MacEachren et al 1999; Gahegan, 2005) and may provide insights and solutions for managing exposure and potential loss. Graphical techniques and the use of geobrowsers such as Google Earth[1] are also being used in a communicative role to engage a variety of different audiences within insurance companies with information about policies, exposure and potential losses (e.g. Lloyds, 2005). Such techniques are

[1] http://earth.google.com/ (we are using version 4.2).

being used by a huge and growing online community[2], have visualization applications in science[3] and also provide scope for geovisualization (Wood *et al,* 2007). In this paper, we focus on one particular geo-browser, Google Earth, which provides access to a rich array of datasets including aerial imagery, roads, administrative boundaries and photographs and, importantly, allows additional data to be added through the well-documented KML format.

Google Earth can be used in three interesting ways. Firstly, its easy to use KML input file format allows anyone to augment Google Earth's rich datasets with 'volunteered' geographical information (Goodchild, 2007). For example, it has been used effectively in relation to catastrophic events (Miller, 2006), widely demonstrating the power of combining datasets spatially to mainstream computer users. Secondly, Google Earth has the flexibility to support a wide range of visual encodings for exploring multivariate spatial datasets; for example, Pezanowski *et al* (2007) extended the user interface to support complex spatial querying. Slingsby *et al* (2007a) discuss how cartographic bias in Google Earth can be identified and, at least in some cases, overcome. Finally, its intuitive user interface and the streaming technology it uses to collect data from remote servers, allows spatial data to be disseminated and presented to wide groups of users.

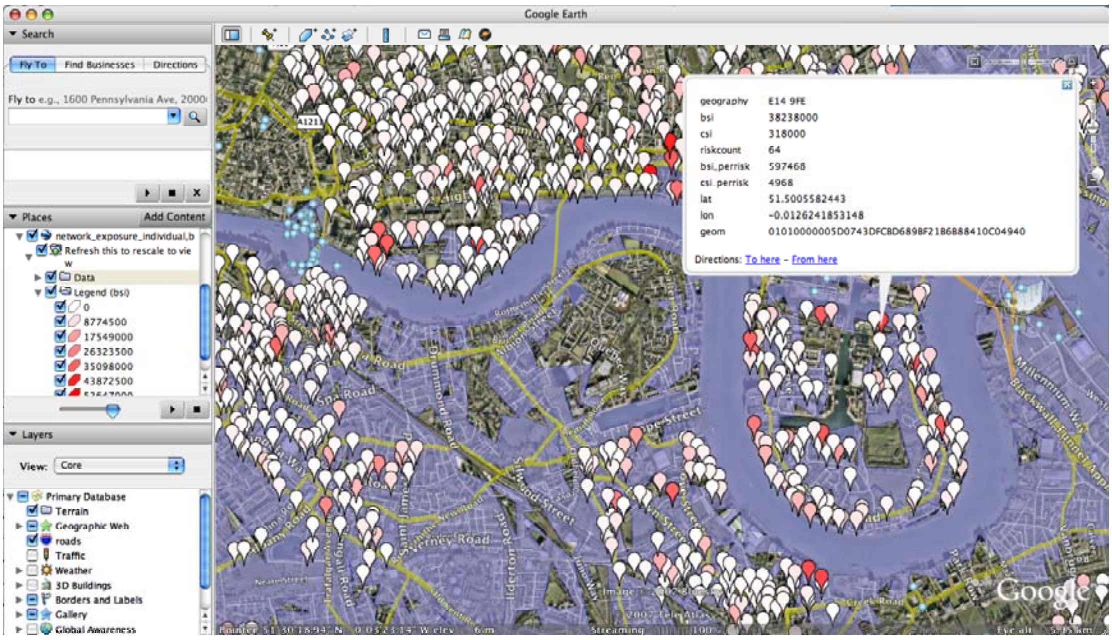


Figure 1: Exposure (plotted as point data) coloured by the building sum insured (see legend), shown with a flooding event with a 500 year return period. Data supplied by Willis Analytics (sums insured are representative only).

In this paper, we provide examples of how Google Earth can be used to interactively explore exposure, catastrophic events and potential loss information to answer questions of relevance to the insurance industry. We hope that some of the ideas presented here will give practical and useful ideas for individuals and organisations working in or with the insurance industry.

---

[2] For example, those listed at http://earth.google.com/gallery/index.html

[3] For example, NASA's Goddard Space Flight Center: http://svs.gsfc.nasa.gov/

Interactive mapping

Mapping

A simple but effective means for visually interpreting spatial data is to map it and compare with other spatially-referenced data. KML[4] is the XML-based markup language used to import spatial data into Google Earth. It allows data to be specified as `<Placemarks>`, each of which has a geometrical description (as points, lines and/or polygons) and a visual appearance (colour, transparency, line thickness, icon). Hypertext descriptions containing further information can be associated and are available on-demand from the Google Earth interface.

Flooding is an example of an event that may be strongly related to potential loss. Mapping exposure with flooding extents and inspecting data for individual data points on-demand may help portfolio managers evaluate geographical relationships between exposure and potential loss and communicate the complexity and spatial variation of this relationship to others (figure 1).

Filtering by attribute

The combination of Google Earth and KML provides the means to encode interactive behaviours for zooming and filtering operations that support Shneiderman's (1996) "visual information-seeking mantra: overview first, zoom and filter, then details-on-demand" (Wood *et al*, 2007). As shown in figure 1, KML can be used to specify details-on-demand behaviour through hypertext descriptions. KML can also be used to specify the filtering of data by attribute, by space and by time.
Filtering by attribute can be achieved by using KML to structure the data into a hierarchical set of `<Folders>` that reflect the attribute structure. These are displayed in the Google Earth "Places" panel, enabling users to turn on and turn off different levels of the hierarchy. Filtering by different combinations of attribute can be achieved by restructuring the hierarchy.

Filtering by space

Google Earth and KML can be used to allow a user to spatially select data and retrieve information based on this selection using the `<NetworkLink>` element. When the user selects data, Google Earth sends the coordinates of the user's selection back to the server as an HTTP request. These can be interpreted by a server-side script[5], which can retrieve spatially relevant results of analyses, encode the results as KML and then return this KML to the user. These may be the results of simple summary statistics or complex GIS or other analyses, run on demand or precomputed.
Figure 2 shows a simple example in which the user selects exposure data by framing it in Google Earth through zooming and panning. The coordinates of the selected area are sent through the `<NetworkLink>` to a server-side script that calculates the average, minimum and maximum building sum insured (BSI) and contents sum

---

[4] http://code.google.com/apis/kml/documentation/ (we are using KML version 2.1).

[5] Perl, ASP, JSP, PHP are suitable server-side scripts which are capable of connecting to databases and producing KML.

insured (CSI) for the selected area. These are returned as KML, and are displayed on the left in figure 2.

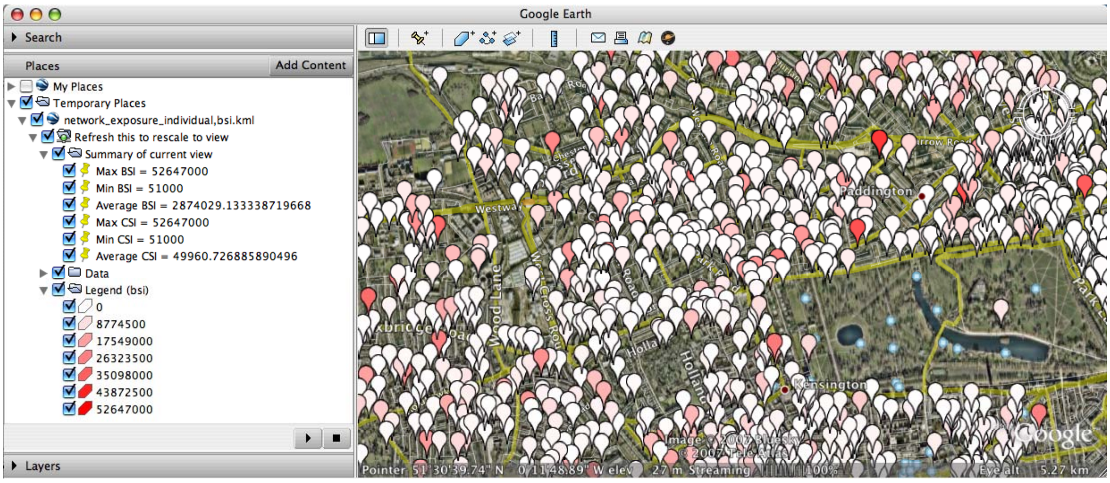


Figure 2: User selection of a spatially-constrained subset of data, and its subsequent summarisation (see statistical summary on the left).

Filtering by time

`<Placemarks>` can be mapped onto a timeline using the `<TimeSpan>` tag. Where `<TimeSpan>` tags are present in a KML file, Google Earth displays an appropriately scaled interactive timeline, enabling the user to filter the display of data by time. The timeline can also be "played" allowing temporal structure to be shown through animation. Figure 3 is an example developed and used by Lloyds, in which the timeline is used to show changes in the value of aggregate exposure over the course of a year. Such data can also be presented as symbols, as more appropriate area symbolism than perspective height or as continuous datasets (Wood and Dykes, 2008).

Figure 3: Changes in aggregate exposure by sum insured for flooding areas in 2006 (January, July and December) by UK flooding region (UK Environment agency). Source: Lloyds (used with permission).

Comparison with ancillary data

In addition to users' own data, Google Earth's built-in aerial imagery provides valuable geographical context for helping users understand the structure of the data in localised areas. In figure 4, the aerial imagery indicates that exposure is unevenly distributed within the postcode sectors shown. Ancillary datasets, such as ground-based photographs or information about local building characteristics, add essential context when the structure of aggregate data is visually analysed. Many useful datasets are provided centrally as Google Earth layers. Others layers may be added using KML.

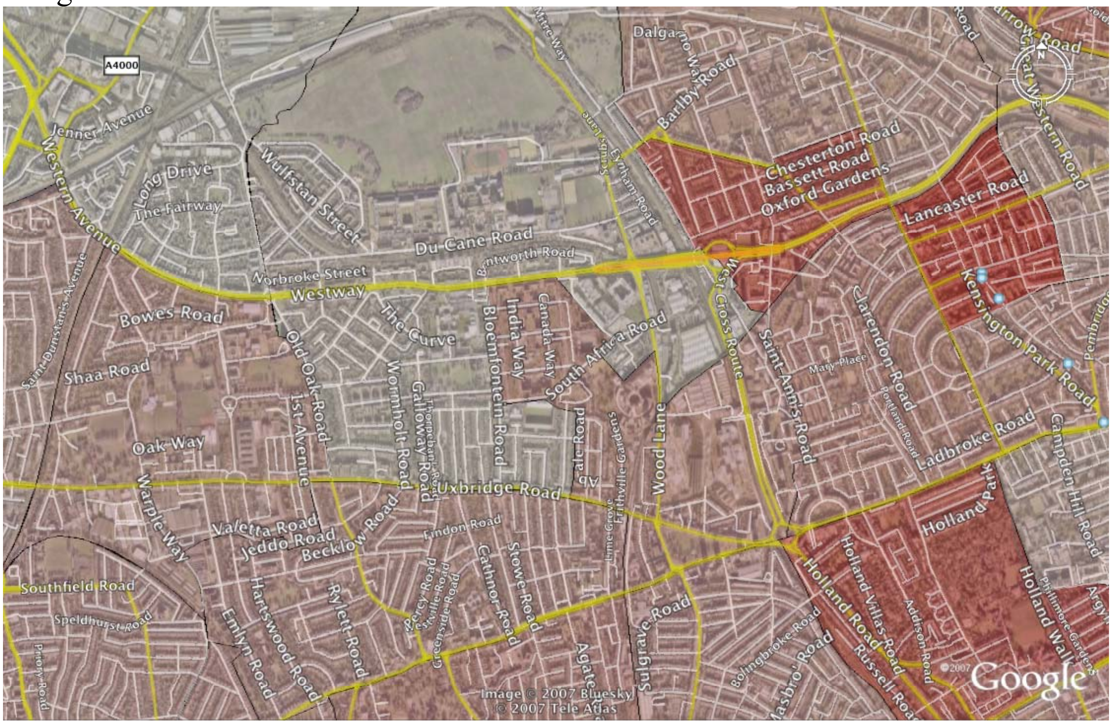


Figure 4: Postcode sectors coloured by average buildings sum insured (normalised by area). Visual comparison with aerial imagery shows that the postcode sector in the upper left is less densely built up than the other postcode sectors, partly accounting for its low aggregate insured value. Data supplied by Willis Analytics (sums insured are representative only).

Spatial aggregation

Patterns in spatial data are scale dependent (Openshaw et al, 1987). However, the aggregation of data also introduces the 'modifiable aerial unit problem (MAUP)', where the patterns may be dependent on the units into which points are aggregated, as illustrated in figure 4. The ability to explore spatial data at different spatial scales and units of aggregation is therefore important. Where the data are disaggregated, there is potential to study the stability of patterns across spatial scales.

Google Earth does not provide built-in spatial aggregating functionality (except for collocated point `<Placemarks>`); however, spatial aggregation can be implemented in a number of ways. It can be performed on demand using a server-side script through the `<NetworkLink>` as described in section 2.3. Alternatively, multiple aggregations can be precomputed and included in a single KML file grouped either into `<Folders>` or mapped to the timeline using the `<TimeSpan>` element. Figure 5 shows the former case, in which the user can select the required spatial aggregation in the "Places" panel. In the latter case, the timeline is used, not to interactively select by time, but to interactively select or "play" (animate) through all the different spatial aggregations.

Figure 5: Four different spatial scales of aggregation (exposure coloured by buildings sum insured) in four different layers. The user can select which layer (spatial aggregation) to view. Data supplied by Willis Analytics (sums insured are representative only).

New views of spatial data

The ability to explore, compare and relate multiple spatial and aspatial views of data are key to data analysis through geovisualization (Haslett *et al.,* 1990; Dykes, 1997, 1998; Andrienko and Andrienko, 1999; Slingsby *et al*., 2007b). KML can be used to describe aspatial and semi-spatial visual encodings and project them onto the surface of Earth. Here, we employ a rectangular cartogram (Krevelda and Speckmann, 2006) and a choropleth map for presenting potential loss by département in France. Potential loss is typically estimated using CAT models that take information on exposure and policies, and output the mean loss estimate and standard deviation resulting from multiple simulated catastrophic events (Grossi *et al*, 2005). Relating client portfolios to model output is a key component in managing exposure and loss. Considering multiple views with different emphases may help in this process.

Figure 6: Rectangular cartogram (squarified) showing CAT model output by French département, sized according to the average potential annual loss and coloured by the standard deviation (left); intermediate stage during transition (middle); map coloured by the standard deviation of annual loss by département (right). Data supplied by Willis Analytics.

Figure 6 (left) shows a squarified (Bruls *et al,* 2000) rectangular cartogram. Each rectangle represents a French département, sized by potential loss and coloured by the standard deviation estimated by a CAT model for a particular event type and a particular portfolio. Rectangles are as square as possible to enable relative sizes to be compared more easily and they are as close as possible to their true geographical position within the constraints of the squarified cartogram (Wood and Dykes, 2008). The cartogram shows that departments with larger potential losses have higher standard deviations around these losses, and that the highest potential losses and standard deviations tend to be in the North and West with the Alps as an outlier. The map in figure 6 (right) shows the geography most effectively and its relationship to the standard deviation (through colour) but does not show the magnitude of potential loss.

There is some evidence that animated transition is a useful way of relating multiple views of data (Heer and Robinson, 2007). We computed transitional stages between our spatial and semi-spatial views and mapped these to the timeline using the `<TimeSpan>` tag – as in section 3 – to allow the user to interactively relate these two views of the same data. Figure 6 (middle) shows an intermediate stage of the transition between the two complementary views. It may be useful to use this technique to compare other types of spatial and more abstract graphical representations.

Conclusions, implications and ongoing work

The relationships between the composition of portfolios and the geographies of exposure and potential losses are important factors in decisions about understanding, accommodating and transferring risk. We have shown how KML and Google Earth can be used for geovisualization in the insurance industry to explore these relationships. This may help analysts develop ideas about spatial patterns in data relating to exposure, risk and potential loss and communicate interesting aspects of these data and findings.

The examples presented here are indicative of cases where analysts and decision makers are using these technologies to share data and functionality to support their work. The techniques described provide plenty of scope for developing novel and bespoke views and interactions for the visual evaluation and synthesis of a range of data types used in the insurance industry with considerable potential benefit.

Acknowledgements

The authors are grateful to Willis Analytics for funding the Willis Research Network, which makes this collaboration between academia and the insurance industry possible. We are also grateful to Trevor Maynard and Paul Nunn at Lloyds for permission to use their example for figure 3.

References

Andrienko, G.L. and Andrienko, N.V. 1999. "Interactive maps for visual data exploration". International Journal of Geographical Information Science 13 (4), pp 355-374.

British National Space Centre / Infoterra. 2001. "Analysis of Earth Observation Uptake by the Insurance Sector." BNSC National Earth Observation (EO) Programme, Sector Studies Programme (BNSC Project OB 3/41/26 – limited access).

Bruls, M., Huizing, K. and Wijk, J.V. 2000. "Squarified Treemaps". http://www.win.tue .nl/~vanwijk/stm.pdf (accessed, 17th December 2007)

Dykes, J.A. 1997. "Exploring spatial data representation with Dynamic Graphics". Computers and Geosciences 23 (4), pp 345-370.

Dykes, J.A. 1998. “Cartographic visualization: Exploratory spatial data analysis with local indicators of spatial association using Tcl/Tk and cdv”. The Statistician 47 (3), pp 485-497.
Gahegan. M. 2005. Chapter 4: “Beyond tools: Visual support for the entire process of GIScience”. Exploring Geovisualization, J. Dykes, A. M. MacEachren and M.-J. Kraak (eds). Elsevier Ltd, 2005. **Goodchild, M.F. 2007.** “Citizens as sensors: the world of volunteered geography”. GeoJournal 69, pp 211–22.
Grossi P., Kunreuther H. and Windeler D. 2005. “An introduction to catastrophe models and insurance.” Chapter 2 in Catastrophe Modelling: A New Approach to Managing Risk. Grossi P., Kuntreuther H. (eds), Springer, New York.
Haslett, J., Wills, G.J. and Unwin, A.R. 1990. “SPIDER: An interactive statistical tool for the analysis of spatially distributed data”. International Journal of Geographical Information Systems 4 (3), pp 285-296.
Heer, J. and Robertson, G. 2007. “Animated Transitions in Statistical Data Graphics.” IEEE Transactions on Visualization and Computer Graphics 13 (6), November/December 2007, pp 1240-1247.
van Krevelda, M. and Speckmannb, B. 2007. “On rectangular cartograms”. Computational Geometry 37 (3), pp 175-187.
Lloyds. 2005. “Lloyds uses Google Earth to prepare for the worst”. http://www.lloyds.com/News_Centre/Features_from_Lloyds/Lloyds_uses_Google_Earth_to_prepare_for_the_worst.htm (accessed, 21st December 2007)
MacEachren, A.M., Wachowicz, M., Edsall, R., Haug D. 1999. “Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods”, International Journal of Geographical Information Science 13 (4), pp 311-334
Miller, C.C. 2006. “A Beast in the Field: The Google Maps Mashup as GIS/2.” Cartographica 41, pp 187-200.
Openshaw, S., Charlton, M.E., Wymer, C., Craft, A. 1987 “A mark 1 geographical analysis machine for the automated analysis of Point data sets”. International Journal of Geographical Information Systems 1 (4), pp 335–358
Pezanowski, S., Tomaszewski, B. and MacEachren, A.M. 2007. “An open geospatial standards-enabled Google Earth application to support crisis management”. Joint CIG/ISPRS conference on geomatics for disaster and risk management, Toronto, Canada, Ontario.
Shneiderman, B. 1996. “The eyes have it: A task by data type taxonomy for information visualizations”, Visual languages, Boulder; CO, IEEE Computer Society Press, 1996.
Slingsby, A., Dykes, J., Wood, J. and Clarke, K. 2007a. “Mashup cartography: cartographic issues of using Google Earth for tag maps”. *ICA Commission on Maps and the Internet, July/August 2007, Warsaw, Poland. pp79-93.*
Slingsby, A., Dykes, J., Wood, J. and Clarke, K. 2007b. “Interactive tag maps and tag clouds for the multiscale exploration of large spatio-temporal datasets”. Information Visualization 07, Banissi, E. (ed), IEEE Computer Society.
Wood, J, Dykes, J., Slingsby, A. and Clarke K. 2007. “Interactive visual exploration of a large spatio-temporal data set: reflections on a geovisualization mashup”. IEEE Transactions on Visualization and Computer Graphics 13 (6), pp1176-1183, November/December 2007.
Wood and Dykes, 2008. “From slice and dice to hierarchical cartograms: spatial referencing of treemaps” (this volume).

Biography

Aidan Slingsby is a research fellow at the giCentre, City University London. He is interested in spatial data representation, methods for the visual analysis of large spatio-temporal datasets and computation methods that support this work. He works part time in the Willis Research Network and part time on other projects.

Jason Dykes is a senior lecturer in geographic information at the giCentre, City University London, with interests in developing and using interactive graphics to explore complex structured data sets. Jason has particular interest in the development and evaluation of geovisualization applications and the statistical analysis of spatial pattern.

Jo Wood is a senior lecturer in geographic information at the giCentre, City University London. He is particularly interested in analysing surfaces and hierarchies in applications ranging from geomorphometry to insurance risk. He is the author of the LandSerf GIS and other software packages in Java.

Matt Foote is Research Director of the Willis Research Network (www.willisresearchnetwork.com), an international group dedicated to researching the key issues facing the global insurance industry. He has 20 years experience in remote sensing, GIS and coastal geomorphology with publications in Transactions in GIS, Geomorphology and the Journal of Coastal Research.

Mike Blom has spent the last 3 years at Willis as a GIS developer, supporting the catastrophe analytics team through development and maintenance of various analysis tools looking at clustering, loss curves, exceedance probabilities, and probability distribution. He is currently finishing an MSc dissertation on natural energy resource distribution throughout the UK.

Using and Evaluating HCI Techniques in Geovisualization:Applying Standard and Adapted Methods in Research and Education

Stephanie Larissa Marsh & Jason Dykes

giCentre, School of Informatics, City University London, EC1V 0HB
Tel. (+ 44 20 7040 8906) Fax (+44 20 7040 8845)
s.l.marsh¦jad7 @soi.city.ac.uk, www.gicentre.org/jad7

KEYWORDS: usability engineering, evaluation, HCI, education, research, geovisualization

1. Introduction and Purpose.

Geovisualization involves the use of interactive software tools to prompt thinking about unknown datasets (MacEachren and Kraak, 2001). Successful use of visualization in research results in 'ideation' – the generation of ideas (MacEachren and Kraak, 1997). Successful visualization in education results in the communication of knowledge and improved understanding (Marsh and Dykes, 2005). Techniques from Human Computer Interaction Design (HCI) are becoming more prevalent in a field that relies upon rapid communication interactions between machine and user. MacEachren and Kraak (2001) describe a wide range of potential benefits of applying HCI methods to the geovisualization domain. They also describe some likely impediments. For example, HCI techniques include evaluations that are costly in terms of time and resources, particularly in the case of user testing and geovisualization may involve a small number of expert users. In most cases, the results of HCI experiments are specific to particular software / task / user configurations – indeed the standard definition of usability refers to specific users, specific goals and specific context of use (ISO 9241-11, 1998) - yet there is a need to generalize the findings of individual experiments understand geovisualization, evaluate the utility of HCI techniques and make geovisualization software more usable (Fuhrmann et al., 2005). The reliability and replicability of HCI evaluations is also questionable.

Geovisualization is not alone in using interactive graphics to explore structured data and draws upon disciplines such as information visualization (see Gahegan *et al*. 2001; Andrienko *et al*. 1999; Tobón, 2003). Many of the impediments and benefits are thus shared. Information visualization usability studies can inform geovisualization usability research, as effective evaluation in both domains is likely to require the adaptation of existing HCI techniques or the development of new paradigms and methods. Yet the dominant relationship between the locations of spatial phenomena and their representation on screen puts additional constraints on geovisualization. There is little or no geovisualization literature critiquing usability in the context of the domain's specific requirements. Figure 1 shows these schematically by differentiating between software that is 'standard', for information visualization and for geovisualization using characteristics relating to users, interaction, tools and layout.

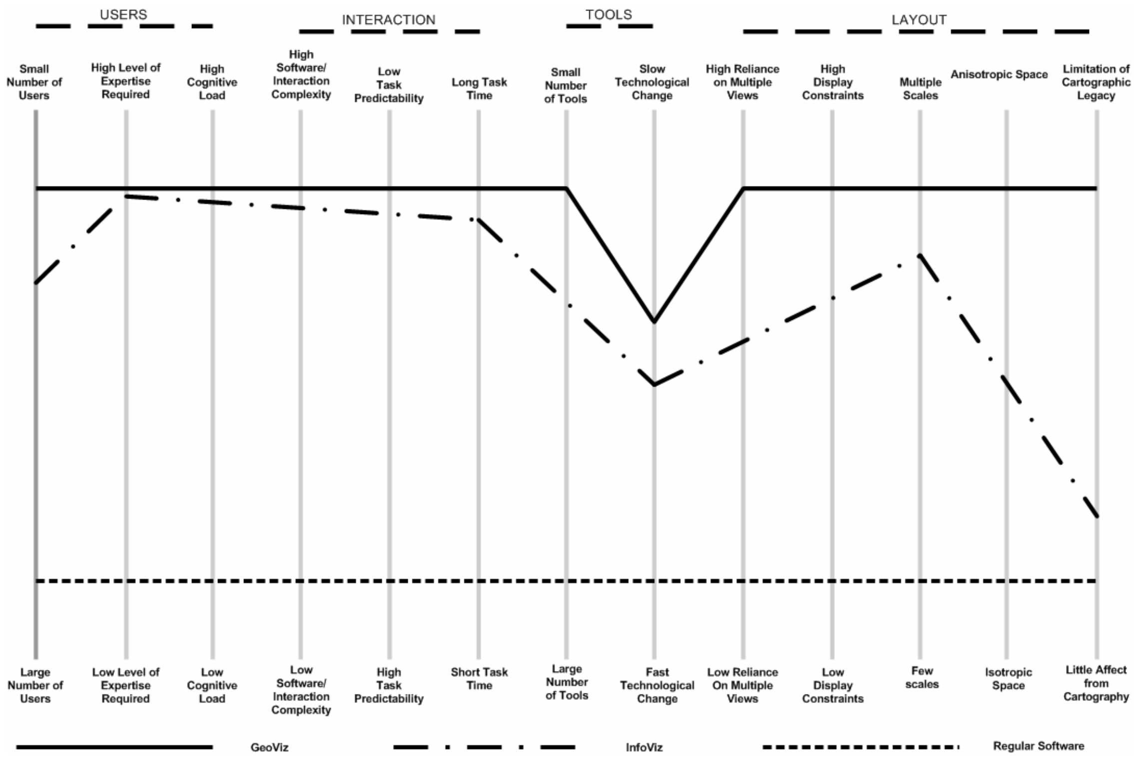


Figure 1. Qualitative representation of differences between geovisualization, information visualization and 'regular' software for which HCI techniques have been developed.

A variety of HCI methods can be utilised to evaluate tools, however little systematic research has been undertaken on the reliability and replicability of these methods. Appropriate evaluation solutions are likely to vary for different contexts making the selection and application of HCI techniques to geovisualization tools complex. Another challenge of applying evaluation techniques to geovisualization is the assessment of the support of ideation. How do we assess whether or not a tool supports ideation, knowledge discovery and hypothesis generation? Can ideation be measured and be an indicator of goal achievement? Such research problems are not considered widely by traditional HCI literature.

However, confidence in these techniques is crucial to the use of HCI in geovisualization and the development of geovisualization applications. Advances in geovisualization design can only be made if the evaluation findings are accurate and reliable. Consequently, we need to know the most efficient, effective and reliable ways to evaluate geovisualization tools, and the circumstances in which these techniques might be valid.

2. Approach.

An experimental framework that enable a variety of data collection techniques to be tested and compared under various conditions was developed, involving eight related experiments. This included both laboratory-based tests in which conditions were controlled and users evaluated '*in vitro*' and evaluations '*in vivo*' where geovisualization software was employed in real tasks (Robinson *et al*., 2005). The framework operated in a linear fashion, enabling us to reflect upon the approaches and subsequently adjusted them, to suit the needs and requirements of geovisualization and the situations in which they occurred. This enabled us to investigate the effectiveness and validity of candidate techniques for geovisualization in a number of contexts and draw general conclusions by triangulating evidence taken from a variety of tests. A number of candidate usability evaluation methods were considered, including: ethnographic observation, retrospective think-aloud, focus groups and questionnaires for experiments conducted in the field; onscreen interaction capture, concurrent think-aloud and task sheets for lab experiments (Buttenfield, 1999; Andreinko *et al*. 2003; Harrower *et al*., 2000; Tobón, 2002; Zaphiris *et al*., 2004; Bekker, *et al*. 2004; Maguire, 1998).

We evaluated geovisualization in research and education with a variety or users undertaking tasks of different type and complexity in a range of conditions. We focused on three aspects of software use in the majority of experiments: usability issues, ideation and interactions that occurred. These factors are integral to the effective use of geovisualization. Quantitative methods were used to analyse the numeric data collected, including ANOVA analysis. Qualitative data was subjected to content analysis to draw out evidence on which to derive results and confidence in them (Silverman, 2001). Triangulation was used to establish the strength of support for our findings.

3. Findings.

Our analysis allows us to make a number of inferences and draw conclusions about using and evaluating HCI techniques in geovisualization by applying standard and adapted methods in research and education. In summary, we have strong evidence that:

1. HCI techniques allow us to capture ideas that are generated in a research context and knowledge being generated in the context of learning through a process of visualization.
2. Different user groups do not perform consistently *in vitro*. *Sampling strategies have an effect on the results generated from methods involving small numbers of users and selection will thus influence some evaluation results in geovisualization.*
3. The inclusion of students in evaluation is only appropriate if evaluating pedagogic tools. *Students are often used as a substitute for experts to increase the numbers of participants in geovisualization evaluations (e.g. Andrienko et al., 2002). In comparative tests, we recorded considerable differences in behaviour. Though this behaviour was not statistically significant due to the small numbers of participants used, the statistics together with qualitative data constitute convincing evidence. Any results derived from studies that use students as expert surrogates should be treated with extreme caution.*
4. HCI techniques and methodologies (standard and adapted) are complementary and not interchangeable. *Alternative sets of methods may produce different results in geovisualization*.
5. Both standard and adapted HCI techniques successfully identified usability issues. *An overview of the results shows that 294 usability issues were identified over 3 in vivo studies, 3 in vitro studies and 1 remote study, involving 2 geovisualization tools and 1 geovisualization demonstrator. Furthermore, 143 issues (49%) were accepted for re-design by two tool developers.*
6. The setting, data collection techniques and user groups employed in the

evaluation influenced results in our comparative studies. *This is particularly true in relation to ideation – the main objective of geovisualization, but also in terms of usability responses and the interactions that occur. The effect is particularly evident when comparing in vitro and in vivo experiments.*

7. Short answer open-ended dataset directed exploratory tasks are more appropriate for facilitating ideation and identifying usability issues *in vitro* than instructional tasks. *Whilst they are less specific than many typical tasks used in HCI such tasks are more typical of geovisualization activity and result in more ideation.*
8. Onscreen interaction capture and think-aloud are the most appropriate data collection techniques in the *in vitro* setting during geovisualization usability evaluations. *These methods are least disruptive to the geovisualization process which involves a significant cognitive load and result in a higher volume of more useful data than alternatives.*
9. When conducting an evaluation with students, discursive and reflective methods yield the most useful and usable data. *Effective discursive methods include mediated focus groups and paired retrospective think-aloud. Reflective methods include structured questionnaires yield. These are particularly useful if scheduled immediately after learning activity.*
10. Satisfaction levels amongst students are affected by levels of training and preparation, motivation, time and resources. *Time allocated for geovisualization learning activity must be adequate for preparation as well as satisfactory task completion. Resources must support the tasks and learning activity fully. This has not been the case in some geovisualization user studies.*
11. Longitudinal *in vivo* studies with experts are important for understanding how geovisualization supports truly complex ideation. *The detailed case study approach may be the most appropriate as geovisualization takes time, requires expertise and the most effective solutions are likely to be data, task and user dependent.*

4. Practical Implications and Conclusions.

We have strong evidence that useful knowledge can be gained by employing HCI methods in geovisualization. Yet we find that experimental results may not be consistent between applications, users, situations and contexts. Our comparative experiments show that evaluation of geovisualization in research and education requires different methodological approaches due to the differing levels of expertise, motivation and knowledge of the participants involved. Thus, it is important that intended users are the subjects of evaluation and a small number of these is preferable to a large number of inappropriate users. Whilst statistically significant results may be achieved in the latter case these may not be relevant if test users do not behave in meaningful ways. Students are not a suitable substitute for experts in order to gain statistical significance. The evaluation of tools intended for the use in education can sample from students and visualization experts, but the evaluation of research tools should sample a range of domain and visualization experts even if the number sampled is small. This implication has a situational caveat, in that students work within their abilities in the *in vitro* environment, perhaps due its artificiality. However, *in vivo*, students will work at level that challenges their abilities, as they are motivated to learn, and this level of activity is essential for visualization to be effective. Visualization experts can simulate students with an intermediate level of expertise within an *in vitro* environment – but this is a potentially expensive way to test, that conflicts with the rationale for using students to simulate experts!

The different behaviours of different user groups in different situations adds complexity to the use of HCI techniques in geovisualization, particularly as more precision can be gained through *in vitro* than *in vivo* experiments. However, data from *in vivo* tests offer more ecological validity, particularly in the case of ideation.

Evaluation frameworks being implemented in education should include methods such as group sessions and questionnaires within both *in vitro* and *in vivo* environments. These techniques elicit comprehensive data on sophisticated visualization software use from students.

There have been several reports of the use of a variety of HCI techniques in geovisualization. However, there is no evidence of studies comparing different HCI techniques and establishing which techniques are most suitable in certain geovisualization circumstances. This research provides empirical evidence upon which geovisualization researchers may select and adapt HCI techniques for developing geovisualization applications and evaluating their use.

Whilst HCI studies in geovisualization are important and beneficial they are limited in their standard forms and may require some adaptation and coordinated use in within a well planned framework to provide meaningful results (Marsh, 2007). If we are to use HCI techniques and methodologies in geovisualization, then we must accept that results are subjective and when applied to small user groups, not statistically significant. However, this does not mean that the results are not scientifically important. The key is not to attempt to gain statistically valid results from geovisualization evaluation by using large numbers of potentially inappropriate users. Rather the focus should be on creating frameworks for conducting *in vivo* longitudinal studies with core users, when capturing data on usability, interaction and ideation. If we want to capture truly in-depth and ecological data, intensive longitudinal *in vivo* studies are essential. Finally, it is worth noting that these conclusions are not specifically geographic in nature; suggesting either that geovisualization evaluation may not differ in nature to that of evaluation in other exploratory domains in which graphical exploration plays a key role, or that the techniques utilised here do not provide the detail required to differentiate between graphical visual interfaces in which layout is constrained by geography and used to represent geographic relationships and those in which it is not.

5. Acknowledgements

The many participants who took part in the tests reported here are gratefully acknowledged and their time effort and cooperation is much appreciated. This includes students of the Masters in Geographic Information at City University London. The work conducted by Fenia Attilakou as part of her Masters thesis is gratefully acknowledged. This paper is based upon:

Marsh, S. L. (2007). Using and Evaluating HCI Techniques in Geovisualization: Applying Standard and Adapted Methods in Research and Education. PhD Thesis. City University.

References

ISO 9241-11 (1998) Guidance on Usability, also issued by the International Organization for Standardization.

Andrienko, G. and Andrienko, N. (1999). Interactive Maps for Visual Data Exploration. *International Journal of Geographical Information Science.* **13(4).** 355-374.

Andrienko, N., Andrienko, G., Voss, H., Bernardo, F., Hipolito, J. and Kretchmer, U. (2002). Testing the Usability of Interactive Maps in CommonGIS. *Cartography and Geographic Information Science.* **29(4).** 325-342.

Bekker, M., Barendregt, W., Crombeen, S. and Biesheuvel, M. (2004). Evaluating Usability and Challenge During Initial and Extended Use of Children's Computer Games. In *HCI 2004*, Vol. 1 (Eds, Fincher, S., Markopoulos, P., Moore, D. and Ruddle, R.). Springer, Leeds, pp. 331-346.

Buttenfield, B. (1999). Usability Evaluation of Digital Libraries. *Science and Technology Libraries.* **17(3/4).** 39-59.

Fuhrmann, S., Ahonen-Rainio, P., Edsall, R. M., Fabrikant, S. I., Koua, E. L., Tobón, C., Ware, C. and Wilson, S. (2005). Making useful and useable geovisualization: design and evaluation issues In *Exploring Geovisualization.* (Eds, Dykes, J., MacEachren, A. and Kraak, M.-J.). Elsevier, Oxford, pp. 553-566.

Gahegan, M., Wachowicz, M., Harrower, M. and Rhyne, T. M. (2001). The Integration of Geographic Visualization with Knowledge Discovery in Databases and Geocomputation. *Cartography and Geographic Information Science.* **28(1).** 29-44.

Harrower, M., MacEachren, A. and Griffin, A. L. (2000). Developing a Geographic Visualization Tool to Support Earth Science Learning. *Cartography and Geographic Information Science.* **27(4).** 279-293.

MacEachren, A.M. and Kraak, M.-J., 1997. Exploratory cartographic visualization: advancing the agenda. Computers & Geosciences, 23(4): 335-343.

MacEachren, A. M. and Kraak, M.-J. (2001). Research Challenges in Geovisualization. *Cartography and Geographic Information Science.* **28(1).** 1-11.

Maguire, M. C. (1998). Usability Methods: Diary Keeping. http://www.lboro.ac.uk/eusc/g_diary_keeping.html. (24/03/05).

Marsh, S. L. and Dykes, J. A. (2005). Using Usability to Evaluate Geovisualization for Learning and Teaching. In *Proceedings of the GIS Research UK 2005 13th Annual Conference* (Eds, Billen, R., Drummond, J., Forrest, D. and Joao, E.). University of Glasgow, Glasgow, pp. 477-483.

Robinson, A. C., Chen, J., Lengerich, E. J., Meyer, H. G. and MacEachren, A. M. (2005). Combining Usability Techniques to Design Geovisualization Tools for Epidemiology. *Cartography and Geographic Information Science.* **32(4).** 243-255.

Tobón, C. (2002). Usability Testing in for Improving Interactive Geovisualisation Techniques. http://casa.ucl.ac.uk/working_papers/paper45.pdf. (2003).

Tobón, C. (2003). Evaluating Geovisualization and User Interaction: an Experimental Design Approach Based Upon User Tasks. Department of Geography, University College London. PhD.

Silverman, D. (2001) *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction, 2.*Sage, London.

Zaphiris, P., Gill, K., Ma, T. H.-Y., Wilson, S. and Petrie, H. (2004). Exploring the Use of Information Visualization for Digital Libraries. *The New Review of Information Networking.* **10(1).** 51-69.

Biographies

Stephanie Marsh was awarded a PhD in Information Science at City University in September 2007. She now works for HSBC.

Jason Dykes was awarded his PhD in Geography at the University of Leicester in December 1999. He now works at the giCentre, City University London.

Analysing the relationship between landscape composition and preference – the combination of GIS and visualisations

Åsa Ode[1], David R. Miller[2]

[1]Dept. of Landscape Architecture, Swedish University of Agricultural Sciences, PO Box 58, SE 230 53 Alnarp, Sweden. Tel. (0046)40415428 Fax (0046)40465442 Email asa.ode@ltj.slu.se
[2]Macaulay Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK

KEYWORDS: Landscape preference, spatial analysis, visualisations, landscape complexity

1. Introduction

The rural landscape provides a valuable resource supporting multiple functions, and changes in the structure of the landscape will mean that supported functions are likely to change accordingly (e.g. Van Eetvelde & Antrop, 2004). With the prospect of climate change the landscape will meet new challenges through for instances changes in agricultural productions and the production of renewable energy, which will have an impact on other functions in the landscape. However, although many drivers of change are global or regional, their manifestation at a local level will vary, and it is at this level these changes will relate to the general population and planners (Antrop, 2004; Dockerty et al., 2005). For management, planning and policy purposes it is desirable to be able to analyse the effects of landscape change. The increased urbanisation is changing our demand on the rural landscape as not only as primary production landscape but also as a service provider for the society at large. These new services include biodiversity, tourism, health and other amenity functions, and are not just limited to the peri-urban landscape. A large extent of these functions is dependent on the landscape as a scenic resource.

Within the VisuLand framework, a range of visual indicators based on theory have been suggested (e.g. Tveit et al. 2006; Ode et al. 2008). The framework is based on nine visual concepts found across a range of literature covering landscape aesthetic, environmental psychology and landscape characterisation. Though put forward in the literature, few of the indicators have so far been through any systematic testing on their relationship with landscape preference. In landscape preference surveys, photographs have traditionally been used as a surrogate for the real environment (e.g. Hagerhall, 1999). However, one of the limitations with the use of photographs in preference surveys is the lack of control of the content of the image that may impact on the observer's perception. With the development of computer visualisations new possibilities of evaluating landscapes under different scenarios has arisen (Daniel and Meitner 2001; Appleton and Lovett, 2005). Ode et al. (in press) means that the use of visualisations in order to test between relationship between preference and landscape composition holds two advantages: i) the use of a controlled environment, enabling changes to be made only to the indicators to be tested in a survey; ii) a direct link between the three-dimensional and two-dimensional data (*i.e.* visualisations can be derived from map-based data).

The concept of complexity is one of the concepts put forward by Tveit et al. (2006) as contributing towards visual quality and for which several indicators has been suggested and applied. However, as put forward for instance by Stamps (2004) there is within preference surveys limited suggestion on how to objectively measure what constitute complexity. Within landscape ecology, a range of measurements for complexity have been developed. These measurements could be divided into:

1. Composition Complexity Indices – different form of diversity, evenness, richness and dominance indices. These do generally lack any spatial references and are therefore mainly focusing on the composition of the landscape and not on the configuration. E.g. Shannon Diversity Index (SDI) and Shannon Eveness Index (SEI).

2. Patch Complexity Indices – different form of indices related to the complexity of shapes, and relation between edge and interior area. This group does also cover degree of complexity regarding variation in these values. E.g. Area Weighted Mean Shape Index (AWMSI), Number of Patches (NumP).

3. Pattern Complexity Indices – different forms of indices to describe the spatial pattern of the landscape. These are both presenting arrangement of polygons as well as for the arrangement of cells of similar class within a grid. E.g. Contagion (CONTAG).

The presented study will together with outlining a method for testing visual indicators, present some initial findings on the linkage between preference and visual complexity indicators.

2. Method

In order to test the linkage between landscape complexity and preference, a methodology was developed combining spatial analysis and visualisations (based on the methodology developed by Ode et al. in press).

The base landscape was an area in the highlands of Scotland, which comprised of a general hilly landscape with few different types of land cover. For this landscape 6 different scenarios of land cover composition was developed based on altering the amount of land cover types present and the distribution between them, creating six scenarios of land cover composition, as presented in Figure 1.

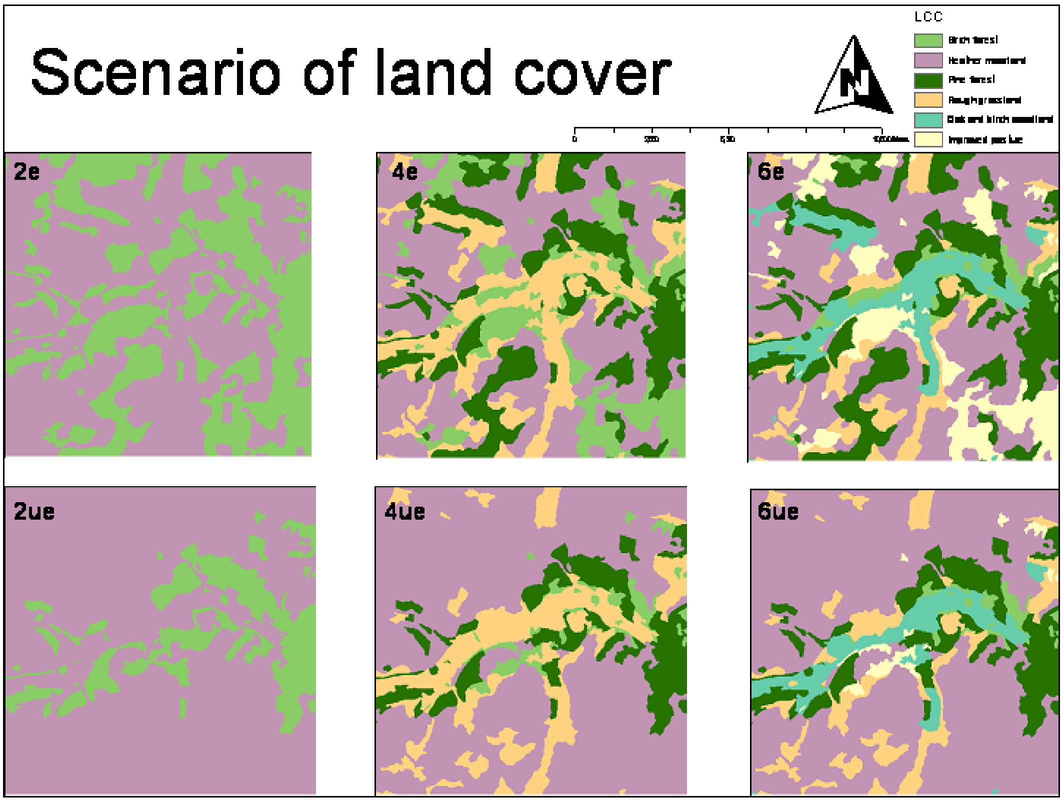


Figure 1. The six scenarios of land cover distribution. The top scenarios are having an equal distribution between open land and forest, while the bottom scenarios are having more of an unequal distribution.

For the six scenarios, indicators of different form of complexity were calculated, with the result as shown in Table 1.

Table 1. Complexity values for selected indicators

Scenario	*SDI*	*SEI*	*NumP*	*AWMSI*	*Contag*
2e	0.66	0.95	71	6.23	42.98
2ue	0.44	0.64	47	4.37	62.54
4e	1.3	0.94	145	3.55	45.00
4ue	0.97	0.7	101	3.20	54.46
6e	1.52	0.85	165	3.20	50.92
6ue	1.12	0.62	117	3.09	64.18

Within the landscape two viewpoints that were representative for the changes in land cover between the scenarios were selected. For these two viewpoints visualisations were generated for all six scenarios, resulting in 12 images. The visualisations were generated using Visual Nature Studio, VNS, with ecosystem created to represent the 6 type of land cover present in the landscape and automatically allocated for the GIS layers showing land cover for each scenarios. Examples of the images generated are presented in Figure 2.

Viewpoint 1. Six land cover types and an even distribution

Viewpoint 2. Four land cover types and an uneven distribution

Figure 2. *Examples of the visualisations generated and used in the survey.*

In order to analyse for preference of the six landscapes, a preference survey was developed using forced choice. The study was run as an internet based survey, where the respondents were asked to choose which image the preferred in a random generated pair of images.

3. Results and discussion

The result of the survey was analysed together with the complexity measurements for the six scenarios using REML in order to establish relationships.

The preliminary results of the analysis, as presented in Table 2, show that all four of the tested indicators has a relationship to preference that is statistical significant. Shannon Diversity

Index is suggested to be the strongest predictor of preference while Contagion is the weakest of the indicators tested for this study. However, all indicators are of the same order.

Table 2. Results from theREML analysis for the five indicators of complexity.

Metrics	*Wald Statistic*	*Chi pr*	*Table of effects*	*Standard error*
SDI	53.10	<0.001	74.82	10.268
NumP	50.19	<0.001	-0.7572	0.10688
AWMSI	46.01	<0.001	-3.176	0.4682
SEI	40.67	<0.001	61.36	9.621
CONTAG	38.04	<0.001	0.5601	0.09082

Both SDI and SEI show a strong positive trend between indicator value and preference, suggesting an increase in preference for landscape with high diversity and evenness. For AWMSI and NumP this relationship, though less strong. This would suggest that an increase in AWMSI and NumP would have a slight negative influence on preference. This slight negative correlation was also found within a previous study on naturalness indicators (e.g. Ode et al. in press), which showed that the highest preference scores was for landscape with a low-medium AWMSI while the images with high AWMSI had a relatively low predicted mean preference score. However, this results is preliminary and further statistical analysis is needed which include a broader range of complexity indicators.

In order to draw any more general results with regards to indicators of complexity and their relationship with preference for landscape there is a need for further research This could include the testing of additional complexity indicators, the running of the study for a range of different landscape types and the investigation into new methods for indicator application, e.g. visual topology.

4. Acknowledgements

The research was conducted during the main authors Post doc period at the Macaulay Institute which was funded through the Swedish Research Council FORMAS.

References

Antrop M (2004) Landscape change and the urbanization process in Europe *Landscape and Urban Planning* **67** pp9-26

Appleton K and Lovett A (2005) GIS-based visualization of development proposals: reactions from planning and related professionals *Computers, Environment and Urban Systems* **29** pp321-339

Daniel TC and Meitner MJ (2001) Representational validity of landscape visualizations: The effects of graphical realism on perceived scenic beauty of forest vistas *Journal of Environmental Psychology* **21** pp61-72

Dockerty T, Lovett A, Sünnenberg G, Appleton K and Parry M (2005) Visualising the potential impacts of climate change on rural landscapes *Computers, Environment and Urban Systems* **29** pp297-320

Hagerhall C, (1999) The Experience of Pastoral Landscapes. Doctoral thesis, *Acta Universitatis Agriculturae Sueciae, subserie Agraria no 182*

Ode Å, Fry G, Tveit MS, Messager P and Miller D (in press) Indicators of perceived naturalness as drivers of landscape preference *Journal of Environmental Management*

Ode Å, Tveit M and Fry G (2008) Capturing landscape visual character using indicators - touching base with landscape aesthetic theory. *Landscape Research* **33** pp89-117

Stamps AE (2004) Mystery, complexity, legibility and coherence: A meta-analysis *Journal of Environmental Psychology* **24** pp1-16

Tveit M, Ode Å and Fry G (2006) Key concepts in a framework for analyzing visual landscape character *Landscape Research* **31** pp229-255

Van Eetvelde V and Antrop M (2004) Analyzing structural and functional changes of traditional landscapes – two examples from Southern France *Landscape and Urban Planning* **67** pp79-95

Biography

Dr Åsa Ode is a postdoctoral fellow at the Dept. of Landscape Architecture at the Swedish University of Agricultural Sciences in Alnarp. Her current research focuses on linking the experience of landscape with the spatial structure of landscape in order to develop landscape indicators.

Estimating land use profiles for diffuse pollution modelling in river catchments

*Paulette Posen[1], Andrew Lovett[1], Michael Hutchins[2], Amelie Deflandre-Vlandas[2], Helen Davies[2]

[1]School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ
[2]Centre for Ecology and Hydrology Wallingford, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB
*Corresponding author: Tel. +44 (0) 1603 592547 Fax +44 (0) 1603 591327
p.posen@uea.ac.uk

KEYWORDS: agricultural census, diffuse pollution, integrated modelling, land use, Water Framework Directive

1. Introduction

Implementation of the EU Water Framework Directive (WFD) (CEC, 2000) calls for a major restructuring of European water management to be completed by 2012. Water management under the WFD, which aims to achieve "good ecological status" in all water bodies by 2015, is based on the natural geographic and hydrological unit of the river basin. Diffuse pollution from agriculture, in the form of nitrates, phosphates and pesticides, is a primary reason for observed concentrations of these compounds being in excess of the upper limits prescribed by the WFD, and reductions in these concentrations will require widespread changes in farm practice.

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 has involved updating an existing diffuse pollution model (Hutchins et al, 2006) to reflect present-day land use profiles and comparing outputs (in terms of nitrate concentrations) from current land use with those modelled from early 1990s land use. The same approach will then be used to predict the possible outcomes of future, policy-driven, land use change.

Combining agricultural 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). For instance, in order to preserve the anonymity of individual farms, agricultural census data for England and Wales are aggregated over the spatial extent of the Office of National Statistics (ONS) geographical unit known as the Super Output Area (SOA), corresponding to statistical ward boundaries. These data have been disaggregated by EDINA at Edinburgh University Data Library, to 2 km and 5 km grid resolution 'Agcensus' datasets, based on a 1 km 'Landuse Framework' which defines the spatial extents of seven discrete land use categories (EDINA, http://edina.ac.uk/agcensus/description.shtml). In order to provide suitable land use profiles for input into a hydrological model, further manipulation is required to interpolate the Agcensus data to the spatial extent of arable land within each hydrological unit. Due to the way in which the raw agricultural census data were aggregated, there is no true land use information against which to assess the accuracy of this areal interpolation, so it is necessary to assess spatial accuracy by examining the variability of the results across a range of spatial scales. This work examines these issues, setting out to identify the range of spatial resolutions at which meaningful estimations of agricultural land use can be made, and to scrutinise the ability to reliably assess change over time.

2. Interpolating Agricultural Census data to hydrological units

The location chosen for this detailed analysis was the River Derwent catchment in North Yorkshire, a subcatchment of the Humber basin, the latter being the main focus of the ChREAM study. The Yorkshire Derwent covers an area of 160,000 ha, 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 urban areas.

Land use was initially divided into six broad categories (Table 1) based on the Centre for Ecology and Hydrology (CEH) Land Cover Map 2000 (LCM 2000) urban and suburban classes, combined with Ordnance Survey digital Meridian 2 Developed Land Use Area (DLUA) data, to update the spatial extent of built-up areas. The LCM 2000 grassland categories do not completely coincide with the grassland categories used in the CASCADE hydrological model (Hutchins et al, 2006) used by CEH, so the LCM categories had to be reclassified to match those used in CASCADE. It was then necessary to distribute 2 km grid resolution Agcensus data within the areas defined as 'arable' and 'setaside', in order to calculate input values for the land use classes required for the CASCADE model (Table 1). Once again, 'arable' and 'setaside' are classified individually in LCM 2000, but these areas were combined for use within the CASCADE model.

Table 1. Land use categories for use in the CASCADE model

Broad land use categories	CASCADE classes & subdivisions
Urban/Suburban	Urban/Suburban
All Other Land	Other Land
Temporary/Permanent Grassland	Temporary/Permanent Grassland
Rough Grazing	Rough Grazing
Woods	Woods
Arable/Setaside	Setaside
	Winter Wheat
	Winter Barley
	Spring Barley
	Other Cereals
	Potatoes
	Sugar Beet
	Peas and Beans
	Oilseed Rape
	Linseed
	Other Crops and Fallow
	Vegetables
	Greenhouses
	Fruit

The 2 km resolution of the Agcensus data is coarse with respect to the HRU, on which CASCADE input is based, these units being, typically, 5-8 km^2 in area. In an attempt to improve the fit of the areal interpolation, this problem was addressed in two ways. Firstly, the HRUs were aggregated into 61 larger spatial units: these aggregated HRUs are used in the

QUESTOR model, which takes the outputs from CASCADE and calculates the effects of in-river processes on agricultural pollutants. Secondly, values from the 2 km Agcensus squares were subdivided into four equal 1 km squares before the interpolation was performed. The interpolated Agcensus values were then scaled to fit the available arable and setaside areas within each aggregated HRU. This process was performed using Agcensus values from 1994 and from 2004, for comparison.

3. Comparison of results across a range of spatial scales

Three different interpolation methods were examined, in order to establish what level of detail was required to give the most representative land use profiles for input to the CASCADE model. These were:

1. 'Point in polygon' method to match 2 km Agcensus grid squares to aggregated HRUs.
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.
3. Areal interpolation of the 1 km squares to get a proportional profile for each aggregated HRU.

The results of these three methods were compared across a range of spatial scales, defined by sub-catchment area (as shown in Table 2). Differences between interpolation methods were also appraised. The values obtained for grassland (in terms of percentage of sub-catchment) are presented in Table 2.

Table 2. Comparison of the percentage areas of grassland derived from the 2004 Agricultural Census, using different interpolation methods at a range of spatial scales.

Sub-catchment	**Sub-catchment area (ha)**	**Percentage of sub-catchment available for grassland**	**Percentage grassland derived from Agcensus 2004 using three different methods**		
			aggregated HRU (proportional)	*1k grid*	*2k grid*
Pickering Beck	7071	23.08	30.06	19.98	27.96
Rye at Ness	9847	24.54	26.79	20.98	52.52
Seven at Normanby	11734	23.51	23.29	23.81	36.20
Derwent at Low Marishes	36133	17.78	21.45	21.49	35.95
Derwent	159480	22.26	23.48	23.45	23.36

Inspection of the derived grassland proportions for the entire Derwent catchment indicates that there is very little variation in results obtained using the different interpolation methods, and that all of the values are close to the area available within the relevant LCM 2000 grassland classes. However, as the spatial scale decreases, much more variation is observed between the results produced by the different interpolation methods, indicating that less confidence can be placed in land use profiles derived for smaller catchment areas. Interpolation based on 2 km grid squares can be particularly problematic in catchments where some grid squares extend well beyond the catchment boundary (e.g. the Rye at Ness, Table 2). In such cases, agricultural activity beyond the catchment boundary may be registered to in-catchment farms, skewing the land use profiles of the catchment.

According to the Agcensus data, the proportional interpolation of grassland for the entire Derwent catchment remained fairly constant between 1994 and 2004, with 24% and 23%

areal coverage, respectively. In contrast, a decline in total arable area was observed, from 48% in 1994 to 39% in 2004. However, these results are not borne out at aggregated HRU level, as shown by plots of 1994 vs. 2004 percentage grassland areas (Figure 1(a)) and arable land (Figure 1(b)). Each point represents one aggregated HRU. Both plots are very scattered, showing no clear relationship between values obtained for each year, although there is a suggestion that there has been a decline in both grassland and arable areas, with arable areas Figure 1(b) showing the most marked decrease. Figure 1(b) certainly suggests that the decline in arable agriculture has been more concentrated in certain parts of the Derwent catchment.

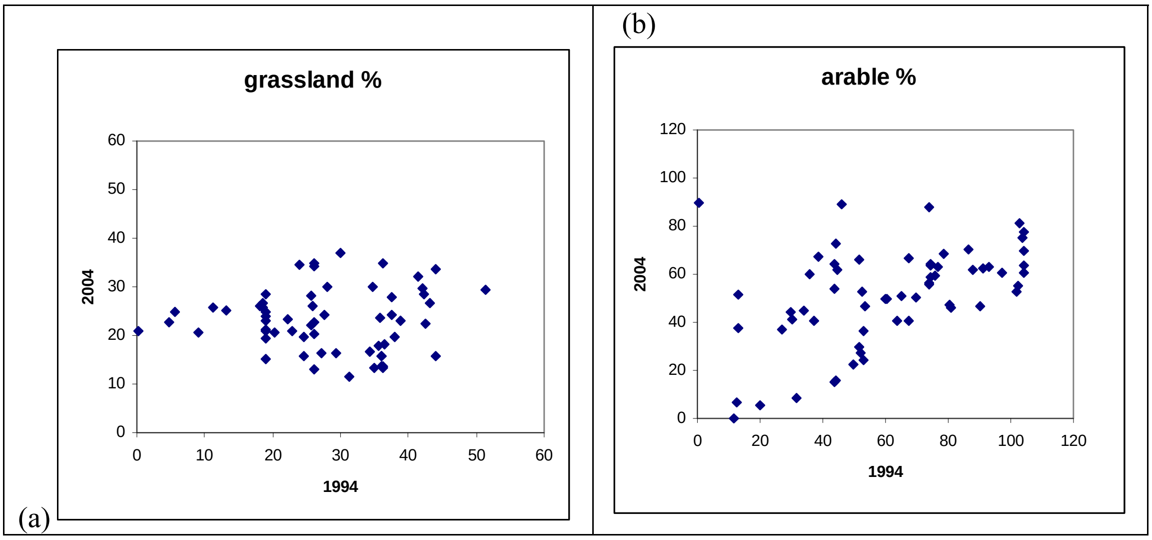


Figure 1. Plots of (a) percentage coverage of grassland 1994 vs. 2004, and (b) percentage coverage of arable land 1994 vs. 2004, per aggregated HRU

This observed variability in results at aggregated HRU level is supported by plots of arable/grassland ratios (Figure 2): although values appear to remain fairly constant across many areas between 1994 and 2004, a great deal of scatter occurs in around one quarter of the 2004 data points, giving rise to large differences between the ratios for the two years in individual hydrological units. However, the implication from many of these scattered data points is that, contrary to the whole-catchment results, there has been an increase in arable farming in certain aggregated HRU areas. Although this may be true, it is more likely that some of these discrepancies relate to changes in the way the census data have been compiled. In 1994 the Agricultural Census was compiled at the parish scale, but post-1999 agricultural data have been compiled at the larger scale of the Super Output Area.

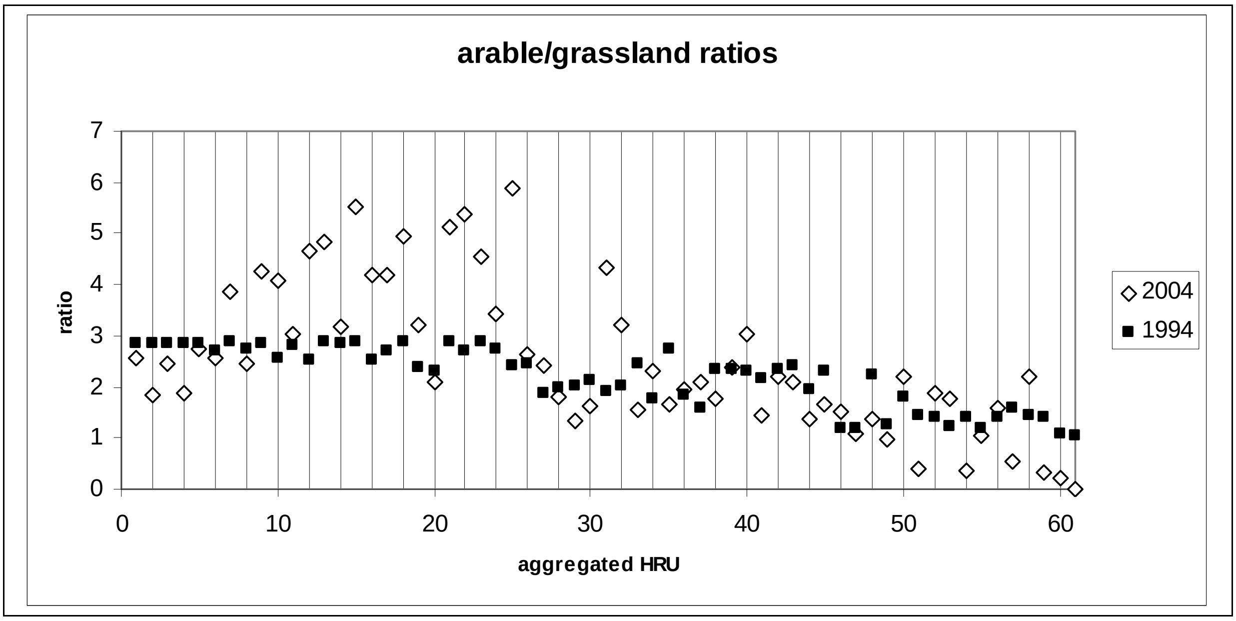


Figure 2. Arable/grassland ratios in 1994 and 2004, per aggregated HRU

4. Conclusions and implications for diffuse pollution modelling

This work 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 hydrological units, depending on the method of interpolation and the scale at which the results are reported. This variability arises for various reasons, 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 this reliability decreases at the scales of smaller hydrological units. It is, therefore, important to exercise caution when deriving such data for use in diffuse pollution modelling, since differences in land use inputs may have a marked effect on model output.

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.

Geddes A, Gimona A and Elston DA (2003). Estimating local variations in land use statistics. *International Journal of Geographical Information Science* 17 pp299-319

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

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. She is currently involved in modelling likely impacts of the Water Framework Directive upon rural land use and farm incomes.

Return of the Stanford Bunny - Definition, Computation and Application of Visual Topology.

Sang, Neil[1-2]

[1] Macaulay Institute, Craigiebuckler, Aberdeen, AB15 8QH / [2] University of Glamorgan

KEYWORDS: Landscape preference, visual topology, landscape indicators

1. Introduction

The visual characteristics of a landscape are widely experienced but, difficult, and controversial to define (Walker 1995). Recognition of their importance, in treaties such as the European Landscape Convention (Council of Europe 2003) and European Council Directive 97/11 (which enacted the Environmental Impact Assessment (EIA) system in Europe (EC 1997)), has led to the need for an 'objective' assessment of their values and the potential impact of changes to them.

Both the scale and displacement forms of the Modifiable Area Unit Problem (MAUP) (Openshaw 1984, Sang et al 2005) occur when one attempts to model landscape value as it appears from a particular viewing point, using data available on a flat map (Sang Ode and Miller 2008). Perspective influences the geometry and scale at which different parts of the map data are seen, and topology is affected when landform masks one part of the data (Sang Ode and Miller 2008). It could be, for example, that a polygon on the map is classed as heather, but the segment of the polygon actually visible in the view is predominantly rough grassland. This limitation in the data may be more significant if that unit is in the foreground than in the background, since it fills a greater percentage of the view.

However, while apparent geometry and scale vary continuously with any small change in viewpoint, the points at which particular landscape units become partially or entirely masked are discreet events, as are those when two land covers appear to become adjacent in the view, or when linear features appear to meet (for example the classic "layered" v-intersections of hills). This makes variance in geometric measurements with different view points difficult to estimate, but also identifies when the connectivity of shapes in the view change and, conversely, when it is locally invariant to change in view point (hence the term 'Visual Topology'(VT)). There is some evidence to suggest that the topology (Egenhofer and Herring 1990, Kinsey 1991) of landscape is cognitively significant (Appleton 1996, Sharif Egenhofer and Mark 1997, Mark and Turk 2003, Mark and Egenhofer 1995). Identifying the areas where VT is locally stable (and hence geometric variance with view point is statistically tractable) could help place the selection of view points on a more objective basis, a particular problem for ensuring legitimate representation in landscape research and planning processes (Appleton and Lovett 2005).

2. Establishing Visual Topological Relationships

There are two basic forms of VT, inter-visibility (along lines of sight as per a standard visibility graph) and 'lateral topology' (which elements appear to be next to each other from a given perspective). They are both, however, mediated by the same elements – horizons and their shadows.

Inter-visibility may be viewed as a topological concept as it describes an information link between the two points. O'Sullivan and Turner (2001) argue that, since the process of determining what locations are visible from where is so computationally expensive, it may instead be pre-calculated and stored as a topological matrix. However the number of viewpoints is limited by computational cost and by concentrating only on the inter-visibility topology, the arrangement of objects cannot be re-constructed from the matrix (O'Sullivan and Turner 2001). In practical terms, this means that because the relationship between the viewer and those objects ascribing the view is not retained, one cannot predict whether a landscape or view point change is likely to affect the inter-visibility graph, and where. Dynamic processes would therefore need to repeat the computationally expensive visibility calculations at each step.

To establish the apparent adjacency between objects in the view (lateral topology) is problematic for traditional ray tracing based visibility analyses as the visible areas are often fragmented on the map, while screen projections are usually "dumb" having lost the relevant classification information, and means to re-connect this information are often slow (Sang Ode and Miller 2005). Thus while existing approaches to visibility analysis (Burrough and McDonnell 1998) hidden surface removal and image rendering (Foley et al. 1990, Purcell et al 2002, Yang et al. 2006) provide the means to map visibility or quickly render objects such as "the Stanford Bunny" for visualisation, they do not provide all the required functionality for landscape analysis.

Both problems arise from the fact that the visibility information has been removed from its spatial context. Rather than building visibility information in a separate map or matrix we propose building the horizon and shadows into the primary data by setting pointers from one to the other.

3. Scanning Horizons

Horizon edges can be detected efficiently on triangulated terrains by comparing the vector normal of each triangle, with the viewing vector (Southerland et al. 1974) (Figure 1). If the vector Normal is less than 90 degrees from the viewing vector, the face must be visible, if over 90 degrees the triangle must be facing away from the viewer. TIN provide implicit topology so which triangles are adjacent to each other can be established (Wang et al 2001). If one triangle is facing the viewer, and its neighbour facing away, then the common edge must form a horizon in the view. The data set which must be projected to establish the relative location of surfaces across horizon boundaries in perspective view, is thus reduced to a limited set of point pairs.

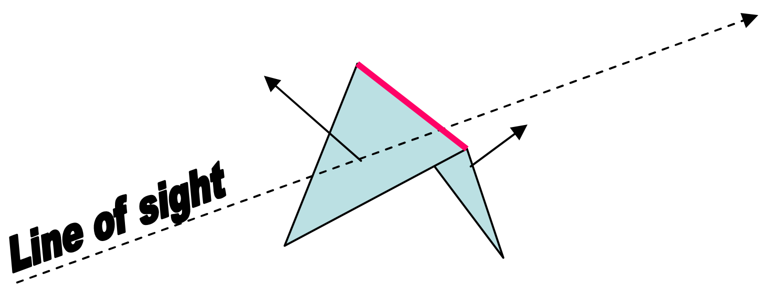


Figure 1. Horizon identification

4. Casting Shadows

Gold, Nantel and Wang (1996) describe an algorithm where by the triangulation may be traversed such that those elements nearer to a particular point are processed first. Applying this to visualisation, when a horizon is identified and projected to screen coordinates (e.g. line 'a-b' in Figure 2), if it falls below existing horizon lines, the edge must therefore be hidden behind a closer feature. If above previously projected features, it forms a new horizon, and if it intersects the existing horizon, this identifies a node in the horizon graph. The lines and edges in the completed horizon graph may then be subjected to Euler's equation (Euler 1735) as a metric of its topological complexity.

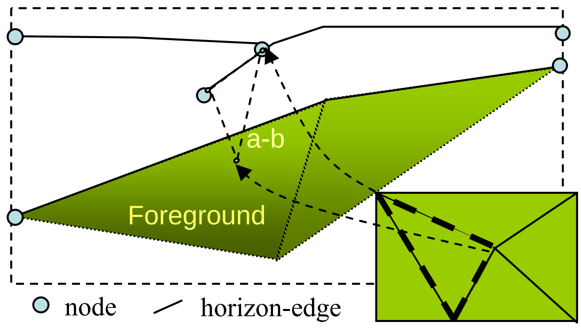


Figure 2. Horizon Graph

The result from the intersection is, in fact, the shadow the more distant horizon casts on the screen intersecting the shadow of the nearer horizon. Since both the projected element and the element it intersects are known at this time, a pointer may be set in the TIN between the two. The more distant elements will, when projected, be smaller than those projected first, so many more distant elements may intersect a nearer one providing a tree of pointers along lines of sight, by which the terrain may be navigated from the horizon to the view point. This avoids the geometric problems of building visual topology from a visibility map of equally sized pixels and retains the classification information in the original TIN.

5. Dynamic Update

Since we are storing the visual topology within the original TIN structure, much can be done to predict when a change in view point or landform might require a new visibility calculation and where. If a change in landform occurs an initial prediction as to its visibility can be made from its position relative to the horizon and shadow elements and then (more precisely) by whether it intersects the plane between them.

If the viewpoint changes incrementally, it may be initially assumed that the shadow will also move only incrementally. It makes sense therefore to initially only test those elements nearest the previous shadow line, to see if the new shadow line is found, and move progressively away from that point. In this manner, a complete new visibility analysis will be avoidable in most cases.

If the change is only regarding attributes, and will not affect visibility, the change in apparent visual relationships can be established directly from the pointers between horizons and shadow edges, making such change update very efficient.

6. Application

A common methodology to relate landscape characteristics and visual preference, is to rank scenes according to what score each receives using indices, developed within landscape-ecology, such as Mean_Shape Complexity (e.g. as edge to area ratio) and Shannon Diversity Index (i.e. the number of patches of different land covers per area) (McGarigal et al 2002; Ode Tveit and Fry 2008). Photographs or computer generated visualisations of the study area are then presented to survey respondents to be ranked according to preference. However, landscapes may rank differently for some indicators, depending on whether the index scores are calculated on the planar map or in perspective view, so changing the direction of any correlation with preference (Germino et al 2001, Sang Ode and Miller 2008). VT may help identify where these effects are significant, and objectively select view points which are typical of those to be found in a landscape region.

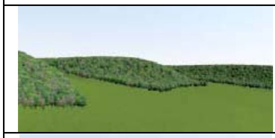
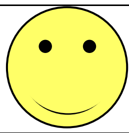
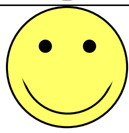
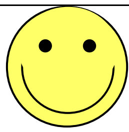
Landscape Image	Indicator: Patch Size	Preference
	10	
	6	
	4	

Figure 3 – Landscape metrics and preference correlation

Figure 4 – Landscape preference measurement with the Virtual Landscape Theatre.

8. Conclusion

The provision of pointers along lines of site to link spatial information with visual topology provides the option to automate operations that model spatial data based on visual effects, and to incrementally update visibility analyses. In particular it provides the potential for:

- More objective selection of viewpoints in landscape planning and research. Based on the mean and variance of landscape metrics in perspective view for regions that are topologically invariant under local viewpoint change.
- Testing of the graph complexity of horizons as an explanatory variable of landscape preference and the facility to map this factor if it proves significant.
- Automatic analysis of local landscape change in perspective view, including framing effects.
- The possibility of efficiency savings in dynamic visibility analysis for some kinds of landscape.
- The possibility of embedding VT for sights of particular significance into the DTM (for example as part of an SDI serving EIA).

The author is aware of exceptions with which this approach does not presently deal. For example where a new horizon arises which was not previously apparent, this cannot currently be detected by local methods, however the advantages the method appears to offer for a qualitatively richer analysis of visual-data suggest the method merits further research. While the intention is to facilitate rural landscape management and planning, it is hoped there may be wider applications, for example urban design and automated route planning, where such perspective information could warn for blocked lines of sight, or predict the most visually interesting route.

9. Acknowledgements

While the author takes sole responsibility for any short comings in this particular paper, the work is part of a PhD supervised by Professor Chris Gold (University of Glamorgan) and Professor David Miller (Macaulay Institute). This work is funded by the Scottish Rural and Environment Research and Analysis Directorate (RERAD), in collaboration with the University of Glamorgan. Visualisations in Figure 3 are by Pernette Messager (Macaulay Institute) for the 'VisuLands' project (EU QLRT-2001-01017).

10. References

Appleton, J., 1996. The Experience of Landscape. Revised edition. New York, Wiley.

Appleton, K., Lovett, A., 2005, GIS based visualisations of development proposals: reactions from planning and related professionals, *Computers, Environments and Urban Systems,* 29, 3, 321-339.

Burrough, P.A., And McDonnell, R. A., 1998, Principles of Geographic Information Systems, Oxford Univ. Press,

Council of Europe, 2003, Presentation of the European Landscape Convention Council of Europe, Strasbourg.

EC 1997 Council Directive 97/11/EC of 3 March 1997 amending Directive 85/337/EEC) on the Assessment of the effects of certain public and private projects on the environment, Official Journal of the European Union, L 073, p.0005 – 0015.

Egenhofer, M., Herring, J., 1990, Categorizing Binary Topological Relations Between Regions, Lines and Points in Geographic Databases. Technical report, Department of Surveying Engineering, University of Maine.
http://www.spatial.maine.edu/~max/9intReport.pdf

Euler, L., 1735, Solutio problematis ad geometriam situs pertinentis, Commentarii academiae scientiarum Petropolitanae 8, 1741, pp. 128-140.
http://www.math.dartmouth.edu/~euler/docs/originals/E053.pdf
English Translation: Newman, J., 1956, World of Mathematics, Paramount Publishing, New York.

Foley, J., van Dam, A., Feiner, S., Hughes, J., 1990, Computer Graphics : Principles and Practice, Addison-Wesley Publ. Reading.

Germino M, Reiners W, Benedict B, McLeod D, Bastian C, 2001, "Estimating Visual Properties of Rocky Mountain landscapes Using GIS" Landscape and Urban Planning 53 71 – 83

Gold, C., Nantel, J., Yang, W., 1996, Outside-in: An Alternative Approach to Forest Map Digitising, International Journal of Geographic Information Systems, 10, 3, 291-310.

Kaplan, R., Kaplan, S., 1989. The Experience of Nature: A Psychological Perspective. Cambridge University Press, Cambridge, UK.

Kinsey, C., 1991, Topology of Surfaces (Springer-Verlag, New York)

Mark, D., Egenhofer, M., 1995, Topology of Prototypical Spatial Relations Between Lines and Regions in English and Spanish, In proceedings, Auto Cart 12, Charlotte, North Carolina, pp. 245-254.

Mark, D., Turk, A., 2003, Ethnophyiography, Workshop on Spatial and Geographical Ontologies, 23rd September (Prior to COSIT 03).

McGarigal, K., Cushman, S., Neel, M., and Ene, E., 2002, FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: www.umass.edu/landeco/research/fragstats/fragstats.html

Ode, Å., Tveit, M.S., and Fry, G., 2008, Capturing landscape visual character using indicators – touching base with landscape aesthetic theory. Landscape Research, **33**, 1, 89 – 117.

Openshaw, S., 1984. The Modifiable Areal Unit Problem. “Conceptual Techniques in Modern Geography”, 38, Geo Books, Norwich, UK.
O’Sullivan, D., And Turner, A., 2001, Visibility graphs and landscape visibility analysis. International Journal of Geographical Information Science, **15**, 221–237.
Palmer J F, 2004, “Using spatial metrics to predict scenic perception in a changing landscape” Landscape and Urban Planning 69 201 – 218
Purcell, T., Buck, I., Mark, W., Hanrahan, P., 2002, ACM Transactions on Graphics. 21 (3), 703-712, (Proceedings of ACM SIGGRAPH 2002). http://graphics.stanford.edu/papers/rtongfx/
Sang, N., Ode, Å., and Miller, 2008, Landscape Metrics and Visual Topology in the Analysis of Landscape Preference, Environment and Planning B, advance online publication, doi:10.1068/b33049.
Sang, N., Birnie, R.V., Geddes, A., Bayfield, N.G., Midgley, J.L, Shucksmith, D.M., and Elston, D.M., 2005, Improving The Rural Data Infrastructure: The Problem Of Addressable Spatial Units In A Rural Context, Land Use Policy, 22, 175-186
Sang, N., Miller, D., Gold, C., subm, Visual Topology : Extending the Nine Intersection Topological Model to Support Analysis of Visual Relationships.
Sang N, Ode Å, Miller D, 2005, “Visual Topology for Analysing Landscape Change and Preference” , in Proceedings for 17th Annual Meeting of ECLAS “Landscape Change”, Ankara, Turkey, 14 - 18th September*,* pp. 176-185
Shariff, Egenhofer and Mark, 1997, Shariff, A. Egenhofer, M. and Mark, D., 1997, Natural-language spatial relations between linear and areal objects: the topology and metric of English-language terms. International Journal of GIS, 11, 3, 215-246.
Southerland, I., Sproull, R., Schumacker, R., 1974, Characteristation of Ten Hidden Surface Algorithms, Computing Surveys, 6, 1.
Walker, G., 1995, “Renewable energy and the public” Land Use Policy 12 49 - 59
Yang, R., Guinnip, D., Wang, L., 2006, View-Dependent Textured Splatting, *The Visual Computer 22(7): 456-467,* http://vis.uky.edu/~wangl/Research/VDTS/VDTS/VDTS.pdf
Wang, K., Lo, C., Brook, G,. 2001, Arabnia, H,. Comparison of Existing Triangulation Methods for Regularly and Irregularly Spaced Height Fields, “International Journal of Geographical Information Science”,15, No. 8, 743- 762.

Biography

Neil Sang is a researcher in Geographical Information Science at the Macaulay Land Use Research Institute (Aberdeen, Scotland). A core theme of interest is to develop methods for measuring and improving the “fitness for purpose” of spatial data, for example through spatial sampling design and spatial data infrastructures. He is currently undertaking a part-time PhD at the University of Glamorgan, for which the particular “purpose” in question is modeling and analysis of visual landscape character from cartographic data.

Assessing bioenergy crop potential in Yorkshire and the Humber using land suitability and multi-objective land allocation models

Patrizia Tenerelli[1], Steve Carver[2]

[1] Dipartimento PROGESA, Facoltà di Agraria, University of Bari
Phone +39 080 544 2869, Fax +39 080 544 2863 Via Amendola, 165 70125 BARI, Italy
email patrizia@tenerelli.it
[2]School of Geography, University of Leeds, LS2 9JT, UK

KEYWORDS: energy crops, land capability, land allocation, landscape planning, land use changes

1. Introduction

The rising of the fossil fuel price and the concern for global changes are the main drivers of the increasing interest in the bioenergy path towards distributed energy generation systems, climate change mitigation and increasing opportunities for the rural development. Biomass applications include heat and electricity generation and transport fuels. They thus have an important role in the future UK overall economy (Elghali et al 2007). However, the bioenergy sector is a complex system made up by different potential feedstock, environmental impacts and technological pathways. Moreover, the development of the sector largely depends on the economic sustainability and social acceptance. Therefore an efficient and sustainable strategy for the exploitation of biomass resources for energy use should be based on an integrated analysis by a multi-disciplinary approach.

Despite the high level of interest in bioenergy, there have been few studies on the sustainability implications of broad-scale biomass exploitation in terms of land use changes and ecosystems resilience. Most of the literature comparing the environmental implications of using bioenergy against fossil fuels points out the greenhouse gas emission issues, while few studies consider the biophysical impact on the ecosystem involved in the biomass production phase (Giampietro et al 1997, Graham et al 1996). The increasing demand for bioenergy is expected to affect land use patterns and it may cause further intensification of agricultural systems. Hence the feasibility analysis of the introduction of energy crops should include an evaluation of the land suitability for the crops under specific pedo-climatic conditions, their competitiveness with traditional crops and natural environments, their impact on agri-landscapes and the implications for biodiversity.

The aim of this work is to analyse future scenarios from the potential supply side through a landscape-based approach which takes into account the agricultural and environmental issues involved in the introduction of novel crops in existing land use patterns at the local scale. The final goal is to asses the reasonable bioenergy potential from local resources and the likely success of bioenergy chains that create the least land-use problems.

2. Background

The biomass-to-energy system has an important spatial dimension due to the dependence of the supply system on agronomic factors and environmental constraints. Compared to other energy sources, biomass has a significant land requirement and high dispersion of resources, resulting in a low energy density per unit of land. Moreover, the trade-off between energy and food production means that the land competition issue is one of the biggest challenges for the sustainable development at the global scale (Ignaciuk 2006, Johansson and Azar 2006, Giampietro et al 1997, Berndes 2006).

For this reason Geographic Information Systems and spatial models are particularly appropriate for the analysis and decision support in agro-energy planning and design of biomass to energy supply systems. In literature numerous GIS approaches are applied for the characterization of geographic variation that may influence biomass availability (Martelli F et al 2002, Schneider et al 2001), costs (Noon and Daly 1996, Voivontas et al. 2001), supplies (Graham et al 2000) and demand as well as the optimal locating of facilities (Freppaz D et al. 2004, Ma J et al. 2005, Edwards RAH 2005, Towers et al. 1997) and biomass collection systems (Masera et. al 2006, Freppaz D. et al 2004).

In this study a multi-criteria GIS-based approach for land suitability and land allocation has been applied to asses the most suitable land-conversion areas in the Yorkshire and Humber region, identifying the key parameters and related criteria for sustainable agri-energy land use.

3. Methodology

The energy crops supply scenario has been built considering both the case of annual and perennial crops. The methodology is composed of a two phase combined spatial model. The first phase aims to apply a land capability model to map the suitable area and the capability level for each energy crop by considering some pedo-climatic factors. The second phase deals with a land allocation model (figure1) which considers some environmental constraints in order to integrate the agro-energy use in existing land use patterns. The models takes into account the need for food and feed crops and assumes that the energy crops never displace natural habitat areas or other land with high ecological value, and that only primarily agricultural land can be converted into energy crop fields. Moreover, the model allocates the different energy crops according to their agronomic requirements and management systems in such way that the less intensive systems are adopted in the areas with higher environmental constraints, such as areas with high erosion risk, high hydraulic risk, areas sensitive to pollution from fertilizers, riparian buffer areas or connection/corridor areas for ecological networks and areas with water scarcity problems. The land allocation model thus addresses nature conservation and sustainable land use needs by creating suitable land use patterns for the integration of agri-energy land uses.

Two different scenarios have been built under different assumptions for energy and environmental policy directions. The first scenarios consider the effect of strong environmental policies other than avoided greenhouses emissions, such as incentives for environmentally-oriented farming, implemented rules for natural resources conservation and regulations for global food security. The second scenario has been built under the assumption of strong bioenergy production incentives with high carbon taxes.

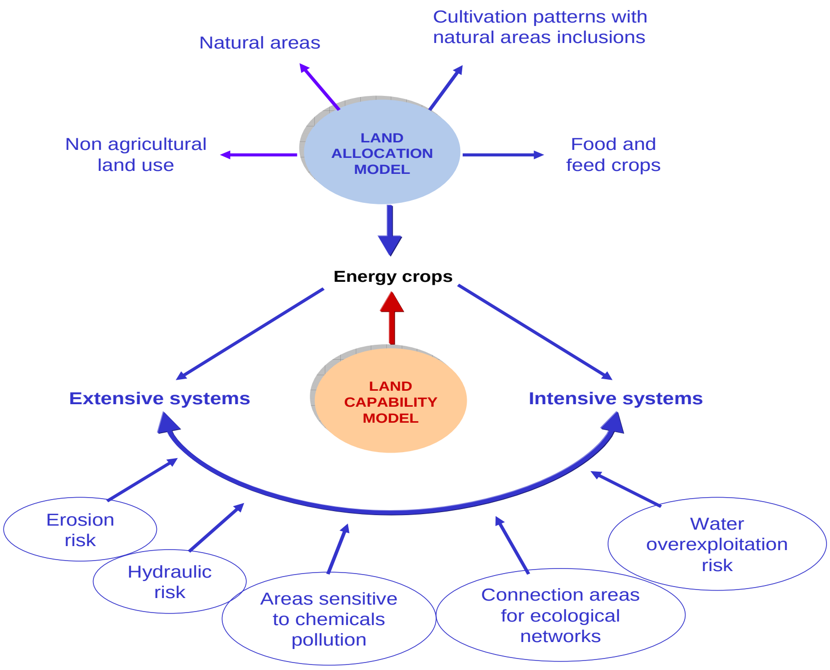


Figure 1. Multi objective land allocation model

4. Results

The results from the GIS-based methodology are represented as maps of suitable areas and locations for energy crops. The estimated potentials under different energy crop penetration scenarios represent the results of the trade-off between the different environmental and energy policy assumptions. The results are aimed at developing the best strategies for the integration of energy crops into existing rural landscapes by selecting promising pathways for sustainable biomass supply and proposing some agri-environment measures to ensure that novel crops do not cause negative impacts. The comparison between a strong energy crop penetration scenario against a lower penetration scenario with less intensive production systems shows the potential positive effects that a right choice of the crop types and cultivation systems may have on the environment in terms of land conservation and balanced management of natural resources. Considering the high biodiversity associated with the UK agricultural land (Hope and Johnson 2003), extensive cultivation systems and integrated landscape planning approach against the risk of land use intensification would provide several benefits in terms of nature conservation in UK.

5. Acknowledgements

This study was undertaken under the auspices of the Marie Curie Early Stage Training Site BIOCONS (European Centre for Biodiversity & Conservation Research). The support through the European Commission's Marie Curie Training Fellowships is gratefully acknowledged.

References

Berndes G (2006). *The contribution of renewables to society*. In: Dewulf, J and Van Langenhove, H (eds.). Renewables-based technology. John Wiley & Sons, Ltd. London, UK.

Edwards RAH, Šúri M, Huld MA, Dallemand JF (2005) GIS-Based Assessment of Cereal Straw Energy Resource in the European Union *Proceedings of the 14th European Biomass Conference & Exhibition Biomass for Energy, Industry and Climate Protection* 17-21 October 2005 Paris

Elghali L, Clift R, Sinclair P, Panoutsou C and Bauen A (2007) Developing a sustainability framework for the assessment of bioenergy systems *Energy Policy* **35(12)** pp 743-768

Freppaz D, Minciardi R, Robba M, Rovatti M, Sacile R, Taramasso A (2004) Optimizing forest biomass exploitation for energy supply at a regional level *Biomass and Bioenergy* (**26)** pp 15-25

Giampietro M, Ulgiati S, Pimentel D (1997) Feasibility of Large-Scale Biofuel Production *BioScience* **47(9)** pp 587-600

Graham R L, Downing M, Walsh ME (2005) A framework to assess regional environmental impacts of dedicated energy crop production *Environmental Management* **20(4) pp** 475-485

Graham RL, Liu W, Downing M, Noon CE, Daly M, and Moore A (1997) The effect of location and facility demand on the marginal cost of delivered wood chips from energy crops: A case study of the state of tennessee *Biomass and Bioenergy* **13(3) pp** 117-123

Hope A, Johnson B (2003) English nature—discussion paper on biofuels *Terrestrial Wildlife Team* www.english-nature.org.uk.

Ignaciuk A, Vöhringer F, Ruijs A and van Ierland EC (2004) Competition between biomass and food production in the presence of energy policies: a partial equilibrium analysis *Energy Policy* **34(10)** pp 1127-1138

Johansson DJA and Azar C (2007) A scenario based analysis of land competition between food and bioenergy production in the US *Climatic Change* **82(3-4)** pp 267-291

Ma J, Scott NR, DeGloria SD, and Lembo AJ (2005) Siting analysis of farm-based centralized anaerobic digester systems for distributed generation using gis *Biomass and Bioenergy* **(28)** 591-600

Masera O, Ghilardi A, Drigo R, and Trossero MA (2006) Wisdom: A gis-based supply demand mapping tool for woodfuel management *Biomass and Bioenergy* **(30)** 618-637

Martelli F, Maltagliati S, Riccio G, Bernetti I, Fagarazzi C, Fratini R, Chiaramonti D, Nibbi L (2002) GIS-based planning tool for Greenhouse gases emission reduction through biomass exploitation in Tuscany *12th European Conference on Biomass for Energy, Industry and Climate Protection* Amsterdam RAI, The Netherlands, 17-21 June 2002

Noon CE and Daly MJ (1996) Gis-based biomass resource assessment with bravo *Biomass and Bioenergy*, **10(2-3)** pp 101-109

Schneider LC, Kinzig AP, Larson ED, and Solorzano LA (2001) Method for spatially explicit calculations of potential biomass yields and assessment of land availability for biomass energy production in northeastern brazil *Agriculture, Ecosystems and Environment* **(84)** pp 207-226

Towers W, Morrice J, Aspinall RJ, Birnie RV, and Dagnall S (1997) Assessing the potential for short rotation coppice in scotland *Proceedings of Environmental Impact of Biomass for Energy, Centre for Agriculture and Environment* Utrecht, The Netherlands, 1997

Voivontas D, Assimacopoulos D, and Koukios EG (2001) Assessment of biomass potential for power production: A GIS based method *Biomass and Bioenergy* **(20)** 101-112

Biography

Patrizia Tenerelli graduated in forestry and environmental science at the University of Bari where she is currently completing her Ph.D.. Her primary interests are rural development and application of GIS for natural resources evaluation. She presently carries out research in the field of land planning for sustainable bioenergy production.

Steve Carver is a senior lecturer in the School of Geography, University of Leeds. His main interests are GIS, environmental modelling and landscape evaluation.

Investigating Spatial Variation of Renewable Energy Potential in the UK

Michael Blom

MSc GIS, Department of Information Science, School of Computing, City University

mike_geomatics@yahoo.com

Keywords: Visualization, Natural, Energy, Spatial, Distribution

1 Introduction

Gordon Brown did it two years ago, and David Cameron did it last March, to great interest. What have they both done? They've both installed means of harnessing natural energy at their private residences. To break opinion down, they've either taken steps towards contributing to a greener future, leading the way in advocating the use of natural energy sources, or they've blindly jumped on the green bandwagon, and become caught up in increasingly popular "eco-cons". Somewhere in between lies the truth, and in finding the truth lies the investigation of the spatial distribution of natural energy.

The statistics are overwhelming and seem to contradict each another more often than agreeing. On December 13th, 2007, *The Times* printed numbers stating that Mr Brown will have recouped his solar panel investment in 100 years, while Mr Cameron is looking at 60 years before he's recouped the cost of his wind turbine (*The Times, 13/12/07*). The next day, the public had picked apart these numbers; the numbers had ignored wind gusts, inflation of energy prices and life spans of the systems (*Political Penguin 14/12/07, R Kyriakides's Weblog 14/12/07*). The varying statistics are in reality a direct result of the wide range of variables inherent in calculating available wind and solar energy.

With such varying landscape in the United Kingdom, the location variable cannot be ignored when considering natural energy production. The estimated energy outputs for a wind turbine on the shores of western Scotland are obviously not going to be applicable to a turbine on a roof in suburban London. Indeed, solar panel customers are warned to de-rate manufacturer's power ratings by 20-40%, based on local atmospheric conditions (*Kemp, p240*). Unfortunately, in order to simplify things for public consumption, small scale natural energy harnessing installations are generally described in very broad terms, which they shouldn't be. With so many variables in flux, attributing a benchmark value to a natural energy system will always be fraught with uncertainty.

This paper investigates a means of advanced geovisualization and user interaction in giving the public a service to determine the potential natural energy that exists in their community, allowing them to see beyond the often inaccurate benchmarked values of small scale solar panels and wind turbines.

2 The Data

2.1 The Stations

The UK Met Office has a system of stations based at sites that are representative of an area up to several tens of km from the station (*Met Office*). In all, the Met Office has almost 60,000 records in their stations table. From this total, 220 stations were used: 141 stations for wind recordings, and 79 stations for solar recordings. Stations were chosen on their recording status being current, and their recordings being made consistently throughout the year. Four wind stations with good readings were removed from the study due to their high elevations, which heavily influenced their readings.

2.2 Recordings at Stations

Stations used in this study had at least 8000 recordings throughout 2006. In ideal circumstances the stations would all have 8760 recordings (24hr x 365 days), and deterministic calculations for 2006 *potential* wind power at each station would be calculated based on hourly wind speed mean averages, and hourly max gust speeds. However, the drawbacks to this technique are numerous; few stations recorded weather stats for every hour of every day, and at none of the stations is gust duration recorded. To deterministically record a site's potential wind power, greater details than the Met Office's hourly recordings would be needed.

3 Means of Calculations for Potential Energy

When considering wind speed distribution, it is more accurate to look at a PDF, or probability distribution function (Barnsley, 2007). Instead of attempting to calculate deterministic wind and solar energy using each individual hourly recording, histograms were created and probabilistic potential energy values were calculated. Analysis using gnuplot allowed fitting of the histogram data to a Weibull curve which allows for the consideration of wind speeds that might occur at less than yearly frequencies by extrapolating values in the curve's tail.

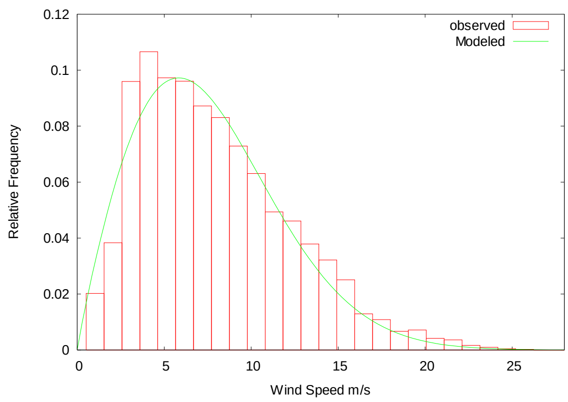


Figure 7: Wind frequency distribution at the UK's highest potential wind energy location, Credenhill, Hereford.

Yearly wind energy output based on mean wind speeds was generated from the histogram. Energy output from wind is proportional to the cube of the wind speed (equation 1); consequently, steady winds produce less power over time than winds prone to gusts (Gipe, 1999). Doubling the wind speed produces 8 times more energy:

Wind Speed (v) = 2 m/s $v^3 = 2*2*2 = 8$
Wind Speed (v) = 8 m/s $v^3 = 4*4*4 = 64$

Ignoring wind gusts implies that less energy can be extracted than it is available in the wind. This extra energy is represented by a factor known as the "energy pattern factor" (Gholam, 2007). This value was found to range from 1.6 to 3.0 in the UK.

The energy potential at each station is calculated by multiplying the energy pattern factor against *P*, where:

$$P = \frac{1}{2}\alpha\rho\pi r^2 v^3 \qquad \text{(equation 1)}$$

Probabilistic solar energy values are more easily calculated than wind, as the Met Office records irradiance energy as power in KW·H. Solar cell efficiency of 15% is applied to yearly values to give resulting potential energy values for each station location.The resulting energy values for each solar and wind station are added as an extra field to the two station tables.

4 Local Weather Stations

In order for a user to get the most accurate idea of potential natural energy resources in their region, the system must take the user's location and consider the network of stations to determine which station from each of the solar and wind station networks is closest. In figure 2, Thiessen polygons help visualize which solar station a location would use.

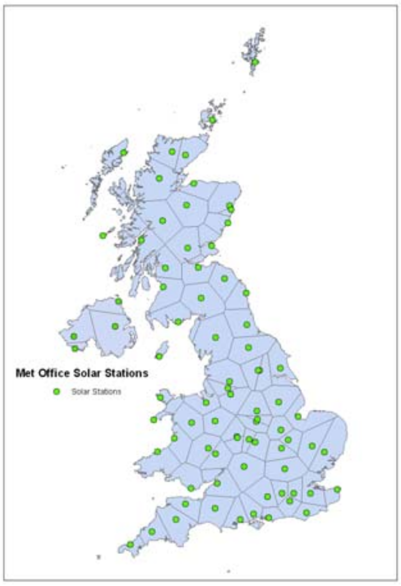


Figure 8: UK solar station regions

The user navigates in the Google Earth geo-browser environment which acts as the interface where the user will identify the location of their residence. Google Earth passes WGSS84 latitude and longitude coordinates to a PHP script. This lat/long value is converted to British or Irish National Grid, to perform distance calculations to determine each of the nearest wind and solar stations.

5 System Functionality

The user is delivered the KML file which opens in Google Earth. The client then enters their postcode in the “find address” toolbar. Within the KML file, a <networklink> tag points towards a PHP script that resides on a server. Apart from a pointer to the PHP script, this <networklink> tag also nests an attribute that determines when the PHP script will be called. In this case, the PHP is called two seconds after the camera has come to a rest, after zooming into the user’s postcode. The <networklink> tag passes the bounding box coordinates of the user’s view to the script, and these coordinates form the basis of the appropriate solar and wind station selections. Once the nearest solar and wind stations are determined, the total KJ output at both locations is queried for. The data is house in a MySQL database, in two separate tables for wind and solar stations.

The energy values returned from the database are wrapped within KML and sent back to the user’s Google Earth browser. These results are displayed in a placemark callout box, styled with text that advises the user of the amount of savings that could have been achieved over the 2006 year.

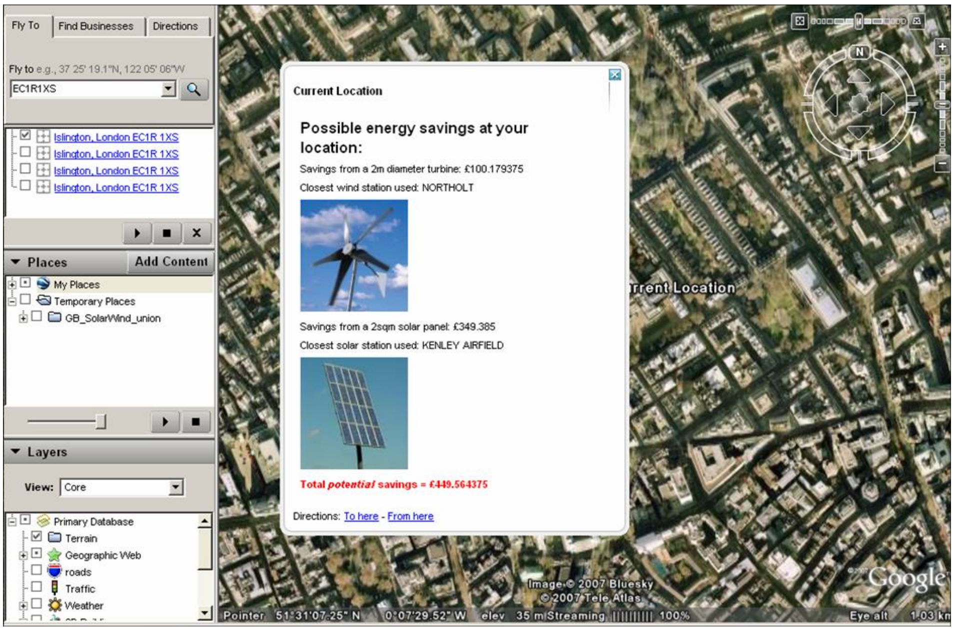


Figure 9: The call-out balloon displays natural energy potential results.

6 Visualizing Energy Potential throughout the UK

The system created in this project allows querying for local information regarding natural energy potential, but it is also beneficial to visualize energy potential across the UK in order to determine trends and potential errors over a larger area.

Potential solar and wind energy values were extruded to create thematic 3d displays to assist in the visualization process. Anomalies can be studied and their possible explanations considered. The polygons shown in figure 2 are shown extruded in figure 4, viewed in Google Earth. ESRI's 3d analyst toolbox contains a *lyr to kml* converter which transforms the shapefile to the open source kml format, viewable in any virtual globe.

To no surprise, the south parts of England provide the greatest return from solar energy panels, while Western Wales contains the best return from wind turbine investment. It was also clearly seen that wind stations at higher elevations reported higher potential energy, and locations along the coast tend to provide more hours of sunshine in addition to generally higher energy-density winds.

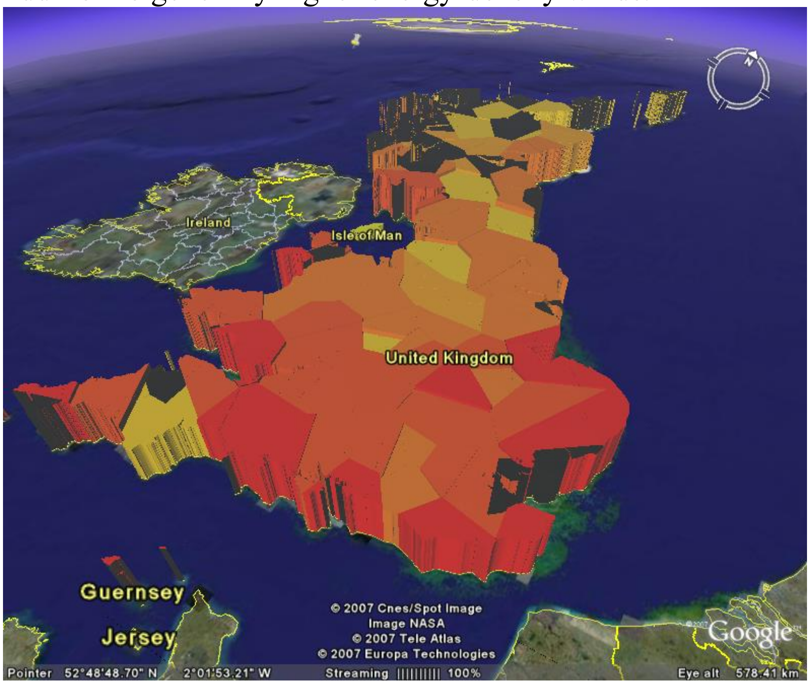


Figure 10 : Potential solar energy throughout Great Britain.

7 Conclusion and Recommendations

Virtual globes provide a lightweight geo-browser to enhance GIS visualization. The benefits of their usability and accessibility have been utilized in this case to provide insightful and enlightening information to the public who are inundated with natural energy statistics that might not necessarily be relevant to their locale. By allowing the user to enter their postcode, natural energy information that is region specific can be obtained to give a better understanding of the estimated amount of natural energy that is available to be harnessed from the wind and sun.

While aggregating all UK locations to 141 wind or 79 solar stations improves localisation of renewable energy information, in order to get the highest level of precision measurements, meteorological instruments would be required installed on individual rooftops throughout the UK to consider neighbouring building structures, nearby trees, and other impediments to wind and solar energies. There is potential here for user generated meteorological data to contribute to even higher resolution datasets.

References

Barnsley, Michael J, 2007. *Environmental Modeling, a Practical Introduction*. (London: CRC Press)

Gholam, Riahy H., 2007. *A New Method for Calculating the Power Output of a Wind Turbine based on Wind Speed Variations and the EPF.* (Mashad, Iran)

Gipe, Paul. *Wind Energy Basics: A Guide to Small and Micro Wind Systems.* (United States, Chelsea Green Publishing Company)

Jowit, Juliette. Home wind turbines 'too feeble'. 2008. *The Weekend Observer,* 06/01/08, page 21.
Kemp, William. H. 2005. *The Renewable Energy Handbook.* (Tamworth, Ontario. Aztext Press)

Kyriakides, Robert. 2007. Gordon Brown's solar panel and David Cameron's wind turbine. 14/12/2007. *Robert Kyriakides's Weblog* [online]. [Accessed 02/01/8]. Available from World Wide Web: http://robertkyriakides.wordpress.com/2007/12/14/gordon-browns-solar-panels-and-david-camerons-turbine/

Penguin, Political. 2007. Disappointed. 14/12/2007. *Political Penguin Blog* [online]. [Accessed 02/01/08]. Available from World Wide Web: http://www.politicalpenguin.org.uk/blog/p,275/

UK Met Office. *Abridged MIDAS Handbook*, Maintained by Desktop and Databases branch. Available from World Wide Web: http://badc.nerc.ac.uk/data/surface/ukmo_guide.html

Unknown author. 2007. Up on the roof – a power struggle. *The Times*, 13/12/2007, page 1

Biography

Michael Blom has most recently worked at Willis at a GIS developer, supporting catastrophe management teams. He has also worked as part of the London Cycling Network. His main interest in GIS is using the numerous emerging geographic technologies to better visualize trends in helping to make more informed decisions.

Discovering Spatially Multiway Collocations

Didier Leibovici[1], Lucy Bastin[2] and Mike Jackson[1]

[1]Centre for Geospatial Sciences, University of Nottingham, UK
[2]School of Engineering & Applied Science, Aston University, UK
Tel. (+44(0)115 846 8408) Fax (+44(0)115 951 5249)
Didier.leibovici@nottingham.ac.uk

KEYWORDS: spatial statistics, co-occurrences, multiway data, marked point process

1. Introduction

Analysing geographical patterns by collocating objects, events or attributes, has a long history in applied science such as ecology or epidemiology. In spatial statistics, this is often associated with marked point processes, (Diggle, 2003). The problem of identifying patterns of co-occurrences, or at least establishing the existence of certain structures at some scales, is usually addressed by plotting the spatial dependence functions (cross-K functions) against distance, and by testing them against complete randomness or other generated stationary processes, (Ripley, 1977); Schlater and Diggle, 2004). This paper will investigate optimum spatial representations of collocations in terms of lack of spatial independence between two categorical variables or more. This can be complementary to the approach described above. The framework, using a tensorial data representation developed in Leibovici and Sabatier (1998); Leibovici (2007) allows more detailed collocation by analysing co-occurrences of higher order than pairs.

2. Measuring Spatial coocurrences,

Let $i = 1, ...I$ be categories of a variable vI, $j = 1, ...J$ be categories of a variable vJ, $k = 1, ...K$ be categories of a variable vK, and let $s = 1, ...S$ be locations where one can record either vI, vJ or vK but also in some case studies all of them and sometimes more than one record, depending on the geometrical object and the semantic associated with the locations. The variables vI, vJ, vK describe a general "event": e.g.**(i)** a person of age i with social class level j, diagnosed for a certain disease k and living at location s; **(ii)** a plant species i, on a soil class j, at location s with annual rainfall **k**, or **(iii)** a crime of type i, at time slot j, in a zone of wealth class k of location s. So we are looking for associations of vI, vJ, vK variables in their spatial co-occurrences, either as multivariate observations on a spatial domain or as already collocated observations of different variables, or both. Clearly **(i)** is multivariate on the persons, **(ii)** is a collocation of different measurements **(iii)** is a mixture of both. One can argue that **(ii)** could be seen as multivariate characteristics of the location.

2.1. Counting pairs, triples

With, $n_{ii'}$, n_{ij}, and n_{ijk} the number of collocations of events or marks i and i', or i and **j**, or **i**, **j** and **k** when the distance of collocating events is implicit, and $n^d ij$ when the distance **d** of collocating events is explicit, a general definition of collocations is, ***Cg***: *a collocation of marks or labels* ***{i, j, k}*** *is recorded if, the distance between the locations* ***{s, s', s''}*** *(which may be equal), all together expressing the labels* ***{i, j, k}***, *is at most* ***d***. The standard way of computing the collocations (here without edge corrections) for two labels is:

$$n_{ij}^d = \sum_S \#_{C(s_i,d)}\{j\} = \sum_S \#_{C(s_j,d)}\{i\} = n_{ji}^d \qquad (1)$$

where *#C(si,d){j}* means the number of "events" or marked locations ***j*** at maximum distance ***d*** from a location labelled ***i***: the number of s_j found in a circle of radius **d** and centre s_i. Adding flexibility about the searched geometry area (***G(s_i,d))***, with **d** being the buffer size of the geometry s_i, and also about the way ***(O)***the occurrence is recorded "within" the geometry (as depending also on the geometry of the ***j*** mark), gives a more general formula:

$$n_{ij}^d = \sum_s \mathcal{OG}_{(s_i,d)}(\{s_j\}) \tag{2}$$

Depending on the choices for ***O*** and ***G(.,d)*** the collocation value may not be symmetrical. When addressing 2nd-order collocations in multitype point patterns, Lotwick and Silverman (1982), generalise cross-Ripley's K function and work with the labels two by two, and Baddeley and Turner (2005), use subsets of categories *e.g.* one from ***vI*** and one from ***vJ***. These methods are used to test for clustering under assumptions for the spatial process such as stationarity and isotropy, which can be alleviated for particular studies (Diggle et al., 2007). Using (dis)similarities of order three, Heiser and Bennani Dosse (1997) demonstrated improved grouping or sorting descriptions and pattern recognition.

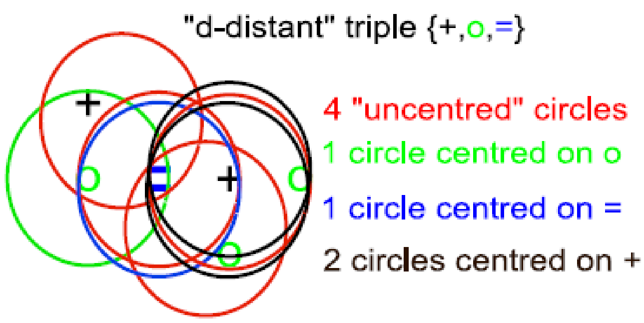


Figure 1. Symmetrical and asymetrical ways of counting collocations of a triple with circles of radius ***d*** (not necessarily verifying definition ***Cg***)

For three or more marked events the standard formula has to be rewritten if symmetry is to be maintained:

$$\begin{aligned} n_{ijk}^d &= \sum_S \#_{C(d)}\{i,j,k\} \qquad (3)\\ &\neq \sum_s \#_{C(s_i,d)}\{j,k\} \neq \sum_s \#_{C(s_j,d)}\{i,k\} \neq \sum_s \#_{C(s_k,d)}\{i,j\} \\ n_{ijk}^d &=_{Cg} \sum_S \#_{C(s_i,d)\cap C(s_j,d)\cap C(s_k,d)}\{i,j,k\} \end{aligned}$$

As seen in Figure 1, the non-location-centred circle method is symmetrical but collocation counts from asymmetrical methods based on marked-location-centred circles are different. In fact a triple, collocation will be recorded if the intersection of the three circles contains the three labels. Different counting methods and interpretations can introduce a range of methods suitable for discovering spatial collocations. Considering **S** itself as a variable instead of marginalising through it, will allow spatial interaction of associations between categorical variables.

These counts may be very low, depending on the study as well as the approach taken: multivariate observations or collocated observations. Can a collocation approach (case **(ii)**) be informative for multivariate observations? Sample sizes needed for a multivariate approach (case **(i)**) are larger! The approach of this paper being multitype, multi-variable, without assumptions on the process distribution, the focus is on a very simple and well known statistic: the χ^2 measure for lack of independence, with the aim of demonstrating its potential use with appropriate tables of counts for multiway collocation.

3. Co-occurrences and tensorial framework

The tensorial framework developed previously is used here to analyse and represent data "summaries" in a similar approach to multidimensional analysis. This framework allows some flexibility about representing or modelling data information within an array of any dimensions, affording new ways to discover spatial pattern. From Figure 2 some two way tables can be analysed by correspondence analysis, (method **CA2OO**). The extension of correspondence analysis to a k-way table provided in Leibovici (2007, 2000) can be used to analyse triple (or more) collocations (method **CAkOO**). This allows us to analyse multiway dissimilarities as in Benanni-Dosse's thesis, (published partially in Heiser and Bennani Dosse, 1997), with generalised unfolding metric multidimensional scaling.

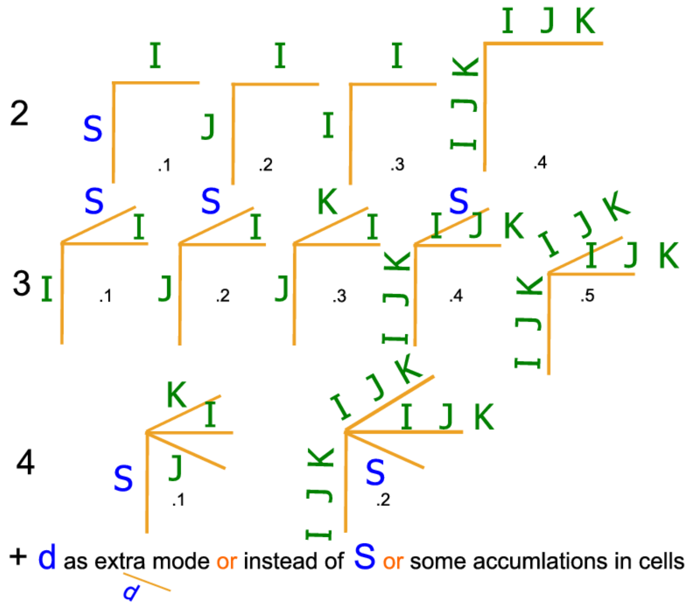


Figure 2. k-way Tables for collocation discovery analysis: all 2-modes, 3-modes, 4-modes possibilities with cell values as counted collocations as explained in the text: *I, J, K* categorical variables, *S* spatial locations, d distance for collocation event.

3.1. Two-way Correspondence Analysis

Correspondence analysis of a two-way contingency table (**FCA**) $p_{ij} = n_{ij}/N$ is achieved by performing the Principal Component Analysis, **PCA**, (or generalised **PCA**) of the following triplet, (Escoufier, 1987); Dray and Dufour, 2007):

$$(D_I^{-1}PD_J^{-1},\ D_I,\ D_J) \tag{4}$$

with D_I and D_J diagonal metrics containing vector margins $p_{.i}$ and $p_{.j}$; where a **PCA** of a triplet
(X, Q, D) with X a data matrix $n \times p$, Q a $p \times p$ metric on the rows (or in the column-space) and similarly D a $n \times n$ metric on the columns (or in the row-space), is generalising a standard **PCA** by diagonalising with Q-normed vectors the matrix ${}^tXDXQ$ equivalent, to the covariance matrix if X is column-centred, $D = 1/nIdn$ and $Q = Idp$, or to the correlation matrix if instead of the identity metric $Q = diag(1/var_1 \cdots 1/var_p)$.
The measure of lack of independence is linked to the analysis by:

$$1 + \frac{\chi^2}{N} = \sum_{ij} \frac{(p_{ij} - p_{i.}p_{.j})^2}{p_{i.}p_{.j}} = \sum_{ij} p_{i.}p_{.j}(\frac{p_{ij}}{p_{i.}p_{.j}})^2 = \sum_{s} \sigma_s^2 \quad (5)$$

where the $\boldsymbol{\sigma_s^2}$ are the eigenvalues. $\boldsymbol{\sigma_0 = 1}$ with components equal to unit vectors in their respective spaces.

3.2. Correspondence Analysis of k-way cOOccurrences

For a **CAkOO** analysis we propose to use a generalisation of correspondence analysis to k-way tables, **FCA-kmodes**, which is a specific **PTA-kmodes** just as **FCA** is a specific **PCA**. The **PTA-kmodes** decomposition of a k-way table was already applied in a spatial context in Leibovici et al. (2007). With similar notations for a three-way table $\boldsymbol{I} \times \boldsymbol{J} \times \boldsymbol{K}$, one performs the **PTA-3modes** of the quadruplet:

$$((D_I^{-1} \otimes D_J^{-1} \otimes D_K^{-1})P, \quad D_I, \quad D_J, \quad D_K) \quad (6)$$

Equation 7 shows that three-way independence can be orthogonally decomposed into deviations from independence for the two-way margins of the three-way table, and a three-way interaction term. Each two-way margin's deviation from independence is reminiscent of (simple) correspondence analysis:

$$\begin{aligned}
\frac{\chi^2}{N} &= \sum_{ijk} p_{i..}p_{.j.}p_{..k}(\frac{p_{ijk} - p_{i..}p_{.j.}p_{..k}}{p_{i..}p_{.j.}p_{..k}})^2 \quad (7)\\
&= \sum_{jk} p_{.j.}p_{..k}(\frac{p_{.jk} - p_{.j.}p_{..k}}{p_{.j.}p_{..k}})^2 + \sum_{ik} p_{i..}p_{..k}(\frac{p_{i.k} - p_{i..}p_{..k}}{p_{i..}p_{..k}})^2 + \sum_{ij} p_{i..}p_{.j.}(\frac{p_{ij.} - p_{i..}p_{.j.}}{p_{i..}p_{.j.}})^2\\
&+ \sum_{ijk} p_{i..}p_{.j.}p_{..k}(\frac{p_{ijk} - \delta_{ijk}}{p_{i..}p_{.j.}p_{..k}})^2
\end{aligned}$$

2-way margins analysis from decomposition of 3-way co-occurrences are not equivalent to 2-way co-occurrence analysis as in correspondence analysis of contingency tables. Nonetheless, considering co-occurrence counts and contingency tables one can symbolically write for a multiway table like 3.3 or 4.1, on Figure 2:

$$\lim_{d \to 0} CAkOO = FCAk \quad (8)$$

4. Examples

Different examples for CA2OO and CAkOO methods will be shown for the presentation using datasets from the literature but also with datasets from recent epidemiological studies:

- *langsing* from **spatstat**
- an epidemiological dataset of persons (5 age groups) located to home postcode, having contracted a bacterium resistant (R) or sensitive (S) to an antibiotic at one of three successive time periods,
- a similar epidemiological dataset where the resistance to antibiotics has been classified into 3 or 20 categories and where patient gender is known.

On each dataset, we compared a range of possible analyses from Figure 2, and will demonstrate the added value of the approach for classical 2[nd] order analysis.

References

Baddeley, A. and Turner, R. (2005). spatstat: An r package for analyzing spatial point patterns. *Journal of Statistical Software*, **12(6)**, 1–42.

Diggle, P., Gomez-Rubio, V. Brown, P., Chetwynd, A., and Gooding, S. (2007). Second order analysis of inhomogeneous spatial point processes using case-control data. *Biometrics*, **63**, 550–557.

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*. Hodder Arnold, London.

Dray, S. and Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22(4),** 1–20

Escoufier, Y. (1987). The duality diagram : a means of better practical applications. In P. Legendre and L. Legendre, editors, *Development in numerical ecology*, pages 139–156. NATO advanced Institute, Springer Verlag, Berlin.

Heiser, W. and Bennani Dosse, M. (1997). Triadic distance models : axiomatization and least squares representation. *Journal of Mathematical Psychology*, **41**, 189–206.

Leibovici, D. (2000). Multiway Multidimensional Analysis for Pharmaco-EEG Studies. Report initiated at SANOFI-RECHERCHE , TR00DL2, FMRIB Centre, University of Oxford, UK.

Leibovici, D. (2007). PTAk: Principal Tensor Analysis on k modes. *Contributing R-package*, version 1.1-16.

Leibovici, D. and Sabatier, R. (1998). A Singular Value Decomposition of k-Way Array for a Principal Component Analysis of Multiway Data, PTA-k. *Linear Algebra and Its Applications,* **269**, 307–329.

Leibovici, D., Quillevere, G., and Desconnets, J.-C. (2007). A Method to Classify Ecoclimatic Arid and Semi-Arid Zones in Circum-Saharan Africa Using Monthly Dynamics of Multiple Indicators. *IEEE Transactions on Geoscience and Remote Sensing*, **45(12)**, 4000–4007.

Lotwick, H. and Silverman, B. W. (1982). Methods for analysing spatial processes of several types of points. *Journal of Royal Statistical Society*, **B(44)**, 406–413.

Ripley, B. (1977). Modelling spatial patterns. *Journal of Royal Statistical Society*, **B(39)**, 172–212.

Schlater, M. Riberio, P. and Diggle, P. (2004). Detecting dependence between marks and locations of marked point process. *Journal of Royal Statistical Society*, **B(66),** 79–93.

Biography

Dr Didier 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 for cross-scales for integrated modelling applications within an interoperable framework chaining web services.

Lucy Bastin is a Lecturer in GIS at Aston University. After a PhD on urban plant metapopulations, and research into fuzzy classification / uncertainty visualisation at Leicester University, she spent 3 years as a GIS software developer. Her current research interests include Web Processing Services for automatic interpolation, and spatial epidemiology.

Pr.Mike Jackson is head of department. He worked for the geospatial industry (QinetiQ, Hutchison 3G, Laser_Scan) and in research for NERC and as investigator for NASA. Mike is non-executive director of the Open Geospatial Consortium Europe, with interests in combining new technologies such as location based services with geospatial intelligence.

Mind the Gap:
Reuniting the Tube network with underlying spatial themes

Craig Foot[1], Pragya Agarwal[2]

[1]Deloitte & Touche LLP, 180 Strand, London
Email: cfoot@deloitte.co.uk
[2]Department of Civil, Environmental and Geomatic Engineering, University College London, London, Email: pagarwal@ge.ucl.ac.uk

KEYWORDS: graph, networks, nodes, spatial, usability

1. Introduction

In this project, graph theory was applied to the analysis of an urban transport network – specifically the London Underground. In modelling the network as a graph, a number of measures become available for use in an analysis of the network structure. The use of these measures will allow a comparison of the system with that of other world cities, as well as providing useful information for the planning of network expansion or the current running and maintenance of the network. A wide range of socio-economic indices also exist that are related to boundaries covering the spatial extent of the network. Using recent research by Agarwal et al. (2006), the project demonstrated an application of graph centrality indices in the context of London Borough boundaries in order to provide a link between a network analysis and data that is applicable to its spatial extent. It is expected that an improved understanding of the spatial structure underlying the tube network will enable enhanced usability of the transport network, support navigation choices in the urban area as well as aid future planning decisions.

2. Methodology

The first stage involved the development of a network model on which to perform the analysis. The aim was to produce a representation of the network in its simplest form – as an undirected graph with edges of equal weight. Social network methods were adapted to be applied in this particular case to a physical network and its underlying spatial structure. Measures of centrality given by graph theory were applied to this model, specifically centrality indices of degree, closeness, and betweenness. The degree of a particular node describes the number of nodes that are adjacent to it (connected by one link, or edge). The concept of betweenness centrality was introduced by Bavelas (1948) and formalised by Freeman (1977) in Equation 1.

$$c_B(i) = \sum_{k \neq j \neq i} \frac{S_{jk}(i)}{S_{jk}} \tag{1}$$

Here, the betweenness centrality of node i can be found as the ratio of shortest paths between nodes j and k that i is situated on to the total number of shortest paths to be found between j and k.

Another measure of centrality describes the how close a node is to other nodes in the network. In the context of transport networks it can be seen as an indicator of how accessible all destinations on the network are from a particular node origin. Sabidussi (1966) defines it as shown in Equation 2.

$$c_C(i) = \frac{1}{\sum_{j \in V} d(i,j)} \tag{2}$$

Here the closeness centrality of a node i is the reciprocal of the sum of all the distances between the origin i and every possible destination on the network represented by j, where j belongs to the set of nodes V. Distance in this sense refers to the graph theoretical distance, or the weight of each edge where weight represents the cost (e.g. distance, time or money) of traversing an edge. In an unweighted graph all edges have a weight of 1.

Data for the project was sourced from the London Tube map and from the online community in the form of contributors to websites such as OpenStreetMap. The positions of Tube stations are available from a number of sources, such as Ordnance Survey (OS) maps or Google Earth placemarks. A tabulated set of latitude and longitude values were located online and subsequently downloaded and imported into a spreadsheet. The locations were considered accurate enough for the purposes of the project, and a visual inspection of a sample of the data was used as a check. To ensure completeness, the downloaded coordinates were compared to the station names given in the station index of the January 2007 Tube Map. A number of versions of this map are available on Transport for London's website, in Portable Document Format (PDF), which allows for simple extraction of the complete list of station names from an official source. The station names were manipulated using a spreadsheet package, and matched to the latitude and longitude positions given in the downloaded dataset. A benefit of this method was that the zone attribute for each station is also included in the map index, and therefore was the first attribute to be added to the dataset. Other attributes can be easily added to the stations using the method of spreadsheet manipulation, and then visualised in a GIS such as MapInfo or ArcGIS.

Figure 1. Visualisation of the geographic patterns in the tube network

Network Analysis software traditionally employed for social network analysis was also investigated as a tool to represent and analyse the network. The values were normalized to the range 0 – 1 in order for future work to compare the Tube graph with other networks. Brandes & Erlebach (2005) was used as a reference for the normalization of the centrality indices. Figure 1 shows the geographic patterns of the London tube network. Connectivity in the tube stations, visualized in figure 1 as nodes, is analyzed as the degree of closeness in the stations. Figure 2 highlights discrete nodes on the graph that have a degree other than two, particularly highlighting the central concentration of high degree nodes and a peripheral group of lower degree nodes that provide a less dense zone of connectivity. This visualisation of the geographic dispersion of the transport network in London makes it apparent that the central parts of the city are better served and connected by the tube network than the outlying urban area.

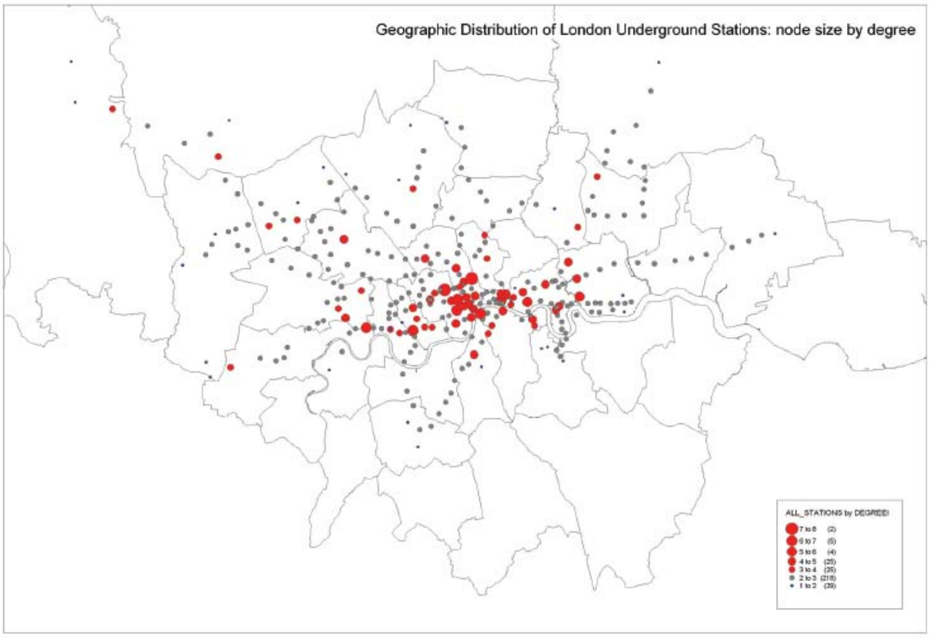


Figure 2. Spatial distribution and connectivity in the tube network

A different approach was also taken to relate the network data to the administrative boundaries. The concepts of geographic dispersion and heterogeneity after Agarwal et al. (2006) were applied to each borough where network data existed. A value for the geographic dispersion of each node was found and averaged by borough. Figure 4 shows an example of the thematic map output that this generated.

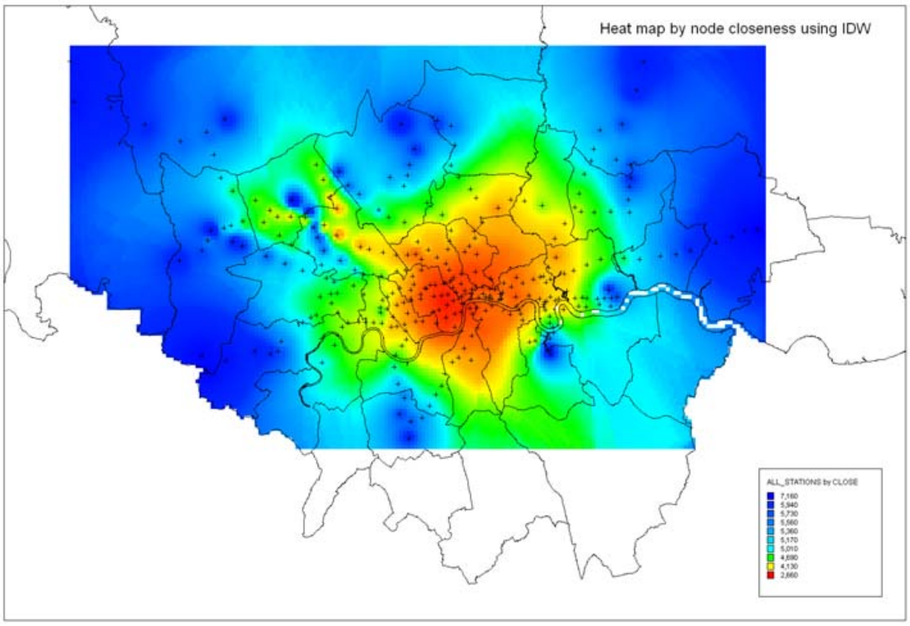


Figure 3. Closeness centrality of Tube stations

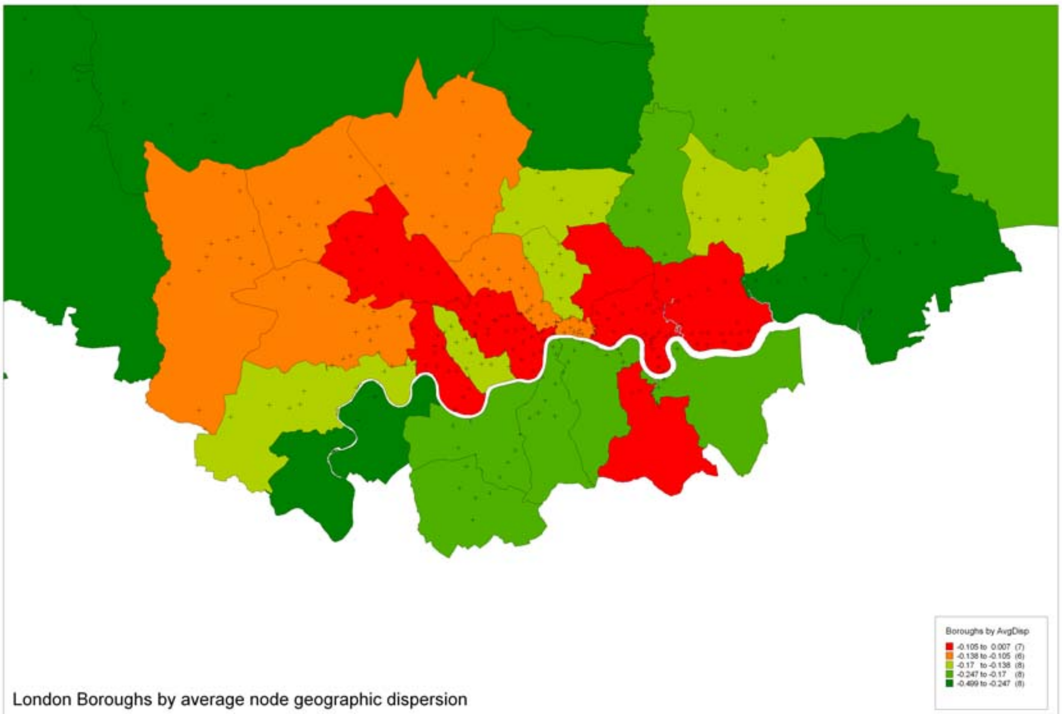


Figure 4. Average node geographic dispersion by borough

3. Analysis

The maps above provided some significant indicators. Firstly, the extent of the most central area can be appreciated as the warmest part of the image, and it is interesting to note that the strongest improvements to the closeness centrality of the system occur when Tube lines work in pairs, with one cutting out stations served by the other. This is illustrated by the fact that the mutual presence of the Metropolitan and Jubilee lines in the North West has produced local hotspots such as Harrow-on-the-Hill, Finchley Road, and Wembley Park. If the graph was weighted by travel time, the same effect would be achieved if lines were to run 'fast trains' that cut out intermediate stations – something that is only possible where the track permits it.

Cooler areas shown in Figure 3 indicate areas that provide the highest travel cost to other nodes on the graph. In the case of the Bakerloo line, poor centrality can be attributed to the high number of edges and few interchanges in this area. This highlights one of the first problems encountered with the over-simplification provided by the model: by not taking into account edge weights the closeness centrality equation is biased against sections of track that have a high station count. In fact, such sections may be closer to the rest of the graph in real terms if they are close together or a high speed is obtained en route. By weighting each edge by 1 it is assumed that the distances between stations, reflected in the travel time, is uniform throughout the network – which for the Tube is not the case. More data is required in order to weight the edges properly. The project's preliminary study was successful in collecting the track length distances between stations in Zone 1, after Jacobs (2002) and future work is going to look at extending the methodology developed here to analyse it.

The heat map for closeness centrality can be viewed as a starting point to highlight London boroughs that are good value for money in terms of travelling on the Tube. Such data could be developed to include a consideration of average house prices so that a query could return the set of boroughs or geographic areas that minimise house price or cost of living and maximise network accessibility. A heat map of the shortest path travel cost to a destination such as the workplace would highlight suitable origin locations that optimise travel cost in time, distance or money. An application of this information would be useful to a commuter who wished to live in an area that minimised the travel cost to work based on a set of constraints such as a house price or fare range.

4. Conclusion

The project has succeeded in providing a model of the London Underground as an undirected graph in its simplest form, providing a foundation dataset that allows future work to develop and reify as required for specific analyses. Data has been collated to provide a set of three centrality measures for all stations on the network: degree, closeness, and betweenness. This enables each station to be given a ranking based on each centrality measure and when joined with the list of geographic station locations, the distribution of centrality can be visualised.

The project has also provided a normalisation for each of the centrality measures so that the network can be compared firstly to networks of other transport modes, such as the London bus or tram network that shares the geographic space, and secondly to transport systems elsewhere in the world. The centralisation of the graph has been derived using normalised node centrality values so that an idea of network vulnerability – gauged in terms of its reliance on central nodes – can be clearly deduced. An additional benefit is that using normalised values to find graph centralisation enables easy, reliable comparison with other networks.

The concept of centrality has been extended to the geographic space by the creation of centrality heat maps. This decision in the development of the project was taken because extending network analysis techniques to a continuous geographic space, or one partitioned by boundaries, enables the application of GIS techniques to explore a wide range of queries that can be built and processed in a GIS, whilst taking advantage of proven data management techniques for spatial data. This approach combines well documented network analysis methods using graph theory with the flexible and extensible environment provided by GIS software to solve questions relating to areas such as location-allocation, access to culture, transport policy, and public security to name but a few. Each heat map consists of a grid of interpolated centrality values using the inverse distance-weighted method of interpolation. This process expands the discrete centrality values associated with the nodes to a more continuous space. The net benefit here is that the visualisation is improved when supplementary geographical contexts are overlaid. The centrality of the geographic space surrounding the graph has been further developed by the use of heterogeneity and dispersion indices to indicate the centrality of a set of partitions of the Greater London area. he example offered in this project was the use of Greater London borough boundaries, but others could have been deployed such as postcode districts, fare zones or isochrones radiating from a central point or destination.

5. References

Agarwal P, Béra R, Claramunt C (2006) A Social and Spatial Network approach to the investigation of research communities over the World Wide Web, *GIScience 2006, Lecture Notes in Computer Science*, 4197: 1-17.

Bavelas A (1948) A mathematical model for group structure, *Applied Anthropology (Human Organisations)*, 7: 16-30.

Borgatti S P (2002) *NetDraw: Graph Visualization Software.* Harvard: Analytic Technologies.

Brandes U, Erlebach T (2005) Network Analysis: Methodological Foundations. *Lecture Notes in Computer Science*, 3418: 1-471.

Freeman L C (1979) Centrality in social networks conceptual clarification, *Social Networks*, 1(3): 215-239.

Jacobs G (2002) *Railway Track Diagrams: 5 England South and London Underground.* TRACKmaps, Bradford-on-Avon. 64 pages.

Sabidussi G (1966) The centrality index of a graph, *Psychometrika*, 31: 581-603.

Biography

Craig Foot completed his BEng in Geospatial and Environmental Information Management at UCL in 2007. He is currently working for Petroleum Services at Deloitte where he assists with GIS development and related data management issues. His primary research interests lie in using relative adjacency and graph theory measures to analyse the effect of networks on demographics, the management of spatial data workflows and GIS usability.

Pragya Agarwal is a Lecturer in Geographic Information Science at University College London. Her primary research interests are in social and spatial networks, personalisation, location-aware systems and ontologies.

Automated schematization using memetic algorithms

Jerry Swan[1], Suchith Anand[1], Mark Ware[2], Mike Jackson[1]

[1]Centre for Geospatial Science, University of Nottingham, UK
[2]Faculty of Advanced Technology, University of Glamorgan, Wales, UK
Tel: +44(0)115 846 8411
Email: Jerry.Swan@nottingham.ac.uk ; URL: http://www.nottingham.ac.uk/cgs/

KEYWORDS: Automated schematization, optimization, memetic algorithms

1. Introduction

Mapping is a way of visualizing parts of the world and maps are largely diagrammatic and two dimensional. There is usually a one-to-one correspondence between places in the world and places on the map, but while there are limitless aspects to the world, the cartographer can only select a few to map (Dorling, 1996). Map generalization is required when there is a need to represent geographic information that is captured at large scale in a derived form at a smaller scale (Buttenfield and McMaster, 1989). The potential applications, and hence importance, of automated map generalization has increased tremendously with the advent of digital geographic datasets. Much work has been carried out in recent years, and considerable progress has been made. This is evidenced by the many academic papers published on the subject (e.g. Weibel, 1995; Weibel and Jones, 1998; Jones and Ware, 2005), the various working groups that have been set up (e.g. ICA Commission on Generalisation and Multiple Representation) and the increasingly advanced and useful generalization functionality now being found in commercial GIS software. However, many tasks associated with map generalization have proven difficult to automate and many research challenges remain.

Generalization by vertex displacement has been investigated using a number of metaheuristic techniques, including simulated annealing (Ware et al, 2003a), genetic algorithms (Ware et al, 2003b) and tabu search (Ware et al, 2002). Schematization by vertex displacement using simulated annealing is the subject of previous research (Anand,2007), In this paper, we examine the application of memetic algorithms to automated schematization in particular for water network modelling application.

2. Automated schematization

Manual generation of schematic maps requires considerable effort by a skilled cartographer, who must undertake an iterative labour intensive approach in generating the schematic output by hand. In computer assisted method, graphics software is used to undertake the schematic drawing by computer using raster maps scanned as input. Tools within the graphics package are used for drawing and editing etc. This again is an iterative process, though better quality results can be obtained much more quickly. In the automatic production of schematic maps vector based source datasets are simply input to suitable algorithm/software application, together within any control parameters that might be required. The basic steps for generating schematic maps are to eliminate all features that are not functionally relevant and to eliminate any networks (or portions of networks) not functionally relevant to the single system chosen for mapping. All geometric invariants of the network's structure are relaxed except topological accuracy. Routes and junctions are symbolized abstractly (Waldorf, 1979).

Elroi (1988) refined the process by adding three graphic manipulations, although implementation details and results were not given. First, lines are simplified to their most elementary shapes. Next, lines are re-oriented to conform to a regular grid, such that they all run horizontally, vertically or at a forty-five degree diagonal. Third, congested areas are increased in scale at the expense of scale in areas of lesser node density. Though Elroi in this paper has listed the theory on schematization, the actual real-life implementation was not given. The first step in the process is line simplification, which can be achieved using an algorithm such as that of Douglas and Peucker (1973). Care must be taken when performing this step to avoid the introduction of topological errors; this can be achieved most easily by making use of topology preserving variants of the Douglas-Peucker algorithm, such as that presented by Saalfeld (1999). Steps two and three are the key components of the process, and their automation has been the focus of previous work by several researchers.

3. The algorithm and software implementation

For a period in the late 1980s, Genetic Algorithms was perhaps the optimization technique of first choice. In (De Jong, 1993), it was shown that GAs were "satisficers" rather than "optimizers", i.e. they are useful for finding "near optimal" solutions, but cannot be relied upon to discover global optima. This finding was anticipated in practice by Grefenstette who augmented a GA by applying local search to each population member prior to evaluation (an analogous method can in fact be credited to (Kaufman, 1967). The results obtained in (Swan et al.,2008) for schematization by vertex displacement provide strong motivation for the application of Memetic Algorithms to the problem.

The generic memetic algorithm as presented below is essentially the same as that of (Grefenstette, 1987).For the schematization-specific genetic algorithm components, we adapt the Displacement Vector Template approach of (Ware et al.,2003) since we are only concerned with mapping vertices to a grid, we are able to represent the neighbourhood of a vertex by an integer value that maps to the grid coordinates. The local search component drives each population member to a local optimum by employing steepest-ascent hill climbing over the nine possible 1-ply grid positions of all the vertices (giving e.g. 900 operations for each iteration for a 100 vertex map).

```
Schematic MemeticAlgorithm():
Begin
Initialise population
Do
    Evaluate each member of the population
    Create new members by crossover and mutation
    Apply local search to new members
    Evaluate the new members
    Replace existing population members with new members
Repeat until termination condition is reached (e.g. time or #iterations )
Return the best population member ever generated as the solution

End
```

Based on the characteristics of a schematic map identified, the schematic map algorithm presented here considers seven constraints. Additional constraints can be added if required.

– Topological: The original network and derived schematic map should be topologically consistent;

– Orientation: If possible, network edges should lie in a horizontal, vertical or diagonal direction;

– Length: If possible, all network edges should have length greater than or equal to some minimum length (to ensure clarity);

– Clearance: If possible, the distance between disjoint features should be greater than or equal to some minimum distance (to ensure clarity);

– Angle: If possible, the angle between a pair of connected edges should be greater than or equal to some minimum angle (to ensure clarity).

– Rotation: An edge's orientation should remain as close to its starting orientation as possible

– Displacement: Vertices should remain as close to their starting positions as possible (Anand, 2006)

The software prototype has been implemented based on the memetic algorithm described. The algorithm is implemented in Java as a Web Feature Service. A Java-client application (Figures 1 and 2) has also been created to act as a workbench for demonstration and research purposes. Water pipeline networks are good examples of a practical application of schematic maps, where for example they can also a useful visualization tool for water engineers to help understand and analyze the hydraulic conditions of a network. Once connectivity details can be generated by the network modellers using this schematization technique it will give help water modellers to visualize mental maps of true connectivity from the derived schematic and this can knowledge can be used when carrying out analysis and design with network modelling. The datasets used for the study are water network datasets from DfID KaR Project R6872.

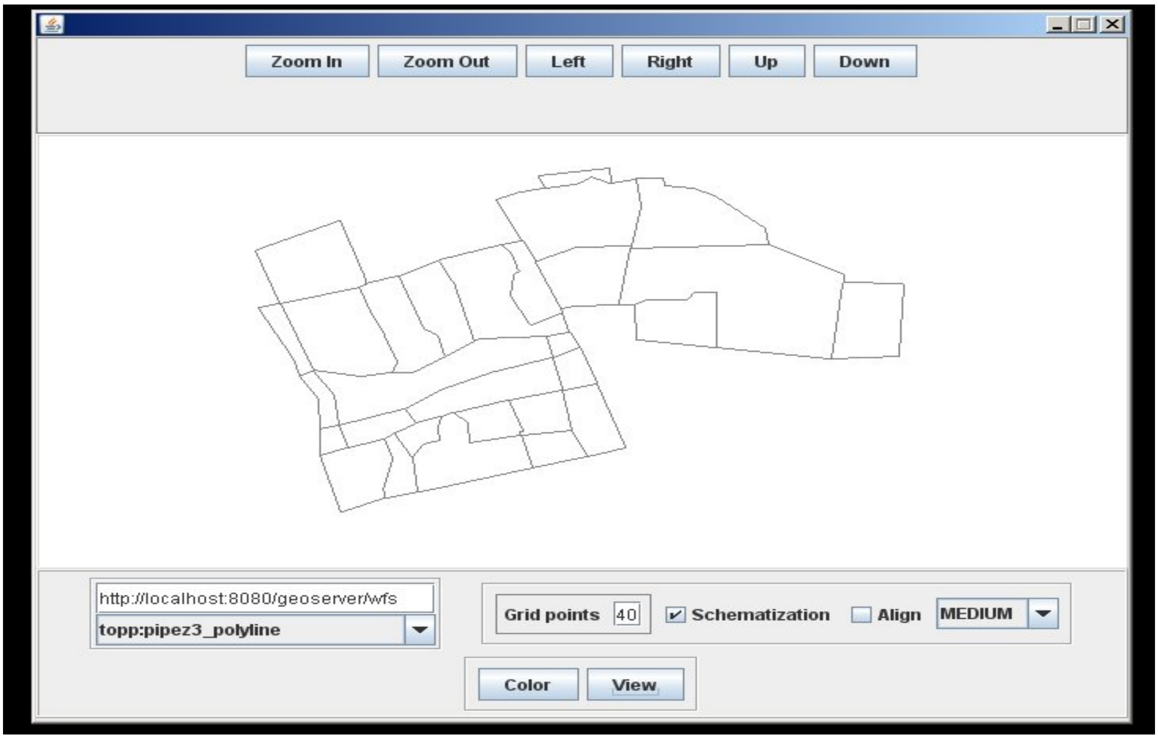


Figure 1: *Client application showing original water network feature set (Dataset Source: DfID KaR Project R6872)*

The client application allows a user to specify the URL of remotely-hosted data from an external WFS source. In the present implementation, input feature geometry is required to be of polyline format – an error message is displayed if the input data does not conform. The user can additionally select the number of grid points and the schematization quality. These parameters are then passed to the schematizing feature service, which obtains the feature collection from the remote data source and transforms them using the above algorithm. When optimization is complete, the schematized features are transferred to the client for visualization.

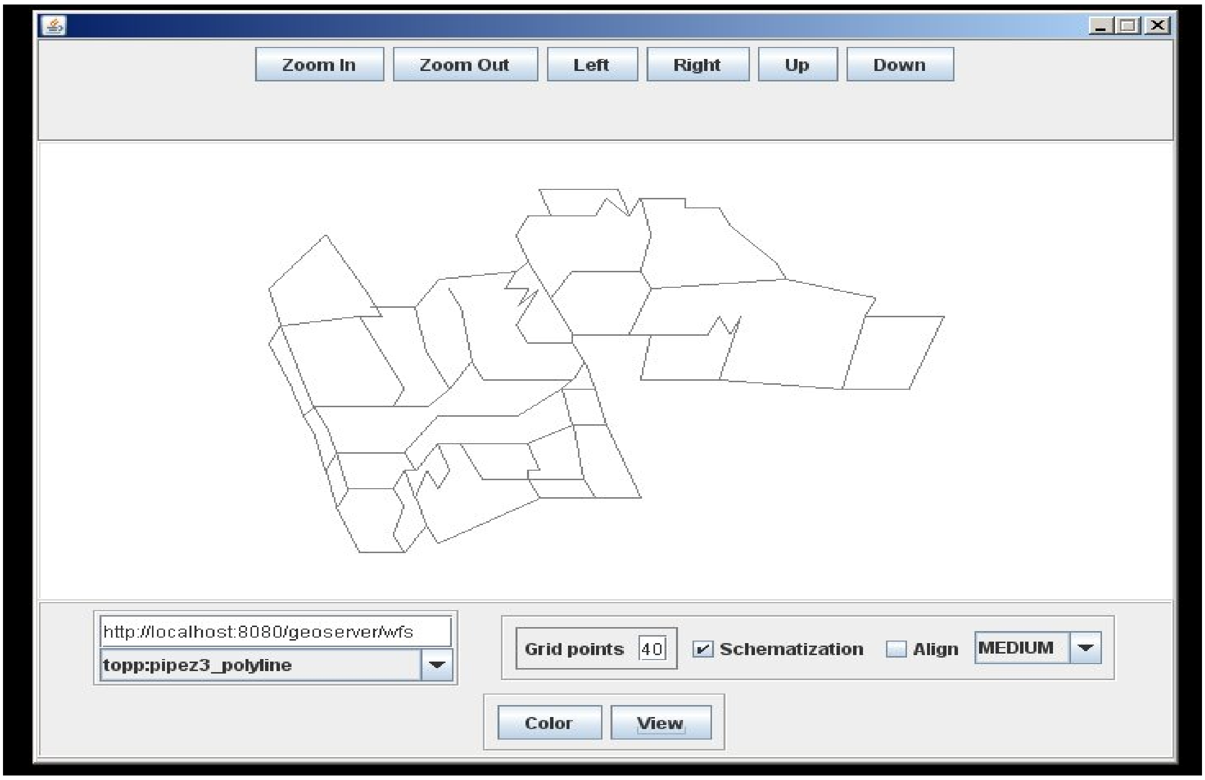


Figure 2: *Client application showing the results of schematized water network highlighting the true connectivity. (Dataset Source: DfID KaR Project R6872)*

4. Conclusions

This paper looks into the application memetic algorithms in generating schematic maps. The results are promising but more work need to be done in refining the process in consultation with water network modellers who are the key beneficiaries of this work. Future work will concentrate on refining the technique through the use of additional constraints to enhance visualization, usability and application in other domains.

5. Acknowledgements

The authors express thanks for DfID KaR Project R6872 for the data used in this work.

References

Anand S. (2006) Automatic derivation of schematic maps from large scale digital geographic datasets for MobileGIS , PhD Thesis, University of Glamorgan

Anand.S, Ware.J.M, Taylor.G.E., (2007) "Generalization and schematization of large scale digital geographic datasets for MobileGIS applications", Book chapter in Drummond, J., Billen, R., Forrest, D. and João, E. (eds), Dynamic & Mobile GIS: Investigating Change in Space and Time, Taylor & Francis publications, pp 161-174

Buttenfield B P and McMaster R B (1991), Map Generalization: Making rules for knowledge representation, Longman, London

De Jong, K. (1975). *An analysis of the behaviour of a class of genetic adaptive systems.* PhD thesis, University of Michigan.

Dorling D, (1996) Area cartograms: their use and creation, *Concepts and Techniques in Modern Geography*, CATMOG, Volume 59

Elroi D S, (1988) Designing a Network Linemap Schematization Software Enhancement Package, *Proceedings of the 8th Annual ESRI User Conference*, Redlands, CA

Grefenstette, J.J. (1987), *"Incorporating Problem-Specific Knowledge into Genetic Algorithms"* in Genetic Algorithms and Simulated Annealing, L. Davis, ed., Morgan Kaufmann, San Mateo, Calif., 1987, pp. 42-60.

Jones C B and Ware J M, (2005), *International Journal of Geographical Information Science: Map Generalisation* Special Issue (Taylor and Francis), Volume 19, Number 8-9, pages 859-997

Kaufman, H. (1967) "An experimental investigation of process identification by competitive evolution," *IEEE Trans. Systems Science and Cybernetics*, Vol. SSC3-1, pp. 11-16.

Swan J, Anand S, (2008) "A comparison of metaheuristics for automated map schematization". In preparation.

Saafeld A., (1999) Topologically consistent line simplification with the Douglas-Peucker algorithm. *Cartography and GIS*, 26(1): 7-18.

Ware, J. M., Wilson, I. D., Ware, J. A. and Jones, C.B., (2002), "A tabu search approach to automated map generalisation", Proceedings of 10th ACM International Symposium on Advances in Geographical Information Systems (ACM), McLean VA, pages 101-106.

Ware, J.M., Jones, C.B. and Thomas, N., (2003a), "Automated cartographic map generalisation with multiple operators: a simulated annealing approach", The International Journal of Geographical Information Science (Taylor and Francis), Volume 17, Number 8, pages 743–769.

Ware J.M., Wilson I.D. and Ware J.A., (2003b), A knowledge based genetic algorithm approach to automating cartographic generalization, Knowledge Based Systems, 16/5–6, pp. 295–303.

Waldorf S P (1979), Schematic Navigational Map Design, *University of Kansas Masters Dissertation*, Lawrence.

Weibel R (guest editor) (1995), *Special issue on automated map generalization*, Cartography and GIS, Volume 22, Number4

Weibel R and Jones C B (guest editors) (1998), *Special issue on automated map generalization*, GeoInformatica, Volume 2, Number 4

Biography

Jerry Swan is Research Fellow and interoperability test bed architect at the Centre for Geospatial Science, University of Nottingham
Suchith Anand is Research Fellow in mobile location based services at the Centre for Geospatial Science, University of Nottingham
J Mark Ware is Reader in the Faculty of Advanced Technology, University of Glamorgan
Mike Jackson is the Director of the Centre for Geospatial Science, University of Nottingham.

Space-time trajectories in a planar and torus-based self-organising map: the importance of eliminating boundary effects

Nick Mount and Dan Weaver

School of Geography, University of Nottingham, Nottingham NG7 2RD
Tel. +44(0)115 951 5428 Fax. +44 (0)115 951 5249
nick.mount@nottingham.ac.uk danofweaver@googlemail..com

KEYWORDS: SOM, self-organizing map, torus, projection, space-time, boundary effects

1. Self-organizing maps and boundary effects

Traditional self-organizing maps (SOMs) are a multi-dimensional clustering method, based on an unsupervised neural network, which maps complex input data to a two dimensional neuron array. Mapping is achieved using a competitive, iterative algorithm, which maps sample input data to individual neurons in the array based on the fit between the multi-dimensional value set (i.e. an n-dimensional vector) in the sample data and the corresponding vectors associated with each neuron (usually using Euclidean distance as the parameter by which fit is assessed). In each iteration, the best matching neuron (BMN) is adjusted to become more like the sample data, with proximal neurons similarly adjusted, though limited by a decay function so that the magnitude of adjustment decreases with increasing distance from the BMN. A Gaussian decay function is used as standard, with the amplitude and spread of the function declining progressively with the number of iterations; so allowing increasingly fine adjustment of the neuron vectors to match the input data.

Given the two-dimensional nature of the neuron array, a two-dimensional Gaussian decay function is required (figure 1) which can is computed using the following formula:

$$f(x, y) = A\exp - \left(\frac{(x - x_0)^2}{2\sigma_x^2}\right) - \left(\frac{(y - y_0)^2}{2\sigma_y^2}\right)$$

where A is the amplitude, x_0 , y_0 is the centre, and σ_x and σ_y are the spreads of the function.

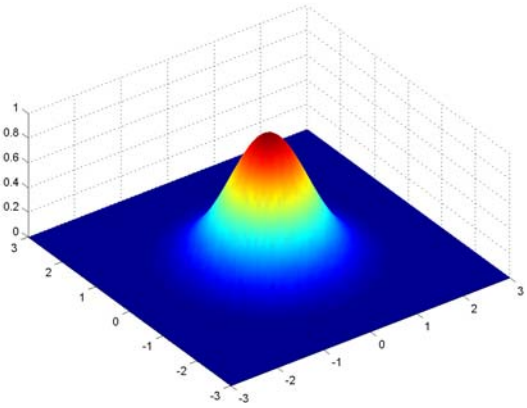


Figure 1. The form of a two-dimensional Gaussian decay function.

The finite nature of the neuron array in the traditional SOM means that neurons close to the boundary have truncated neighbourhoods, with the truncation being greatest in the corners of the neuron array. Indeed, the magnitude of the boundary effect is proportional to the number of neurons in the array and the spread of the decay function. Consequently, when the decay function is centred on boundary neurons, a smaller number of neurons are adjusted relative to decay function centred on a neuron in the middle of the array (see figure 2). Kiang et al., (1997) show that it is possible for up to 75% of the possible adjustments to be lost due to the array boundary.

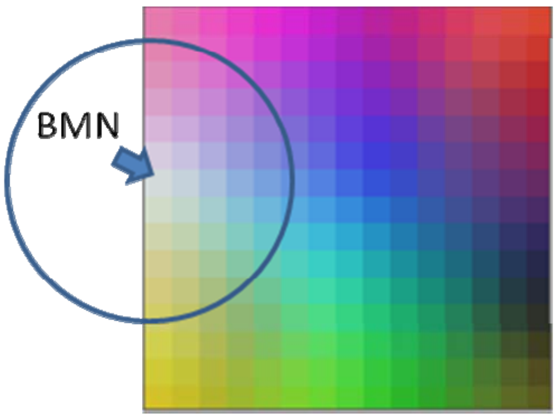


Figure 2. A 15 x 15 SOM with the location of the BMN shown together with the limit of the decay function. The boundary location of the BMU means that, compared to a BMN located in the middle of the SOM, half the number of neurons are located under, and adjusted by the decay function.

The impact of this truncation is a reduction in the performance of the SOM, with divided clusters common, particularly at opposite boundaries and corners of the SOM. The impact can be shown by mapping a colour palette (figure 3) containing the full range of RGB values to a SOM array. A correctly functioning SOM will maintain the proximity relations of the colours within the palette in the mapped output. However, figure 3 shows the impact of the SOM boundaries, which have caused two separate clusters for red-dominant and blue-dominant hues, incorrectly mapping the proximity relations in the input image.

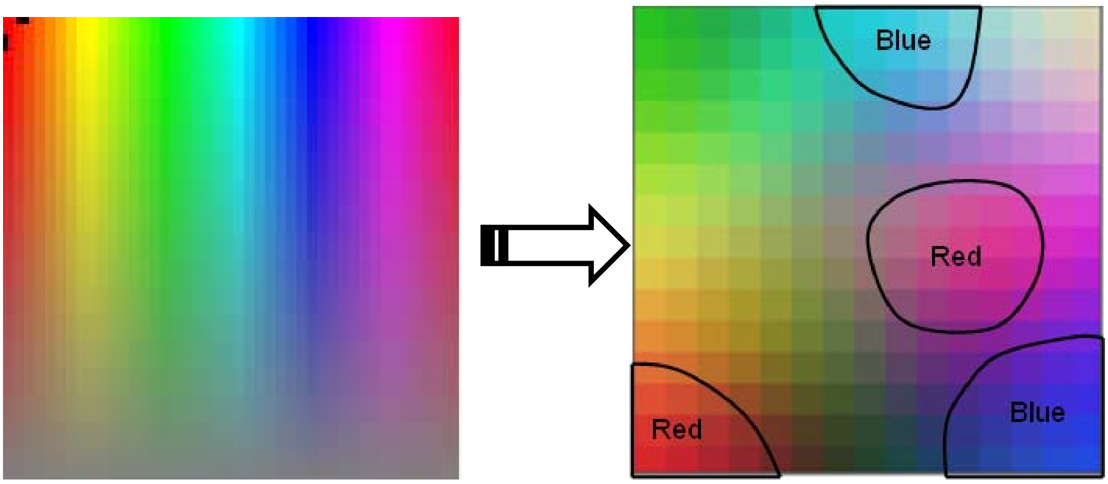


Figure 3. A SOM used to cluster red, green and blue colour bands from a palette. Boundary effects have caused inconsistencies in the SOM's proximity relationships.

2. Boundary effects in spatio-temporal self-organizing maps

Boundary effects become particularly problematic in spatio-temporal self-organizing maps (e.g. Weaver and Mount, 2007). Here the data used to train the SOM are sampled across multi-temporal input data with the weight vectors against which the SOM learns implicitly including a temporal dimension; resulting in the evolution of a spatio-temporal SOM (Kohonen, 1995). SOM locations to which input data for each time period have been mapped can then be linked together so that a set of vectors describing the space-time trajectory can be extracted (figure 4). These vectors represent the changes in the multiple dimensions of each sampled datum through time, with high-magnitude vectors indicating large-scale changes, and low-magnitude vectors indicating only minor change. By comparing the set of vectors associated with a data set, the SOM can then be used to differentiate processes encoded within the data (e.g. Skupin and Hagelman).

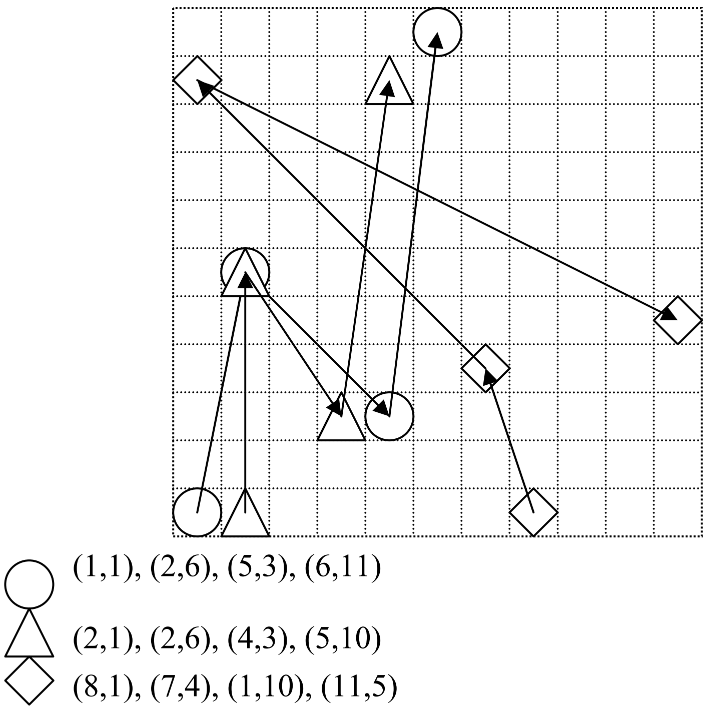


Figure 4. Space time trajectories for three samples in an 11x11 spatio-temporal SOM. The

data represented by the circle and triangle vary similarly through time. The datum represented by the diamond has a very different response in its parameters through time.

The divided clusters caused by SOM boundaries will increase the magnitude of the vectors starting and / or ending in the divided clusters; erroneously suggesting high-magnitude processes where none exist and making impossible to compare them with other vectors. As the SOM process is dependent on a random-sampling strategy, it is impossible to identify which trajectories will be affected *a priori*, making it impossible to identify and remove problem trajectories as a post-processing operation. Given the importance of boundary effects on space-time trajectories, it is surprising that examples in the geographical literature (e.g. Skupin and Hagleman, 2005) fail to explicitly recognise or address the problem.

3. Solving spatio-temporal SOM boundary problems

The solution to the boundary problem requires the removal of the boundaries via the projection of the SOM onto a three-dimensional surface. A torus (figure 5) offers a relatively simple projection as it is, in effect, a dual-axis wrapping of the planar SOM array, which removes all boundaries by connecting opposing boundaries. From a programming perspective, it requires computation of the complete set of x and y coordinates of neurons to be modified around the BMN for each iteration, irrespective of whether these coordinates actually exist within the SOM array, and knowledge of the extent to which each x and y coordinate exceeds the SOM limits.

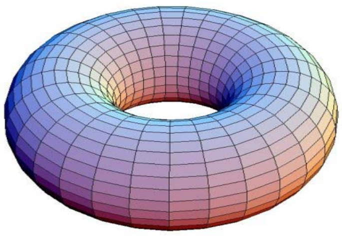


Figure 5. A torus-based projection of a planar surface produced by a dual axis wrapping of the planar surface.

For example, consider the case of a 10 x 10 SOM array, an iteration in which a BMN neuron is found to be located at location (10,5) and a decay function spread that requires adjustment of the surrounding 8 neighbours to the BMN (see figure 6). Neurons existing at (11,4), (11,5) and (11,6) should be adjusted but these exceed the SOM array extent in the x-axis by 1. Therefore, neurons at (1,4), (1,5) and (1,6) are adjusted, thereby wrapping the planar SOM around its y-axis.

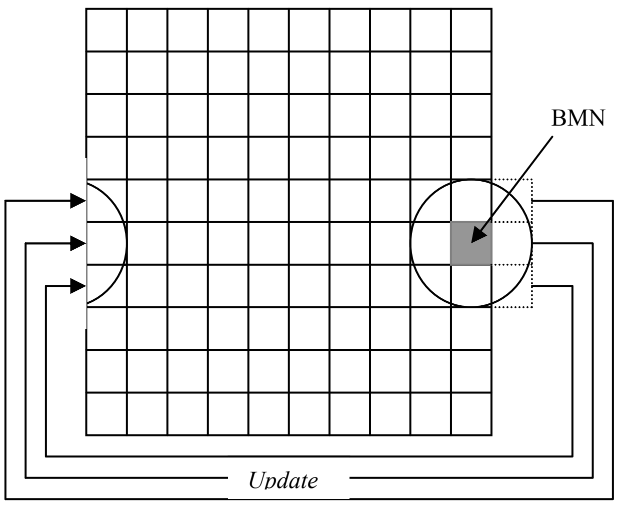


Figure 6. Creating a torus-based SOM by wrapping axes.

4. An experiment to exemplify the impacts on space-time trajectories of a torus-based and planar SOM.

The geographically-integrated, spatio-temporal, self-organising map (GISTSOM) software (Weaver and Mount, 2007) being developed in the School of Geography at the University of Nottingham offers the ability to compare space-time trajectories for planar and torus-based SOMs, identifying erroneous trajectories created as a result of boundary effects.

To investigate the impact of SOM boundaries on space-time trajectories, GISTSOM has been used to assess the space-time trajectories in two RGB images, derived from a colour palette (figure 7). Palette(0) is modified to palette(1) via the application of a 'swirl' effect, with its origin on the image midpoint. As a result, the difference in the RGB values in pixels close to the centre of the two images (image coordinate (149,138)) will show little variability in their RGB values across between palette(0) and palette(1), whereas those pixels further away from the origin will show greater variability. When mapped to a spatio-temporal SOM, pixels in the central portion of the image should have low-magnitude trajectories, whist those away from the centre should have higher-magnitude trajectories.

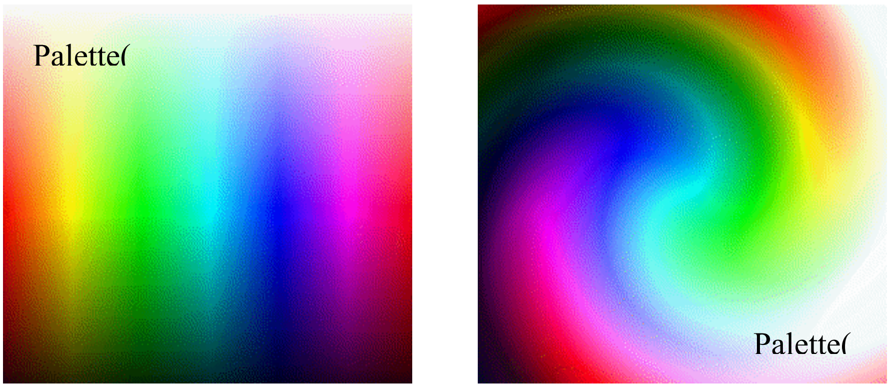


Figure 7. The two palettes using in the SOM trajectory comparison. Palette(1) is identical to palette(0) except for the fact that a swirl effect has been applied with its origin at the image centre point.

The input images were mapped to both a planar and a torus-based 15x15 SOM. In both cases a Gaussian decay function with a start variance of 13 and a linear decay / iteration relationship in both the amplitude and spread parameters was used. The SOM was trained using 2000 iterations. To exemplify the impact of the boundaries, the trajectory lengths associated with 10 pixels at increasing distance from the image centre were computed for both SOMs (figure 8).

Given the nature of the input data, a correctly functioning SOM will show a monotonic pattern of increasing trajectory length with distance from the image centre, with the rate of trajectory length change being approximately stable. No large, sudden increases in trajectory length should occur. However, it is clear from figure 8 that, whilst the torus-based SOM is able to replicate this ideal, the planar SOM is unable to reproduce this pattern, being neither monotonic nor stable in the rate of trajectory change. This is due to a boundary effect associated with the sample at 65 pixels from the image centre. The boundary forces a long trajectory at this sample point, and those at 78 and 91 pixels, causing a rapid increase in the trajectory length and preventing its monotonic increase.

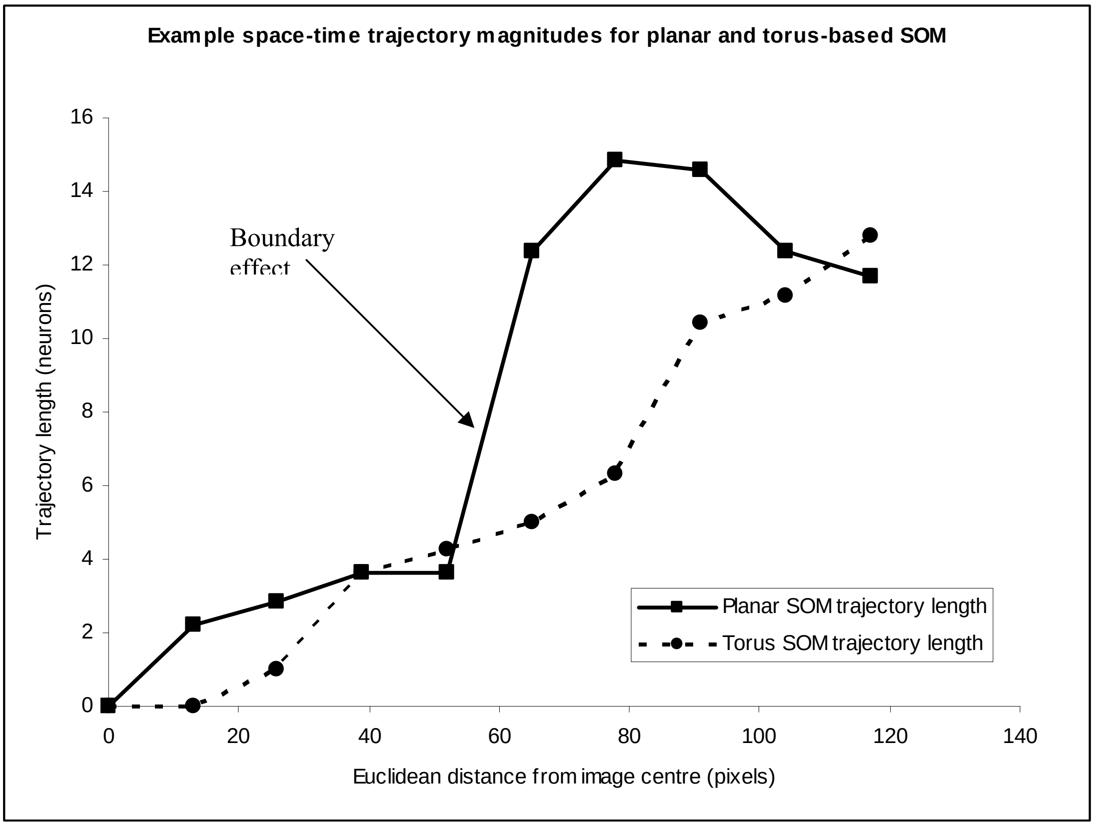


Figure 8. A comparison of the trajectory lengths associated with 11 sample data points in a planar and torus-based SOM with the impact of the boundary effect highlighted.

5. Conclusions

The example analysis presented here demonstrates the impact of planar SOM boundaries on trajectory lengths. Planar SOMs result in increased trajectory lengths where neuron clusters are split across the boundaries, erroneously indicating high-magnitude changes in the space-time structure of the input data where none exists. Where SOMs are used to elucidate space-time trajectories, a projected SOM must be used, with a torus offering a relatively simple and effective projection.

6. References

Kiang, M.Y., Kulkarni, U.R., Goul, M., Philippakis, A., Chi, R.T., Turban, E. (1997). Improving the effectiveness of self-organizing map networks using a circular Kohonen layer. *Proc. 13th Hawwaii Int. Conf. on System Sciences*, 521-529.

Kohonen, T. (1995). Self-organizing maps. Springer, Berlin.

Skupin, A., Hagelman, R. (2005). Visualizing demographic traectories with self-organizing maps. *Geoinformatica*, **9**(2), 159-179.

Weaver, D., Mount, N.J. (2007). The use of Kohonen mapping for the elucidation of space-time trajectories of multiple parameters: potential applications in fluvial geomorphology. *Proceedings of the GIScience Research UK 15th Annual Conference, University of Maynooth*.

Biographies

Nick Mount is a lecturer in GISc at the University of Nottingham and Dan Weaver completed Nottingham's MSc in GISc in 2006 and is now a Research Associate in GISc. The GISTSOM software is being developed with the aid of a New Researcher Fund grant and is due for Beta release in June 2008.

Mapping uncertainty in perceptions of landscape ‘wildness’

Alexis (Lex) Comber[1], Steve Carver[2], Steffen Fritz[3], Robert McMorran[4], Justin Washtell[2]

[1]Department of Geography, University of Leicester, Leicester, LE1 7RH, UK,
Tel +44 (0)116 252 3812, Fax +44 (0)116 252 3854
ajc36@le.ac.uk;
[2]School of Geography, University of Leeds, Leeds, LS2 9JT
[3]International Institute for Applied Systems Analysis, Schlossplatz 1, A-2361, Laxenburg, Austria
Tel +43 (0)2236 807 0, Fax: +43 (0)2236 71 313
[4]Centre for Mountain Studies, Perth College-UHI, Perth, PH1 2NX, UK
Tel +44 (0)1738 877761, Fax +44 (0)1738 877018

KEYWORDS: Uncertainty, wildness, Fuzzy, Dempster-Shafer, Bayes

Introduction

There is much interest in identifying the landscape features that contribute perceptions of ‘wild’ areas, such as are found in the Cairngorms National Park (CNP) area. Understanding public perceptions of the features that contribute to ‘wild-ness’ allows decisions about the future management of such areas to be responsive to that opinion and safeguards against inappropriate development or changes in land-use. A recent survey on behalf of Scottish Natural Heritage (SNH) and CNP evaluated public perceptions of wild places held by a cross-section of Scottish residents including a subset of living within CNP boundaries (SNH, forthcoming)

This work is part of ‘perception mappings’ study which has sought to use the data collected from a recently commissioned survey of public perceptions of wild places in Scotland (SNH, forthcoming).
SNH and CNPA commissioned a market research study to evaluate public perceptions of wild places amongst a representative cross-section of Scottish residents and a subset amongst those living within the boundaries of the Cairngorms National Park (CNP). The study identifies the level of support for wild places and whether the views of those who live within CNP match the population of Scotland as a whole. A total of 1,304 face to face interviews were conducted (1,004 across Scotland and 300 with residents of the CNP).

The project, of which the work reported here is a part, has developed a GIS-based methodology, which incorporates the results of the perception study and maps the attributes contributing to perceptions of wildness. The mappings were developed using GIS-based multi-criteria evaluation (MCE) and fuzzy mapping methods, using fuzzy membership functions derived from the results of the perception survey to parameterise existing digital datasets. The mapped data describe the four principal attributes that contribute to wildness in the CNP as defined by SNH, namely

- perceived naturalness of land cover;
- absence of modern artefacts;
- rugged/physically challenging terrain;
- remoteness.

The attribute maps were combined into a single wildness map for the CNP using MCE/fuzzy methods as in Figure 1. This allowed the relative priorities derived from the perception study to be reflected in the wildness map without the need for deterministic criteria or sharp boundaries for defining what is considered wild and that which is not.

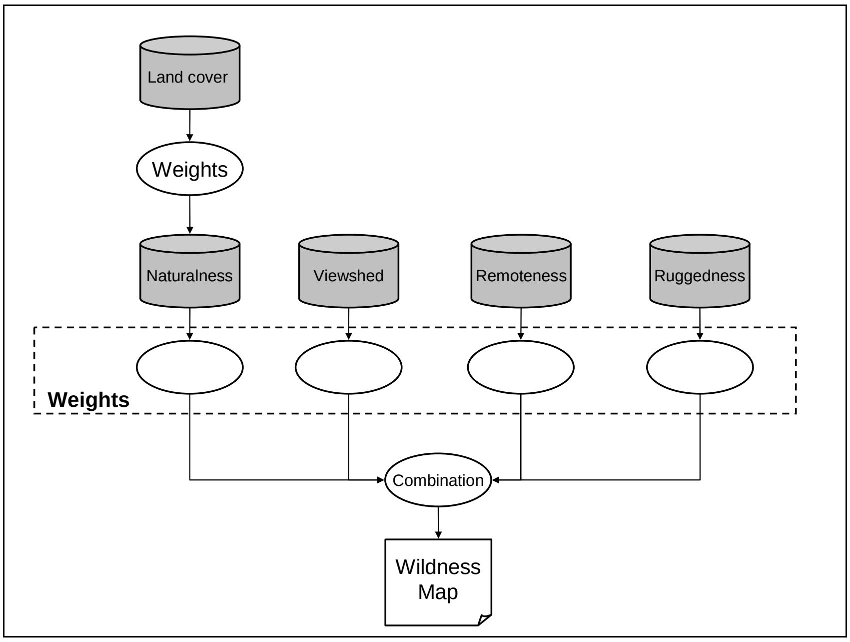


Figure 1. The overall schema showing how the data are parameterised by weights and combined to generate wildness mappings

In many cases, a range of values were identified in the perception survey for specific landscape features. As an example, one of the locations that the public associated with wild areas in Scotland was 'Woodland/forest', identified by 86% of the CNP residents and 83% of the respondents across Scotland. Whilst, 19% of CNP residents and 28% of respondents across Scotland identified 'Forest / woods / trees' as features that make an area wild. In order to generate the maps of wildness a single set of weightings were derived from the various survey data and applied to the features identified as contributing to wildness. This required decisions to be made about the final weights or membership functions that were used to relate each data feature (e.g. woodlands in land cover data) in order to generate the final fuzzy perception mappings.

The objective of this work was to explore i) the influence of different weightings derived from the perception survey on the mapped outcomes, and ii) the effect of using different formalisms for combining the weightings. Different formalisms for combining belief (or *evidence* or *weights* or *membership functions*) have different conceptual approaches that underpin them. Fuzzy Set Theory uses a *MIN* operator. Bayesian Probability assumes the weights or Beliefs to be a series of independent outcomes (which they may not be). Any evidence that does not support the hypothesis is allocated ¬Belief (that is there no 'don't know' option). Dempster-Shafer operates under the assumption of conjunctive evidence and provides clear indications of the combined belief *plus* the associated uncertainty. In this work we applied Dempster-Shafer to combine the weights

2. Background

The alternative to a Boolean map is a one where the assumption is that any location can always belong to a finite number of classes up to the number of classes being mapped over an area. Fisher (1997) identified 4 possible reasons that would justify the conceptualisation of land cover as suitable for classification by fuzzy sets, which included situations where classified areas are *Intergrades, merging* into each other over space or where the boundaries between different areas are not crisp. In this work the classes are the four attributes of wildness defined above and any location is represented by a matrix of values recording the degree to which that location belongs to each class. The overall fuzzy membership to each class is defined by applying the minimum operator in fuzzy logic to the parameterised data representing the wildness features

Bayes' theorem computes the probability of an hypothesis or event, *h* given the evidence, *e* in support of that event, $P(h|e)$:

$$P(h \mid e) = \frac{P(h) \times P(e \mid h)}{P(e)} \tag{1}$$

where:

$P(h|e_i \,\&\, \ldots \,\&\, e_m)$ is the posterior probability of hypothesis h given evidence e_i to e_m;
$P(h_i)$ is the prior probability;
$P(e_1|h_i)$ is the likelihood

Dempster-Shafer can be considered as an extension to Bayesian statistics which contains an explicit description of uncertainty, plausibility. It assigns a numerical measure of the weight of evidence (mass assignment, *m*) to *sets of hypotheses* as well as individual hypotheses. It does not consider the evidence hypothesis by hypothesis as Bayes' theorem does, rather the evidence is considered in light of the hypotheses. A second piece of evidence is introduced by combining the mass assignments (*m* and *m*') using Dempster's rule of combination, to create a new mass assignment *m''*. Dempster's rule of combination is defined by:

$$m''(C) = \sum_{\substack{i,j \\ A_i \cap B_j = C}} m(A_i) \times m'(B_j) \tag{2}$$

Dempster-Shafer accommodates explicit representations of uncertainty, plausibility, which equates to belief plus uncertainty. Therefore a weak belief in a hypothesis does not imply a strong belief in its negation. One of the weaknesses of Dempster's rule is that it can favour a class which has low mass in two data sets over any class that has a high mass in only one data set. The classic example is that of the two doctors, one of which is 90% certain the patient has disease A and 10% disease B; the other 90% convinced over disease C and 10% disease B. DS will give 100% support for disease B, even though neither doctor thought it likely (although this can be overcome by the use of alternative fusion rules).

3. Results

The data for each layer (remoteness, visibility, ruggedness, naturalness) were using a cumulative distribution function:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2} \tag{3}$$

where *x* is the original score, μ the group mean and σ the group standard deviation.

The results, to be presented at the conference, are a series of mapped outcomes, describing the different attributes contributing to wildness in the CNP area (perceived naturalness of land cover, absence of modern artefacts, rugged / physically challenging terrain and remoteness). The results of combining this information use Dempster-Shafer are compared with Bayesian and Fuzzy approaches. The results provide CNP management with an alternative mapping of perception and with uncertainties in the weightings to augment the final deliverables of the perception mapping project. These include alternative mappings of the importance of wild places in Scotland, perceptions of wild places and the location of wild places in the Cairngorms National Park.

References

SNH, (forthcoming). Public Perceptions of Wild Places and Landscapes in Scotland, Report No. XXX (ROAME No. F06NC03), SNH, Edinburgh.

Biographies

Steve Carver is a Senior Lecturer at the University of Leeds with over 20 years experience in the field of GIS and multi-criteria evaluation and interests in wild land, landscape evaluation and public participation. Steve has worked extensively on the development of wild land mapping and evaluation methodologies.

Steffen Fritz works on the Forestry Program at the International Institute for Applied Systems Analysis in Laxenburg, Austria and is involved in an EU-funded project Geo-bene assessing the socioeconomic benefits of earth observation. He has worked and published extensively on wild land mapping. From 2001 – 2006 he was at the JRC working on the Global Land Cover 2000 database and the GEOLAND project.

Lex Comber is a lecturer in GIS at the University of Leicester with research interests in land cover analysis, managing ontological uncertainty in geographical information, the application of different uncertainty reasoning formalisms for combining weighted evidence from spatial data.

.

Robert McMorran is completing his PhD research at the Centre for Mountain Studies, Perth College-UHI. His research is based on the opportunities and challenges of multi-functional forest management in the Cairngorms region. The PhD incorporated in-depth interviews with forest managers, a postal questionnaire survey of the regions landowners and GIS analysis of forest cover and ownership.

Justin Washtell is a PhD student in the School of Computing's Natural Language Processing Group at the University of Leeds with research interest in the modelling of natural systems. Recently he has been working with Dr Carver on developing voxel-based viewshed transforms for rapid landscape assessment and real-time viewshed calculations.

In Search of Spatial Literacy: the Nottingham Sputnik Model

Jim Nixon, Robert J. Abrahart, Claire Chambers

Centre for Geospatial Science, University of Nottingham, Nottingham, NG7 2RD
Tel. (+44) 0115 951 5428 Fax (+44) 0115 951 5249
Email jim.nixon@nottingham.ac.uk
www.nottingham.ac.uk/cgs

KEYWORDS: Spatial Literacy, Spatial Skills Test, Nottingham Sputnik Model

1. Introduction

Tate et al. (2005) highlight the need to establish 'benchmark measures and diagnostic tests' of spatial literacy. This paper addresses such concerns by evaluating an existing test of spatial skills, recognising that skills and literacy may not be the same thing. It presents a new model of spatial literacy that was developed from end-user feedback. This conceptual model is then used to pinpoint the relationship between GIS and spatial literacy. The selected test is both broad enough, and designed with sufficient geographical expertise, to permit the exploration of end-user results and to support some initial steps towards providing a working definition of spatial literacy. The fallacy of confusing measuring instrument with concept is nevertheless acknowledged.

Spatial literacy is a complex issue (Hernandez et al., 2006, King, 2006). It is also something of a 'holy grail' and many have sought to define it. Goodchild (2006) exemplifies the difficulties of this challenge by defining spatial literacy in broad-brush terms as 'a set of abilities related to working and reasoning in a spatial world', a general definition at best. Albert and Golledge (1999) refer to three standard subdivisions: spatial orientation, visualization and relations. The National Research Council (2006) has recently proposed a substantial framework that encapsulates different mechanisms related to 'spatial thinking' including (i) consistent spatial thinking, (ii) being able to represent spatial concepts and (iii) possessing spatial knowledge. Thus spatial literacy remains an elusive concept: it means different things to different people and cannot be compartmentalised into a simple description. However, an agreed model for spatial literacy could be used to inform educational policy and teaching guidelines. The Association of American Geographers has responded to the need for teacher training material by developing a *Teacher's Guide to Modern Geography*. This project is funded by FIPSE (US Department of Education). The primary aim of the teacher's guide is to improve the preparation of pre-service teachers. Jongwon Lee, under the auspices of this programme has constructed a test to assess teachers' mastery of the guide's content and skills: the so-called 'spatial-skills test' (SST) (Association of American Geographers, 2007). This twenty item, standardised test measures performance on a variety of tasks based on many aspects of spatial thinking and interpretation, including orienting on a street map, comparing map information to graphic information, choosing the optimum location based on several spatial factors, imagining a slope profile based on a topographic map, correlating spatially distributed phenomena, overlaying maps, and recognising spatial data types (point, line, or polygon) based on verbal descriptions.

In this paper, we present a 'first pass' evaluation of SST. This test was evaluated from the point of view of 'testing the test'. The test was adapted so it could be delivered in a controlled environment and scored electronically. The test was digitised by using the images form the online version and using HTML and JavaScript to compute the scores and deliver the test. Impressions of the test were then gathered from students. In response to this feedback we propose a set of modifications to the initial test and options to be explored in subsequent developments. Next, from our experiences, we present a tentative model of spatial literacy based on emergent concepts and constructs within the test. This model is then used to identify the relationship between GIS and spatial literacy.

2. Method

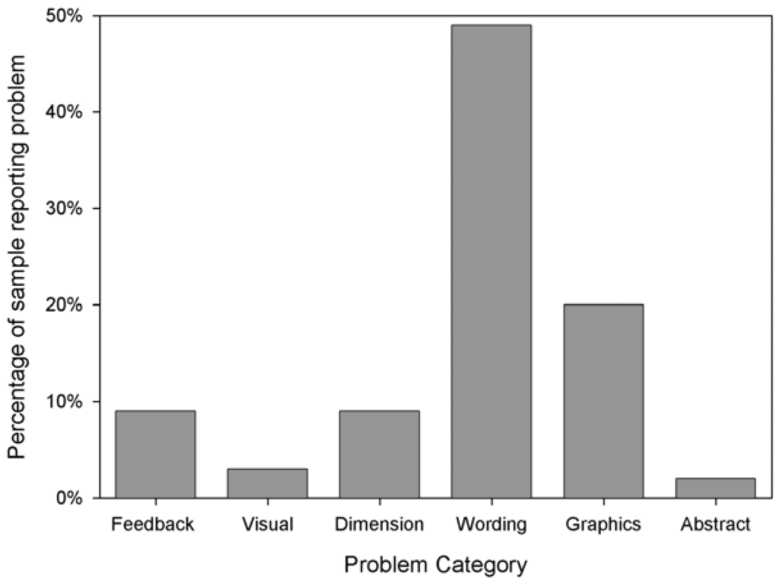
A first-year cohort of geography undergraduate students were asked to complete the test. Following the test, any criticisms students had were recorded. A completely unstructured approach was taken to elicit student opinion. Students were asked to write their criticisms down candidly. Ninety-eight students returned their opinions to the experimenters. Categories were developed in response to student feedback. The opinions were then sorted into categories. The analysis is shown in Figure 1. Criticisms are shown as proportions of the full sample size, grouped into problem categories. The problem categories were developed from the criticisms. Definitions of the problem categories are shown in Table 1. Throughout this paper, individual questions are referred to as Q1, Q2, etc.

Table 1 Definition of each problem category

Problem Category	Definition
Feedback	Immediate feedback preferred by the student
Visual	Domination of questions which test visuospatial skills
Dimension	Lack of clarity as to what skills the test is testing
Wording	Poor or ambiguous question wording
Graphics	Poor or ambiguous question graphics
Abstract	Question highly abstracted and lacking geographical context

3. Analysis

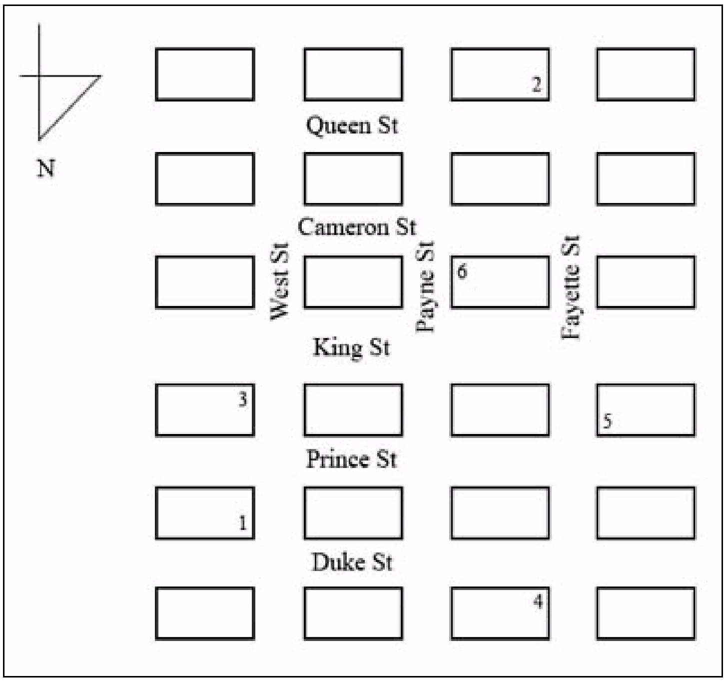
Figure 1 Number of students reporting problems in each category as a percentage of the total sample returned



With reference to Figure 1, a larger proportion of the participants noted problems with question wording and low resolution or ambiguous graphics associated with the test. Sentence structure and instances of contradictory wording, particularly with regard to the spatial correlation questions, confused many participants affecting their ability to answer the questions. For example, the lead in to Q10 and Q11 indicates that participants should find one positive and one negative spatial correlation. The individual questions then require participants to find two positive correlations. In Q1 and Q2, there are semantic ambiguities. Other participants remarked on low-resolution graphics preventing effective answering of questions. Many participants also wanted immediate feedback; in our implementation this was not delivered. Given the electronic delivery of the test this should pose no problem in the future. Following reporting of these problems, a logical appraisal of the test was undertaken and a set of potential modifications were developed. In developing this list it was important to consider the range of software packages which are available to assist the researcher to electronically deliver the content of tests. In this case Question Mark Perception™ (QMP) authoring software was selected as the basis for selecting potential modifications. This package allows maximum flexibility in the processing and presentation of results. Innovative interaction styles can be specified increasing the level of student engagement in a way that static HTML, or form based authoring software cannot emulate.

Substantial improvements could be made to the interactive elements of the test, increasing the clarity of the questions and addressing student concerns over the graphics and layout of the test. Additionally changes to the wording could be made where deemed unclear. Figure 2 shows the stimuli for Q1 and Q2 on the SST. In this question students are required to complete a sequence of spatial operations and report their new location on the map. A clearer, more familiar north-arrow could be used. The question involves the concept of a block, in which a block is represented by rectangles on a map. The rectangles are separated by streets. The start and end locations are represented by numbers placed in the corner of the rectangles. It is not clear whether movement in terms of blocks represents corner-corner or block-block movement and where the streets fit into the conceptual movements demanded of the question. The concept of movement by block is currently ambiguous and needs to be defined ensuring that spatial dimensions are being tested, as opposed to the chance matching of a linguistic concept. Additionally the method of interaction could be changed. Users could move an object around the grid and click enter when the object is on the correct block removing the need for finish points to be specified. This is in contrast to the drop down box provided in the original version; users could now interact directly with the material. A number of other questions could be adapted in this way (e.g. Q4, 5, 9) to reduce ambiguity and afford a more direct method of interacting with the test. Many questions used US style concepts and language. Q19 highlights another ambiguity in wording. In the original question the answer depends on the concept of a bus route. If this concept includes only the route, this can be described as a line. If the concept is expanded to include bus stops, points and lines are a more appropriate response. In the original question, both responses are available but only one is recorded as correct.

Figure 2 SST stimuli for Q1



Many students remarked on the more obvious issues with the test; layout, interaction, ambiguity. Smaller numbers of students noted more subtle problems with the test. Some parts of the test present abstract stimuli such as the shape matching and overlay questions (Q13 - 16). These questions were unpopular with students as they tended to be more difficult and more aligned with psychological notions of spatial ability, lacking the geographical context that could anchor the reasoning required in other questions. Other students noted the emphasis on visuospatial skills as opposed to knowledge of spatial concepts or linguistic spatial reasoning. Most questions in the test contain visual stimuli, on which judgements must be made. This kind of criticism moves more into the 'what are we testing' debate and highlights an issue with the test as it stands.

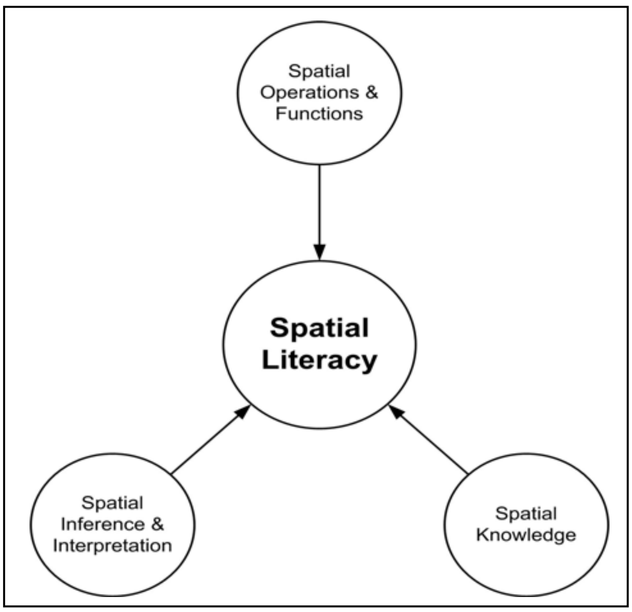
An evaluative framework with which to approach this question is the psychometric perspective, the cornerstones of which are reliability, sensitivity and validity of a test instrument (Kline, 1993, Cohen and Swerdlik, 2005, Kline, 2005). Several psychometric tests have been developed which examine very specific spatial skills. For example, mental rotation (Shepard and Metzler, 1971), perspective taking (Hegarty and Waller, 2004), attitude towards own sense of direction (Hegarty et al., 2006). These tests were all developed in the psychometric tradition: each test contains a number of items that measure a single construct or concept; mental rotation, perspective taking or attitude. These tests have the benefit of higher reliability due to many similar items that are designed to test the same spatial skill. Statistical techniques are then used to assess the test items, weeding out weaker questions and ensuring that all questions are measuring what they should be. In this way empirically derived measures of the reliability, sensitivity and validity of the test can be formally established and published.

SST in its current form does not sit easily within the psychometric approach but we can base our assessment of its merits within this perspective. The test has many different kinds of item measuring many different kinds of abilities. The majority of items focus on visuospatial abilities requiring spatial operations and perspective shifts (Q6 - 9). Other parts require interpretation and inference from spatial data (Q3, 10, 11). In the final part of the test (Q17 - 20) factual, or declarative, knowledge is tested. Students are required to demonstrate knowledge about the building blocks of GIS: points, lines and polygons. This kind of knowledge relies more on educational background rather than spatial skills per se. The small number of items testing each component of spatial literacy means that reliability may be an issue with the test. For example, a basic lack of knowledge about GIS would result in light penalties for a student answering the final questions in the test which relate to this knowledge (Q17 - 20). Many more questions deal with spatial operations and functions. A student weak in this area of spatial literacy would incur heavy penalties and perhaps this is the correct balance. It is up to the geographic community to decide what is important. Longer tests with more questions testing the same skill will give a more reliable measure of that skill. Again, given the small number of items, the sensitivity of each question requires exploration. Not unreasonably, we want our students to be spatially literate. Some spatial skills are innate: others, we hope can be developed or improved with training or education. To know whether such outcomes are possible and if we are achieving or failing in our expectations will require a trusted diagnostic testing instrument. Spatial cognition, experience, ability or knowledge in combination or individually, may all be under the microscope at any given point during the SST. However what do incorrect answers to specific questions mean in terms of spatial literacy? In the following section, we propose a conceptual model of spatial literacy based on fundamental building blocks around which subsequent developments can be proposed.

4. The Bigger Picture

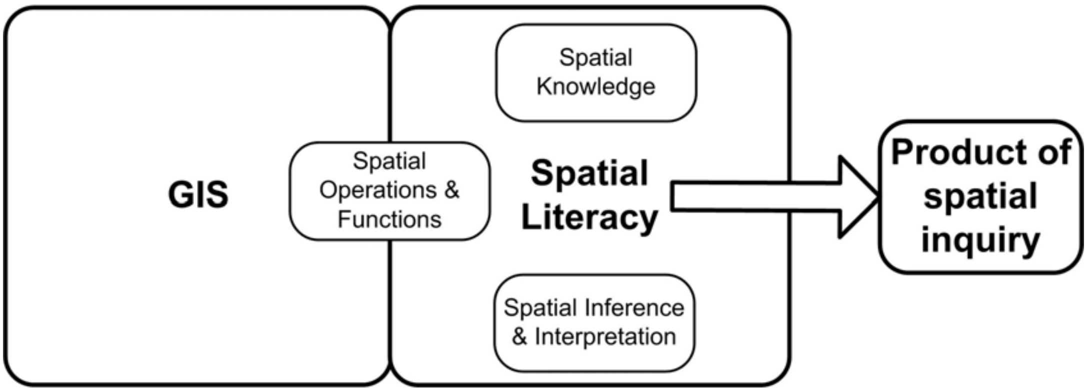
The initial suggested changes to the test will provide a sharper tool upon which to base future research into the validity and reliability of that instrument and permit further inferences to be made about spatial literacy. However, in broader terms, our experiences suggest a more radical overhaul of the test and to support such developments we propose the conceptual model depicted in Figure 3. This is the Nottingham Sputnik Model of Spatial Literacy (NSM). Spatial literacy is shown as a latent concept. 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 test hence their inclusion in our conceptual model. Taken together we propose a tentative definition. Spatial literacy is composed of the three spokes which when converged form the 'raw ability' that allows the individual to manage and interpret geospatial data and materials.

Figure 3 Nottingham Sputnik Model



There remains a need to frame spatial literacy as related to GIS with particular reference to tertiary education. The proposed model is flexible enough to be able to handle this question without adding undue complexity to its structure. Figure 4 shows how GIS can be seamlessly positioned within the model. GIS feeds into the 'spatial operations and functions' part of the model. GIS software routinely handles many spatial operations on large datasets or performs complex queries that would simply place too much burden on human memory or cognition to be implemented in any other way. Spatial literacy is required for the correct selection of operations and their subsequent interpretation. Our conceptual model would suggests that the quality of output, be that in a student assessment or real-world application, would depend on the degree of spatial literacy possessed by an end-user.

Figure 4 Extended Sputnik Model



5. Conclusions

New tools and testing instruments should always be evaluated and the results of that process acted upon. It is easy to fall into the trap of simply reproducing the paper version of any test online without considering the increased flexibility that different interaction styles and layouts can give. Delivering a test electronically has many benefits. Very large sample sizes can be tested at the same time or in different locations, limited only by the availability of computers or internet connections. Automated delivery of the test can be highly controlled; individual questions can have time limits allocated to them or extra time can be provided. Machines can be locked down in 'kiosk mode' to prevent other programmes being run and referred to. For the researcher, the benefit of immediate access to captured data including automated processing, storage and analysis while avoiding drudgery of error-prone data entry by hand is enormous. Students can benefit from instant feedback if appropriate and a wide range of dynamic interaction styles can be used generally eliciting positive user response. We have described certain changes that should be made: it is nevertheless difficult to keep abreast of rapid developments in smarter software.

Returning to the psychometric perspective and SST. In its current form, the test has high face validity; it appears to be measuring something to do with geographical skills. This is to be expected since the test was developed by geographers. However, the construct validity of the test remains a matter of concern. Construct validity refers to the alignment of that test with some theoretical framework or concept. We have proposed spatial literacy as the latent construct, embedded within the test, itself composed of different skill and knowledge spokes. Using and understanding spatial skills testing could shed further light on the extent to which these separate components interact.

Without a great deal of further work, we may never be sure quite where or when or indeed what skills are being used. NSM integrates different aspects of SST while handling the conceptual structure of spatial literacy. The models listed in the introduction (Albert and

Golledge, 1999, Goodchild, 2006, National Research Council, 2006) all describe diverse aspects of spatial literacy. The concepts proposed in these earlier publications can be related to NSM: such models both (i) dovetail with its structure and (ii) are described by it. This paper offers a fresh perspective that is presented for discussion. It is believed that spatial literacy, or at the very least its covariates, reside somewhere in the SST. Having sharpened the instrument and developed the conceptual model, we now set to explore matters further.

Acknowledgements

The authors would like to thank Michael Solem, Director of Educational Affairs at the Association of American Geographers, and the Association of American Geographers as a whole for permission to use their test materials.

Special thanks must also go to the University of Nottingham students who participated in our experiment. Their time and effort is much appreciated.

References

Albert, W. S. & Golledge, R. G. (1999) The Use of Spatial Cognitive Abilities in Geographical Information Systems: The Map Overlay Operation. *Transactions in GIS,* 3(1), pp7-21.

Association of American Geographers (2007) Spatial Skills Test, Available at URL: http://www.aag.org/tgmg/sst.htm [Accessed 3rd December 2007]

Cohen, R. J. & Swerdlik, M. E. (2005) *Psychological Testing and Assessment,* Boston, McGraw Hill.

Goodchild, M. (2006) The Fourth R? Rethinking GIS Education. *ArcNews,* 28(3), pp429-431.

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T. & Lovelace, K. (2006) Spatial Abilities at Different Scales: Individual Differences in Aptitude-Test Performance and Spatial-Layout Learning. *Intelligence,* 34(2), pp151-176.

Hegarty, M. & Waller, D. (2004) A Dissociation between Mental Rotation and Perspective Taking Spatial Abilities. *Intelligence,* 32(2), pp175-191.

Hernandez, B. P., Jarvis, C. H., Wellens, J. & Tate, N. J. (2006) What do we Mean by "Spatial Literacy" in a GIS Context? IN PRIESTNALL, G. & APLIN, P. (Eds.) Proceedings of GIS Research UK (GISRUK) 14th Annual Conference, pp206-209. 5th - 7th April Nottingham, UK

King, H. (2006) Understanding spatial literacy: cognitive and curriculum perspectives. *Planet,* 17(2), pp26-28.

Kline, P. (1993) *The Handbook of Psychological Testing,* London, Routledge.

Kline, T. J. B. (2005) *Psychological Testing: A Practical Approach to Design and Testing,* London, Sage Publications Inc.

National Research Council (2006) *Learning to Think Spatially,* Washington D.C., National Academy Press.

Shepard, R. N. & Metzler, J. (1971) Mental Rotation of Three-Dimensional Objects. *Science,* 171(3972), pp701-703.

Tate, N. J., Jarvis, C. H. & Moore, K. E. (2005) Locating Spatial Thinking In Teaching Practice. *Computers, Environment and Urban Systems,* 29, pp89-91.

Biography

Jim Nixon is a research student examining human factors issues surrounding the perception and communication of different geospatial representations over mobile communications. Jim holds an undergraduate degree in psychology and a masters degree in human factors. He is a graduate member of both the Ergonomics Society and the British Psychological Society and is working towards professional registration.

Developing spatial literacy in secondary education: GIS practicals for key stage 3

Alexis (Lex) Comber[1], Matt Buxton[2], Claire Jarvis[1], Jane Wellens[3], Phil Wood[4]

[1]Department of Geography, University of Leicester, Leicester, LE1 7RH, UK,
Tel +44 (0)116 252 3812, Fax +44 (0)116 252 3854, ajc36@le.ac.uk;
[2]Djanogly City Academy, Nottingham, NG7 6ND, UK
[3]Staff Development Centre, University of Leicester, Leicester, LE1 7RH, UK
[4]School of Education, University of Leicester, Leicester, LE1 7RF, UK

KEYWORDS: GIS, secondary education, software, functionality

1. Introduction

Geographical Information System (GIS) technologies have been identified as 'one of the three most important and emerging scientific fields (Gewin, 2004). Whilst GIS and geographical information science are important in their own right, it has been shown that they can support students learn a range of more generic skills. For example, spatial reasoning tasks undertaken within GIS have potential to enhance maths skills; Lehmann and Juling (2002) found that increased spatial abilities are mirrored by increases in algebraic and arithmetical problem solving. Shin (2006) found that the development of GIS skills in the classroom helped to improve students' geographic knowledge. This improvement was found to be gradual but had sequential learning benefits in other areas. Further, Battersby et al. (2006) evaluated student understanding of the concept of map overlay, a set theoretical construct, and showed interaction between learning geospatial knowledge and complex thinking strategies.

The recognition of the importance of GIS as a vehicle for different skills, including spatial literacy and mathematics through representations of space and constructs such as set theory, Boolean logic, triangulation, co-ordinate systems, algebra and geometry, is reflected in the changing schools curricula for geography. The Qualifications and Curriculum Authority (QCA) in England has made explicit recommendations that encourage the delivery of GIS in schools from 2008/9 (Table 2).

Table 2. Qualifications and Curriculum Authority (QCA) recommendations for the delivery of GIS in schools from 2008/9

Stage	Recommendation
Key Stage 3	Students will *'use varied resources, including maps, visual media and geographical information systems'*(QCA, 2007a)
Key Stage 4	Students are to *'become more independent learners through the use of geographical skills, new technologies and enquiry to serve them in their future'* (QCA, 2007b) Teaching will embrace *'use of new technologies including GIS, to assist geographical investigation'*
A Level	Students will *'use modern information technologies, including geographical information systems as appropriate to the content'* (QCA, 2006)

However, despite these recommendations, as yet few (if any) GIS software packages designed for use in schools offer little more functionality than a data viewer. Whilst this is may be currently unproblematic, as curricula change (see below) and as students become more literate (the Google Earth generation) the teaching and learning resources needed will also change. Current software and teaching resources are not up to scratch in terms of true GIS functionality yet it is these missing analysis functions that arguably provide the most opportunity for the transfer of generic skills.

2. GIS in schools

There are many GIS software packages specifically aimed at schools and there are a number of online resources available to teachers. These include lesson plans, summaries of learning outcomes, handouts for students, exercises, etc. Similarly all of the software products aimed at schools come with a set of practical exercises, plus spatial data (typically current and historical OS raster data local to the school with some aerial photography) to augment teaching. From a GIS perspective (rather than a secondary educational perspective), the practicals that are available to teachers are very limited in terms of the GIS concepts they teach and the spatial literacy they promote. Typically, a series of lessons and practicals involves:

- Exploring a map – typically a raster data of an OS product at 1:25,000 or 1:50,000 scale);
- Identifying where students live on the map;
- Tracing a line with a mouse to create a line feature that describes a route you often take (this is usually 'my route to school');
- Measuring the distance using a distance tool;
- Selecting some data that fits some criteria (but note that, crucially, there is no functionality to create new data from the selected features);
- Printing the map for inclusion in a coursework folder.

Our concern is that the GIS software available to and targeted at secondary educational do not have a deep level of GIS functionality – the tasks listed above can all be performed in a standard image viewer such Adobe Photoshop by an intelligent user. The tasks do not foster spatial literacy in students except at a very rudimentary level. In the terms of the NRC Framework (2006, p40-48), spatial tasks relating to the representation of features are facilitated, as are some limited means of comparison between static features, but the *transformational* aspects of a full GIS are not represented. Currently this is unproblematic since the use of GIS within a broader geography curriculum has to address the fact that many students arrive at secondary school unable to use spatial information at all. Furthermore, anecdotal evidence suggests that geography and spatial learning at primary school level are possibly the worst taught elements in the curriculum, certainly according to Ofsted (Ofsted, 2002).

Much early secondary GIS teaching has necessarily been based around a simple thread that it is expected will develop as the student progresses. Thus GIS teaching has been focussed to date on the development of spatial skills based on very simple ideas, such as reading spatial information (e.g. helping students to critically understand choropleth maps). From this starting point, students progressively develop more skills and capacity. Simple GIS functionality, such as the above, provides a stepping stone in this development, introducing the idea of layers, the connectivity between mapped data and the real world and acts as a useful prelude to the true functionality of a professional GIS. To these ends, and to their credit, much software designed for schools has particularly easy to use interfaces when compared to more serious contenders, resulting students being able to feel familiar in the software environment after only a short period of time. However, the current limitations of GIS software for schools (many will not even allow a simple distance selection or overlay operation) and the limited aspirations of the resources provided to teachers, who owing to their own time constraints are likely to accept what is offered to them, preclude a deeper development of spatial literacy in secondary students.

A further dimension of the problem being addressed in this work is that few people who work on a day to day basis with GIS theory or geographical information science have been involved with the development of software and teaching resources for schools: typically, software developers have a background in IT and secondary education, or in the teaching of secondary geography. They are not GIS experts.

3. Method and results

In order to address this specific issue the Department of Geography at the University of Leicester developed and ran a set of 5 GIS practicals for Year 10 (Key Stage 4) with staff at Djanogly City Academy, Nottingham in September and October 2007. This collaborative venture involved the development of practical classes which brought together the GIS specialists knowledge and understanding of spatial literacy issues and the teachers understanding of the curriculum requirements and students prior knowledge and abilities. The practicals used data for the Nottingham area, and adopted Digital Worlds V2 software. This software has a true GIS functionality not found in most commercially available schools GIS software.

The practicals were specifically aimed at developing students' spatial awareness, the nature of spatial data layers and their understanding of GIS and spatial analytical techniques including spatial overlay, spatial queries and the use of sub-setting to create new data layers. These learning objectives were taught and developed using a series of discrete, self contained tasks:

- Understanding maps, and the concept of layers, using paper maps and tracing paper;
- Viewing data layers in a GIS and performing simple analyses by querying data, understanding thematic attributes and measuring distances;
- Classification and thematic queries through analysis, display census data, classification of data, normalization of data e.g. by area, querying by attribute;
- Spatial queries using distances, buffers, applying rules to generate queries to select layer objects, relating criteria to data, using multiple layers (MCE).

The practicals were merged into an established timetable, within which it was possible to find five 1-hour slots without causing disruption to the existing demands of the GCSE curriculum.

Each of the students in the class completed a learning diary at the end of each session. This aspect of the work was incorporated since the educational research literature strongly suggests that self-monitoring is a useful process for improving learning progress (Coutinho, et al., 2005; Brookhart, 2001). The structure for the reflective diaries covered specific aspects of the teaching and learning experience, asking the respondents on each occasion to describe 1) what they learnt, 2) what they have not learnt yet. They offered the opportunity to 3) ask specific questions and to 4) describe their biggest difficulties with the subject. The diaries also 5) capture information on learning or teaching progress (e.g. successes and disappointments). The information from the reflective diaries was collated and analysed to identify the positive and negative aspects of the teaching and learning experience through the case studies. Analysis showed that:

- Students enjoyed the practicals
- Students commented that exploring spatial data using a GIS gave them a better understanding of data (e.g. what are contained in rows and fields in a database) and existing applications e.g. web maps, gaming environments
- Students were fascinated by being able to 'get their hands dirty' with the spatial data that underpins many applications with which they were already familiar – Google Maps etc;
- Students were interested in the socio-economic patterns of their City, e.g. where the wealthy people live and expressed an interest in examining other types of census data;

A number of other issues also emerged from the learning diaries:

- A one hour slot is too short. The students wanted longer to examine the data and to develop their familiarity with the software;
- Running the software and loading data was too slow on laptops that were connected to the network through a wireless connection. Many students complained about these problems.

In the future the practicals will be run in a time slot with less pressure and on workstations hardwired to the network We shall be look to develop other applications relating to different curricula foci, for example practicals supporting existing field courses in Edale, climate change work, site location and looking at local demographics. There are further opportunities to develop mobile GIS applications, allowing students to embed their own data and information (e.g. photos) into a local and more participatory notion of GIS.

The focus of the practicals has been to try to use GIS as a means to improve the spatial literacy of students, whilst enhancing their spatial skills and their implicit understanding of the spatial world they live in. We believe that such 'GIS skills' are important in their own right, as well as being a means to an end for the QCA. One of the major issues encountered by this work for the development of student spatial literacy and their ability to use GIS techniques is the time and resources available to teachers for CPD training. Consequently one of the unexpected benefits of this work was the 'on the job' training received by the school teachers involved as they are under increasing time pressures with little resources set aside for their CPD. This highlights a wider issue: resources are needed to ensure that teachers' CPD keeps pace with fast-evolving technologies such as GIS.

References

Battersby, S.E., Golledge, R.G. & Marsh, M.J., (2006). Incidental learning of geospatial concepts across grade levels: Map overlay. *Journal of Geography*, 105 (4): 139-146

Brookhart, S.M., (2001) Successful students' formative and summative uses of assessment information. *Assessment in Education* 8, 153-169.

Coutinho, S., Wiemer-Hastings, K., Skowronski, J.J. & Britt, M.A., (2005) Metacognition, need for cognition and use of explanations during ongoing learning and problem solving. *Learning and Individual Differences* 15, 321-337.

Gewin, V., (2004). Mapping Opportunities. *Nature*, 427: 376-377.

Lehmann W. & Juling, I. (2002). Spatial reasoning and mathematical abilities - Independent constructs or two sides of the same medal? *Psychologie in Erziehung und Unterricht*, 49 (1): 31-43.

NRC (National Research Council) (2006) *Learning to Think Spatially.* Washington: The National Academes Press.

OFSTED, 2002. Primary Subject Reports 2000/01:Geography (HMI 360). Office for Standards in Education, London

QCA, 2007a Geography: programme of study for key stage 3 and attainment target, http://curriculum.qca.org.uk/uploads/QCA-07-3334-pGeography3_tcm6-400.pdf?return=http%3A//curriculum.qca.org.uk/subjects/geography/index.aspx [available on line 17th December 2007]

QCA, 2007b. Draft GCSE subject criteria for geography (Consultation version), http://www.qca.org.uk/libraryAssets/media/qca-07-3148_geography_draft_GCSE_crit.pdf [available on line 17th December 2007]

Shin, E.K., (2006). Using Geographic Information System (GIS) to improve fourth graders' geographic content knowledge and map skills. *Journal of Geography*, 105 (3): 109-120.

Noisy School Kids: Using GPS in an Urban Environment

Duncan Whyatt[1], Gemma Davies[1], Marion Walker[1], Colin Pooley[1]
Paul Coulton[2] and Will Bamford[2]

[1]Department of Geography, Lancaster University, LA1 4YQ
Tel: +44 1524 510239 Fax: +44 1524 510269
d.whyatt@lancaster.ac.uk http://geography.lancs.ac.uk/
[2] Department of Communication Systems, InfoLab21, Lancaster University, LA1 4WA

KEYWORDS: School Journey GPS Scatter Air Pollution Personal Exposure

1. Introduction

The Global Positioning System (GPS) was initially developed by the US Department of Defence in the 1980s. Civilian use of GPS was primarily restricted to navigation, with the accuracy of the signal downgraded through Selective Availability (SA). Since May 2000, however, SA has been disabled, allowing civilian users to obtain positional fixes of much higher accuracy. This has led to improvements in existing applications such as in-car navigation systems and the development of novel, new applications such as pay-as-you-drive car insurance (Norwich Union, 2007) and electronic baby-sitting. Parallel developments in mobile computing have led to the advent of Location-Based Services (LBS) in which the information the mobile user receives is geographically dependent (see Brimbicombe, 2007, for a review of LBS).

GPS are not, however, without their problems. It can take a GPS receiver several minutes to obtain an accurate positional fix when first activated in a new geographic location (the 'cold start' problem). Accuracy may also be compromised by poor satellite configuration or signal blocking by buildings and large trees. Pertinently, one of the most common causes of poor reception comes from placing receivers in pockets or bags.

This paper describes a project in which 30 school children used GPS within a customised mobile phone application to capture their routes to and from school using a variety of different forms of transport. These data will ultimately be integrated with modelled estimates of traffic-based air pollution to determine individual estimates of personal exposure. Prior to this, however, the accuracy and consistency of positional data needs to be assessed. This paper describes some of the problems encountered when trying to clean individual GPS tracks prior to use in exposure assessment using a combination of spatial and temporal criteria.

2. Methodology

Thirty school children (aged 12-13) were each provided with a mobile phone running a customised application ('GeoBlog') designed to automatically capture the geographic position of the user at 1-second time intervals through Bluetooth connection to a GPS unit. The children were asked to turn on the GPS units prior to commencing their journeys to school (in order to obtain positional fixes) and to wear the units around their necks in order to maximise signal capture. A similar methodology was adopted by Duncan et al (2006) in their study of school journeys in Rockhampton, Australia. Upon arrival at school the children were asked to turn off the equipment. Similar instructions were provided for the homeward journey.

Four one-week periods of data capture were organized throughout 2007. The children were asked to use the mobile phones to take pictures and create text messages of things that ‘interested’ them on their journeys to and from school. Follow-up interviews were conducted with each child after each period of data capture in order to gain a deeper insight into factors affecting the choice of routes and associated photographs and texts. These findings will be reported in more detail in Pooley et al (2008).

3. Results

Initial visual assessment of the GPS tracks generated by the children over the four week period suggested that 654 / 1080 routes were sufficiently complete to merit further use in pollution exposure assessment. Conversely, 426 routes contained little or no positional data.

Figure 1. Sample Routes

The usable tracks reveal incredibly detailed records of journeys to and from school. In addition to the complexity of the routes themselves, these data can also be used to determine variations in speed of travel (by mode of transport), common routes and social spaces where children gather.

However, the usable tracks are problematic and cannot be used in exposure assessment without further processing. Some of these problems relate to the definition of the start and end points of the journey (home and school). Others relate to variations in accuracy and completeness between these two locations. A selection of these problematic cases are illustrated in Figure 1.

Figure 1a) depicts significant scatter around the home address, some signal scatter and drop-out en route to the bus stop, further scatter at the bus stop then a clean signal once travelling on the bus to school. Figure 1b) depicts significant scatter at the school as a consequence of leaving the GPS unit turned on in the classroom. Figures 1c) and 1d) show the difference between a high quality track produced by a child travelling to school in a car and a medium quality track produced by a child walking to school. The latter shows significant signal dropout as a consequence of dense tree cover along the footpath.

In theory, it should be possible to use a combination of spatial and attribute queries to clean up the start and end of such routes. In practice, however, this is difficult to achieve. Duncan et al (2006) use a buffering approach to accept or reject routes relative to home and school addresses. In our case, application of buffers is problematic since the children may be outside either location for a significant period of time (e.g., in the school playground) and these data need to be retained for exposure assessment.

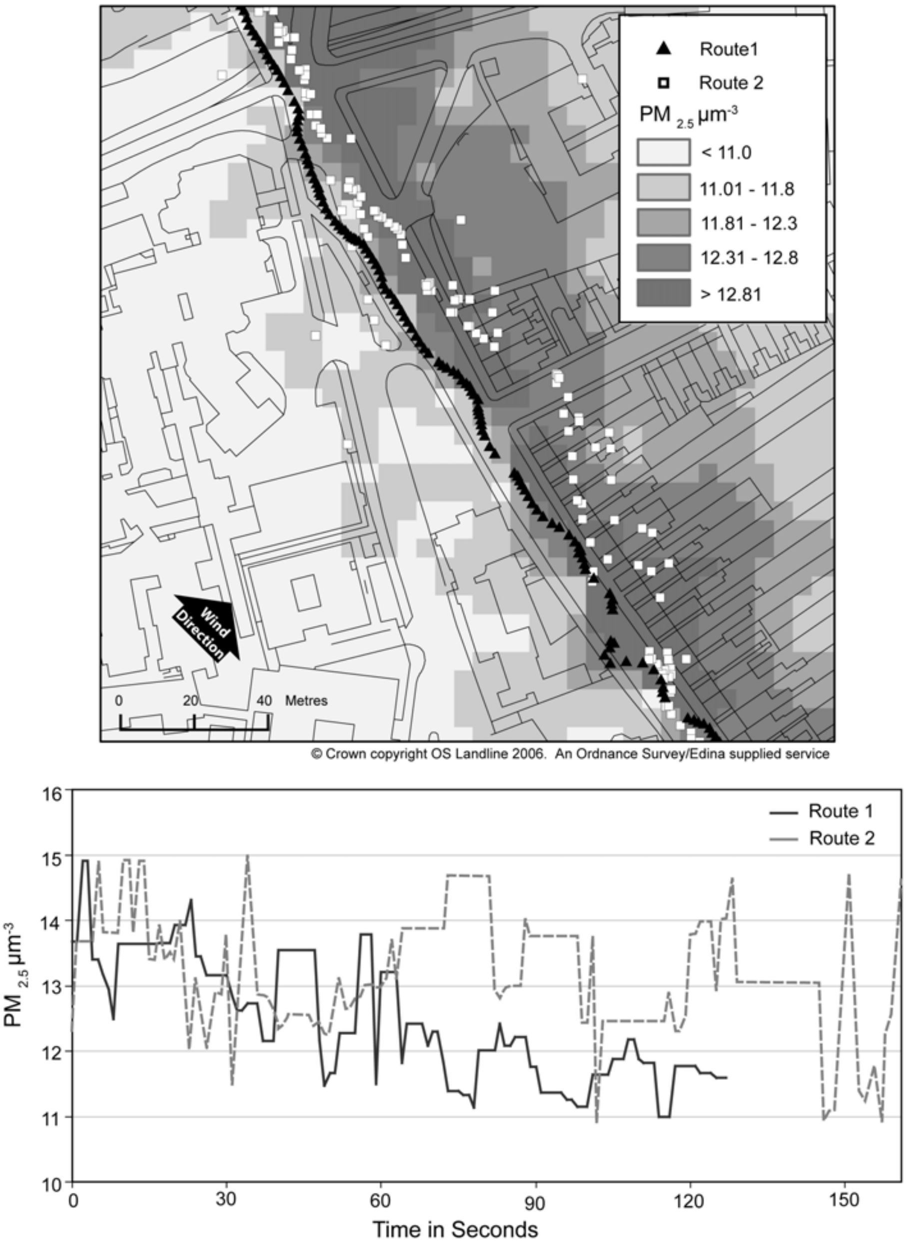
Du et al (2007) use a combination of speed and position to determine ‘dwell times’ which, if greater than some pre-defined threshold (‘time gap’) are taken to represent the end of a journey. This approach may be valid for vehicular based journeys, however, does not seem transferrable to journeys based on foot or bicycle. For example, some children leave home early in the morning, wait at bus stops, travel to school, then wait for considerable periods of time in the school grounds before entering the school buildings (Figures 1a and 1b). Implementing a rule based on the start of the school day is equally problematic. The first bell rings at 8:40am but our data suggests that some children are a considerable distance from school at this time.

In addition to location, speed and bearing the GPS units used in this study provide information on the number of satellites and the horizontal dilution of position (HDOP). One might expect that selecting positions on the basis of a large number of visible satellites and low HDOP would yield optimal results. In our experience, however, the opposite often applies, with continuity in route only possible through the use of positions based on less than four satellites and a high HDOP.

4. Discussion

In order to be useful for personal exposure assessment routes need to be accurate (low degree of scattering) and complete (few gaps between points). Examples of calculated pollution exposure from high and medium quality routes are shown in Figure 2. The higher quality route runs adjacent to a heavily trafficked road and on this occasion is generally upwind of high [modelled] pollution concentrations. The medium quality route, in contrast, is much more poorly defined with many data points falling within the downwind zone of high pollution concentrations. The resulting exposure profiles reflect this difference.

Figure 2. Variations in Personal Exposure (High and Medium Quality Route Data)



5. Conclusions

This project has generated a large amount of locational data which can potentially be used for a wide variety of purposes in addition to exposure assessment. However, these data are noisy. With over 600 potential routes at our disposal it is essential that we develop an automated means of cleaning these data. Attempts to date have highlighted the complexity of the task ahead. Each route presents a unique combination of spatial and temporal circumstances that cannot be addressed through the simple application of rules. This presentation will report on progress made to date and encourage suggestions of alternative approaches from conference participants.

6. Acknowledgements

This project was funded by ESRC small grant RES-000-22-2023. We would like to thank Lancaster City and Lancashire County Council for the provision of data and the staff and pupils of the school participating in this study

References

Brimbicombe, AJ (2007). Location-based Services and Geographic Information Systems in *The Handbook of Geographic Information Science.* Blackwell Publishing.

Duncan, MJ and Mummery, WK (2007). GIS or GPS? A Comparison of Two Methods for Assessing Route Taken During Active Transport. *American Journal of Preventative Medicine* **33** (1) pp51-53.

Du, J and Aultman-Hall, L (2006), Increasing the Accuracy of Trip Rate Information from Passive Multi-Day GPS Travel Datasets: Automatic trip end identification issues. *Transportation Research Part A* **41** pp220-232.

Norwich Union (2007). Pay as you drive Car Insurance. http://www.norwichunion.com/pay-as-you-drive/

Pooley, CG, Whyatt, JD, Walker, M, Davies, G, Coulton, P and Bamford, W (2008), Understanding the School Journey: Integrating Data on Travel, Risk and Environment. In preparation for *Area.*

Biography

Duncan Whyatt is a Lecturer in GIS with interests in air pollution modelling at regional and local scales.

Gemma Davies is a GIS Officer who provides support for teaching and research within the Department of Geography.

Marion Walker was the Research Assistant on the project. She has a PhD in Educational Research and Human Geography and interests in school choice, qualitative methodologies and education policy.

Colin Pooley is Professor of Social and Historical Geography. His research focuses on changes within British society over the past 200 years with particular reference to migration, mobility and health.

Paul Coulton is a Senior Lecturer in Communication Systems and head of mobile game research at Lancaster University. He has been made a founding member of Forum Nokia Champion – a recognition and reward programme that honours a select group of mobile developers from around the world.

Will Bamford is a PhD student in the Department of Communication Systems studying under the supervision of Paul Coulton. He designed and wrote the GeoBlog application used in this study.

Learning to read the landscape: Interactive Visual Tools for exploring spatial thinking

Gary Priestnall[1], Jack March[2]

[1,2]School of Geography, The University of Nottingham, Nottingham, NG7 2RD, UK
Tel. +44 (0) 115 9515443 Fax. +44 (0) 115 95429
Gary.Priestnall@nottingham.ac.uk
www.nottingham.ac.uk/~lgzwww/contacts/staffPages/gary

KEYWORDS: Visualisation, Virtual Environments, Spatial Thinking, Education, Interactivity

1. Introduction

The ability to associate a perspective '3D' view of a landscape with a map-like representation is one tangible component of spatial thinking as described by Golledge (2003). It is relevant to the understanding of how people read maps, how they interpret photorealistic visualisations of terrain models in various contexts including virtual field trips (Whitelock & Jelfs, 2005) and also has implications for the design of graphical aids to navigation and orientation (Augmentra, 2008). This paper will report on the development of an interactive game-like application designed to explore some of these issues and also to couple as a pre-fieldtrip orientation device in advance of a Geography fieldtrip held in Cumbria, NW England. The application is one of several being developed as part of a project entitled 'GeoSpatial Widgets' funded by the Visual Learning Lab (VLL) but also closely aligned to the aims of SPLINT (SPatial Literacy IN Teaching). Both the VLL and SPLINT are Centres for Excellence in Teaching and Learning funded by the Higher Education Funding Council of England (HEFCE).

The first application to be developed through the Geospatial Widgets project is *Locata*, an interactive web-based application which explores the ability of students to associate 3D landscape views with a 2D representation of the same region. The main aims of this application are:

- To explore the ability of students to reconcile 3D views of an area with an equivalent 2D map representation of the same area.
- To introduce students to Digital Terrain Models as representations of a region they will visit on a field trip, whilst familiarising themselves with the major physical features of that area.
- To explore the effectiveness of different 2D representations in supporting the user to locate a 3D view.
- To collect detailed log data recording the timings of each user interaction and the performance in relation to locating various categories of landscape view.

This paper describes the GeoSpatial Widgets project and in particular the *Locata* application including the software development environment, design issues and the potential significance of the log data being collected.

2. The 'GeoSpatial Widgets' project

The VLL aims to research, develop and raise awareness of effective visual learning practices and technologies. The GeoSpatial Widgets project has its origins in a VLL-funded project within the School of Chemical and Environmental Engineering at Nottingham called *Visual LabWidgets*. The aim of this project was to develop a series of virtual, visual laboratory experiments that could be accessed by a wide range of users. These Visual Labs were designed to both test and develop users' 2D and 3D visual and spatial skills while at the same time providing training in fundamental science and engineering concepts as well as strategic thinking and problem solving skills. A number of computer-based interactive applications where developed using the multimedia authoring environment Macromedia® Director®. Although the virtues of interactive multimedia applications for enhancing learning have been promoted for some time (Najjar, 1998) some researchers would argue that the situation is less clear cut (Ainsworth, Bibby and Wood, 2002). The need for very careful design and an appreciation of the contexts of use of such applications are vital (Goldman, 2003). The Director® environment helped to allow the implementation of good design ideas *gaming* styles. This package was also used due to its ability to deliver powerful 3D graphics over the internet via the free, easy to install, Adobe® Shockwave® Player web-browser plugin. Developing predominantly web-delivered applications was considered an essential part of the project, providing an increased potential for student access. Examples of these Visual LabWidgets include a virtual polymerisation plant, a leaking pipe system, and a 3D noughts-and-crosses game (Figure 11).

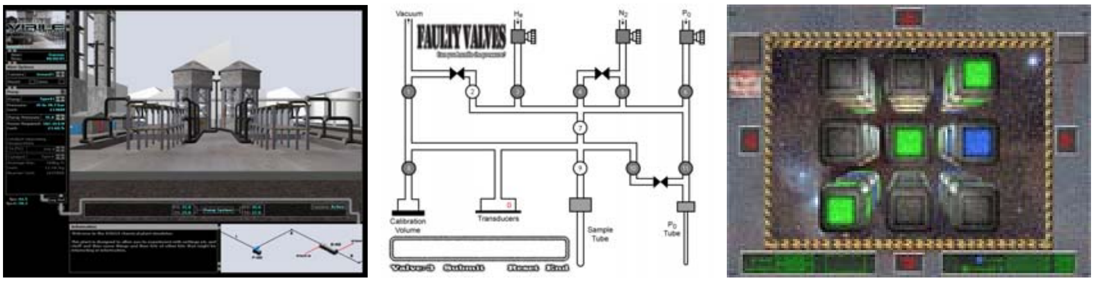


Figure 11 - Visual 'LabWidgets': Virtual Polymerisation plant (left), Fault Values leaking pipe system (centre), and 3DOX (right)

The *GeoSpatial Widgets* project utilises the same interactive visual framework as *Visual LabWidgets* but in a more overtly geographical context. Applications are being developed around a series of case studies focussing on aspects of spatial thinking, *Locata* being one such application.

Clark *et al* (2008) suggested that by focussing on a specific aspect of spatial thinking such as the interpretation of topographic maps it may increase our understanding of how to develop spatial skills tests, which incidentally are increasingly being developed using multi-media techniques.

A review of the relevance of cognitive and learning sciences to spatial thinking in geoscience education by Ishikawa and Kastens (1995) revealed the potential importance of 'constituent strategies' employed by users to locate themselves on a map. By capturing log data from the *Locata* application

3. The 'Locata' application

The landscape chosen for the first version of the *Locata* application is an area around Keswick, Cumbria, NW England. A radar Digital Surface Model (DSM) draped with colour aerial photography formed the virtual model from which perspective views could be rendered by the user within the Shockwave 3D environment.

Figure 2 shows the overall process flow of the application. Users are asked some simple questions about themselves and their familiarity with map reading, 3D computer applications and the geographical area via the start-up screen, before being presented with the main interface, shown also in Figure 2. Users are presented with a series of 3D views (upper right of the interface) selected at random from 100 pre-defined view locations generated from the Digital Terrain Model and attempt to drag a location/direction icon over the map to place the observer at a location where they believe the view was generated from.

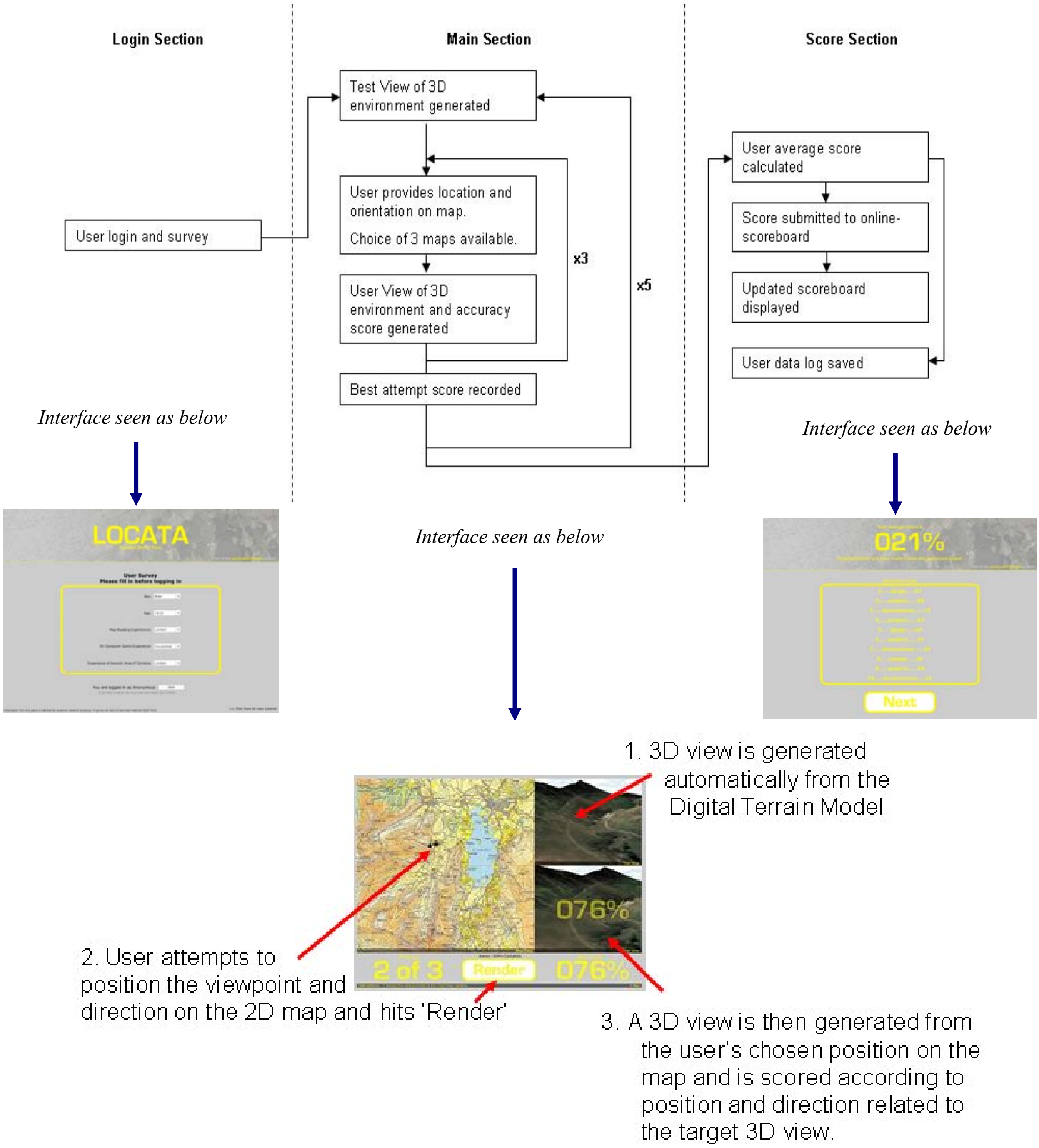


Figure 2. Flow diagram for the Locata application.

The user can select a different map backdrop at any time; Figure 3 illustrating two of the alternative 'map-like' representations currently offered in an attempt to explore their effectiveness to portray terrain in this context. When the user clicks 'Render' a 3D view is generated from the DTM from the user's current chosen viewpoint. A percentage score is generated and displayed on the basis of how well the position and orientation matches that of the target view. After three attempts the next random view is displayed, currently the application shows the user 5 different views. The average score for the session is displayed on a web-based high score table following a typical gaming model to provide an additional means to engage users in 'playing the game'.

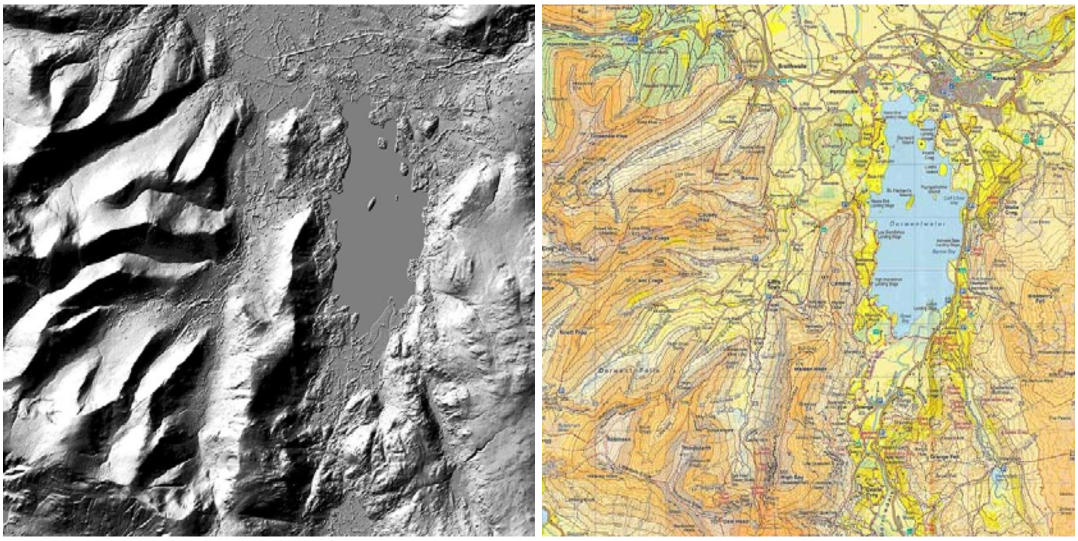


Figure 3 – The hill-shaded terrain model (left) and the Harvey's Mountain Map (right) used in Locata**.**

In addition to the online scoreboard facility, *Locata* also remotely records data from each game played in the form of a text-based log file. An example of such a file is shown in Figure 4. The information recorded includes the survey data supplied by each user at the beginning of *Locata*, which is encoded to facilitate data searching and aggregation. The location and orientation of each test-view (TV), the reference number of the test-view, and the location orientation and score of each of the users three attempts (UV) are recorded for each of the five test-views. The user's choice of map during the exercise (OS, R, HS) and the user's final average score is also recorded (AS). The 100 test views had previously been classified into types based upon aspects of their composition and landscape context with a view to utilising this information in any subsequent analysis of the log data.

15:36:48	06/11/2007	lgzgip1M2D3S4S5S6			
00:00:00	TV	11.13	3.06	-171.25	4
00:00:13	R				
00:00:33	UV	11.89	39.02	-162.47	10
00:01:23	UV	8.38	11.74	-108.92	43
00:02:13	UV	24.09	-14.33	69.71	14
00:02:29	TV	35.82	38.41	123.11	96
00:04:03	UV	14.18	37.5	114.53	31
00:04:18	UV	26.37	33.84	114.53	58
00:05:22	UV	34.15	34.3	92.38	68
00:05:34	TV	-1.83	-19.21	-24.9	86
00:07:26	UV	-43.6	15.24	0	1
00:08:33	UV	44.51	-26.22	0	3
00:09:24	UV	44.51	-26.22	0	3
00:09:34	TV	19.05	-0.3	-54.86	40
00:10:09	UV	13.11	-9.6	-66.37	55
00:11:27	UV	11.28	2.44	-72.25	62
00:12:02	UV	12.2	6.55	-79.11	55
00:12:26	TV	1.83	-10.06	39.8	70
00:13:17	UV	-3.66	8.99	-124.5	19
00:13:56	UV	-7.16	0.46	-124.5	27
00:14:26	UV	-16.92	-18.44	-90	25
00:19:32	AS	40			
15:56:20	06/11/2007				

Figure 4 – Example of log data.

The purpose of *Locata* included pre-fieldtrip orientation and to introduce year 1 students to digital terrain data, however there is clear potential to analyse the log data being captured to address questions such as:

- *Are certain classes of views easier, and quicker, to locate?*
- *Which type of 2D representation proved most popular in supporting observer placement?*
- *Do patterns emerge over the course of a session, or multiple sessions, in terms of the choice of 2D representation or overall performance?*
- *What were the characteristics of individuals who performed well?*
- *Are patterns evident in performance before and after a fieldtrip to the area?*

The application has been tested with a large number of users during the development phase, the large majority being geography students. An early qualitative assessment suggests *Locata* is popular, and in some cases addictive, although it will be properly deployed during the Spring semester of 2008 with year 1 undergraduates, both before and after a fieldtrip to the centre of the region featured. One observation made during early testing is that when students worked in pairs they displayed very animated and geographically interesting commentary on where they thought the views were located and why, and this may be something to explore alongside a more rigorous analysis of the log data.

4. Ongoing work within the GeoSpatial Widgets project

A configurable version of *Locata* has been developed whereby photographs or landscape paintings from known points on the ground can be used as the test views instead of the computer generated views from the terrain model.

A number of other applications are also being developed as part of the GeoSpatial Widgets project. Amongst these is a virtual orienteering application called *Virtualeering*, which simulates some of the basic characteristics of the sport of orienteering (McNeill, Cory-Wright and Renfrew, 1998). This application is based on the same 3D environment used in *Locata* but utilises different game mechanics and learning objectives. Within *Virtualeering* the user is able to navigate around the 3D virtual environment from a first-person perspective, and is required to visit a series of 5 waypoints which are marked on a variety of different map types (Figure 4).

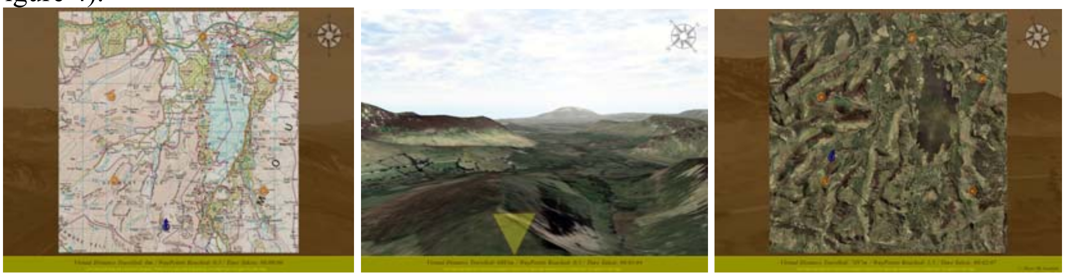


Figure 4 - Virtualeering initial map page (left), 3d environment with waypoint (centre) and photo map with path and location marked (right).

The user must plan their own path to visit each of the waypoints in the quickest time possible. The user's location and orientation within the environment can also be revealed on a map but this incurs a time penalty. As with *Locata* there is an on-line scoreboard facility as well as detailed data logging to provide in-depth information on each user interaction with the 2D and 3D versions of the environment used within the application.

Other applications explore different aspects of spatial thinking, including a real-time overland flow modelling environment and a 3D representation of neighbourhood cells in a water flow direction algorithm dynamically linked to lines of Visual Basic code as they are executed, this being deployed on a Masters level programming module.

7. Conclusions

A game-like web-based application has been developed which introduces students to digital terrain data, familiarises them with a fieldtrip location but also engages them the process of interpreting various 2D landscape representations in building up a picture of the landscape. The log data captured during each session is stored, and along with audio or video recording offers potential for further investigation into this quite specific aspect of spatial thinking.

Acknowledgements
This work was made possible through funding from the Visual Learning Lab based in the School of Education, The University of Nottingham. Digital data for the development of the Cumbria virtual model were obtained from Intermap (the DSM) and Getmapping (aerial photography). The 2D representations used within the Locata application utilised Ordnance Survey mapping through Edina and the Harvey's British Mountain Map: The Lake District (Reproduced with permission from scanned paper mapping, ©Harvey 2008)

References

Ainsworth, S. E., Bibby, P., and Wood, D. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences*, 11(1), 25-61.

Augmentra Ltd (2008) Viewranger [online] Available at: http://www.viewranger.com/ [Accessed 7th January 2008]

Clark, D., Reynolds, S., Lemanowski, V., Stiles, T., Yasar, S., Proctor, S., Lewis, E., Stromfors, C., and Corkins, J. (2008) University Students' Conceptualization and Interpretation of Topographic Maps, *International Journal of Science Education*, 30:3, 377 – 408

Goldman, S. R. (2003). Learning in complex domains: when and why do multiple representations help? *Learning and Instruction*, 13(2), 239-244.

Golledge, R. G. (2003) Thinking Spatially. In *Directions Magazine.* Available online at http://www.directionsmag.com/article.php?article_id=277&trv=1v [Accessed 7th January 2008]

Ishikawa, T. and Kastens, K. A., (2005) Why Some Students Have Trouble with Maps and Other Spatial Representations *Journal of Geoscience Education* Vol 53, No 2, pp.184-197

McNeill, C., Cory-Wright, J., and Renfrew, T. (1998) Teaching Orienteering (Second Edition) Harveys / Human Kinetics

Najjar, L. (1998). Principles of educational multi-media user interface design. *Human Factors*, 40(2), 311-323.

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, Nos. 1/2, pp.121–131.

Biography

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. Jack March is a Research Fellow working on the HEFCE-funded Centre for Excellence in Teaching and Learning – Visual Learning Lab project 'GeoSpatial Widgets', developing game-like visual interfaces to support teaching in learning.

The Fine Scale Spatial Dynamics of the Greater London Housing Market

Andrew T. Crooks[1], Duncan A. Smith[1], Margarethe Theseira[2]

[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, duncan.a.smith@ucl.ac.uk, http://www.casa.ucl.ac.uk

[2]GLA Economics
City Hall, The Queen's Walk, London, SE1 2AA, England
Tel: +44 (0) 207 983 4198
margarethe.theseira@london.gov.uk, www.london.gov.uk/mayor/economic_unit

KEYWORDS: House prices, GIS, Agent-Based Modelling, London

1. Introduction

This paper outlines our current research exploring the fine scale spatial dynamics of the Greater London housing market. Residential property modelling has many applications for informing planning policy from understanding local housing submarkets through to being used as a proxy for measuring the impact of public investment (for example on transport improvements) within specific areas.

In the London context, population and economic growth have created acute pressures on housing. London's population is projected to increase by around 1.14 million to 8.71 million by 2026 (GLA, 2006a) whilst employment is projected to increase by 912,000 jobs between 2006 and 2026 (GLA Economics, 2007). The Further Alterations to the London Plan has set targets of building over 30,000 new homes per year to address this increase in demand. To meet sustainable development and regeneration aims growth is being directed to Inner and East London (Figure 12), with high density development concentrated at areas with good public transport accessibility (see GLA 2004, 2006b for further information).

One established method for the spatial modelling of the housing market is the hedonic approach, where house prices are considered to be a capitalisation of the stream of services provided by the property (Gibbons et al, 2005). These services include internal facilities such as shelter and heating as well as the external opportunities of location, such as access to amenities, jobs and education.

Influences on house prices vary depending on the scale of analysis. For example the expanding jobs market in the South East is driving housing demand at the broader scale and has produced record prices across the entire region. There are also influential factors at very localised scales. For example, the quality of the housing stock can vary widely at small spatial scales in London, and has so far been difficult to model due to a lack of data (Atisreal, 2005). Environmental amenity is also highly valued in the housing market and varies at local scales.

This research will explore the influence of these fine scale factors on the housing market. At present the research is work in progress and the model is under development. The remainder of this paper will discuss our initial research steps.

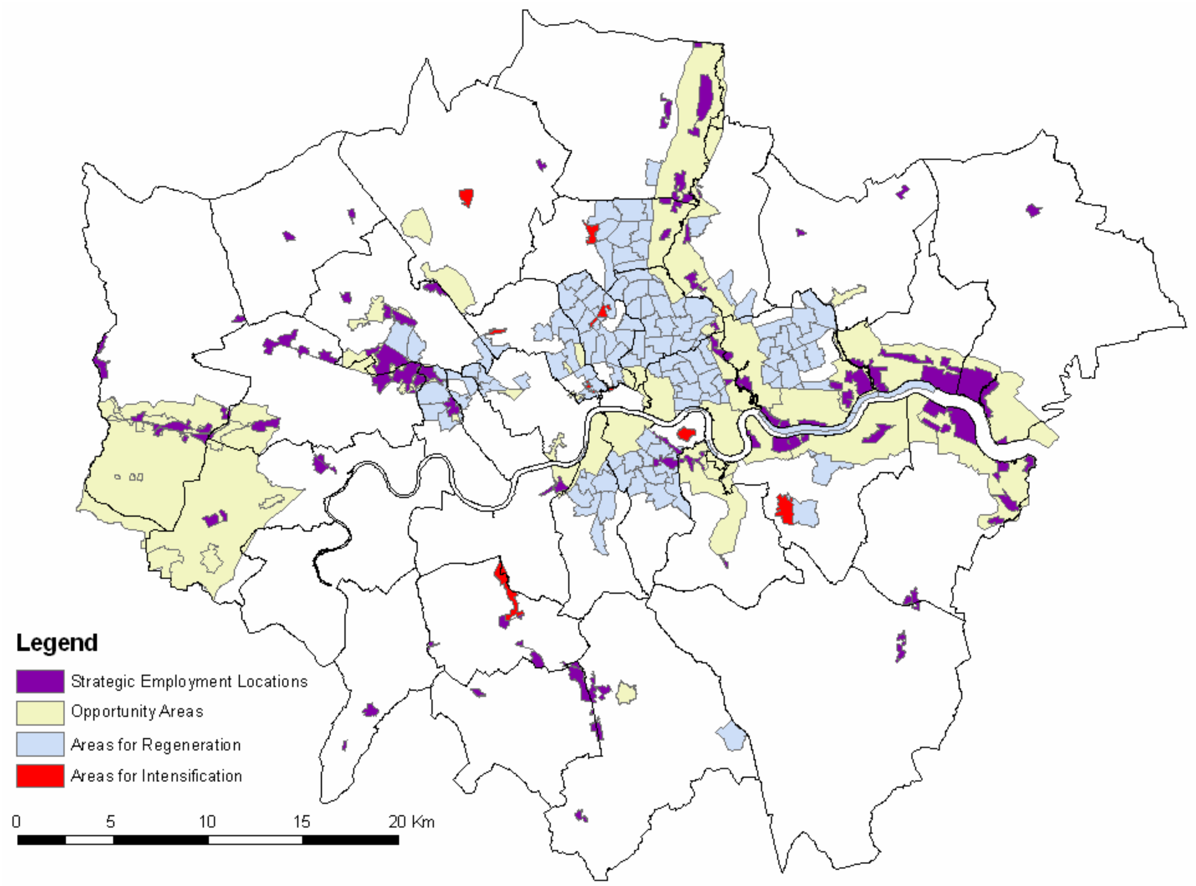


*Figure 12. **Areas designated for change within the London Plan***

2. Methodology

In this research we are aiming to model the most significant influences on the housing market at fine spatial scales. Significant factors identified in recent research (Atisreal, 2005) include:

- Housing quality
- Access to services (jobs, education, local amenities)
- Public transport accessibility
- Crime and deprivation
- Environmental quality

This is in addition to macroeconomic and monetary policy set at a national level, and local taxation.

Within this research we utilise a 3-D GIS/CAD model of London, referred to as Virtual London (Batty and Hudson-Smith, 2005). This is a digital model of all building blocks within Greater London. The dataset has been created from two main sources of data: first vector parcel files part of Ordnance Survey's MasterMap[6] which code all land parcels and streets to at least one meter accuracy; and second a data set of buildings heights constructed from InfoTerra's[7] LIDAR data.. MasterMap's Address Layer 2[8] can be used to link the Virtual

[6] http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/

[7] http://www.infoterra.co.uk/

London data at building level to any data set with postal addresses. The Virtual London dataset will be used to calculate a number of indicators relevant to housing quality (density, private garden space and social housing proportion) and environmental quality (access to public green space, quality of views using building heights and proximity to environmental features such as the Thames).

Combined with the above model we utilise CACI's PayCheck[9] dataset for gross household income and CACI's StreetValue[10] dataset for residential property price, both at the postcode unit level to extrapolate average house price and income. Data is further extracted from the census, Land Registry (e.g. the number and date of sales, the price, the type of property and whether or not it is a new build) and the Greater London Authority (GLA) (e.g. Public transport accessibility levels, population catchments by public transport within 45 generalised minutes[11]). These datasets, once combined provide a unique fine scale information of London giving more accurate modelling options for specific locations.

3. First Steps/Preliminary Results

As this is work in progress it is felt more appropriate to highlight how the above datasets provide us with an understanding of the London housing market. Further analysis is needed before any broad conclusions can be made. For this reason we will focus one some aggregate analysis before focusing on one area of London, that of East London's Docklands.

Simple aggregate analysis of tells us little about the London housing market. For example, the mean house-price when the data was cleaned to remove postcodes with zero population within London is £326,701 with a standard deviation of £229,779. Table 3 clearly demonstrates that comparing residential property prices against selected variables highlights that there is great variation within London. This is further supported when trying to compare one variable against house prices using a linear regression, low R^2 values are seen a for all independent variables apart from household income. This suggests there is much geographical variation as shown in the GLA's 'London's Housing Submarkets' report (GLA Economics, 2004) and more specialist analysis is needed (which we are currently working on). For this reason we are focusing on the fine scale analysis of selected areas. This is to which we now turn.

Table 3. ***Summary statics of house prices at the unit postcode in London against a selection of variables***

Test		Household income (£)	Public transport accessibility levels	Population catchments by public transport within 45 min	Straight Line Distance (m) to Place of Education (School)				Straight Line Distance to Tube Station (m)	Straight Line Distance to Train Station (m)	Straight Line Distance A Road (m)	Straight Line Distance to River (m)
					Nearest school of any type	Nursery	Primary	Secondary				
Mean		39265	4.72	1309585	249	2637	378	838	2372	972	290	5878
Stand Deviation		10370	2.03	748351	169	2447	234	456	2843	641	311	4491
R^2		0.201	0.026	0.042	0.002	0.002	0.019	0.005	0.018	0.005	0.001	0.006
Coefficients	Value	9.93	18306.75	0.06	61.56	-4.19	135.69	34.35	-10.87	25.32	-16.78	-3.93
	Std. Error	0.05	293.67	0.00	3.58	0.25	2.56	1.32	0.21	0.94	1.94	0.13
T Value		190.76	62.34	79.83	17.21	-17.00	53.09	26.02	-51.67	26.95	-8.63	-29.35

[8] http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/layers/addresslayer2/

[9] http://www.caci.co.uk/msd.asp?url=cm-data-paycheck.htm

[10] http://www.caci.co.uk/msd.asp?url=cm-data-streetvalue.htm

[11] Generalised time is the total time spent on a journey, weighted to account for traveller preferences.

The Docklands has been the setting for dramatic urban change over the last 25 years, with large scale market led redevelopment of industrial land creating localised clusters of new housing and resulting in sharp social contrasts. Figure 2 illustrates the distribution of house prices in 2007. High price areas appear to cling to the river, due to new upmarket high density apartment developments at waterfront locations, such as Shadwell and Limehouse (the historic upmarket residential location of Greenwich is also visible, whilst new housing developments at Greenwich Peninsula are not yet complete). Clearly data on new build housing and the value of river views will be significant in modelling property market changes in this area.

Inland from the Thames are areas of much larger social housing estates, such as Peckham and Poplar, which fall into the lowest band of house prices. The new waterfront developments are often adjacent to smaller areas of 1960s and 1970s social housing, as for example on the Isle of Dogs where the social housing appears as small clusters of lower than average prices. Figure 3 shows this in detail, with upmarket new build along the waterfront and social housing tower-blocks behind. Both housing types can be at high densities, implying that relationships between density and price are dependent on the context of housing age and quality.

Market trends towards luxury waterfront housing risk exacerbating social polarisation in Docklands. To minimise social division the 2004 London Plan (GLA 2004) has set the target of a 50% proportion of affordable housing in new developments.

There is no clear relationship between house prices and proximity to public transport interchanges. Research provides evidence of the influence of public transport accessibility on house prices within a kilometre of interchanges (Gibbons and Machin, 2005) but this is by itself not a sufficient condition for higher priced housing (Figure 2).

Figure 13. ***East London / Docklands House Price Surface Map.***
Source: CACI StreetValue data.

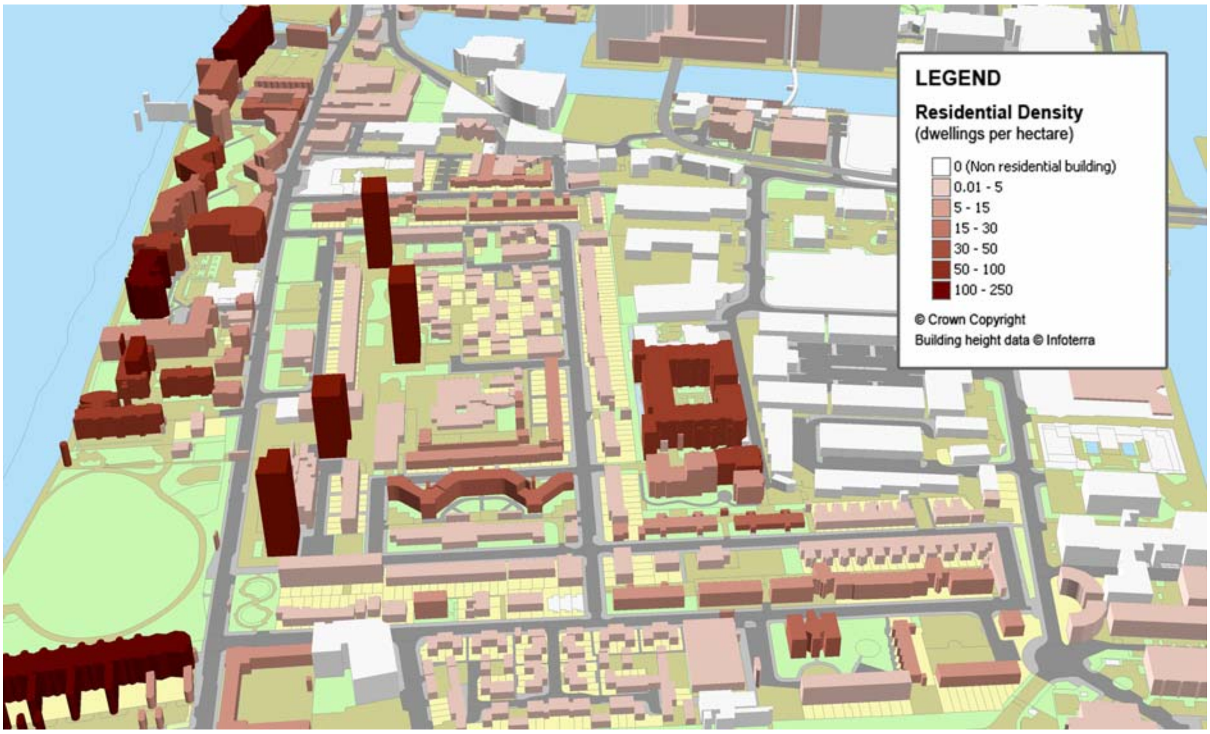


Figure 14. ***Isle of Dogs Housing Density 3D Map.***
Source: Virtual London.

4. Further Work

Fine resolution data, potentially allows us to create models exploring the small scale dynamics of the London housing market from the individual perspective thus furthering our understanding of this complex issue. Agent-based modelling (ABM) is inherently suited to such a study as it allows the representation of an heterogeneous population with individual agents having different behaviours and characteristics. These agents interact locally to form more complex aggregate patterns. Some recent examples of such housing models include Bossomaier *et al.* (2007), Torrens (2007) and Torrens and Nara (2007). There is also a growing interest in combining ABM and GIS (see Castle and Crooks, 2007). However many of the ABM applications currently utilising geospatial data do so using a cellular space representation of reality which involves populating regular cellular space with agents of one or several kinds which can migrate between cells. While agent-based models created using the cellular partition of space have provided valuable insights into urban phenomena especially as they can capture geographic detail, they miss geometric detail. This area is critical to good applications but is barely touched upon in the literature (Batty, 2005) with a few exceptions, (e.g. Benenson *et al.*, 2002). The ability to represent the world as a series of points, lines and polygons allows the inclusion of geometry into the modelling process, therefore allowing for different sizes of features such as houses, roads and so on to be portrayed. This is currently being achieved through the utilisation of the data sets described above. For example, using MasterMap TOIDS (Topographic Identifier) to represent individual buildings; MasterMap's Address layer to populate these building with a number of units and assigning individual agents to each of these units.

A prototype model is currently being developed focusing on residential location and dynamics within London (Figure 15). This model is loosely coupled with GIS and explores vendor/buyer behaviour whereby the agent's decisions/behaviour of where to locate is affected by spatial attributes of actual land-parcels.

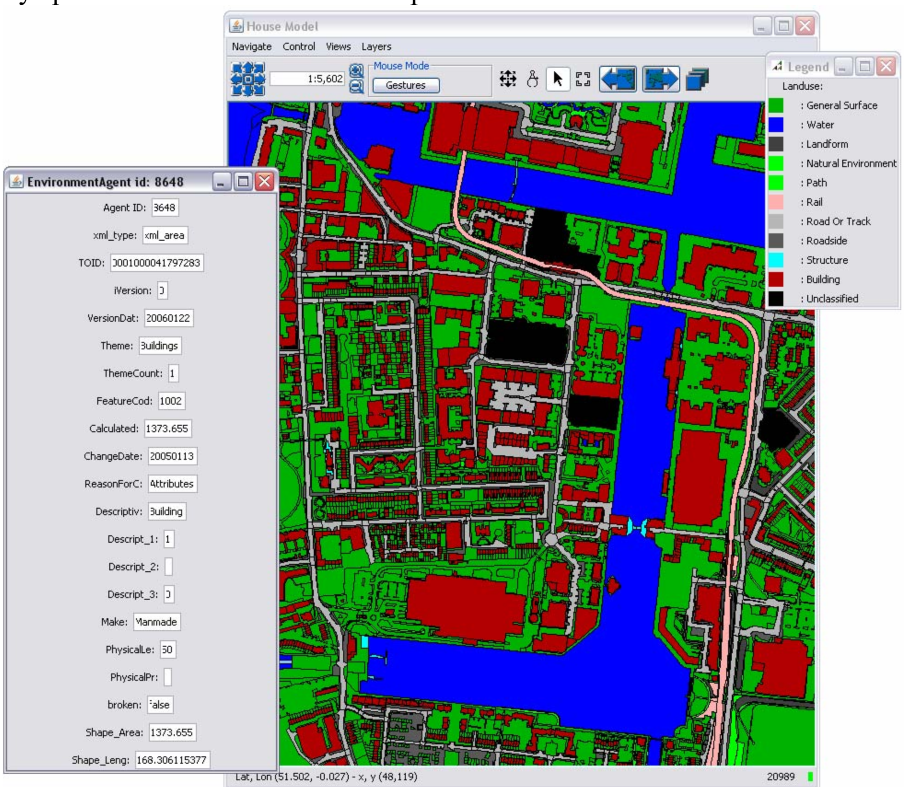


Figure 15. ***Prototype agent-based model (showing the built environment dimensioned to the geography of the Isle of Dogs, Tower Hamlets). ©Crown Copyright.***

5. Conclusion

This paper has outlined how we are building up a detailed understanding of the Greater London housing market. To enhance this understanding three avenues of research are being explored. The first is building a comprehensive dataset of the London housing market. Access to temporal datasets such as sales data at the individual level is key to this research and we are currently acquiring such a dataset. This potentially allows to the quantification of the impact of infrastructure in certain locations (e.g. Gibbons and Machin 2005; Atisreal, 2005). Furthermore, detailed information about who lives in places, for example information from the Electoral Role is needed. This will lead to a detailed picture of the social class and ethnicity of areas (Mateos 2007). Secondly several advanced spatial analysis techniques, for example, geographical weighted regression (Fotheringham *et al.,* 2002) are being applied to the data. Thirdly, the development of an agent-based model which allows us to model residential locational decisions with agents rules loosely based on information from the data analysis. It is anticipated that these combined approaches will lead to a greater understanding of the London Housing Market for informing policy options by the GLA.

6. Acknowledgements

The authors would like to thank Alex Singleton for providing a detailed school dataset and introducing us to CACI's datasets.

References

Atisreal (2005). *Property Value Study - Assessing the Change in Values Attributable to the Jubilee Line Extension*, Transport for London, London.

Batty M. (2005), *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals.* The MIT Press, Cambridge, MA.

Batty M and Hudson-Smith A (2005). Urban Simulacra, *Architectural Design*, 75 (6) pp42-47.

Benenson I Omer I and Hatna E (2002). Entity-Based Modelling of Urban Residential Dynamics: The Case of Yaffo, Tel Aviv, *Environment and Planning B: Planning and Design*, 29(4) pp 491-512.

Bossomaier T Amri S and Thompson J (2007). Agent-Based Modelling of House Price Evolution, *Proceedings of the 2007 IEEE Symposium on Artificial Life (CI-ALife 2007)*, Honolulu, HI, pp. 463 - 467.

Castle CJE and Crooks AT (2006). *Principles and Concepts of Agent-Based Modelling for Developing Geospatial Simulations*, Centre for Advanced Spatial Analysis (University College London): Working Paper 110, London, UK.

Fotheringham SA Brunsdon C and Charlton M (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships* Wiley, Chichester.

Gibbons S and Machin S (2005). Valuing rail access using transport innovations, *Journal of Urban Economics*, 57(1) pp 148-169.

Greater London Authority (GLA) (2004). *The London Plan: Spatial Development Strategy for Greater London.* Greater London Authority, London, UK.

Greater London Authority (2006a), DMAG Briefing 2007/32 – GLA 2006 Round Demographic Projections, Greater London Authority, London, UK

Greater London Authority (GLA) (2006b). *Draft Alterations to The London Plan: Spatial Development Strategy for Greater London.* Greater London Authority, London, UK.

GLA Economics (2004). London's Housing Submarkets. Greater London Authority, London, UK.

GLA Economics (2007). *Working Paper 20- Employment projections for London by sector and borough*. Greater London Authority, London, UK.

Mateos P Webber R and Longley PA (2007). *The Cultural, Ethnic and Linguistic Classification of Populations and Neighbourhoods using Personal Names*, Centre for Advanced Spatial Analysis (University College London): Working Paper 116, London, UK.

Torrens PM (2007). A geographic automata model of residential mobility. *Environment & Planning B* 34 (2) pp 200-222.

Torrens PM and Nara A (2007). 'Modelling Gentrification Dynamics: A Hybrid Approach', Computers, *Environment and Urban Systems*, 31(3) pp 337-361.

Biography

Andrew Crooks is a GLA Economics Research Fellow in Urban Systems at the Centre for Advanced Spatial Analysis (CASA) at UCL. He has recently completed a PhD entitled "Experimenting with Cities: Utilizing Agent-Based Models and GIS to Explore Urban Dynamics." His research interests are structured around issues effecting world cities specifically focusing on London.

Duncan Smith is in his second year of his PhD at CASA who holds an ESRC CASE award sponsored by London Connects for researching '3D GIS for e-planning and e-democracy' which he is researching in the context of CASA's Virtual London project.

Dr Margarethe Theseira is manager of GLA Economics (a unit that conducts economic analysis for the Greater London Authority Group). She is experienced in working with GIS in both the public and private sector. She obtained her geography degree from Liverpool University in 1994 and her PhD from Cranfield University in 2001.

Visualising London's Suburbs

Catherine Emma Jones[1], Muki Haklay[1],
Sam Griffiths[2] and Laura Vaughan[2]

[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
ucfncej@ucl.ac.uk, www.sstc.ucl.ac.uk

[2]The Bartlett School of Graduate Studies, University College London, 1-19 Torrington Place, London, WC1E 7HB
Tel. +44 (0)20 679 1833 Fax + 44 (0)207 9161887
www.sstc.ucl.ac.uk

KEYWORDS: Suburban Geography, Geo-visualisation, GIS, Space Syntax, Web Mapping

1. Introduction

Historically, London's suburbs are, "as old as the city itself" (Ackroyd 2000, p.727) but it is only recently that they have begun to assert themselves on the policy agenda alongside the metropolitan centre. In June 2007 the London Assembly's Planning and Spatial development Committee released a report noting the difficulties predicted for London's suburbs (GLA 2007). However, future development plans for London remain dominated by policy focusing on its core urban area despite evidence of economic stagnation, social exclusion and increasing congestion in the suburbs (Potts et. al., 2007). This is symptomatic of a planning culture that seems reluctant to acknowledge the distinctive contribution of the suburbs to the quality of urban life overall (Kochan 2007). The unspoken assumption is that the suburbs are able to 'look after themselves'. Yet if the challenges facing London's suburbia are not addressed they might face a significant decline.

The suburbs have a greater presence in the academic literature than in the policy debate. Bourne (1996) identified no fewer than ten common interpretative and explanatory approaches to the suburbs from a wide range of socio-economic and cultural perspectives. To name just three, the suburbs have been theorised as an efficient mechanism for capitalist accumulation (Harvey 1999: 122-123), historicised in terms of people's increasing preference for a suburban lifestyle (Clapson 1998) and portrayed as a rural arcadia (Rowley 2006: 195-208). The perspective adopted here, emphasises the historical nature of suburbs as nucleated settlements, addressing the potential of these places to continue as centres of social and economic vitality within emerging polycentric metropolitan regions. This approach is suggested by the morphology of London itself which is widely recognised as a city comprised of villages and sub-centres (Hillier 1999; GLA 2002).

In both the policy and academic arenas there is a need for the provision of more accessible information on Greater London's suburban town centres, to improve understanding of their distinctive characteristics as places to live and work within an extended metropolitan region. The EPSRC Successful Suburban Town Centres (SSTC) project at UCL seeks to meet this need by profiling town centres according to their socio-economic activity, morphological characteristics, commuting patterns and typical modes of transport use. The project draws on a range of methods for the spatial analysis of social and economic activities at various scales. The visualisation of urban form using spaces syntax methodology is of particular interest, enabling the spatial structure of streets and the layout of the buildings to be compared with information about the people who live and work in suburban centres.

This paper describes one phase of the project: the development of an internet based geo-visualisation tool which facilitates visual comparisons of 20 suburban centres. The centres are located in London's outer suburbs, between the M25 and the north/south circular – London's inner and outer orbital roads, chosen because they underwent considerable development during the inter war periods (1919 to 1939). This represents the earliest major period of growth of those London's suburbs (Whitehand and Carr, 2001). The tool enables the user to explore a variety of map themes at consistent scales, enabling local knowledge about the suburban environment to be compiled using a comparative method of transitions to discover patterns within and between centres. The motivation behind the development of the geo-visualisation tool known as the town centre profiler tool was two-fold: firstly, to enable the development of a series of hypotheses to direct the analytical phase of the project, and secondly, to provide local planners with an enriched picture of the local neighbourhood and its suburban structure.

2. Methodology

2.1. Town Centre Profiler Development

The project's interest in uses of the town centre meant attention was paid to activity-generating land uses such as offices, shops and community facilities. A range of activity-generating mapping themes were identified from a diverse range of data sources, listed in Table 1 and 2. The data were processed automatically using MapBasic scripts to produce consistent geovisualisations across all the centres in terms of the scale of the area, ranges, colour scheme and the background map, which in turn ensured they could be directly compared to each other. Thematic maps produced for raw counts were classified according to the shape of the distribution using geometric or arithmetic progressions, percentage maps used a diverging colour palate (Brewer 2005). The aim was to reduce the false assumptions that can be derived from poorly drawn maps, whilst noting the documented limitations associated with using administrative data unit (Openshaw, 1984).

Table 1. Socio-economic mapping themes and associated data sources

Data	Description	Areal Unit	Source
Socio-economic data			
Car Ownership	Car or van ownership per household	Census output area	Census Table KS17
Commuting patterns	Contains journey-to-work flows within and between output areas in England	Census output area	2001 Census: Special Workplace Statistics (Level 3) Data were modified via a process known as *Small Cell Adjustment Methodology (SCAM)*. The SCAM process is a disclosure control mechanism. Cells with a 'small' value (0 to 3) are adjusted and given either a 0 or 3 to project individuals (ONS, 2001)
Infrastructure	Railways, Stations (Tube and Rail), A, B and minor roads, motorways and car parks	Polylines and points	Ordnance Survey Meridian and Address Layers 2
Method of travel to work	Usual population aged 16 to 74 and their method of travel to work. The method of travel to work is for the	Census output area	Census Table UV39 (Resident population)

Data	Description	Areal Unit	Source
	longest part, by distance, of the usual journey to work. The data was mapped for town as place of work and place of residence		Census Table UV37 (Workplace population)
Functional activity land uses	Address Layer 2 is a product derived primarily from the Royal Mail postal address file (PAF), whereby each postal address (delivery point) has been allocated a unique reference and national grid reference. Supplementary to these data Address Layer 2 also incorporates information about geographically derived address locations that do not have specific postal addresses	Points based on national grid coordinates (x,y).	Ordnance Survey Address Layer 2
Socio-economic classification of occupation	Data for population aged 16 to 74 by their socio-economic classification, the replacement in the most recent census for social class based on occupation	Census output area	Census Table UV31 (Resident population) Census Table UV76 (Workplace population)

Table 2. Space Syntax mapping themes and associated data sources

Space Syntax measures The axial map is used in space syntax analysis to represent and analyse all open public space as a continuous spatial network in order to measure how well connected each street space is to its surroundings. This is done by taking an accurate plan of a built up area and drawing the set of least and fewest lines that cover all the open space ensuring that lines intersect where adjacent spaces are contiguous. Space syntax analysis computes all the lines in the network according to their relative depth from each other. Depth increases with the number of changes of direction between lines, see Hillier, B. (2007).			
Choice	Segment analysis takes each axial line and breaks it into segments at the intersections between axial lines. Segment analysis is concerned with the angular properties of graphs by calculating the relative straightness (least angular deviation or 'angular depth') of *each* segment from *all* other segments in the system. Choice is calculated by counting the number of times each segment falls on the shortest path between all pairs of segments within a selected distance-radius where 'shortest path' refers to the path of least angular deviation or straightest route through the system.	Segments	The maps are sections extracted from the segment map of the Greater London area bounded by the M25.

The script enabled the production of more than 1000 map images across 12 different mapping themes (each with a number of sub themes) for each of the 20 case studies, with a selection of different backgrounds. The static image outputs then became the primary composites of the internet profiling tool.

2.2. Development of internet based profiling tool

The principal purpose of the maps was to generate phenomena to assist in hypothesis exploration and formation. This stands opposed to the conventional approach of using maps to disseminate existing knowledge (MacEachren, 2004). This combined with the interdisciplinary nature of the project team meant that cartographic visualisations needed to be produced for exploratory analysis, but without reliance on GIS technology expertise, therefore interactive web mapping tools were not appropriate for the users. Due to the large number of visualisations it was not suitable to print them out for closer inspection; the volume of images would make this unwieldy. Neither was it realistic to view the images using an image viewer as this would reduce the flexibility the user would have. The most appropriate method to view the output images was considered to be a customised profiling tool for dissemination on the internet which would provide access to non-GIS users.

To develop the tool, a number of functional requirements were identified, with importance placed on usability. Firstly, the user has the ability to select a town centre and then browse through the mapping themes (vertical transitions). Secondly, the user must be able to select a mapping theme and then compare it across different town centres (horizontal transitions). This requirement was imperative because the project is interested in sets of relations across centres, as well as within. For this reason, it was decided to enable the user to view the chosen map theme as a slideshow which automatically scrolls through all of the town centres. The third key requirement was to separate the menus for the socio economic mapping themes and the space syntax menus, because they present very different types of information to the user. The fourth and final requirement was to enable the profiling application to be loaded full screen, for a resolution of 1280 by 1024, in the web browser (latest versions of Firefox and Internet Explorer); meaning it could be seen as an independent application, ensuring the size of the map images could be optimised sufficiently, to allow proper evaluation of the information. Noticeably, the decision on the resolution and implementation of the application breaks established guidelines for website development (Nielsen, 1999). However, the decision to break away from these standards meant we could to utilise the advantages of the browser's environment, whilst ensuring that our geo-visualisation tool remains effective.

The website was developed using a mixture of client side JavaScript’s, server side PHP scripts, and a template HTML page. The resultant user selection compiles the interactive html webpage. A screenshot of the developed town centre profiling tool is included in Figure 1. The use of the browser's environment for the dissemination seemed natural, because it is easy to implement and allows the dissemination of the final output over the internet beyond the specific use of the project.

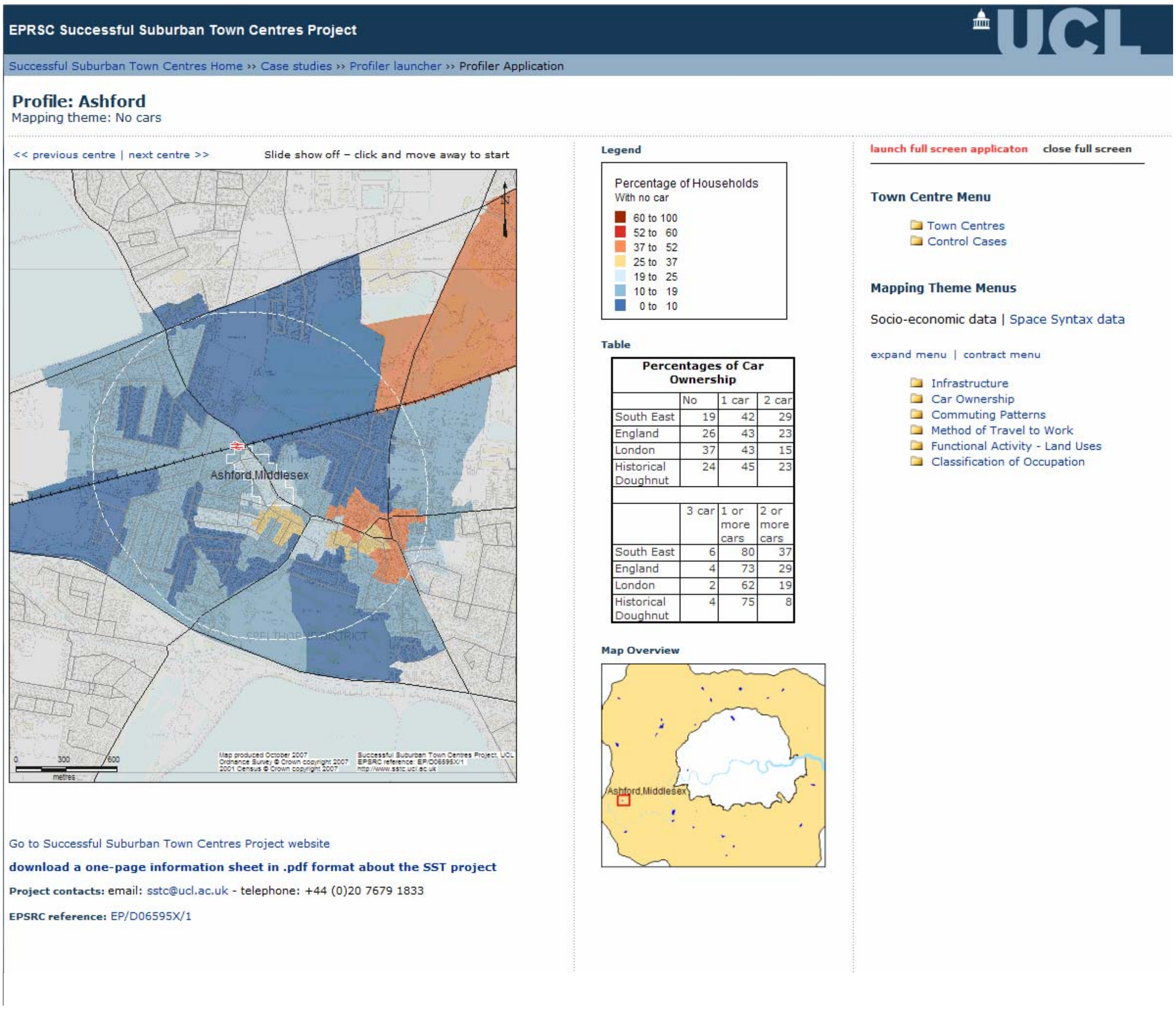


Fig. 1: Screenshot of interactive town centres profiling tool

3. Discussion

The building of local knowledge is essential to a project of this type. A tool such as the one described provides a mechanism for building knowledge by compiling abstract data sources into a coherent and structured interface. With the rise in Web 2.0 and its user friendly mapping facilities such as Google Maps or Microsoft Virtual Earth, the use of static map images may be considered by some to be outmoded. However, whilst the town centre profiler is not providing the high level of interaction with the map it has a number of advantages over using the recently emerging interactive web mapping interfaces.

Creating the maps as images which are then made available to users enables the GIS expert to maintain control of the content, design and scale at which the visualisations are produced. This means they are not merely visually aesthetic but portray relevant information (Monmonier, 1996). A primary motivator for its development was to enable non-GIS users to visually critique and explore a large quantity of information in a user friendly manner. As David Unwin noted in 2005, “In software, the idea of user friendliness is usually equated with interface design, but the concept should also be used in relation to how well the tools provided map into the users perceptions and expectations of what needs to be done with them” (page 683). This tool was built specifically for a clear purpose.

The simplicity of the application provides one advantage to the user and their quest for knowledge. The tool is essentially an interactive atlas of maps, where the menu options are a multidimensional index. The maps themselves cannot be changed, they are static, and this ensures the user considers in detail the actual content of the maps and their meaning. The profiler carefully considers the end use of the maps and the types of functionality required to make them useful, whilst limiting functionality found in a traditional GIS such as panning, zooming, classifying themes and changing colours. This is because the interaction of the map itself would detract away from the content and meaning for non expert users. This is especially important as our project team comprised of researchers from various academic backgrounds. These reasons, combined with the large number of centres, scattered across a regional spatial extent and diverse range of mapping themes, provided the impetus for the slideshow functionality with its gradual transitions enables horizontal comparison across each of the town centres. This enables similarities and differences to be visually examined. Such transitions in a web GIS would be messy and difficult to implement and control, and introduce more complexity than necessary into the system.

What is more, because the images are all produced at the same scale, the application controls the extents of the map making it much easier for the user to move around the different centres; ensuring all centres and map themes conform to conventional standards and by corollary this encourages sensible interpretations and hypotheses to be developed. The profiler was used successfully in a number of project workshops to provoke thought and facilitate potential ideas to be investigated in depth for each of the 20 centres. We were able to reinforce one of the initial hypotheses that suburban centres are not merely sleepy residential commuter towns reliant on the urban centre. The visualisations showed a wider range of activity generating land uses prevalent in our centres and their local hinterland. Additionally we were able to hypothesise that local patterns of movement and integration play an important role in generating activity. The visualisations have enabled the project team to demonstrate a need to understand suburbs as ‘multi-dimensional’ places characterised in part by the complexity of their relations with other places, rather than by a particular relationship to a single central place.

6. Acknowledgements

The research reported in this paper forms part of a 36 months study funded by the EPSRC (Engineering and Physical Sciences Research Council, start date 01/10/06. EPSRC reference EP/D06595X/1). Kind thanks to the Ordnance Survey for granting permission to publish the profiler in the public domain which will be available through http://www.sstc.ucl.ac.uk/

References

Ackroyd P (2000). *London: the biography*, London, Chatto and Windus.

Bourne L S (1996) Reinventing the suburbs: old myths and new realities. *Progress in Planning* 46.3: 163-84.

Brewer CA (2005). *Designing better maps, a guider for GIS users*, Redlands, California, ESRI Press.

Clapson M (1998) *Invincible green suburbs, brave new towns: social change and urban dispersal in postwar*, England, Manchester: Manchester University Press.

Harvey D (1989). *The urban experience.* New York, London, John Hopkins University Press.

Hillier B and Iida S (2005) *Network effects and psychological effects: a theory of urban movement,* 5th International Space Syntax.

Hillier B (1999) Centrality as a process: accounting for attraction inequalities in deformed grids. *Urban Design International* 4: 107-127.

Kochan A (2007). *Achieving a suburban renaissance: the policy challenges*, Town and Country Planning Association for the Joseph Rowntree Foundation, London

GLA (Greater London Authority) (2002) *A city of villages: promoting a sustainable future for London's suburbs.* Report by URBED with the TCPA, London: GLA.

GLA (Greater London Authority) (2007) *Semi-detached: reconnecting London's suburbs*, London Assembly Planning and Spatial Development Committee for the GLA, London, GLA.

MacEachren A (2004). *How maps work*, New York, The Guilford Press.

Monmonier M. (1996). *How to lie with maps*, Chicago, The University of Chicago Press.

Nielsen J (1999). *Designing web usability: the practice of simplicity*, Indianapolis, New Riders Publishing.

Office for National Statistics (2001) *Census: special workplace statistics (Level 3)* [computer file], ESRC/JISC Census Programme, Census Interaction Data Service, University of Leeds and University of St. Andrews

Openshaw S (1984). The modifiable areal unit problem, Norwich, United Kingdom,Geo Book,

Phelps N A, Parsons N, Ballas D , Dowling A. (2006) *Post-suburban europe: planning and politics at the margins of Europe's capital cities.* Basingstoke UK, New York US, Palgrave Macmillian,

Potts G, Falk N and Kochan B (2007) *London's suburbs unlocking their potential, A report accompanying the London Assembly's enquiry into London suburbs*, London, BURA & URBED

Rowley T (2006). *The English landscape in the twentieth century*. London, New York, Hambledon Continuum.

Whitehand J W R and Carr C M H (2001). *Twentieth-century suburbs: a morphological approach*, London, Routledge.

Biography

Kate (Catherine) Jones has an MSc in Geographical Information Systems from UCL and spent three years working on a Knowledge Transfer Partnership between Camden Primary Care Trust (PCT) and UCL Department of Geography implementing GIS and Geodemographic tools for health care. She will shortly be submitting her doctorate to the Department of Geography at UCL. She is currently a research associate on the SSTC project at UCL.

Mordechai (Muki) Haklay is Senior Lecturer in Geographical Information Science at the Department of Civil, Environmental and Geomatic Engineering, UCL. He has been working on socio-economic analysis using GIS, novel analyses techniques with GIS and usability *aspects of GIS. He has published papers on these subjects in Area, International Journal of Geographical Information Science, and the Journal of Environmental Management.*

Sam Griffiths has expertise in space syntax, urban morphology and historical methods. His academic background is in urban history and he has recently completed writing-up his doctoral thesis on the urban transformation of nineteenth-century Sheffield. He is currently working in the Bartlett School of Graduate Studies as a Research Fellow on the EPSRC Towards Successful Suburban Town Centres project.

Laura Vaughan is Senior Lecturer in Urban and Suburban Settlement Patterns and Co-Director of the MSc Advanced Architectural Studies, Bartlett School of Graduate Studies, UCL, She has written widely on urban and suburban settlement patterns, with a focus on space syntax as a method for fine grain analysis of built form. Currently she is Principal Investigator on an EPSRC grant - 'Towards Successful Suburban Town Centres: a study of the relationship between morphology, sociability, economics and accessibility'.

Modelling residential living space for individual properties using digital infrastructure and remote sensing data: a Cardiff case study

Scott Orford[1], Jonathan Radcliffe[2]

[1]School of City and Regional Planning, University of Cardiff
Glamorgan Building
King Edward VII Avenue
Cardiff, Wales. CF10 3WA
Tel. +44 (0)29 20875272 Fax +44 (0)29 20874845
orfords@cardiff.ac.uk
http://www.cf.ac.uk/cplan/contactsandpeople/stafflist/m-r/orford-scott-dr-overview.html
[2]radcliffej@cardiff.ac.uk
http://www.cf.ac.uk/cplan/contactsandpeople/stafflist/m-r/radcliffe-jonathan-mr-overview.html

KEYWORDS: Housing, Living space, OS Mastermap, LiDAR, urban modelling

1. Introduction

Planning is a key element to the success of sustainable urban environments and successful planning relies on access to good quality information. In strategic and spatial planning, this information is often derived from models and simulations of urban systems and hence there is a real need (and demand) for well-founded, data-rich environments in which to develop and test these models (Batty, 1995). With respect to models and simulation of housing markets, British cities are hindered by a lack of detailed structural attribute data at the level of the individual property. A key issue is the lack of data on living space as this is the single most important structural attribute in determining house price (Orford, 2000) and is the most important variable in any housing market model. The British Valuation Office Agency (VOA) holds the most extensive set of property attribute data in the UK, but the vast majority of these data are held on paper records (VOA, 2002) making them both difficult and costly to maintain and access. Researchers and policy makers often have to rely on a sample of properties for their models, typically without data on living space, and this can lead to problematic results – for example, the issues associated with using beacon properties in council tax valuations or the uncertain and frequently contradictory forecasts of the state of the national housing market.

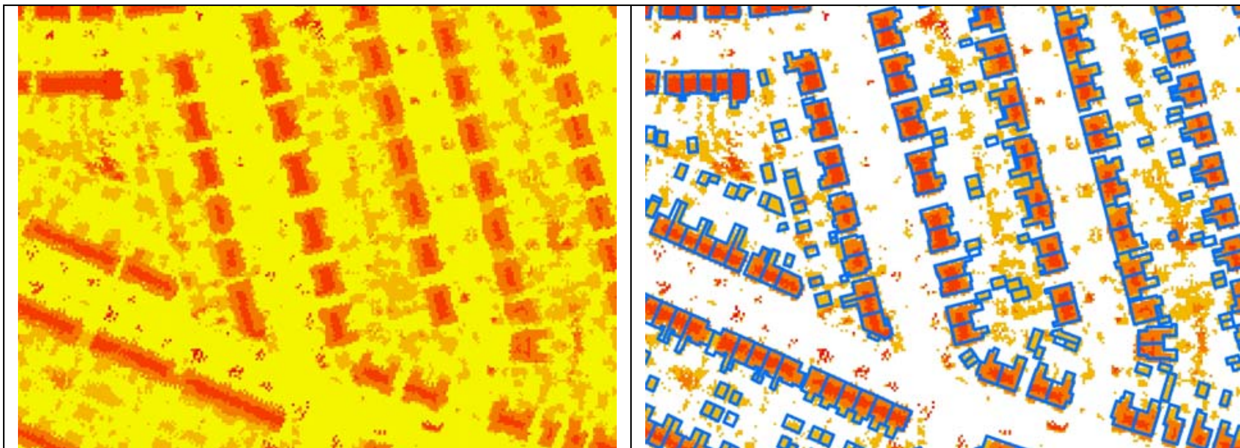
This lack of property level structural attribute data may be resolvable, however. The past decade has seen the continuing development of geographic information handling technologies and the emergence of rich new sources of both GIS and remote sensing data. Commentators such as Longley (2002; 2003) have begun to examine how these developments are improving not only the digital data infrastructure of many cities but also the opportunities for undertaking urban modelling and urban analysis. The research agenda in urban GIS is now increasingly dominated by the development of integrated databases, methods for their analysis and models relevant to fine-scale urban geographies. Traditional sources of ancillary information such as the Census of Population and address and population registers are becoming enhanced by their integration with these data sets. The precision of these data and the frequency with which they are updated makes them increasingly useful as a source of micro-scale structure of the built environment in the absence of more detailed data (e.g. Longley and Mesev, 2000). This paper explores the opportunities provided by some of these new data sources for deriving property level structural attribute data. In particular, the paper presents research on methods for deriving modelled estimates of living space for individual properties within a small area of Cardiff, Wales and analyses the success of these estimates for a sample of properties. The paper outlines how these estimates can be used in further research to improve housing market analysis.

2. Integrating digital infrastructure data with LiDAR data

Previous research by Orford and Radcliffe (2007) demonstrated how property level structural attribute data, in this case dwelling type, can be derived from Ordnance Survey's (OS) digital framework data product, OS Mastermap, It also highlighted errors in the modelling process that could be rectified by the integration of remotely sensed data relating to building height (Rottensteiner, 2003). This research on dwelling type has since been developed in order to provide a methodology for deriving modelled estimates of residential living space. Four principal data sources are used:

- OS Mastermap for Cardiff: topographic and address layers
- Environment Agency (EA) LiDAR data
- Estate agent data for a sample of properties in the study area
- 2001 Census data for Output Areas

In order to estimate living space within a particular property, it is necessary to have data on the building outline and the number of storeys in order to calculate an approximate measure of floor area. It is also important to capture any internal variation in building height across the building outline in order to estimate living space as accurately as possible. For instance, if a single storey extension on a property is treated as a two storey structure, then living space will be over-estimated. The research covers a small area in Cardiff that has been selected based on the findings in Orford and Radcliffe (2007). The OS Mastermap data provides the individual addresses of all the properties in this area, classified into residential and non-residential types, and also provides the building outlines for these addresses. The EA LiDAR data provides height data at a one metre resolution for the area– see Figure 1 which shows LiDAR and OS Mastermap data part of the study area.



Lidar data revealing building heights	Building outlines and height greater than 0.5 metres.

Figure 1. EA LiDAR and OS Mastermap data for part of the study area

In order to capture internal variations in building height within individual properties, the OS Mastermap building outline was sub-divided into rectilinear polygons using a custom written GIS script. Building heights were then calculated for each polygon within a building outline using a weighted average of the LiDAR data and then this was converted into a measure of the number of storeys – see Figure 2. In some cases, the LiDAR data also identified extensions that had not been captured by the OS Mastermap building outlines. In such cases the building outlines were re-drawn to include these extra structures.

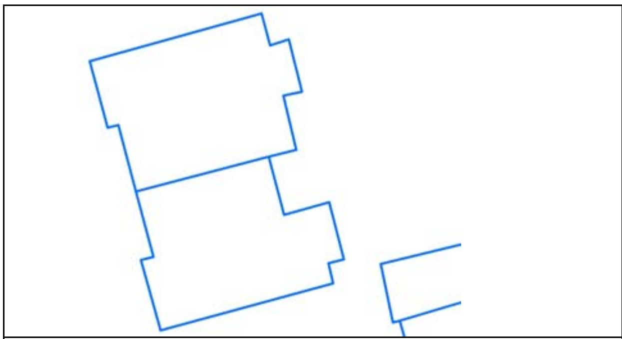
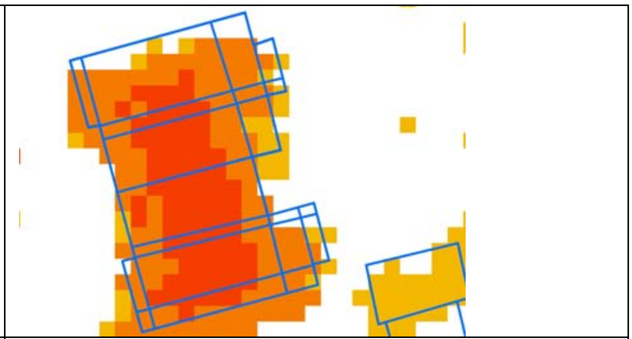
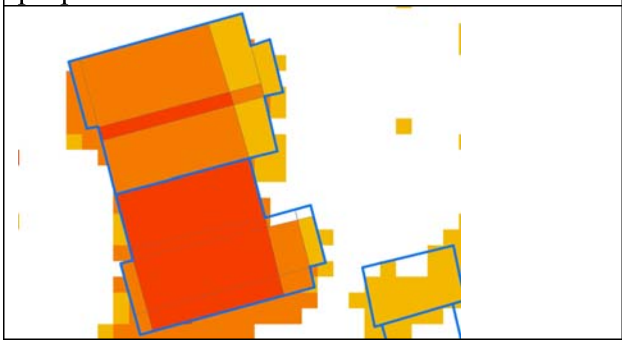
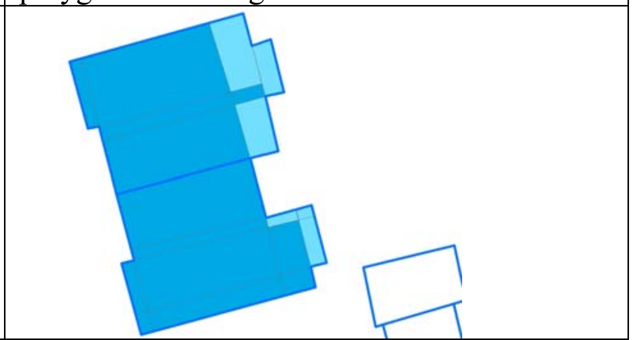
	
Building outlines showing two semi-detached properties	Building outlines subdivided by rectilinear polygons with height data
	
Estimated height data for internal building outline polygons	Estimated storey for each internal building outline polygon showing rear elevation single storey extensions

Figure 2. Calculating building height and number of stories for internal polygons within building outlines

3. Deriving estimates of residential living space

In order to derive estimates of living space based on building outlines and height data, Estate Agent data for a sample of properties in the study area was integrated into the GIS. These data included information on dwelling type, number of storeys and the total floor area of rooms. A regression model was estimated for the Estate Agent sample of properties which specified living space (as measured by floor area) as a function of structural attributes and a measure of location (see Equation 1):

$$\mathbf{LS = f(DT, NS, ABO, LN)} \tag{1}$$

LS = Living space as measured by floor area (Estate Agent)
DT = Dwelling type (Estate Agent)
NS = Number of stories (Estate Agent)
ABO = Area of building outline (OS Mastermap)
LN = Local neighbourhood (OS Mastermap/Census)

The model was then used to estimate living space (LS) for every single property in the study area using the measures of DT, NS, ABO and LN derived from OS Mastermap and EA LiDAR data. These modelled living space estimates were validated against floor area calculations from a sample of Estate Agent properties which had not been included in the calibration of the regression model. By categorising the types of properties and locations in which modelled living space estimates were similar to those calculated using Estate Agent floor area data and those in which there were substantial differences, the utility of the methodology in estimating living space in different urban contexts is explored. These include problems with the digital framework data such as the building outlines being incorrect (e.g. missing extensions); EA LiDAR data not being sensitive enough to capture the internal variation in building heights within a property and both data sources unable to capture particular types of internal spaces such as attics, basements, cellars, integral garages and certain layouts that have an important effect upon living space. There are also issues with the accuracy of Estate Agent data and the measures of floor area and living space. However, the research suggests that these problems have particular spatial configurations within the built environment and the paper provides some means of resolving them.

4. Conclusion

In this paper we presented a method for deriving living space estimates for a sample of individual properties in Cardiff. By doing so, we have moved towards developing a data-rich environment for modelling micro-scale structure in the built environment. The aim of future research is to scale up the study area to the whole of Cardiff. With the integration of sales price data from the Land Registry, the GIS will provide a laboratory for the development and testing of housing market theories through models and computer simulations of the Cardiff housing market.

References

Batty M (1995) Planning support systems and the new logic of computation, *Regional Development Dialogue* **16(1)** pp1-17

Longley PA (2003) Geographic Information Systems: developments in socio-economic data infrastructures. *Progress in Human Geography* **27** pp114-121

Longley PA (2002) Geographical Information Systems: will developments in urban remote sensing and GIS lead to 'better' urban geography? *Progress in Human Geography* **26** pp231–239

Longley PA & Mesev V (2000) On the measurement and generalisation of urban form, *Environment and Planning A* **32(3)** pp473-488

Orford S & Radcliffe J (2007) Modelling UK residential dwelling types using OS Mastermap data: a comparison with the 2001 census, *Computers, Environment and Urban Systems* **31(2)** pp206-227

Orford S (2000) Modelling spatial structures in local housing market dynamics: A multi-level perspecitive. *Urban Studies* **13(9)** pp1643-1671

Rottensteiner, F. (2003) Automatic generation of high-quality building models from Lidar data, *IEEE Computer Graphics and Applications* **23(6)** pp 42-50

VOA (2002) *CTR(E) IA 111202 Data quantity / quality audit – property attributes data*, at http://www.voa.gov.uk/publications/CouncilTaxIas/021211-ctre-ia.htm

Acknowledgements

This research has been supported by the Royal Geographical Society (with The Institute of British Geographers) with an EPSRC research grant. The authors would like to thank the Ordnance Survey and the Environment Agency for providing OS Mastermap and LiDAR data under license.

Biography

Dr Scott Orford is a lecture in GIS and Spatial Analysis in the School of City and Regional Planning, Cardiff University. His research interests include housing market analysis and modelling socio-economic processes.

Mr Jonathan Radcliffe obtained an MSc GIS with distinction from University of Glamorgan. He is currently a research associate for the Wales Rural Observatory at Cardiff. He has recently started a part time PhD investigating the use of visualization tools in planning and policy in Wales.

Spatiotemporal interaction modelling and simulation for transport system based on a geosensor network

Seong K. Choi[1], Pragya Agarwal[1]

[1] Department of Civil, Environmental and Geomatic Engineering, University College London, London WC1E 6BT
Tel. (+44) 20 7679 2730 Fax (+44) 20 7380 0453
{s.choi, pagarwal}@ge.ucl.ac.uk

KEYWORDS: Interaction, Geosensor network, Simulation, Transport system

1. Introduction

The transport system is a crucial infrastructure in the urban area characterized by its spatiotemporal phenomena. With the advent of Information and Communication Technology (ICT), the transport system has been evolving towards an Intelligent Transport System (ITS) based on the communication components providing traffic information to moving vehicles in real time. The problem is that current GIS data models are limited in handling moving objects and the events between them.
In the transportation field, dynamic segmentation and link-node topology have been used to manage transport infrastructures and to provide navigation functions; however a suitable model is needed to represent mobile objects (Goodchild 2000). The goal of this research is to develop a reliable conceptual framework for the GIS data model to support ITS and its communication component. In particular, the emphasis is set on communication between vehicles and road facilities under the assumption that they can interact with each other as a node of a geosensor network.

2. Literature review

Recently, considerable research efforts have been aimed at the representation of dynamic geographical phenomena as well as interaction with each geo-object in the spatiotemporal phenomena in real time. All of them provide valuable techniques to extend traditional GIS data models for supporting spatio-temporal phenomena in diverse application domains.
Several researchers have tried to represent temporal variation in the geographic domain. Hornsby and Egenhofer (2000) proposed the identity-based model which can model changes of objects based on their own identity. The concept of geospatial lifelines was also proposed to present moving objects as ordered points in space and time (Hornsby & Egenhofer 2002). This research provided the multi-granularity model for moving objects using approximations of lifelines and it may be useful for processing in very large datasets. Recently, Goodchild et al. (2007) proposed a model of dynamic geo-objects through time using three conditions: movement (stationary or moving), shape (elastic or rigid), and internal structure (uniform, evolving).
Ontological approaches have been developed for geographical objects, events and processes as well as dynamic representations for moving objects (Worboys 2005). Gernon and Smith (2004) presented a framework for geographical representation of objects, events, and processes using SNAP-SPAN ontology. Moreover, ontological views have been used to examine the semantic relationship between different geographical categories (Agarwal 2005).
Geosensor network analysis has been used to support dynamic phenomena with monitoring and interaction based on a Mobile Ad-hoc Network (MANET) in real time (Nittel et al. 2004).

This research showed that interaction with moving geo-objects as a sensor node can contribute to reduce their redundancies throughout information dissemination related to their locations and settings using geosensor networks.. In order to support ad-hoc shared-ride trip planning using geosensor networks, simulation was performed based on grid environment (Winter & Nittel 2006). They showed that mid-range negotiation is more effective and efficient than unconstrained negotiation and short-range negotiation for communication among sensors. Raubal et al. (2007) considerably reduced communication costs for ad-hoc shared-ride trip planning with different way-finding strategies using time geography in a real street network environment. .

3. A scenario for interaction modelling and simulation

This research extends on the other current research in geo-sensor networks by proposing a model for an ubiquitous geospatial environment in which different nodes can communicate with each other. The scenario employed for developing a model is an intelligent transport infrastructure with dynamics among the geo-objects focused on a data model and an interaction model (Figure 1) in which vehicles and road facilities collaborate with each other within a geosensor network. .

A driver is going to Manchester to attend a conference, so he inputs his destination into an intelligent vehicle. The intelligent vehicle will find a suitable route to the destination and interact with other objects along the route. When the vehicle joins the highway, it becomes a member of a stream of cars. It would then interact with remaining cars in the stream to maintain vehicles speeds and intervals. In addition, the first car of the stream will keep communicating with road facilities and signal controllers to recognize various and dynamic road situations such as road pavement, and car accidents.

Now, twenty cars are leaving the stream on the highway to go to the conference site in Edinburgh. These cars are a member of smaller stream going to the same destination. The first car communicates with the signal controller to inform that cars are member of the stream, so that the signal controller deals with cars like a long lorry. This means that the controller will not change the signal until all the cars have passed. Finally, they arrive at their destination.

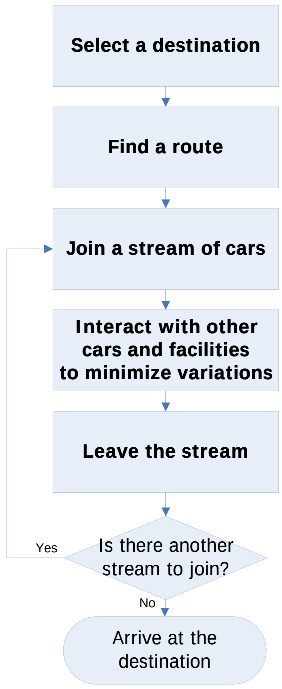


Figure 1. A scenario within which vehicles and road facilities collaborate and communicate with each other within a ubiquitous information space

Based on the above scenario which was made under the assumption that vehicles and road facilities are nodes of a geosensor network, methods such as conceptual modelling and ontologies will be used to model categories and communication networks in a real-time ubiquitous information infrastructure in the transportation domain. Also, agent-based simulations will be used to model specific parts of communication, and test categories and their interactions.

4. Spatio-temporal interaction modelling

The basic underlying theoretical idea of the model is that interaction among geo-objects in real time can have positive influences on transport system itself. Therefore, we divided the setting of the scenario into three parts in the view of service consumer, service provider, and

other environmental parameters. And then, we tried to find relations between these categories based on determined variables, predictable variables, and unpredictable variables.
First, objects and events of this setting were categorized into infrastructure, traffic flow, and others. Traffic flow category can be characterized as moving objects and its events, while most of infrastructure category was populated by services based on their location. In addition, there is another environmental category which affects infrastructure and traffic flow.
Second, variables of spatiotemporal setting were divided into three parts – determined variables, predictable variables, and unpredictable variables (Figure 2). Determined variables relate to known information before the journey such as origin, destination, pavement status, planned events and maintenances. Predictable variables describe variables which can be anticipated like the weather condition or traffic signal, but it can change. Unpredictable variables can not be known before its occurrence such as car accidents, signal failure, and unexpected road closure.

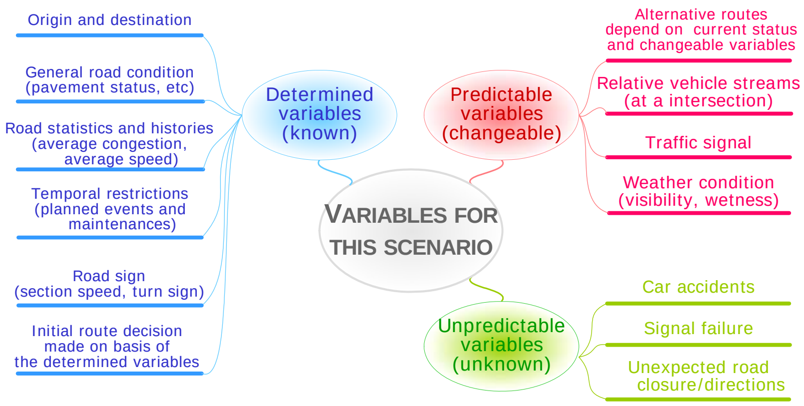


Figure 2. Three kinds of variables for this scenario (determined variables, predictable variables, and unpredictable variables)

Third, each variable is matched to an appropriate semantic category. The transport infrastructure as well as the traffic flow contains determined, predictable, and unpredictable variables (Figure 3). Basically, most variables are related to the transport infrastructure and moving objects (the traffic flow) on the infrastructure. However, there are other variables such as weather condition and emergencies which are unpredictable, and they have an effect on this setting directly and indirectly.

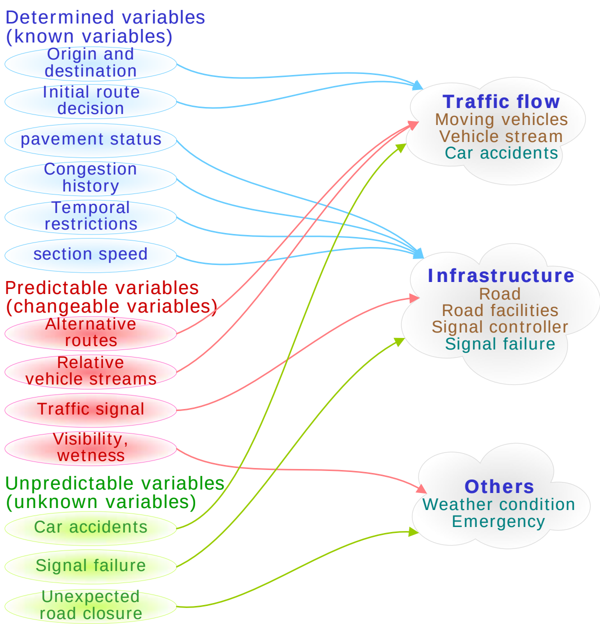


Figure 3. Relations between three kinds of variables and categories

Lastly, we extracted six types of interaction among categories because these variables belonging to each category can affect other variables based on their relations (Figure 4). Although through analysis it was deduced that there are nine types of two-way communication, this research focused on six interactions which have a direct influence on infrastructure and traffic flow.. The infrastructure and the traffic flow can have four-ways effects (infrastructure to infrastructure, traffic flow to traffic flow, infrastructure to traffic flow, traffic flow to infrastructure). Other environmental variables also have an influence on the infrastructure and the traffic flow (two types: other variables to infrastructure, other variables to traffic flow). For example, some precipitation may cause a wet road condition so that vehicles reduce their speed. Also, it can evaluate congestion and intelligent signal controller may try to find a suitable signal pattern for the congestion.

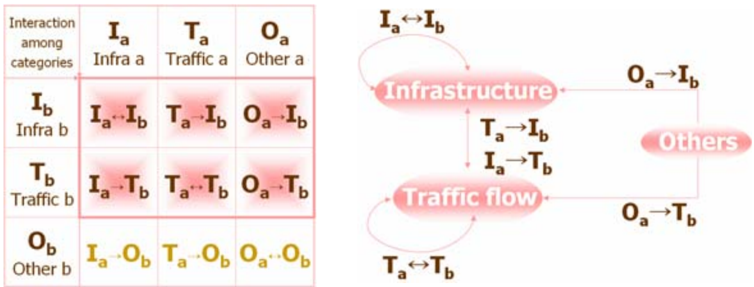


Figure 4. Direct interaction with infrastructure and traffic flow among sensor networks
for transport infrastructure

5. Development of a simulation environment

Further to development of the theoretical model and ontological deployment, we implemented a simulation environment relating to six types of interaction which can reduce various temporal variations during the journey on the road network. The simulation focused on Traffic flow-to-Infrastructure interactions ($T_a \rightarrow I_b$, $I_a \rightarrow T_b$ of figure 4) between vehicles and traffic signal controllers. If signal controllers can make signal schedule considering road status about the traffic flow in real time, traffic efficiency will be improved. The simulation environment was developed with a vector-based grid network to consider scalability towards real street networks.

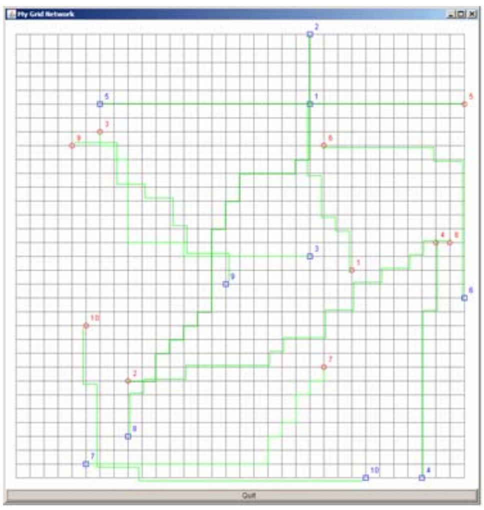


Figure 5. Simulation environment (grid network)

We conducted three types of simulation with different settings relating to signal controllers to measure average travel time from departure to destination: travel without signal, travel with signal; and interactive travel with signal. When we consider the signal, we have to calculate delay time at signals so that three kinds of time domain are needed: absolute time (STimer class), accumulated travel time of each vehicle (SCar class) and circular time for signal controller (SSignal class). If a vehicle arrived at an intersection as soon as the signal turned red, the delay of the vehicle at the intersection is maximized. Signal schedule can be made flexible to minimize the delay throughout the interaction between vehicles and signal controllers. This is the purpose of the simulation to examine whether or not interactions minimize the delay at intersections. For the multi-agent based simulation, several agent classes are created (sample shown in Figure 6). There are many dependency relations among classes so that it can be argued that the simulation for interaction is well directed.

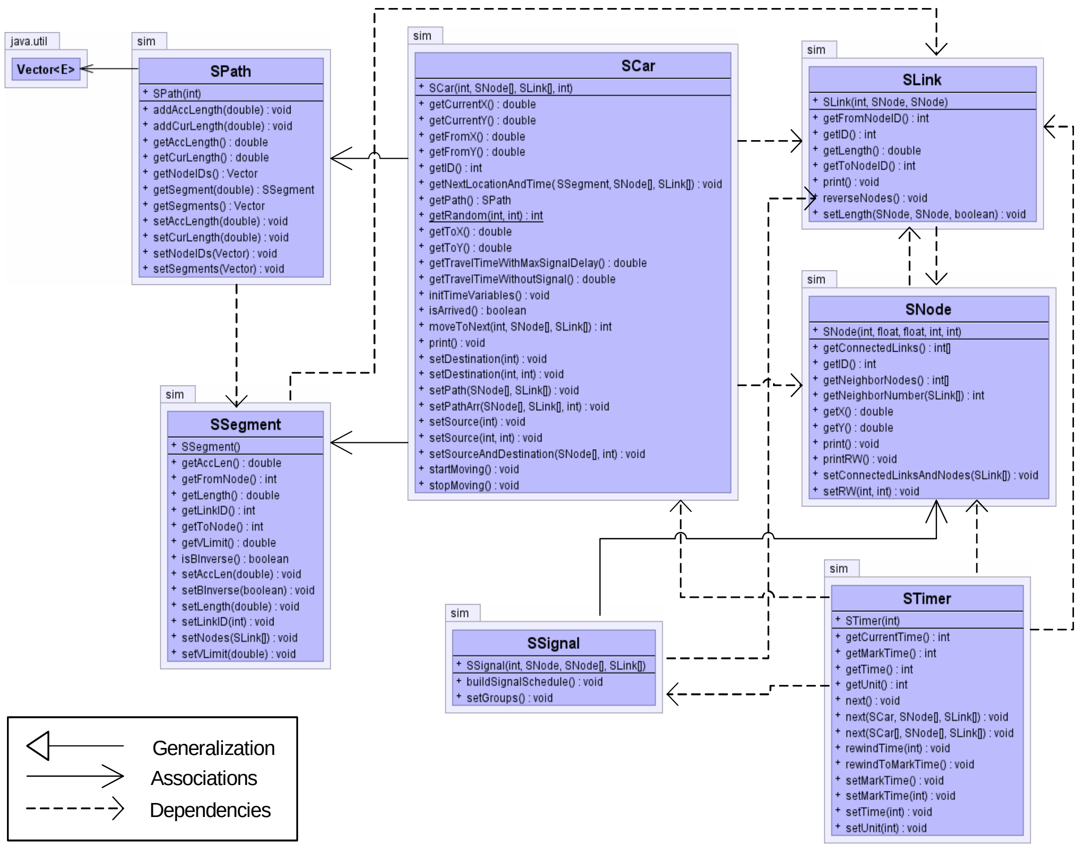


Figure 6. Static class diagram for the simulation

6. Conclusion and future work

This research has been successful in categorizing and modelling interactions between vehicles and road facilities in a road network based on geosensor network and in developing a simulation environment to conduct an experiment for interactions to improve transport system. This research has laid the foundations for future work to examine the efficiency of the simulation of interactions. In order to execute the simulation, some parameters such as number of cars, signal time schedule, length of each segment have to be further examined and defined. In this research, the simulation focused on Traffic flow-to-Infrastructure interactions ($T_a \rightarrow I_b$, $I_a \rightarrow T_b$ of figure 4) between vehicles and traffic signal controllers. However, because

there are four other interactions, the simulation environment will be expanded for more experiments.
This research is aimed at modelling spatio-temporal interactions in real time, and will support research in making geo-objects more interactive and intelligent. Other scenarios besides the one discussed in this paper will be considered in future work, where interaction model may be more useful than representation model. In order to implement GIS functions in a geo-sensor network and ubiquitous information space, ontological methodologies will be useful. For future work, this research will aim for an ontological framework of transport system to support richer simulation and implementation.

References

Agarwal, P. 2005, "Ontological considerations in GIScience", *International Journal of Geographical Information Science*, vol. 19, no. 5, pp. 501-536.

Grenon, P. and Smith, B., 2004, SNAP and SPAN: towards dynamic spatial ontology. *Spatial Cognition and Computation*, 4, pp. 69–103.

Goodchild, M. F. 1997, "What is Geographic Information Science?", *NCGIA Core Curriculum in GIScience*, http://www.ncgia.ucsb.edu/giscc/units/u002/u002.html, posted October 7, 1997.

Goodchild, M. F. 2000, "GIS and Transportation: Status and Challenges", *GeoInformatica*, vol. 4, no. 2, pp. 127-139.

Goodchild, M. F., Yuan, M., & Cova, T. J. 2007, "Towards a general theory of geographic representation in GIS", *International Journal of Geographical Information Science*, vol. 21, no. 3, pp. 239-260.

Hornsby, K. & Egenhofer, M. J. 2002, "Modeling moving objects over multiple granularities", *Annals of Mathematics and Artificial Intelligence*, vol. 36, no. 1-2, pp. 177-194.

Hornsby, K. & Egenhofer, M. J. 2000, "Identity-based change: a foundation for spatio-temporal knowledge representation", *International Journal of Geographical Information Science*, vol. 14, no. 3, pp. 207-224.

Nittel, S., Duckham, M., & Kulik, L. 2004, "Information dissemination in mobile ad-hoc geosensor networks", *Geographic Information Science, Proceedings*, vol. 3234, pp. 206-222.

Nittel, S., Stefanidis, A., Cruz, I., Egenhofer, M., Goldin, D., Howard, A., Labrinidis, A., Madden, S., Voisard, A., & Worboys, M. 2004, *Report from the First Workshop on Geo Sensor Networks*, *SIGMOD record, Special Issue on "Sensor Network Technology Infrastructure, Security, Data processing, and Deployment".*

Raubal, M. et al. 2007, Time geography for ad-hoc shared-ride trip planning in mobile geosensor networks. ISPRS Journal of Photogrammetry & Remote Sensing, doi:10.1016/j.isprsjprs.2007.03.005

Winter, S. & Nittel, S. 2006, "Ad hoc shared-ride trip planning by mobile geosensor networks", *International Journal of Geographical Information Science*, vol. 20, no. 8, pp. 899-916.

Worboys, M. 2005, "Event-oriented approaches to geographic phenomena", *International Journal of Geographical Information Science*, vol. 19, no. 1, pp. 1-28.

Biography

Seong K. Choi is a first year doctoral student in the department of Civil, Environmental and Geomatic Engineering (UCL). He holds a BSc and an MSc in Geographic information Science from the department of Geoinformatic Engineering (Inha University in South Korea). His PhD research focuses on the interaction modelling and simulation of transport communication in a geo-sensor network.

Pragya Agarwal is a Lecturer in GIScience and Director of MSc (GIS) in the department of Civil, Environmental and Geomatic Engineering (UCL). Dr Agarwal's primary research interests are in ontologies, semantic inter-operability, ubiquitous geospatial environments and personalisation.

Origin-specific models for analysing commuting flows in Northern Ireland: scale effects and other problems

C. D. Lloyd, I. G. Shuttleworth, G. Catney

School of Geography, Archaeology and Palaeoecology, Queen's University, Belfast, BT7 1NN
Tel. +44 (0)28 9097 3478; Fax +44 (0)28 9097 3212
Email: c.lloyd@qub.ac.uk

KEYWORDS: commuting, spatial interaction, origin-specific models, Poisson regression

1. Introduction

Understanding commuting flows and their characteristics has obvious economic significance and, as such, has been the subject of much academic research. A large proportion of existing research is based on official data sources such as censuses of population. This analysis entails the application of a spatial interaction modelling framework to explore the size of commuting flows between different wards (administrative zones) in Northern Ireland. Initially, global unconstrained and constrained models and the residuals from these models were explored (Lloyd et al., 2007). In the present paper, an origin-specific model is used to model flows from each individual location. The analysis explores the relationship between commuting flows and a host of explanatory variables including the distance from the origin zone, number of workers by ward of workplace (destination), the absolute difference between the percentage of Catholics at the origin and destination wards, a measure of employment opportunities at the origin and destination wards and finally a measure of destination competition. Selected results are summarised and interpreted after which some problems with this approach are then considered and some ideas for future work outlined.

2. Data and methods

This paper makes use of data provided as outputs from the 2001 Northern Ireland Census of Population. The main focus is on counts of workers moving between wards. There are 582 wards and thus the flow matrix comprises 582 by 582 entries (338724 possible interactions). The small cell adjustment procedure, adopted prior to release of data from the 2001 Census of Population, converted small counts to values of zero or three (see Duke-Williams and Stillwell, 2007 for an assessment of the effects of small cell adjustment on interaction data). The sparsity of the matrix of flows is apparent from the fact that some 87% of the flows have a value of zero. Some 6.4% of flows have a value of three, and this is a function of small cell adjustment. The maximum flow value is 1727. The explanatory variables are listed and described in Table 1. The road distances between Output Areas (OAs; nested within wards) were provided by the Northern Ireland Statistics and Research Agency (NISRA) and the averages of these distances between wards were computed for the purposes of spatial interaction modelling.

Variable name	Description
DistKM	Road distance between wards in KM
WrkTotal	Number of workers by ward of workplace
CathDiff	Absolute diff. between origin Catholic % and destination Catholic %
EmpScoMult	Origin and destination employment access scores multiplied together
Centrality	Destination competition measure

Table 1. Variable names and definitions. Diff. is difference.

The division of the population of Northern Ireland into two groups, Protestants and Catholics, and the study of the degree to which members of the two groups live and work together has received much attention in academic research (see, for example, Smith and Chambers, 1991). The absolute difference between the origin ward and destination ward Catholic percentages is included to enable exploration of the degree to which commuting flows are segregated. The employment accessibility score is defined as the level of job opportunities at location j estimated as the number of jobs in a given area down-weighted in proportion to the distance of the area from location j (Coombes and Raybould, 2000). Centrality is a measure of destination competition as defined by Fotheringham and O'Kelly (1989).

Aggregate interactions between places are often explored in a spatial interaction modelling framework (Haynes and Fotheringham, 1984; Fotheringham and O'Kelly, 1989). This paper makes use of Poisson regression (see Flowerdew and Aitkin, 1982, for an introduction) to model flows from each origin ward to all destinations; this is termed an origin-specific model (Fotheringham and O'Kelly, 1989). For a given ward, the flow to all wards from that ward is the dependent variable while distances from that ward to all other wards and the other variables listed in Table 1 are the independent variables.

3. Results

The minimum and maximum origin-specific coefficients (for the 582 sets of values) are summarised in Table 2. The signs of the coefficients are as expected. That is, all of the distance coefficients are negative, indicating that an increase in distance corresponds to a reduction in the size of flows. The coefficient for the number of workers at the destination ward is positive – the number of workers (the number of filled jobs) is, not surprisingly, an attraction. The coefficients for the absolute difference between the origin ward Catholic percentage and the equivalent figure for the destination ward range from large negative to large positive values. In words, for some origins if the religious composition of a destination ward is different to that of the origin ward then that destination is likely to attract less workers than destination wards with more similar compositions to the origin ward. The employment score multiplier coefficients also vary – there are large positive values in Belfast and neighbouring areas. This is probably because there are more extensive employment opportunities in Belfast than in other (particularly rural) areas (if there are many large flows between an origin and destinations with similarly large employment opportunities then this would result in large positive values of this coefficient). The coefficients for centrality have negative values in some locations and positive values in others.

Variable name	Minimum	Maximum
DistKM	-4.466	-0.406
WrkTotal	0.559	2.061
CathDiff	-8.000	4.584
EmpScoMult	-4.829	7.253
Centrality	-0.333	0.605

Table 2. Origin-specific model coefficients.

The distance coefficient for the origin-specific model is mapped in Figure 1. Distance decay values tend to be larger in the rural west. This suggests that individuals in (for example) the rural west are deterred from travelling large distances, whereas those in urban areas such as Belfast are deterred to a lesser degree; this is clearly counter-intuitive. The competing destinations model of Fotheringham (Fotheringham and O'Kelly, 1989) has been offered as a potential solution to this problem. However, the model employed here is a competing destinations model – a centrality variable has been included. A variety of characteristics of the flows were explored in an attempt to better understand this problem of counter-intuitive distance coefficients.

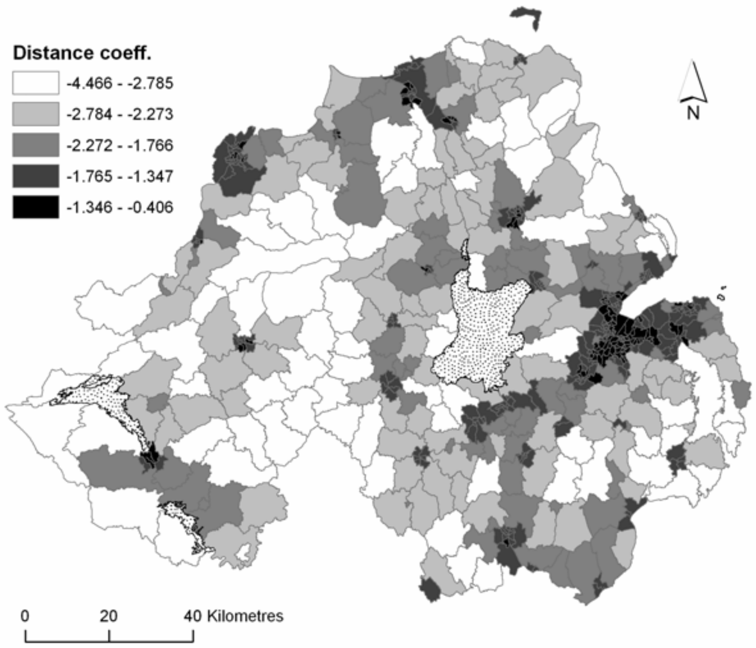


Figure 1. Origin-specific model distance coefficient. Northern Ireland Census of Population data — © Crown Copyright. Reproduced under the terms of the Click-Use Licence.

Figure 2 shows mean distance commuted from each ward against internal flows as a percentage of all flows from a ward. The relationship between the two variables is positive and this is counter-intuitive. The plot points to a key problem with the kinds of data employed in this analysis. In large rural wards often many people live and work in the same ward. The internal distance in this ward (generated using OA centroids, as described above) may be larger than many distances in the neighbourhood of a small urban ward. In effect, using these data, the commuting distance assigned to people living in large rural wards is always larger than the minimum commuting distance in most (if not all) urban areas. The destination choices available for residents of rural areas are likely to be, simply because of large zone sizes in these areas, at larger distances than those choices available to residents of urban areas.

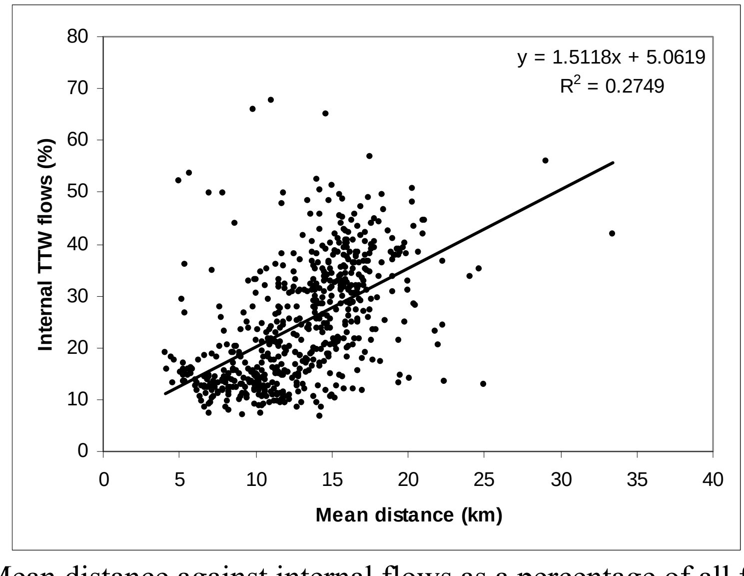


Figure 2. Mean distance against internal flows as a percentage of all flows from a ward.

Figure 3 shows competing destinations origin-specific model distance coefficients against internal flows as a percentage of all flows from a ward. This plot suggests that, on average, residents of areas with a large percentage of internal flows are less likely to travel large distances to work. This seems intuitive; with the plot in Figure 2, this indicates that the size of zones (not surprisingly) has a major impact on results.

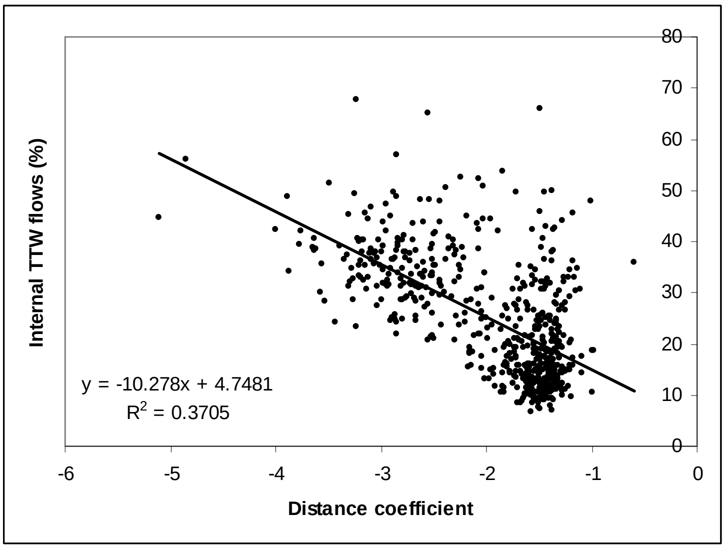


Figure 3. Distance coefficient against internal flows as a percentage of all flows from a ward.

4. Conclusions and future work

This analysis raises a number of issues which must be addressed to make best use of the available data for understanding commuting flows in Northern Ireland. Working with large areas is unavoidable in sparsely-populated areas – counts over small areas may not be released for confidentiality reasons. Various attempts have been made to make use of information on intra-zone flows to help overcome this problem. The use of geographically-weighted approaches (see Nakaya, 2001) may also offer benefits. Such approaches will be explored. In addition, this paper makes use of only origin-specific models. A logical extension of the work would be to make use of destination-specific models to allow exploration of propulsiveness of origin wards, rather than attractiveness of destination wards as examined here.

5. Acknowledgements

The Northern Ireland Statistics and Research Agency (NISRA) are thanked for the provision of 2001 Northern Ireland Census of Population data. The ward-to-ward road distances were calculated using data provided by NISRA.

References

Coombes M and Raybould S (2000) Policy-relevant surfaced data on population distribution and characteristics *Transactions in GIS* **4** pp319–342.

Duke-Williams O and Stillwell J (2007) Investigating the potential effects of small cell adjustment on interaction data from the 2001 Census *Environment and Planning A* **39** pp1079–1100.

Flowerdew R and Aitkin M (1982) A method of fitting the gravity model based on the Poisson distribution *Journal of Regional Science* **22** pp191–202.

Fotheringham AS and O'Kelly ME (1989) *Spatial Interaction Models: Formulations and Applications*. Kluwer Academic Publishers, Dordrecht.
Haynes KE and Fotheringham AS (1984) *Gravity and Spatial Interaction Models*. Sage Publications, Beverly Hills.
Lloyd CD, Shuttleworth IG and Catney G (2007) Commuting in Northern Ireland: Exploring spatial variations through spatial interaction modelling. In Winstanley AC (Ed.) *GISRUK 2007. Proceedings of the Geographical Information Science UK Conference*. National Centre for GeoComputation, National University of Ireland Maynooth, 11th–13th April 2007. Maynooth: National University of Ireland Maynooth, pp258–263.
Nakaya T (2001) Local spatial interaction modelling based on the geographically weighted regression approach *GeoJournal* **53** pp347–358.
Smith D and Chambers G (1991) *Inequality in Northern Ireland*. Clarendon Press, Oxford.

Biography

Chris Lloyd is Lecturer in GIS at Queen's University, Belfast. His research interests include spatial data analysis, the analysis of Census data, and historical geography.

Ian Shuttleworth is Lecturer in Human Geography at Queen's University, Belfast. His research interests include Census analysis, labour markets and labour mobility, and the political economy of transnational migration.

Gemma Catney is a PhD research student in Geography at Queen's University, Belfast. Her research interests are in migration, residential segregation and population change, and in methodologies for contemporary empirical research in population and social geography.

The Accuracy of Gridded Digital Elevation Models Interpolated to Higher Resolution

Stephen Wise, Ed Hanna, Tristram Irvine-Fynn

Department of Geography, University of Sheffield, Sheffield, S10 2TN
Tel. (+44) 0114 222 7940 Fax (+44) 0114 279 7912
s.wise@sheffield.ac.uk, www.shef.ac.uk/geography

KEYWORDS: DEM, RMSE, artefacts, error, resolution

1. Introduction

The issue of the accuracy of Digital Elevation Models has received a good deal of attention in recent years as made clear by the review of Fisher and Tate (2006). The present work was motivated by the need to interpolate a coarse resolution global DEM to a finer resolution for the purposes of climatic modelling of the Greenland ice sheet (Hanna et al 2007). The question of how best to interpolate an exisiting DEM to a higher resolution has received only limited attention (Rees 2000, Kidner 2003) but the issue is one of general interest because

- It is commonly required in global modelling where some of the input datasets may only be available at relatively coarse scales.
- When interpolating from point or line data sources the spatial resolution of the input data is commonly less than that of the final DEM. For example, the British Ordnance Survey produce a 10m DEM from contours at 5m vertical interval. In areas where the slope of the land is less than 14 degrees, the contours are at least 20m apart i.e. twice the DEM resolution.

Both Rees (2000) and Kidner (2003) used the same methodology to explore this issue. An initial DEM is resampled to produce a grid of points at a lower resolution. These points are then used as the input to produce a DEM at the original resolution. For the points which are dropped in the resampling, a comprison of the original value and the interpolated value gives an estimate of interpolation error at that point. These can be used to estimate the Root Mean Square Error of elevatoin ($RMSE_{elev}$), the standard measure of DEM error.

$$RMSE_{elev} = \sqrt{\frac{\sum\left(z_p - z_o\right)^2}{n}} \qquad (1)$$

where Z_p = predicted elevation
Z_o = observed elevation

However, since there are a large number of such points, and they are distributed in space, this opens up the possibility of exploring a number of general issues relating to interpolation error and to the use of $RMSE_{elev}$ as a measure of error. This paper presents the results of some of the early work on this. Three questions are addressed briefly

1. Rees (2000) suggested that there was little benefit in using non-linear interpolation for interpolating to higher resolutions but Kidner (2003) demonstrated that non-linear methods produced lower RSMEelev values. However both authors only considered the case of doubling the DEM resolution, and only considered RMSEelev – here consideration is given to a wider range of resampliigs and to other measures of

quality.

2. $RMSE_{elev}$ has been criticised as a measure of error on a number of grounds (Wise 1998, 2000). One of these is that since small errors in elevation may lead to large errors in surface derivatives such as slope and aspect it may be a poor predictor of these errors. This question is explored here.
3. Since the number of points from which to estimate RMSE directly is usually small it is not possible to consider the spatial pattern of error or the possible link between error and the nature of the terrain surface, such as terrain roughness. Some initial observations on both these points are made here.

2. Study Area and Methodology

In order to consider the possible effect of terrain roughness an area was selected in the Cairngorm Mountains of Scotland (Figure 1) which covered both the rugged mountain terrain and the low lying areas near the coast. A subset of the Ordnance Survey PANORAMA 50m DEM was selected with 1024 points in both X and Y. This number of points allows reduction of the resolution in both X andY by factors of 2,4,8,16 and 32. In all cases the initial corner points are retained as data points after resampling so that interpolation back to the original resolution never involves extrapolation outside the area of the input data points.

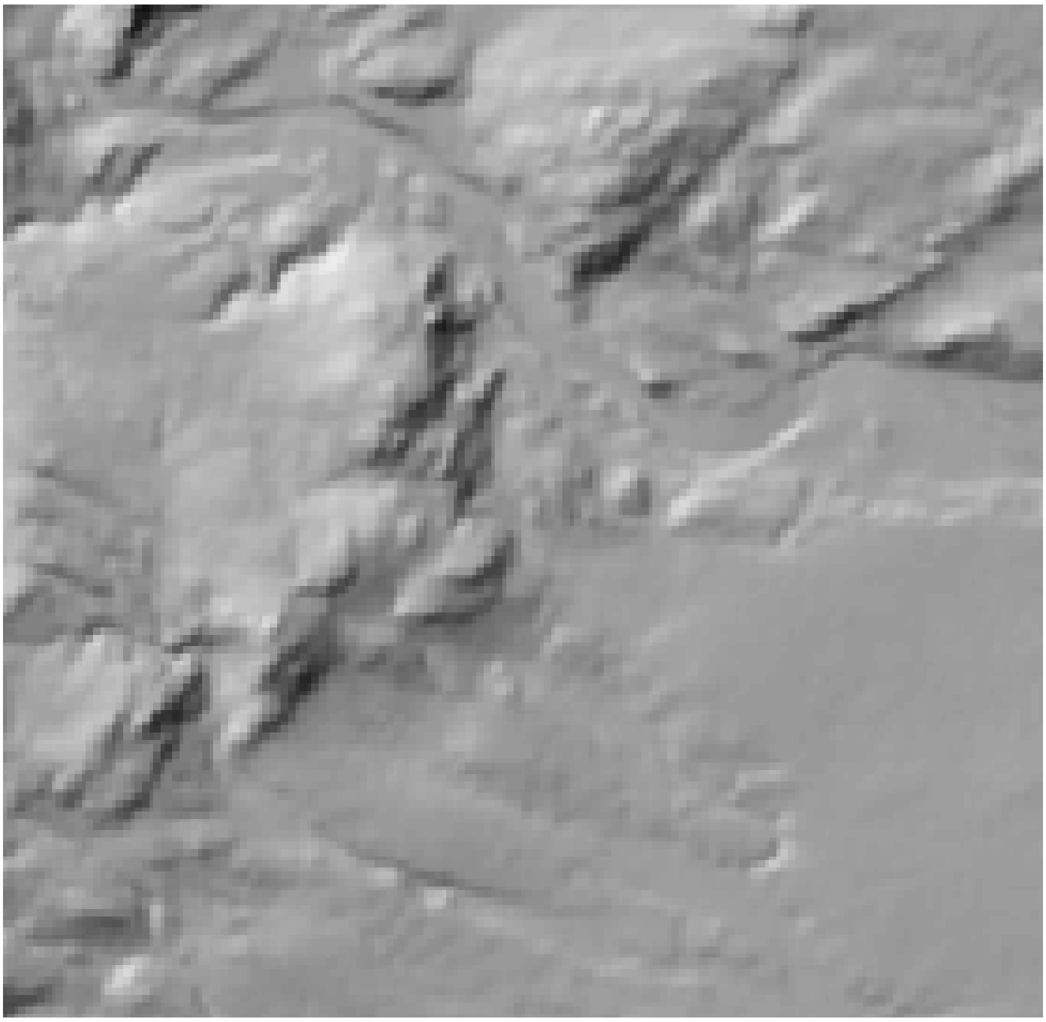


Figure 1. Study area in the Cairngorms of Scotland. The lower left corner of the area has OS coordinates 300000,735000 and the DTM has 1025x1025 pixels at 50m resolution.

Six commonly-used interpolation methods were used in the study – the two letter abbreviations are used in the presentation of some of the results

1. Bilinear interpolation (BL)
2. Inverse Distance Weighting (ID)
3. Radial Basis Function. The Spline option was selected after some initial trials (RS)
4. Spline (SP)
5. Local Polynomial (LP)
6. Topo to Raster. This implements Hutchinson's ANUDEM method and in ARC/INFO was called TopoGrid (TG)

The first four are exact interpolators, which honour the original data points. The last two are approximate interpolators which do not guarantee to honour the original data points. The first method was implemented using a purpose written program – the others were run from ArcGIS 9.2.

3. Results

Table 1 presents summary statistics on RMSEelev for all six methods and the full range of resamplings. For simplicity RMSE is calculated for all data points, including those which were retained after resampling, where for exact interpolators the error will be zero by definition. As Table 1 shows, restricting the caclulations to those points which were not input points has a negligible effect on the results.

Type		BL Exact	ID Exact	LP Approx	RS Exact	SP Exact	TG Approx
32	rmse	51.09	54.16	52.20	52.85	51.94	
	rmse-nd	51.12	54.19	52.23	52.88	51.97	
	mean	0.67	0.65	0.62	0.78	0.74	
	max	342.56	325.96	326.62	377.36	375.65	
16	rmse	28.21	33.21	28.43	26.96	26.61	
	rmse-nd	28.27	33.28	28.49	27.02	26.67	
	mean	-0.25	-0.24	-0.25	-0.31	-0.36	
	max	225.60	244.37	222.45	276.45	259.50	
8	rmse	12.81	17.56	13.10	10.42	10.35	15.17
	rmse-nd	12.92	17.70	13.21	10.50	10.44	15.28
	mean	0.02	0.02	0.02	0.01	0.01	-2.28
	max	135.56	147.84	133.40	164.21	158.14	350.89
4	rmse	5.06	7.89	5.25	3.82	3.77	7.24
	rmse-nd	5.23	8.15	5.43	3.95	3.89	7.37
	mean	-0.01	-0.02	-0.01	-0.01	-0.01	-0.75
	max	82.62	105.52	84.19	69.16	67.42	223.36
2	rmse	1.96	3.11	2.09	1.58	1.59	5.29
	rmse-nd	2.27	3.60	2.41	1.83	1.83	5.38
	mean	0.0	0.0	-0.01	0.0	0.0	-0.41
	max	55.0	67.53	57.51	51.86	50.37	362.19

Table 1. Elevation error statistics for different interpolation methods at different levels of resampling. The values are as follows: rmse – rmse for all points; rmse-nd – rmse for points

not falling on an original data point; mean – mean error; max – maximum value of absolute error. Note: The TOPOGRID algorithm failed to process the 16 and 32-fold resampled sets.

The results do not support the suggestion that non-linear methods always perform better than simple bilinear interpolation in terms of $RMSE_{elev}$. At low levels of resampling two of the spline-based methods (RS and SP) perform best while the third (TG) performs worst. Simple bilinear interpolation consistently produces lower RMSE values than IDW. As the resampling level increases (i.e. as the number of input data points decreases) the difference between the methods reduces.

However visual inspection of hillshade images for the different DEMs shows marked differences between the results (Figure 2). Bilinear interpolation produces very strong linear artefacts, which become more pronounced as the resampling level increases. Spline-based methods on the other hand produce smooth DEMs even when resampling has greatly reduced the number of input data points.

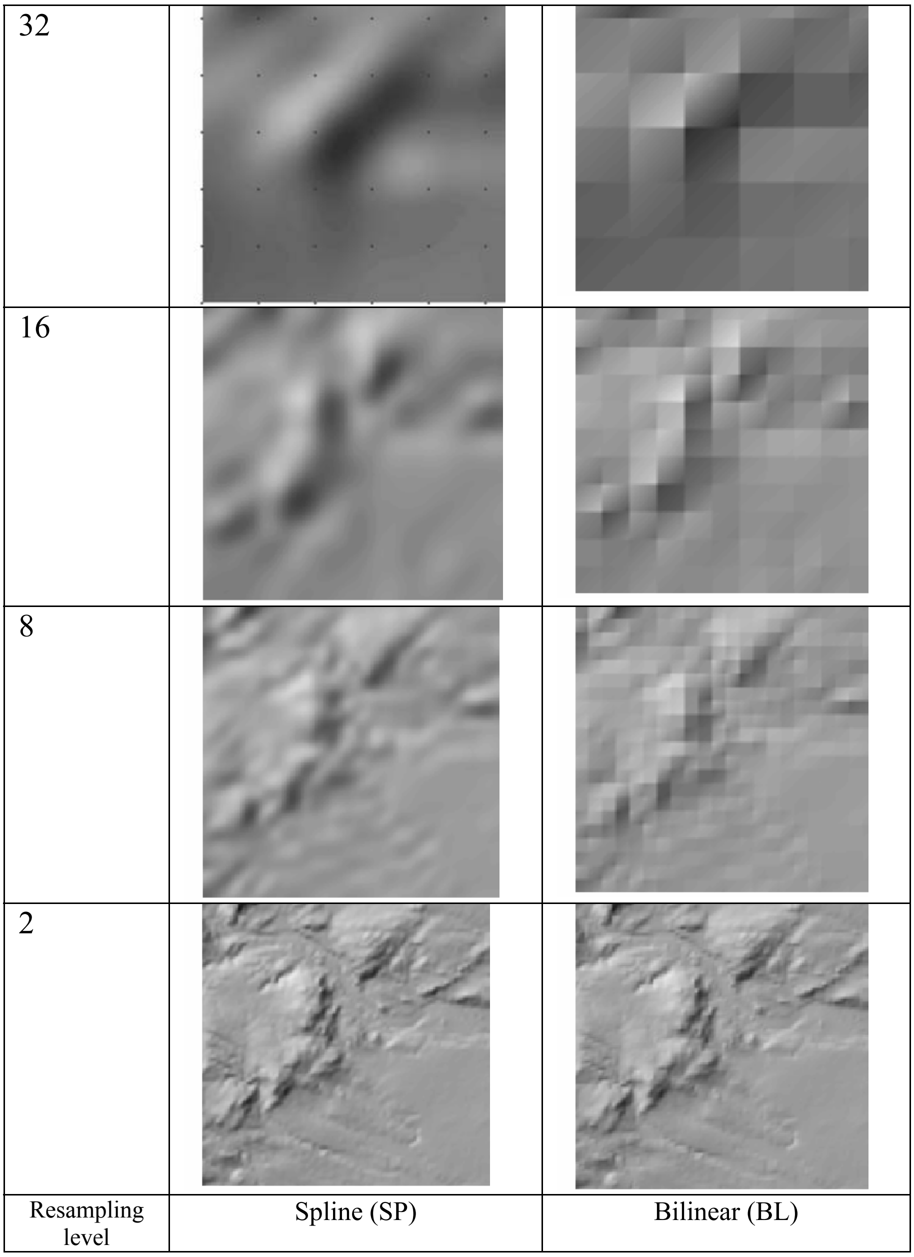


Figure 2. Hillshade images for 2 of the interpolation methods, for selected resampling levels.

By estimating RMSE of elevation and of gradient and aspect for all DEMs (ie. all methods at all resampling levels) it was possible to see how well RMSEelev predicts overall error levels in derived surface values. As Table 2 shows, the relationship is very strong in the case of gradient and aspect but very weak in the case of curvature. This suggests that , despite its shortcoming, $RMSE_{elev}$ may in fact be a reasonable predictor of error in some derived characateristics.

Terrain characteristic	R^2	Best fit line (x = RMSE Elevation)
Gradient	0.974	y = 1.9869Ln(x) - 0.5333
Aspect	0.960	y = 17.418Ln(x) + 3.7808
Curvature	0.211	y = 0.0147Ln(x) + 0.2447

Table 2 Relationship between RMSE of elevation and RMSE of selected terrain characteristics.

There has been a good deal of interest in the spatial characteristics of DEM error (Fisher 1998, Hunter and Goodchild 1997) with numerous authors proposing models of DEM error based on spatial autocorrrelation. Table 3 shows the results of calculating Moran's I on the error values in all DEMs. At low levels of resampling (i.e. when input data volumes are high) the values vary greatly between methods and are often rather low, suggesting that a global autocorrelation function may be a poor model of DEM error in these cases. As the resampling level increases, autocorrelation in the error increases, which is to be expected as most pixels in the final DEM are being interpolated from the same input points as are their neighbours.

	BL	ID	LP	RS	SP	TG
32	.98	.98	.98	.98	.98	
16	.97	.98	.97	.97	.97	
8	.93	.95	.93	.91	.92	.94
4	.78	.84	.80	.72	.73	.88
2	.40	.66	.48	.35	.35	.84

Table 3: Spatial autocorrelation of error in elevation measured using Moran's I.

The results show a consistent relationship between terrain roughness and error, with higher error values in the more rugged upland areas (Figure 2). Table 4 shows the correlation coefficients between Standard Deviation of Slope (a measure of terrain roughness) and RMSEelev. When roughness is calculated from the original, full-resolution DEM (i.e. the results labelled STD1) it can be seen that it is a good predictor of error, even at high resampling levels. In most applications of course, the only data available will be the lower resolution data. When this is used to calculate roughness the results are still quite good at low levels of resampling, but fall away as the resolution of the input data becomes too coarse to estimate roughness reliably.

		BL	ID	LP	RS	SP	TG
32	STD1	0.63	0.53	0.62	0.63	0.64	
	STD32	0.22	0.11	0.26	0.51	0.41	
16	STD1	0.78	0.70	0.78	0.73	0.72	
	STD16	0.58	0.55	0.58	0.39	0.47	
8	STD1	0.91	0.85	0.91	0.85	0.86	0.87
	STD8	0.76	0.81	0.78	0.58	0.58	0.73
4	STD1	0.93	0.95	0.94	0.85	0.85	0.87
	STD4	0.87	0.90	0.88	0.80	0.79	0.75
2	STD1	0.90	0.94	0.91	0.86	0.87	0.78
	STD2	0.89	0.92	0.89	0.87	0.87	0.75

Table 4. Correlation between RMSEelev and terrain roughness. STD1 is the Standard deviation of slope measured from the original DEM (i.e. at a resampling of 1). STD32 etc is the STDslope measured from the DEM interpolated from the data resampled at level 32 etc

4. Conclusions

This paper has presented a summary of some of the findings of a study of DEM error. The main conclusions which may be drawn are

1. From the evidence presented here, spline-based methods produce DEMs which have low RMSE values and also have a smooth appearance.
2. Depsite its acknowledged limitations. $RMSE_{elev}$ may in fact be a reasonable predictor of error in some surface derivatives.
3. The spatial pattern of DEM error may not always be modelled by a simple spatial autocorrelation function and models of DEM error should possibly consider how error is related to the characteristics of the terrain itself.

5. Acknowledgements

The Ordnance Survey DEM data used for this work was obtained from the EDINA Digimap service and is Crown Copyright.

References

Fisher, P. 1998. Improved modeling of elevation error with Geostatistics. Geoinformatica 2 (3) 215-233.

Fisher, P. F., Tate, N. J. 2006. Causes and consequences of error in digital elevation models. Progress in Physical Geography 30 (4) 467-489.

Hunter, G. J., Goodchild, M. F. 1997. Modeling the uncertainty of slope and aspect estimates derived from spatial databases. Geographical Analysis 29 (1) 35-49.

Kidner, D. B. 2003. Higher-order interpolation of regular grid digital elevation models. International Journal of Remote Sensing 24 (14) 2981-2987.

Rees, W. G. 2000. The accuracy of Digital Elevation Models interpolated to higher resolutions. International Journal of Remote Sensing 21 (1) 7-20.

Wise, S. 2000. Assessing the quality for hydrological applications of digital elevation models derived from contours. Hydrological Processes 14 (11-12) 1909-1929.

Wise, S. M. 1998. The effect of GIS interpolation errors on the use of DEMs in geomorphology. In: S. N. Lane, Richards K.S. and Chandler J.H. (Eds) Landform Monitoring, Modeling and Analysis. John Wiley and Sons, Chichester.

Biography

Stephen Wise is a Senior Lecturer in GIS and author of a successful textbook 'GIS Basics'. His main research interests are the analysis of error propagation in DEMs and geomorphometry. He is a member of the GISRUK National Steering Committee (and was chairman and organiser of GISRUK 2002 in Sheffield) and is an Associate Editor of Computers and Geosciences.

Ed Hanna is a Senior Lecturer in Climate Change. His main research interests are in the link between climate change and the mass balance of ice-sheets. He is a Fellow of both the Royal Meteorological and Royal Astronomical Societies, has served as a Council member of the British Astronomical Association, and is an Editorial Board member of WEATHER (RMetS).

Tristram Irvine-Fynn is a PhD student working on water storage and the hydrological evolution of polythermal glaciers.

Assessment of DTM quality: a case study using fine spatial resolution data from alternative sources

Michal Gallay[1]

[1]School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, BT7 1NN, Belfast, United Kingdom
Tel: +44 (0)28 9097 3929, Fax: +44 (0)28 9097 3212
Email: mgallay01@qub.ac.uk, web address: www.qub.ac.uk/schools/gap/

KEYWORDS: digital terrain model, elevation, error, floods

1. Introduction

This research examines differences among digital terrain models (DTMs) generated from elevation data which were acquired using four modern technologies. The objective is to explore the properties of the DTMs and their errors. All the data covers the same area and was analysed in the form in which it was provided. The differences are demonstrated through analysis of the absolute errors of elevations and by the performance of the DTMs in modelling a flood event.

2. Research background

2.1 Landscape visualization and the data

Three dimensional visualization of landscape is becoming increasingly popular. There is a high demand for digital models of the Earth's surface in the public domain and thus by commercial companies too. Depending on the application, the users are interested in bare earth surface models (DTMs) or models which include the surface of the vegetation canopy and man-made structures (digital surface models, DSMs). For both types of modelling, three dimensional samples of the surface elevations are needed. Thus, one can term models based on elevation as digital elevation models (DEMs).

Nowadays, elevation data can be collected by a variety of methods including ground survey and remote sensing techniques. Developments in the latter allow for rapid collection of fine scale data with a high degree of accuracy. DEMs can be easily created from secondary data sources such as topographic maps. Furthermore, there is also a wide range of ready-to-use elevation data available for anyone to purchase and use. This is especially true for Great Britain, where DEMs derived from light detection and ranging (LiDAR), interferometric synthetic aperture radar (InSAR), photogrammetry and digitizing of contours are available on a national scale at a fine resolution (i.e., a point every 2 - 10 m). It is just a matter of choice which model to use. Or is it?

2.2 Digital terrain modelling: A piece of cake or a complex process?

Undoubtedly, digital terrain modelling is a complex process. The effects of the decisions involved propagate from the very first stage to the final product which is the DEM and from there to applications which make use of the DEM. Thus the quality of a DEM also determines the properties of the products derived from this DEM (e.g., slope angle, slope aspect, flow direction). A wide range of authors such as Wise (2007), Wechsler and Kroll (2006), Fisher and Tate (2006) and Chaplot et al. (2006) refer that the latter are often in more demand than the DEM itself because they are key components of models predicting floods or soil erosion. Furthermore, a DEM designed and generated for one specific purpose may not suit different applications. Key factors driving the decision process comprise the purpose, the available time and the arising costs. These determine subsequent decisions on how to sample the terrain, what method will be utilized, what way the data will be processed and the representation and data format to be chosen for the DEM. Above all this the factor of scale including spatial resolution and the complexity of terrain, controls the quality of a DEM.

From the user point of view it is important to understand what output properties of the DEM and the DEM derived products can be expected when various types of elevation data are used as inputs. Thus, developing an index or strategy for measuring sensitivity of outputs to various data types used as inputs would provide an invaluable resource to a wide range of users.

2.3 The problem

A plethora of papers deals with specific issues in digital terrain modelling although complex studies from the point of view of the user are rather rare. Many papers concern performance of various interpolation procedures when the same data are used as the input. For example, Carrara et al. (1997) compared different DEMs derived from digitized contours, Lloyd and Atkinson (2002) tested three kriging approaches with LiDAR data. Absolute height accuracy of the DEMs or comparison of two data acquisition methods is also a very popular topic and can be found in Baltsavias (1999) or Mercer (2001).

Several researchers have called for assessment of impacts of DEM properties on derived products and their further application (Fisher and Tate 2006, Chaplot et al. 2006, Raaflaub and Collins 2005). For example, Wise (2007) assessed six contour based DEMs and their effect on the outputs of the hydrological modelling with TOPMODEL. Other works look at the DEM error propagation into morphometric properties (Oksanen and Sarjakosi 2006, Raaflaub and Collins 2005), slope stability (Haneberg 2006) or viewshed (Fisher 1998).

This paper attempts to contribute to the discussion by exploring elevation errors of DEMs based on four different acquisition techniques together. This research is a part of a doctoral project which focuses on assessment and quantification of the effect of differences between the output DEMs and DEM derived products when various types of elevation data are used as inputs. The change in the input data is considered from the following viewpoints: i) acquisition method, ii) data processing method (filtering or interpolation) and iii) terrain complexity. It is very fortunate that there is available an extensive range of overlapping data of different origins and collected for various types of landscape. The DEMs and DEM derived products are analysed on the basis of summary statistics and variograms. In addition, error modelling is also undertaken.

3. Datasets and methodology

All four datasets used were DTMs covering an area of a mountainous region of the Lake District, Cumbria. The site is approximately 1500 by 1500 metres in extent. The DTMs were generated from data collected by airborne remote sensing with LiDAR (© Environment Agency), InSAR (© Intermap) and digital photogrammetry (© GeoPerspectives). The fourth DTM was interpolated from contour lines generated by analogue and digital photogrammetry (Ordnance Survey © Crown copyright). All DTMs are publicly available products and were provided in the form of square grids of 2, 5, 10, and 10 metre cell-size, respectively. They were projected in the OSGB36 coordinate system using Ordnance Datum Newlyn.

The data acquisition method and the spacing sizes of the original DTMs indicate that the terrain was sampled at different spatial scales, for example in more spatial detail by LiDAR (2 x 2 m). Despite this fact, the DTMs were regarded as representing a similar spatial scale and to keep the comparison as consistent as possible, all the DTMs were re-sampled to the same spatial resolution of 5 metres.

Raw LiDAR points were also available for this research and were used to produce a ground truth DEM. This data come from the same survey mission as the data from which the LiDAR DTM was created. The vertical accuracy of the points on flat land is expected to be less than 15 centimetres (Environment Agency, 1997 cited in Marks and Bates, 2000). The recorded values represent the last returns of a laser pulse. In cases where the laser light penetrated down to the ground, these elevations should be the most accurate samples of the ground surface amongst the DTMs analysed. While there are limitations to this assumption, the last return points were used as a ground truth in this study.

A DEM was interpolated from the points with inverse distance weighting and the cell-size of 5 metres to match the resolution of the resampled DTMs. This DEM was used as the ground truth and was subtracted from the four DTMs (5 x 5 m) and so difference surfaces of the same spatial resolution were generated (Figure 1). All the data were statistically analysed and visualized in R (© The R Foundation for Statistical Computing), ArcGIS™ (© ESRI), and LanSerf (© Jo Wood).

4. Results and discussion

The data, from which the resampled DTMs were generated, are samples of the same surface on a very similar scale. The impacts of data currency and land-cover at the time of the data collection are not considered. To obtain a first impression of the data the first step taken was visualization which is generally agreed to be the fastest means of DEM assessment. Here, shaded relief maps (Figure 1) and 3D views (Figure 2) are presented. Although it is very difficult to judge which of the DTMs is of the highest quality in any sense, at least the obvious unnatural terrain features can be immediately recognized (Figure 1 a-d, Figure 3).

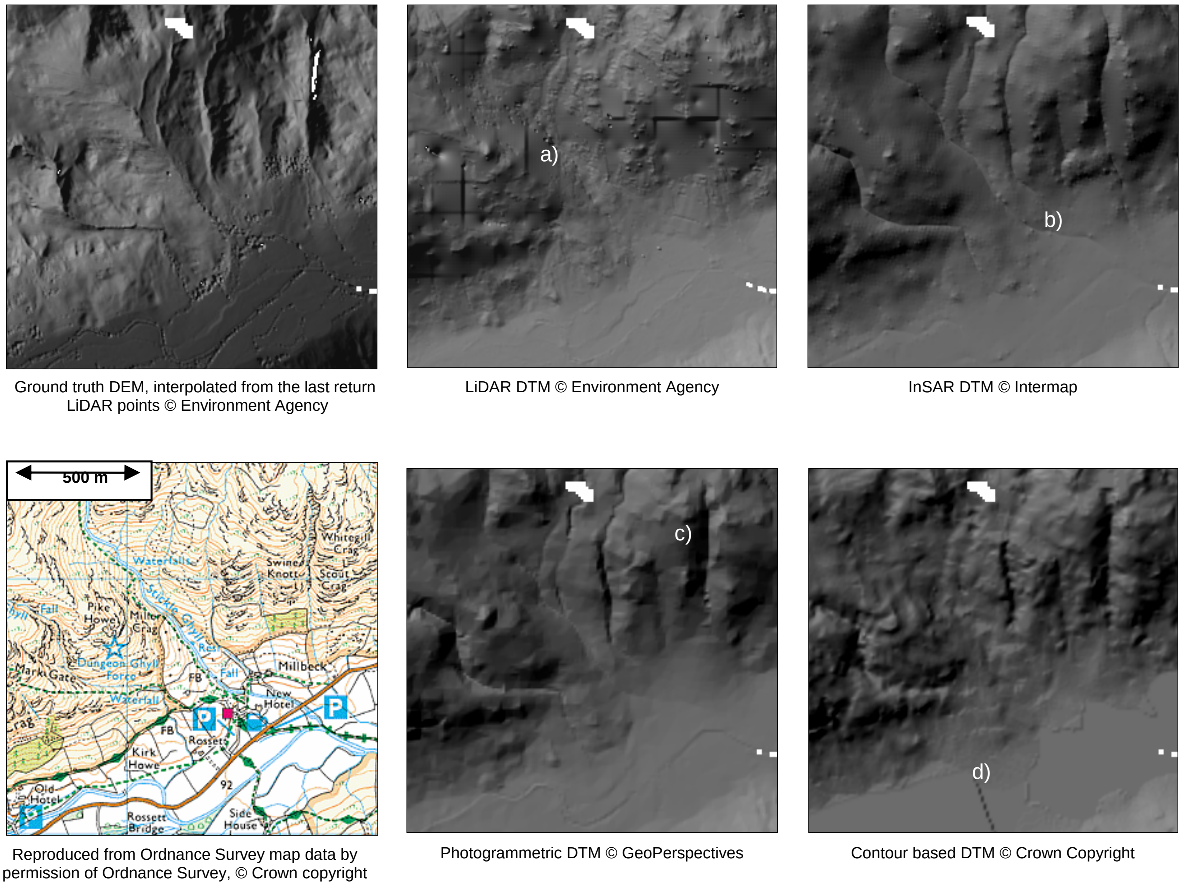


Figure 1: Shaded relief maps of the datasets used and a topographic map of the study area. Spatial resolution of the input DEMs and the shaded relief maps is 5 metres, the light source position is southeast and 15 degrees above horizon. Note the non-natural surface features such as a), c) planar smoothing of the slopes, b) regular pattern of shallow depressions, d) terraced steps in the valley.

Assessment of DTM accuracy was conducted through calculation of differences with respect to the DEM interpolated from LiDAR points of last return. The difference surfaces and their statistics show that the LiDAR DTM is not the most accurate, perhaps counter to what might be expected (Table 1, Figure 2). Despite the residuals being below the measurement error (less than 15 centimetres) within the flat valley floor (Figure 2), the areas of steep slope or areas with trees and buildings exhibit errors in the order of metres. The reason is the method of filtering the raw LiDAR data to get the DTM. The remaining DTMs are not markedly different when looking at the statistics of the error although

Table 1. Errors of the four DTMs with respect to DEM interpolated from LiDAR points of last return.

DTM origin	Min	Max	1.Qrt	3.Qrt	IQR	ME	MAE	SDE	RMSE
LiDAR	-27.92	63.00	-0.36	0.35	0.71	0.98	2.64	6.40	6.47
InSAR	-27.66	38.82	-1.12	0.58	1.69	0.05	1.87	3.69	3.69
Photogrammetry	-32.89	32.89	-1.29	0.81	2.10	-0.43	1.70	2.62	2.65
Contour based	-33.12	33.12	-1.96	1.28	3.24	-0.38	2.18	3.02	3.04

Min – minimum error, Max – maximum error, 1.Qrt – lower quartile, 3.Qrt – upper quartile, IQR – inter-quartile range, ME – mean error, MAE – mean absolute error, SDE – Standard deviation of error, RMSE – root mean squared error.

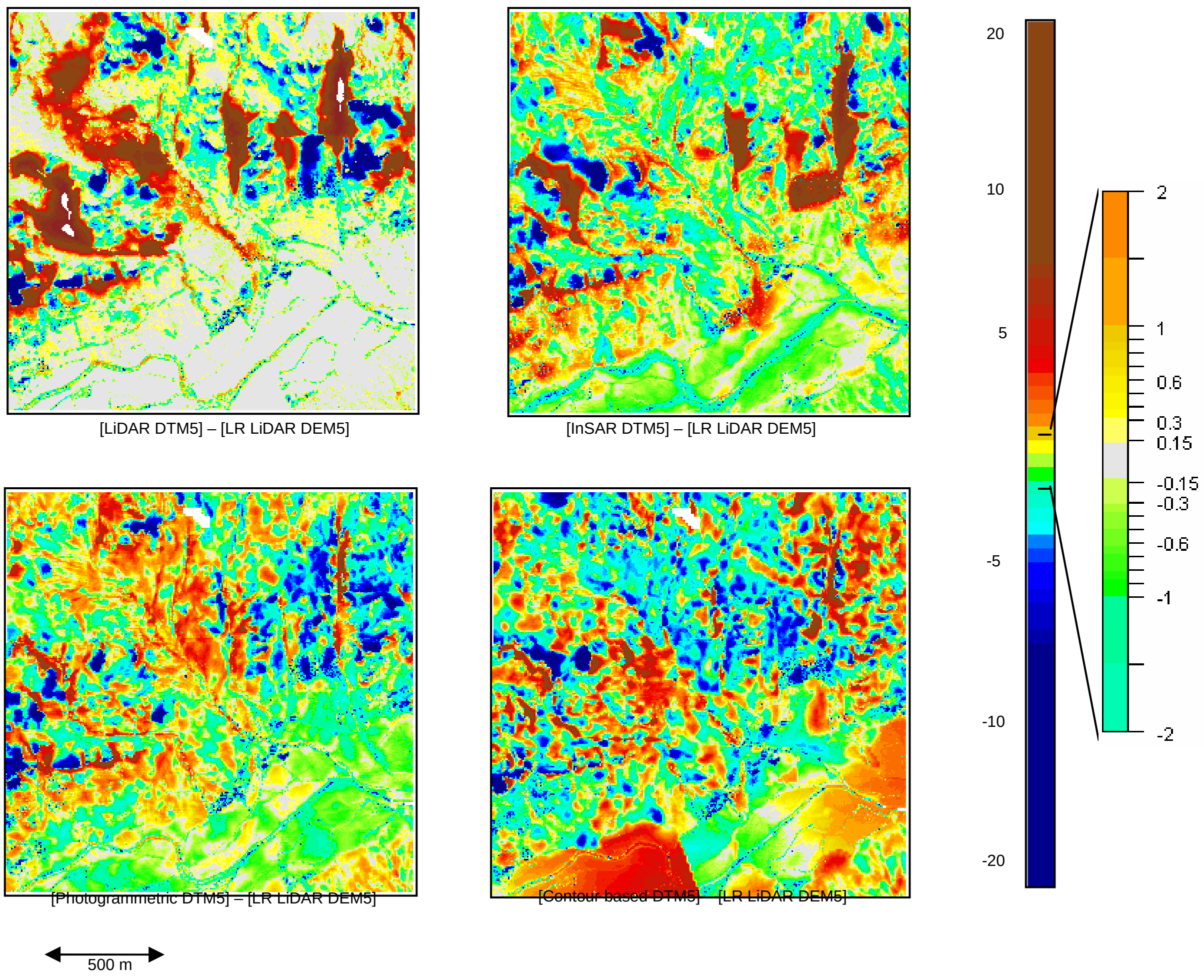


Figure 2. Elevation difference surfaces of the DTMs. DEM interpolated from the last return LiDAR points (LR LiDAR DEM) was used as the ground truth which was subtracted from the DTMs. Spatial resolution of all DTMs and the LS DEM was 5 metres and the units are in metres.

the spatial patterns of the errors differ. One thing in common for all four datasets is that the error patterns are not random but spatially correlated. This has implications for more sophisticated approaches to modelling the uncertainty in DEM data which is recognized by many authors including, for example, Hunter and Goodchild (1997), Fisher and Tate (2006), Haneberg (2006) and Wechsler and Kroll (2006). If only the RMSE was used to guide DTM selection, which is usually the only measure of error given to the user, one would not choose the LiDAR DTM. Instead, selection of the the InSAR or the photogrammetric DTM, for which RMSEs are smaller, would be more likely. However, knowledge about the spatial distribution of errors is important and a single statistic is likely to be a poor guide to DTM quality.

Looking at the flood scenarios in figure 3, one can observe that the LiDAR, photogrammetric and contour based DTMs produce very similar outcomes. If the users had the opportunity to see the outputs of the DTM application prior to purchasing the data, they would choose the cheaper option if all data sources produce similar results. In reality, it would be necessary to buy and try all possible DEMs. Of course, nobody can usually afford this luxury.

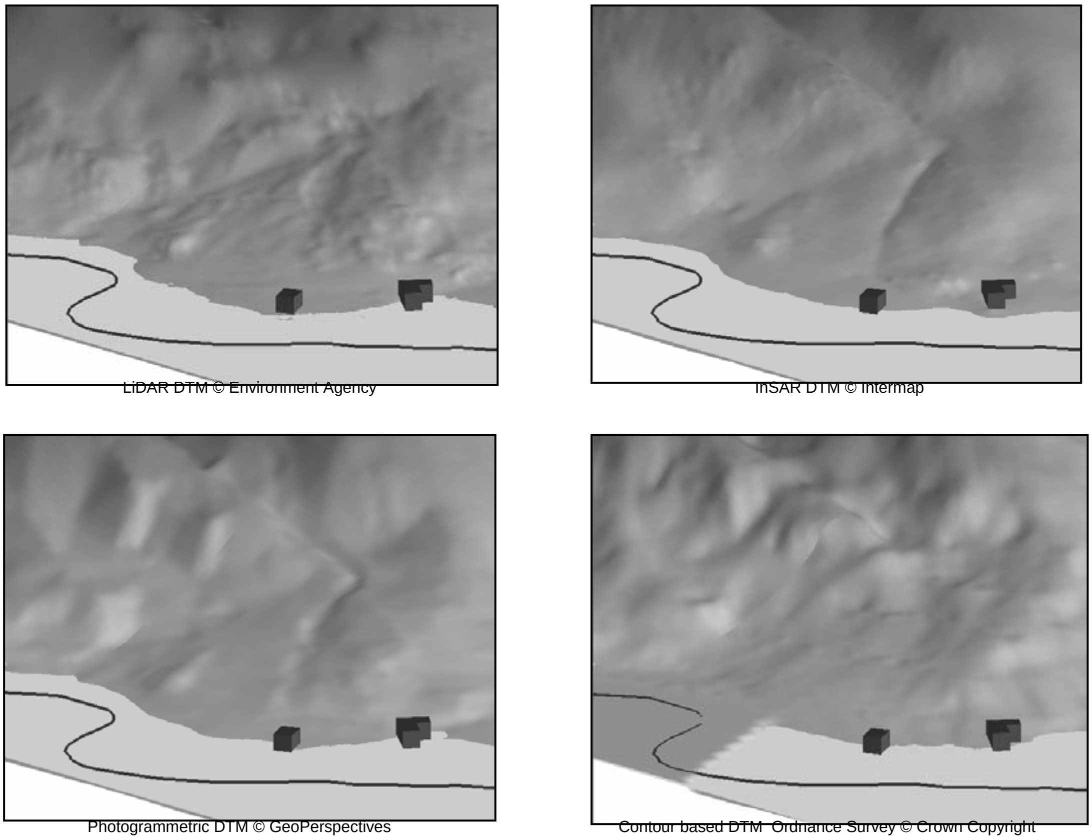


Figure 3. Hypothetical flooding scenario of the valley. The originally provided DTMs were resampled to 5 x 5 metres. The flooded zone is about 500 metres in width. The view angle is from the southeast.

5. Conclusions

The results presented here suggest that factors such as the purpose of the DTM are often more important than its absolute accuracy. Even if the latter matters most, it depends on how the accuracy is quantified. This simple case study could be challenged in various ways, however, it does provide evidence that challenges the ways in which digital elevation products are used.

Acknowledgements

I would like to thank the European Social Fund for funding my study as well as Dr. Chris Lloyd and Dr. Jennifer McKinley for supervising the project. Last, but not least, I thank the companies and institutions which provided the data for this research, namely: Environment Agency UK, Intermap Technologies UK Ltd., Infoterra Ltd., Ordnance Survey Great Britain.

References

Baltsavias, E. P. (1999). A comparison between photogrametry and laser scanning. ISPRS Journal of *Photogrametry & Remote Sensing,* ***54***, pp.83-94.

Chaplot,V., Darboux, F., Bourennane, H., Leguédois, S., Silvera, N., Phachomphon, K. (2006). Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density. *Geomorphology*, ***77(1-2)***, pp.126-141.

Fisher, P. (1998). Improved Modeling of Elevation Error with Geostatistics. *GeoInformatica*, ***2(3)***, pp. 215-233.

Fisher, P. F., Tate, N. J. (2006). Causes and consequences of error in digital elevation models. *Progress in Physical Geography,* ***30(4)***, pp.467-489.

Haneberg, W.C. (2006). Effects of digital elevation model errors on spatially distributed seismic slope stability calculations: an example from Seattle, Washington. *Environmental & Engineering Geoscience,* ***12(3)***, pp.247-260.

Hunter, G. J., Goodchild, M. F. (1997). Modeling the Uncertainty of Slope and Aspect Estimates. *Geographical Analysis,* ***29(1)***, pp.35-49.

Lloyd, C., Atkinson, P., M. (2002). Deriving DSMs from LiDAR data with kriging. *International Journal of Remote Sensing,* ***23(12),*** pp.2519-2524.

Marks, K. and Bates, P. (2000): Integration of High Resolution Topographic Data with Floodplain Flow Models. *Hydrological Processes,* ***14(11-12)***, pp.2109-2122.

Mercer, B. (2001). Combining LIDAR and IfSAR: What can you expect? *Photogrammetrische Woche 2001*. Heidelberg: Wichmann Verlag. pp.227-237.

Raaflaub, L. D., Collins, M. J. (2005). The effect of error in gridded digital elevation models on the estimation of topographic parameters. *Environmental Modelling & Software,* ***21(5)***. pp.710-732

Wechsler, S. P., Kroll, C., N. (2006). Quantifying DEM Uncertainty and its Effect on Topographic Parameters. *Photogrammetric Engineering and Remote Sensing,* ***72(9)***, pp.1081-1090.

Biography

Michal Gallay completed his undergraduate studies in Geography and Cartography (2003) and MSc. in Physical Geography and Geoecology (2005) at the Comenius University in Bratislava, Slovakia. Currently, he is a third year PhD. student at Queen's University Belfast working on the assessment of methods for acquiring and processing digital elevation data. His main areas of focus encompass DTM generation and applications as well as spatial analysis in GIS.

Developing an automatic process for hydrology generalization: combining a vector GIS topography model with DTM-derived flow direction data.

Marshall J. Heap[1]

[1]Birkbeck College, University of London.
Contact postal address: School of Geography
Birkbeck, University of London
Malet Street
London, WC1E 7HX, UK
Tel: +44 (0) 20 7631 6473 Fax: +44 (0) 20 7631 6498
Email: heapmj@gmail.com

KEYWORDS: GIS, hydrology, generalization, vector, DTM

1. Introduction

This project introduces a new automatic process methodology for hydrology generalization that aids cartographic representation of hydrology by matching an appropriate stream order classification to map scale.

In a departure from established *stream burning* techniques (Saunders 1999), the developed process performs hydrology generalization by *burning in* flow data from the DTM to the vector topography layer rather than the opposite way around.

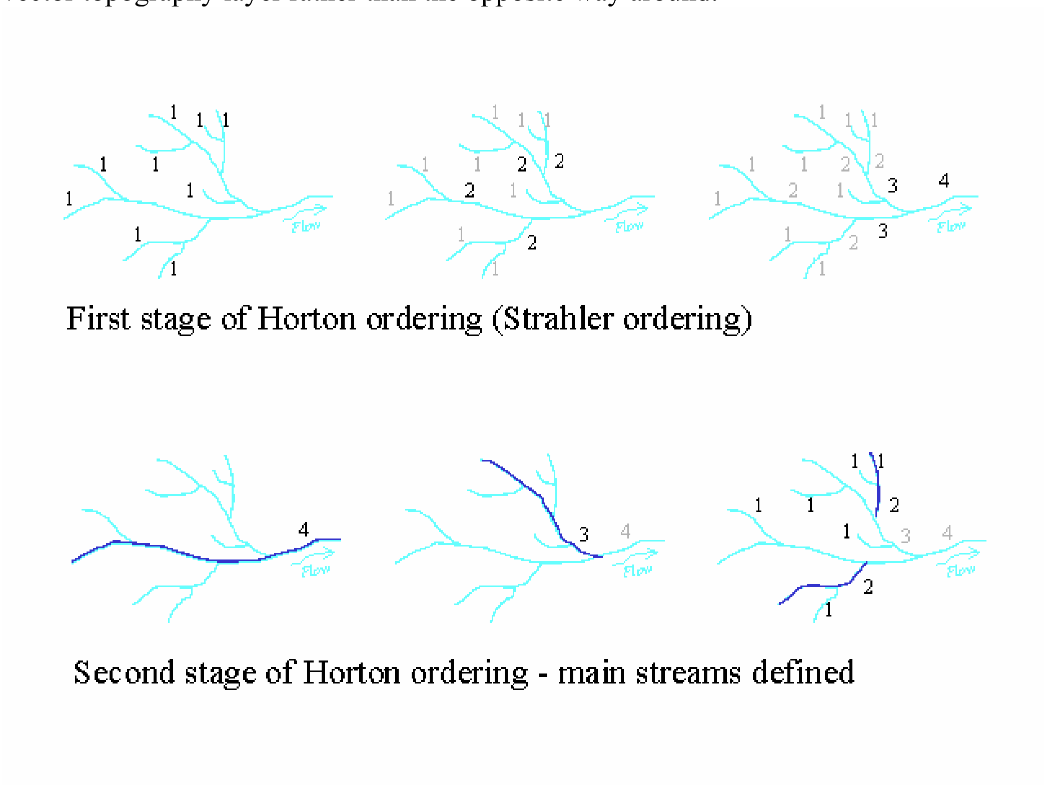


Figure 1 – The two stages of Horton ordering

Hydrology generalization is based on Horton ordering (Horton 1945), the first stage of which is Strahler ordering (Strahler 1957), a downstream operation where an outflowing stream order is one higher than the order of two similarly ordered inflowing streams (Figure 1). The second stage is an upstream operation from the river network's estuary. At confluences, the mainstream continues along the highest Strahler order-inflowing stream. Where inflowing streams have the same Strahler order, the mainstream chooses the one farthest from its source. Tributary mainstreams are then similarly defined.

Stream ordering requires flow direction data. The raster (GIS) model has been extensively used for hydrology generalization with the 'D8' method defining flow direction (Chang 2005). However, common raster-based problems are; that the minimum stream size is constrained by the raster resolution and poor performance of the basic algorithm in flat areas (Chang 2005).

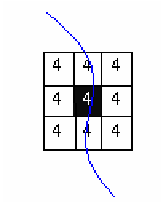


Figure 2 – Stream burning in a flat area

Stream burning vector hydrology into the DEM solves these problems (Saunders 1999); first by facilitating stream delineation and second by forcing flow through raster cells corresponding to the river network in flat areas (Figure 2). Nevertheless, *stream burning* limitations include; an inability to handle off-stream lakes and marine features, difficulty in handling river splits, large datasets and high data processing times (Saunders 1999). Consequently, this project's automatic process for hydrology generalization was designed to overcome these limitations.

2. Methodology & Development

2.1 Methodology

OS topographic vector hydrology data depicts watercourses using lines and/or polygons, without any distinction between rivers and lakes, relative importance or flow direction. Any method of hydrology labelling has to be able to cope with numerous problems that this single fact engenders. In this, the 'burning in' approach developed has significant advantages over previous methods.

Initially, the experiment used a 50 tile 10-metre resolution DTM that was pre-processed to remove voids and pits. Then an automatic process was developed as a software application consisting of nine plug-ins to Vivid Solutions Inc's Java © Unified Mapping Platform (JUMP). Figure 3 describes the overall process and task sequence.

Step 1: Create data schema and import appropriate topographic features.

↓

Step 2: Create river networks from intersecting features excluding marine features.

↓

Step 3: Import elevation values of water feature end nodes from the DTM.

↓

Step 4: Locate and eliminate removable flat river features from the dataset.

↓

Step 5: Strahler order downstream from source features using imported elevation values.

↓

Step 6: Horton order upstream from river network estuary using Strahler order feature values.

↓

Step 7: Develop visualization algorithms allowing user to select Horton order & width generalization.

Figure 3 – process and task sequence

2.2 Development

Steps 1, 2 and 3 of Figure 3 placed the source data in a state in which Strahler/Horton ordering (steps 5 and 6) could be performed. Step 7 was required to allow users to easily visualize and compare the Horton classes generated. The significant numbers of flat features in the dataset lead to poor initial results. Consequently, step 4 was added to improve these results. Figure 4, illustrates the design of the key RemoveFlatsPlugIn that implemented step4.

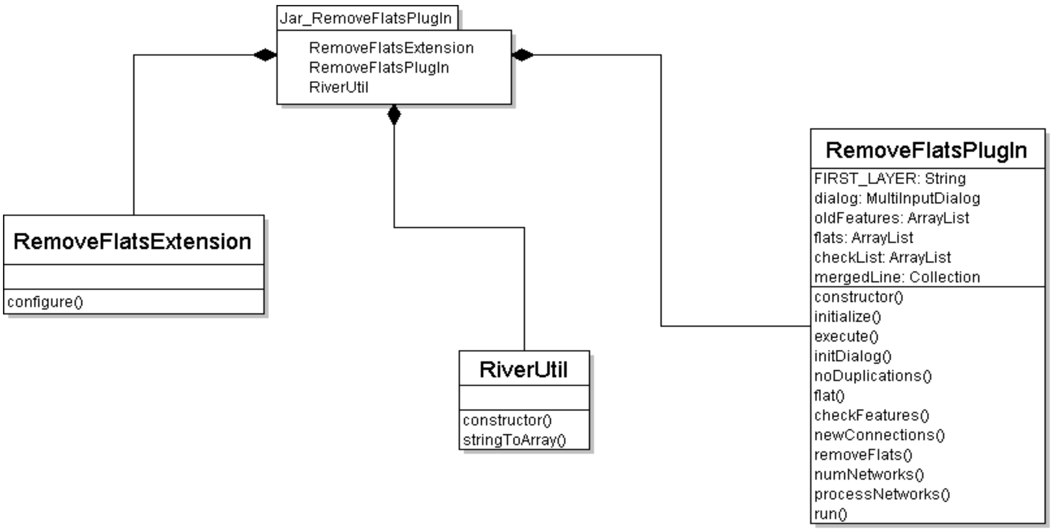


Figure 4 – Class diagram for Jar_RemoveFlatsPlugIn, step 4 in Figure 3.

The removeFlats() method of the RemoveFlatsPlugIn provides most of this PlugIn's functionality and Figure 5 below shows the Java code used for this method.

```
    public void removeFlats(TaskMonitor monitor, List rFeatures,
FeatureCollection r)
   throws Exception {

        oldFeatures = new ArrayList();
        flats = new ArrayList();
        for (int i = 0; (i < rFeatures.size()) && !monitor.isCancelRequested();
                i++) {
            Feature aFeature = (Feature) rFeatures.get(i);
            Object ob = aFeature.getAttribute("INFLOW_HT");
            String rString = ob.toString();
            int heightFirst = Integer.parseInt(rString);
            Object bob= aFeature.getAttribute("OUTFLOW_HT");
            String yString = bob.toString();
            int heightLast = Integer.parseInt(yString);
            if (heightFirst == heightLast)
            {
                flats.add(aFeature);
            }
       }
       for (int ii = 0; (ii < flats.size()) && !monitor.isCancelRequested();
                ii++) {
       // MJH - bFeature is flat
       Feature bFeature = (Feature) flats.get(ii);
       Object obj = bFeature.getAttribute("CONNECTIONS");
       String aString = obj.toString();
       ArrayList cons = RiverUtil.stringToArray(monitor, rFeatures, aString);
```

```
ArrayList openEnds = new ArrayList();
boolean looseEnd = false;
boolean attached = false;
for (int iii = 0; (iii < cons.size()) && !monitor.isCancelRequested();
         iii++) {
Feature cFeature = (Feature) cons.get(iii);
looseEnd = false;
attached = false;
for (int iiii = 0; (iiii < cons.size()) && !monitor.isCancelRequested();
         iiii++) {
Feature dFeature = (Feature) cons.get(iiii);
 if (!(dFeature.equals(cFeature))&& !(flat(cFeature)) //MJH - only
 attach to non flat feature
&& !(checkFeatures(cFeature))&& !(checkFeatures(bFeature)))
{
if (cFeature.getGeometry().touches(dFeature.getGeometry()))
{
attached = true;
}
else if (!(cFeature.getGeometry().touches(dFeature.getGeometry())))
{
looseEnd = true;
}
}
} /** end for second cons **/
if (looseEnd && !attached)
{
 openEnds.add(cFeature);
 oldFeatures.add(bFeature);
}
} /** end for first cons **/
if (!(openEnds.isEmpty()))
{
Feature eFeature = (Feature) openEnds.get(0);
LineMerger lineMerger = new LineMerger();
lineMerger.add(eFeature.getGeometry());
lineMerger.add(bFeature.getGeometry());
mergedLine = lineMerger.getMergedLineStrings();
Iterator it = mergedLine.iterator();
LineString lineString = (LineString) it.next();
eFeature.setGeometry(lineString);
eFeature.getGeometry().geometryChanged();
}
} /** end for flats **/
for (int v = 0; (v < oldFeatures.size()) &&
!monitor.isCancelRequested();
         v++) {
Feature oldFeature = (Feature) oldFeatures.get(v);
r.remove(oldFeature);
}
mergedLine.clear();
} /** end removeFlats **/
```

Figure 5 – The removeFlats() method used by the RemoveFlatsPlugIn.

The Java code for the above method executes the following task sequence:

i. Creates a list of flat features in the dataset (with identical inflow/outflow heights).

ii. For each flat feature, finds touching candidate sloping features whose geometry could be extended to include the flat feature.

iii. Uses the LineMerger Java class to extend the geometry of chosen candidate sloping features.

iv. Removes from the dataset flat features whose geometry has been incorporated into another feature.

3. Results

Figure 6 depicts a DEM of the area on which the test river hydrology is draped.

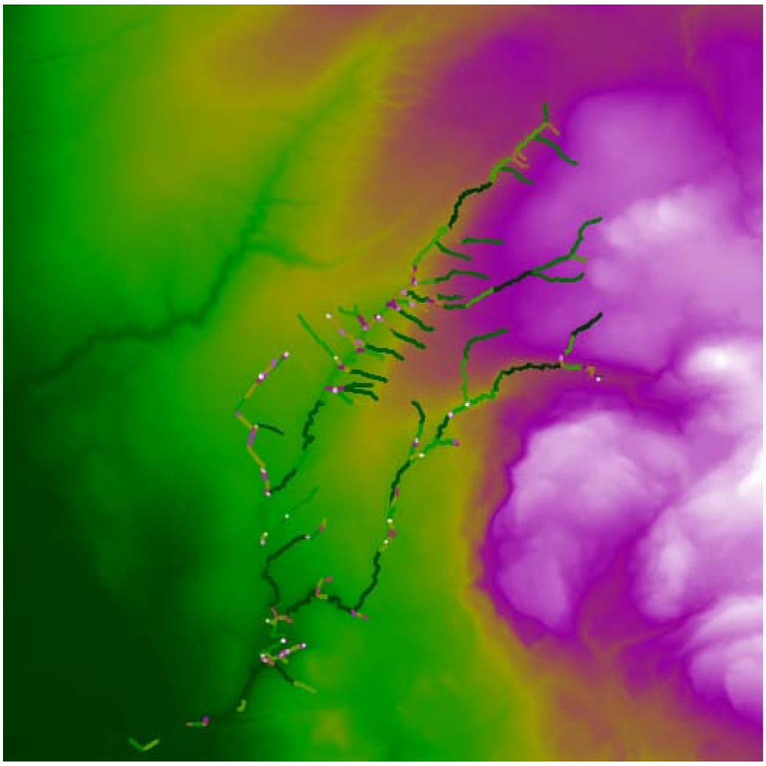


Figure 6 – Test river network (image generated by LandSerf © Jo Wood).

Figure 7 produced by the project's visualization algorithms, depicts Horton class delineation of the test river network. Examination at higher resolution validates that accurate class delineation also occurs at fine scales.

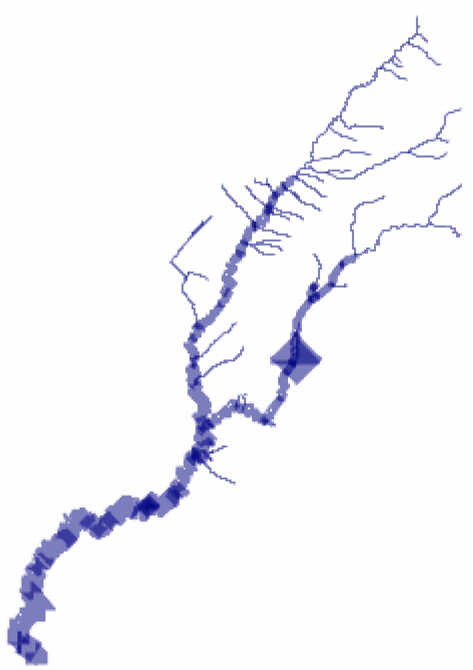


Figure 7a -Horton classes >=1

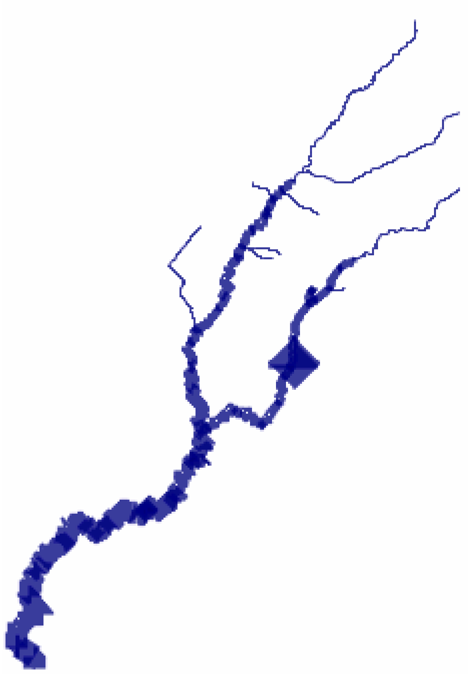


Figure 7b -Horton classes >=2

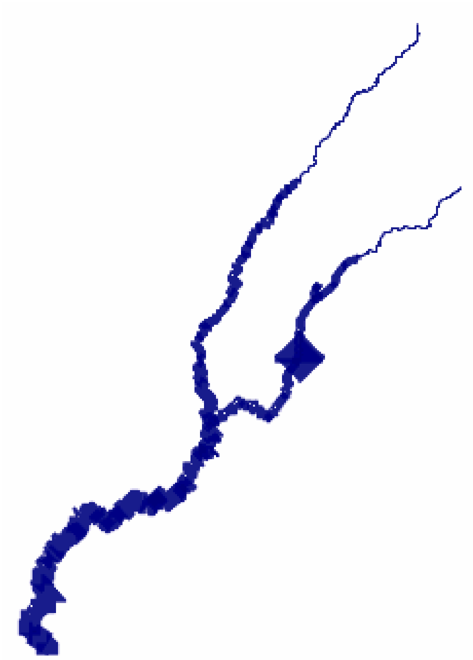


Figure 7c - Horton classes >=3

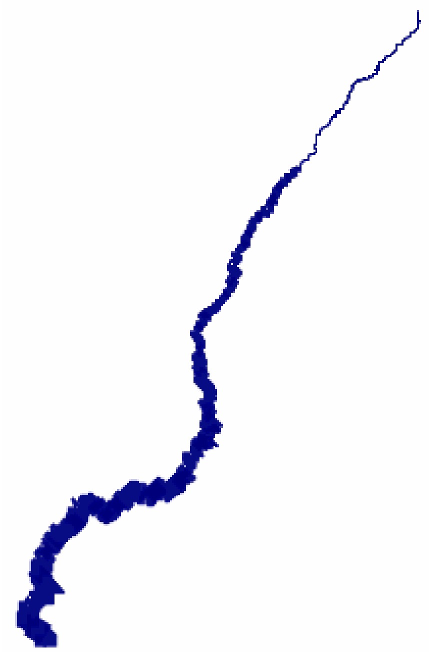


Figure 7d - Horton class 4

3.1 Problems encountered

Flat features are problematic as they can cause incorrect Strahler ordering when the algorithm cannot deduce whether the feature adjoins an inflowing or outflowing stream.

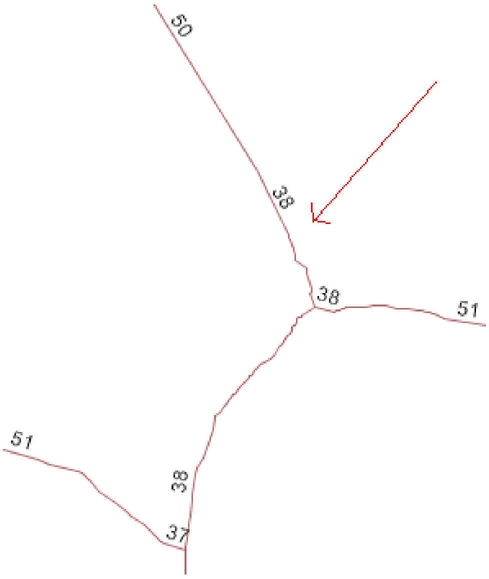


Figure 8a – INFLOW_HT

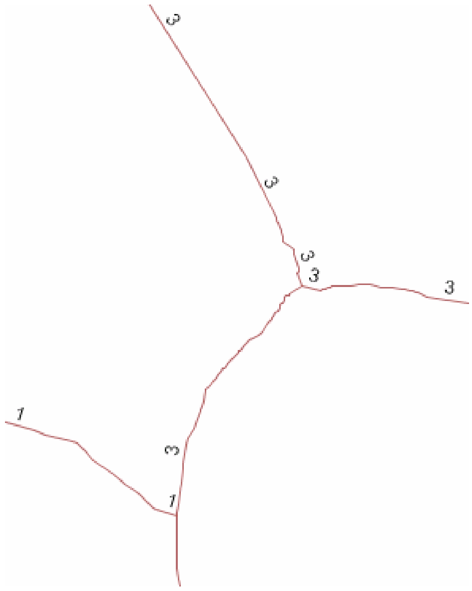


Figure 8b – STRAHLER class

Figure 8c – INFLOW_HT

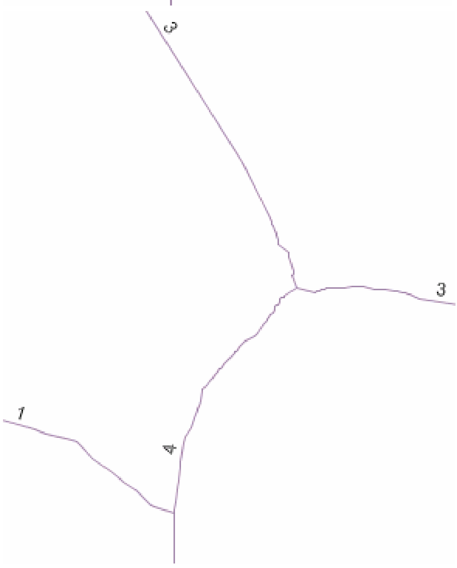


Figure 8d – STRAHLER class

In Figure 8a, the two upper streams are inflowing considering inflow height but the indicated feature is flat (inflow/outflow height = 38 metres). A right-hand incoming Strahler order() process, assumes this flat feature isn't inflowing and stops (because it wont proceed uphill). Consequently, the outflowing stream below, does not receive a Strahler class increment (Figure 8b). Processes must verify two identical Strahler order inflowing streams before an outflowing stream's Strahler order is incremented. This dilemma sparked the removeFlatsPlugIn development (Figure 3, step 4).

Applying removeFlatsPlugIn first, removes the offending flat feature and extends the geometry of the contiguous upper non-flat feature (Figure 8c). Now, the order() process can see the extended feature is inflowing and therefore increment the outflowing stream Strahler order (Figure 8d).

Figure 9a Dataset_Test1 class >=4

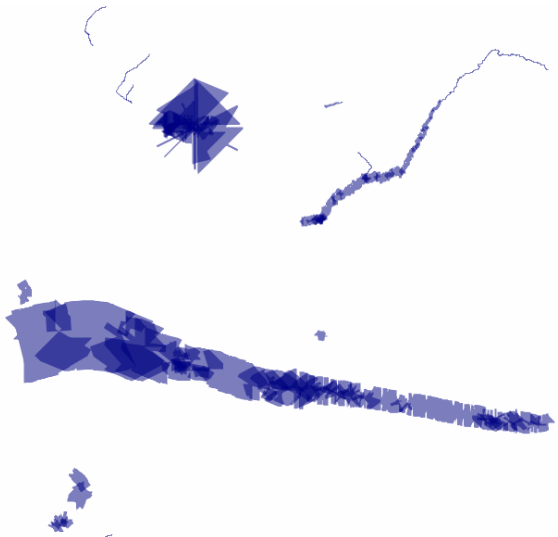


Figure 9b Dataset_Test2 class >=4

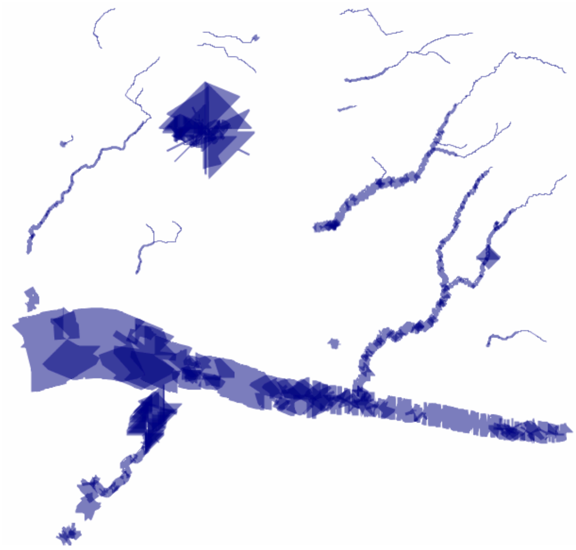


Figure 9c Dataset_Test1 class>=3

Figure 9d Dataset_Test2 class >=3

At dataset level, Figure 9a & b show differences in Horton class 4+ delineation where Figure 9a omits the test river network because, unlike Dataset_Test2, Dataset_Test1 was not pre-processed with the RemoveFlatsPlugIn. In Figure 9c & d, careful scrutiny reveals once more that Dataset_Test2 shows some improvements in Horton classification.

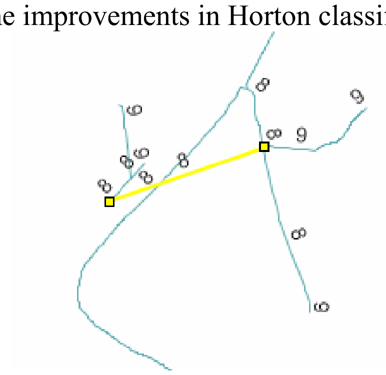


Figure 10 – feature inflow elevations in metres

The incorrect candidate link example (Figure 10) illustrates that candidate links crossing another feature's geometry violate dataset integrity and DTM-imported elevation data could help define candidate links. There is much research regarding topographic candidate links.

For example, Regnauld & Mackaness (2006) consider *proximity* and *continuity* in developing candidate links and whether they lie within a buffered talweg. However, this project design offers an alternative approach to talwegs. For example, starting from one candidate link node, DTM points falling within a 10 metre buffered candidate link line could be progressively interrogated to validate slope direction. Further development work will be required to develop reliable candidate links.

3.2 Data processing issues

The StrahlerOrder and HortonOrder PlugIns completed on Dataset_Main but the removeFlatsPlugIn crashed after an hour processing river network #1 (with 5,610 features). Probably, this failure was memory-related. Increasing physical computer memory (from the 0.5GB used) to 1 GB+ is a simple solution. Alternatively, re-designing removeFlatsPlugIn. with recursive calls to remove flat features may increase algorithm efficiency by completely consolidating chains of flat features.
Analysis of the 30-river networks sample from Dataset_Test2.shp revealed a total of 53 errors i.e. an average of 1.8 per network divided between data and algorithm errors in the ratio 20:33 (Figure 11). Total errors increase with Horton class complexity.

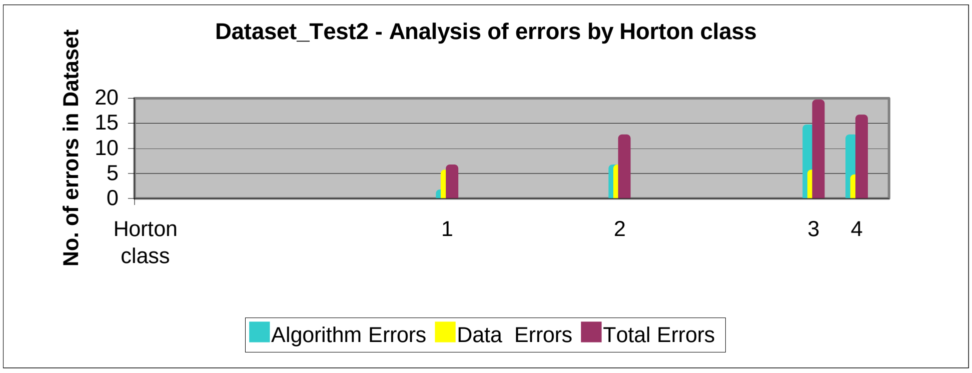


Figure 11

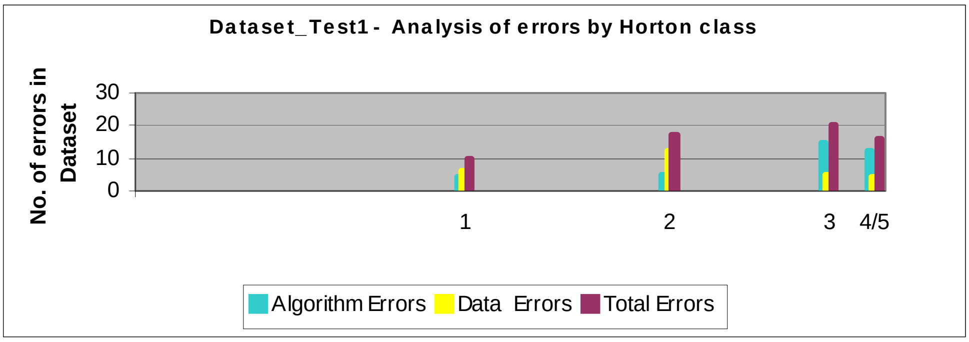


Figure 12

The 10 additional Dataset_Test1 errors shown in Figure 12 are mainly of data origin due to non pre-processing of this dataset with the RemoveFlats algorithm.

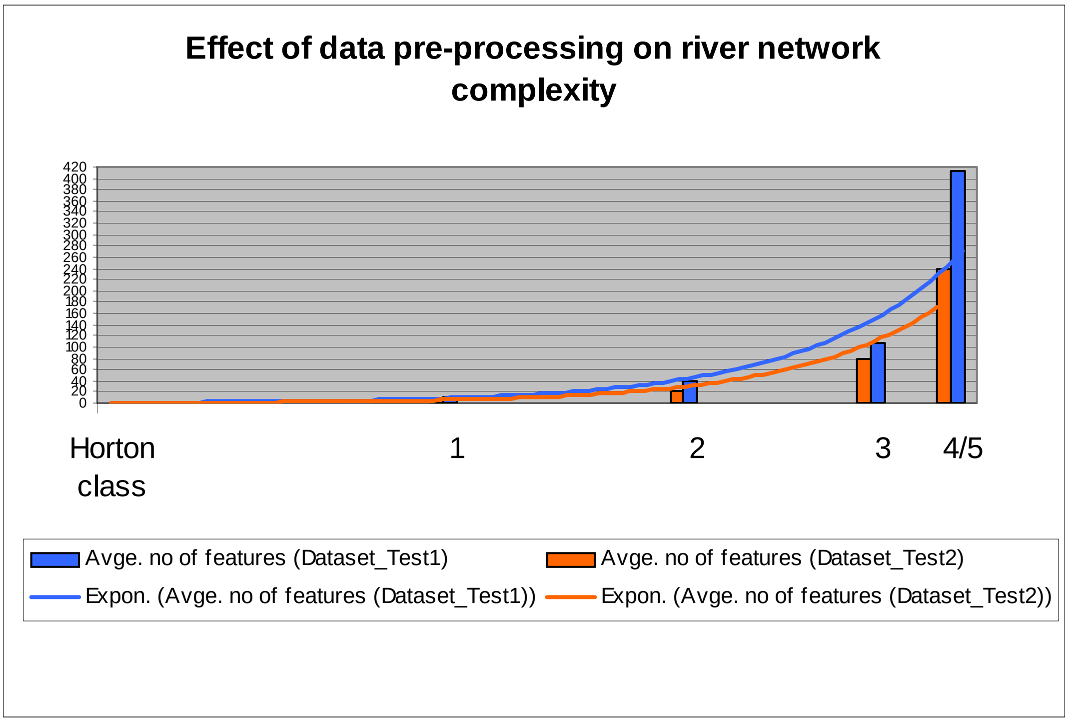


Figure 13

Figure 13 shows the relative increase in quantitative feature complexity between the two datasets due to the presence of more flat features in Dataset_Test1.

Table 1

Dataset_Main sizes			
Dataset	Size in MB's	No. of features	No. of networks
Dataset_Main.shp	39.5	44,111	1,571
DTM.shp	13.0	N/A	N/A

A limitation of raster-based hydrology generalization is large datasets. In this case, such a solution would have potentially run to 690 megabytes (i.e. (50 DTM tiles x 13MB) + 39.5MB of topographic data). By contrast, this project's *DTM burn* approach used only 39.5MB.

Table 2

Dataset_Main processing times	
PlugIn	Dataset_Main
Data	2 sec
RiverNetworks	2 hours 39 min 23 sec
ImportDTM	50 x 44 min = 36 hours 40 min
RemoveFlats	N/A
TOTAL DATA PRE-PROCESSING TIME	39 hours 19 min

Although the total data pre-processing time incurred on Dataset_Main was excessive (Table 2), dividing the topographic dataset set into 'virtual' tiles, each of whose boundaries correspond to its DTM counterpart should produce substantial time savings for the ImportDTM PlugIn i.e. 44 min versus 36 hours 40 min.

These results were subject to the +/- 2.5 metres uncertainty of the DTM elevation data used.

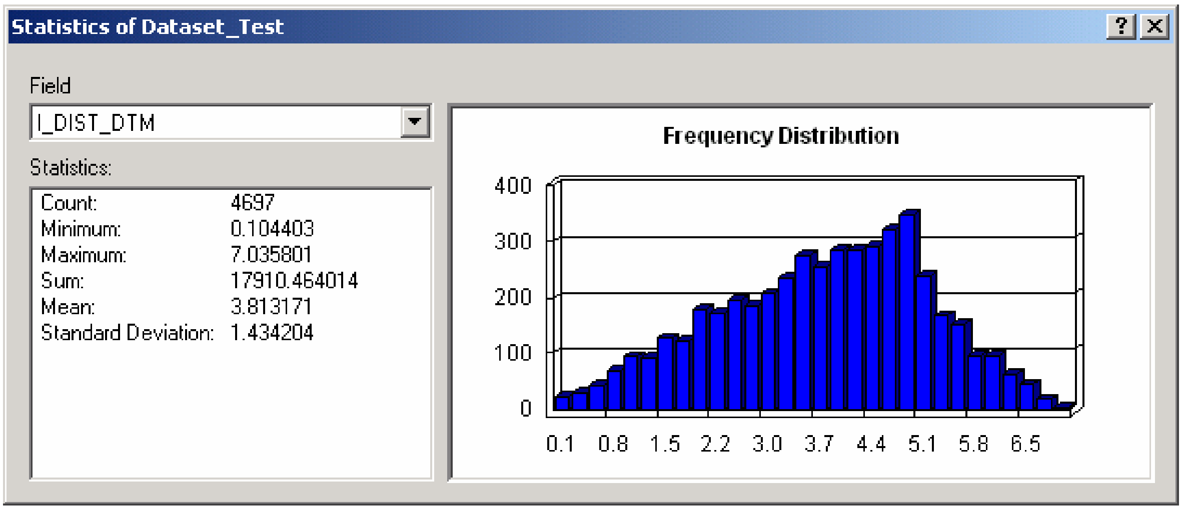


Figure 14 – Distance between feature inflow height & DTM (ESRI ® ArcMap™)

Further data uncertainty was introduced because feature node elevation values came from the nearest point in the DTM (Figure 14). However, except for steep slopes, feature inflow/outflow elevation values are unlikely to have varied by more than a metre from the true value.

4. Conclusions

While the accurate Horton ordering of the test river network was encouraging, the developed process proved qualitatively more effective on the test river network than at dataset level. Nevertheless, although classification errors increased with Horton class complexity at dataset level, there were few errors suggesting that the overall quality of the algorithms used is fairly good.

The use of recursive calls to Strahler/Horton order the main dataset's 1,250 km^2 in 4 minutes proved particularly efficient.

Key actions needed for qualitative improvement at the dataset level are:

1. Future development using 1 GB+ of computer RAM.
2. Re-design of the RemoveFlatsPlugIn using a recursive call based solution.
3. Experimentation with higher precision and resolution DTM's.
4. Use of *fuzziness* for defining complex lake feature gradients.
5. Division of the topographic set into *virtual* tiles to reduce data pre-processing time.
6. Utilization of feature inflow/outflow elevation data in developing *candidate link* algorithms.
7. 'Stepping' the mainStream() method in the HortonOrderPlugIn to follow algorithm processes and determine the cause of incorrect Horton class errors.

These actions will aid a more complete qualitative dataset comparison between the established *stream burning* and *DTM burning* approaches to hydrology generalization.

A demonstrated benefit of the *DTM burning* approach is massive file size savings. Potentially, there are significant pre-processing and processing timesavings. In contrast to *stream burning*, *DTM burning* handled river splits and both in-stream and off-stream lakes as well as marine features. However, *flat feature chain consolidation* is perhaps the strongest argument in favour of the *DTM burn* technique for future development of automated hydrology generalization. First, because this is a more verifiable method of determining flow direction in flat areas than the *grid cell force through* used in *stream burning* and second because of the greater *reach* offered over the D8 method.

5. Acknowledgements

The author acknowledges and thanks the following individuals and organisations for their contributions to this project:

Name	Affiliation	Contribution
Ordnance Survey		For providing topographic and DTM data.
Unwin, David	Birkbeck College, University of London	For marking feedback on the author's M.Sc. dissertation based on this project and feedback on this extended abstract.
Vivid Solutions, Inc		For free use of their JUMP GIS application & access to source code.
Voudouris, Vlasios	Birkbeck College, University of London City University, London London Metropolitan University	For supervising the author's M.Sc. dissertation and for feedback on this extended abstract.

References

Chang, K.S. (2005). *Introduction to Geographic Information Systems, 3rd Edition*. McGraw Hill.

Horton, H. (1945) Erosional development of streams and their drainage basins *Bulletin of the Geological Society of America*, 56, pp. 275-370.

Regnauld N. & Mackaness W.A. (2006) Creating a hydrographic network from its cartographic representation: a case study using Ordnance Survey MasterMap data *International Journal of Geographic Information Science*, 20, pp. 611-631

Saunders, W.K. (1999) *Preparation of DEMs for Use in Environmental Modeling Analysis*. In proceedings of 1999 ESRI User Conference, July 1999 (San Diego, CA:ESRI). Retrieved March 06, 2007, from
http://gis.esri.com/library/userconf/proc99/proceed/papers/pap802/p802.htm

Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. *Trans. of the American Geophysical Union*, **38**, pp. 913-920.

Biography

After gaining a B.Sc. Computing (1st class) from the Open University (2005), the author applied this knowledge in achieving a M.Sc. with Distinction in GIScience from Birkbeck College (2007). This extended abstract is derived from the author's M.Sc. dissertation project. The author hopes to develop his GIS interests at research degree level.

Using GIS to Reproduce Landscape Evolution in Active Volcanic Environments

Amii Darnell[1], Andrew Lovett[1], Jennifer 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: geomorphology, volcanic hazard, digital elevation model, GPS, georeferencing

1. Introduction

GIS has the potential to benefit the study and management of active volcanic environments. A GIS can be used to compile risk maps, develop evacuation plans, monitor an ongoing or impending eruption, and through correct land use planning can be used to minimise risk (Pareschi et al, 2000). However data acquisition from an active volcanic system can be challenging and field work can be dangerous. For example steam and ash emanating from a volcanic dome creates cloud cover and can hinder the gathering of quality satellite data. Furthermore prioritising the set up and maintenance of a GIS during a crisis can be a daunting, and perhaps for many people, an unmanageable task.

The Soufriere Hills Volcano currently (February 2008) poses a sustained threat to the inhabitants of the Caribbean island of Montserrat. The volcano has been in an eruptive state for over 12 years, intermittently ejecting ash and other volcanic debris. The tropical environment with its frequent heavy rainfall mobilises this volcanic material into mudflows, or lahars. These lahars are rapidly moving mixtures of sediment and water that are extremely hazardous to entities in their path. Furthermore on descent, sediment flux through valley erosion and deposition influences lahar rheological (flow) behaviour and also changes the surrounding geomorphology. Recreating these terrain changes should enable the spatial extent of future flows to be predicted.

Lahars are gravitational flows and therefore primarily dependent on topography. However flow behaviour is also influenced by lahar composition at source and sediment flux, and frictional parameters on descent etc. As a first step towards a spatial hazard assessment, topographic data are used to direct flow routing models. More complex parameters are then added to represent such features as vegetation in the area studied. This paper presents initial findings on the difficulties of digitally representing the evolution of terrain when studying active volcanoes, and presents a research agenda for future work. First the methodology will be discussed.

2. Methodology

2.1 Study area

The hazardous nature of lahars is especially relevant in the Belham River valley region of Montserrat where anthropogenic activity is high. In periods of low volcanic activity the valley is used for industrial extraction (sand and gravel), recreational activities and for travel from the north to the south of the island. However it continually acts as a conduit for lahars.

At this stage in the research only the lower Belham valley is being studied in detail, as it incorporates a lot of topographic variation and can be used to test the methods employed.

2.2 Digital elevation model

A representation of the valley surface is the primary GIS layer over which a lahar simulation flows. This takes the form of a digital elevation model (DEM). In late 2006 the active nature of the volcano meant available satellite photographs showed abundant cloud cover and were therefore unusable. Elevation data needed to be gathered in the field through a roving Global Positioning System (GPS). Data were acquired when work in the valley was safely permitted by the authorities. (To detect changes in valley morphology a second and successive DEM will also be required; a DEM for late 2007 is under construction).

Initial analysis revealed the GPS trace (Figure 1) inadequate to reflect micro changes in topography that affect lahar flow direction (Darnell 2007). Thus the topographic data were supplemented by oblique aerial photography and field knowledge. Using ground control points photographs were georeferenced and then channels and vegetation blocks were digitised (as breaklines and polygons respectively). Two alternate representations of the 2006 topography were developed: a primary DEM using GPS points only and a secondary DEM incorporating the supplementary digitised information. Both have a resolution of 6 m.

Next (for each DEM) the ArcGIS module COST WEIGHTED DISTANCE was used to generate a cumulative cost surface. The cost raster was the DEM as constructed above and therefore only had a topographical element. For flow routing the neighbourhood search was limited to a 3 x 3 pixel matrix, resulting in eight possible single flow directions. The SHORTEST PATH module was used to demonstrate the preferential flow path from a considered start point to an arbitrary end point. This was repeated for different start and end point locations.

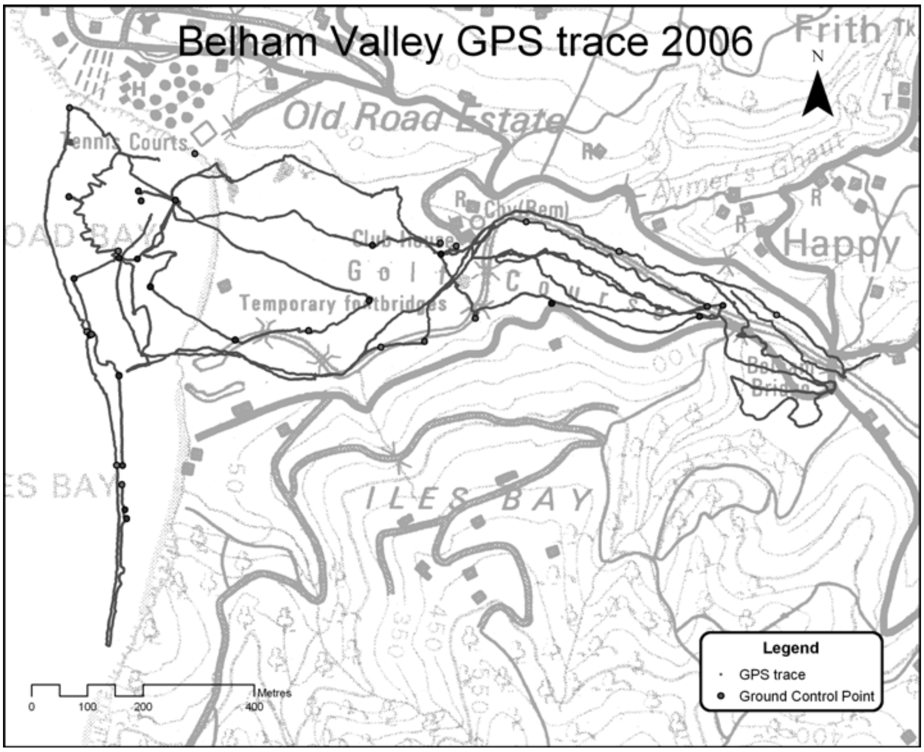


Figure 1. Distribution of GPS data and the availability of ground control points for georeferencing (appear as a line due to point density) (Darnell 2007). The base map

shows the pre-eruptive thematic state of the valley (courtesy Montserrat Volcano Observatory)

3. Initial results

For a given start and end point, the simple single-flow-direction algorithm produced a different output dependent on the input DEM. The different responses of the DEMs are shown by an example in Figure 2. The shortest path given by the secondary DEM (Path B, Figure 2b) is much more direct than that from the primary (GPS only) DEM (Path A, Figure 2a).

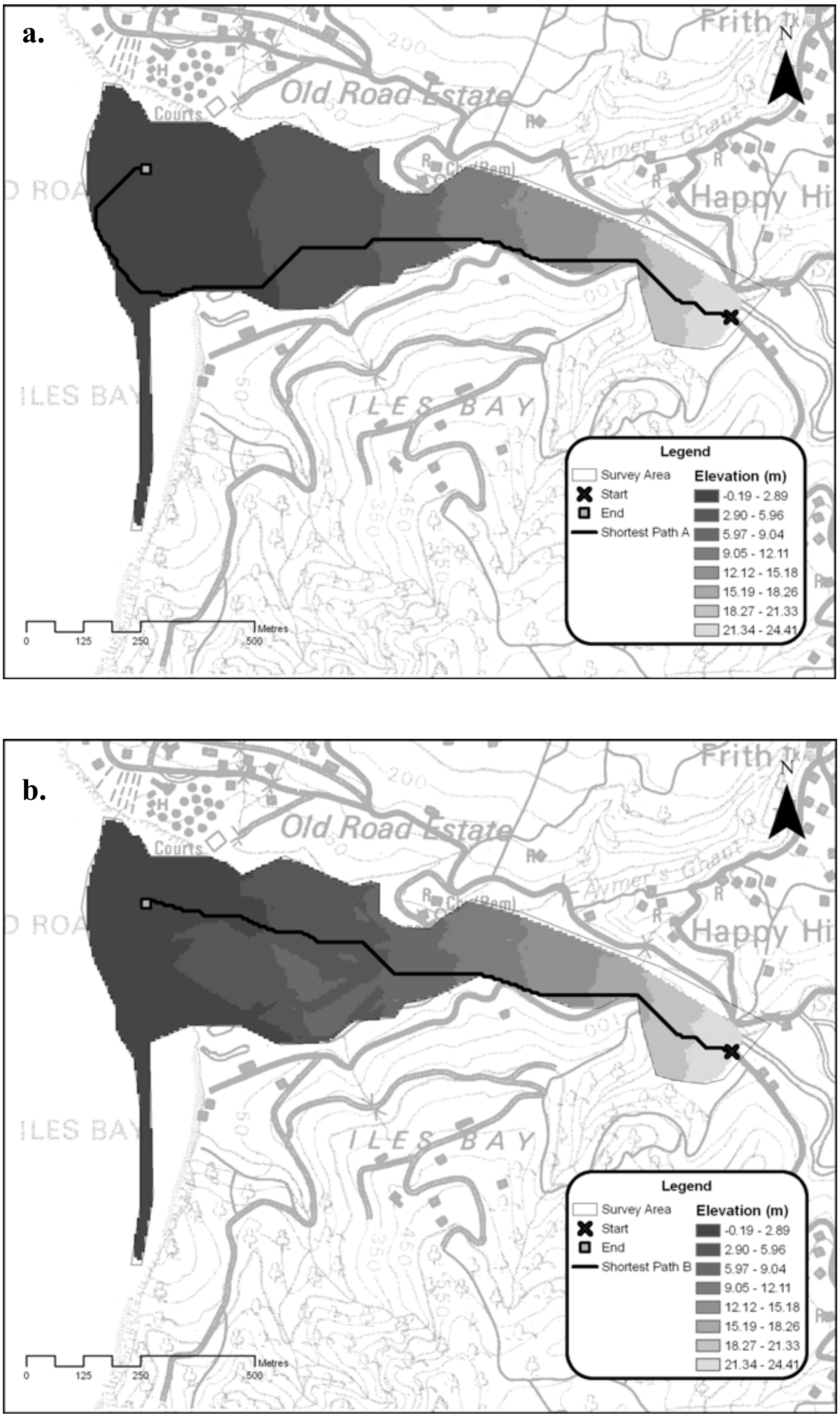


Figure 2. Shortest flow path: a) for Primary DEM, b) for Secondary DEM

The relevance of the additional information can also be demonstrated by overlaying the georeferenced photographs with elevation measurements from *Secondary DEM* and displaying the flow paths (Figure 3). The inconsistency of Path A with field observations is evident when the route intersects a major block of vegetation (as highlighted). Path B from the secondary DEM follows the observed main channel.

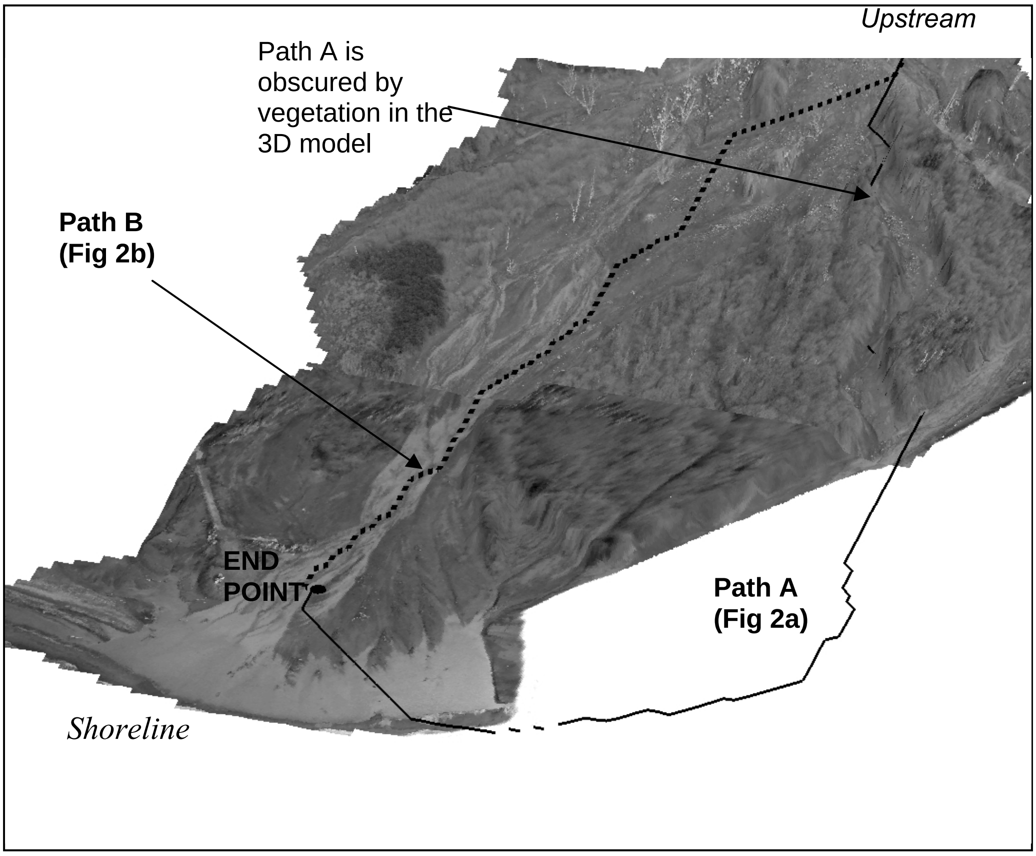


Figure 3. 3D Georeferenced (oblique) aerial photographs of the lower Belham Valley showing the unrealistic nature of Path A

4. Sensitivity analyses

Two digital representations of the Belham Valley surface have been produced and initial analyses of a simple hydrological model has identified the main flow path, thereby suggesting Secondary DEM is more accurate. However, DEMs are approximations of topographic reality and therefore prone to error and uncertainty in the results. Due to the nature of the survey area the GPS rover needed to be carried in a rucksack. This inherently induced some error due to the motion of walking. It was estimated to be a maximum of 0.1 m. Using this figure as a guide GPS sample points that were resolved to a precision greater than 0.1 m were deleted from the data. However each data point is subject to error within this range. Furthermore the interpolation process (as GPS points were converted to Triangular Irregular Network, then to DEM) will have contributed to uncertainty. Currently simple uncertainty analyses are being performed to test DEM sensitivity to error. Propagation to the simple hydrological model will also be studied.

To do this a global measure of accuracy will first be estimated based on the above information. Next ModelBuilder in ArcGIS will be used for simple Monte Carlo-type simulation. This will assess the implications of uncertainty by applying a standard deviation as a measure of the variability in elevation and assuming errors are normal distributed. However this method doesn't consider the spatial dependence between errors (spatial autocorrelation). Adopting a grid-cell methodology similar to Hunter et al (1995) an appropriate pattern of spatial dependence will be found. An adjusted (error) noise file will then be added to the source data. Multiple equally probable realisations of the error (and DEMs) will be output. Simulations can be used to combat uncertainty as follows: (1) to highlight areas of the valley susceptible to changes in parameter values (e.g. elevation or slope); (2) to assess the likelihood of a cell's membership of a particular class (i.e. probability of being part of the main channel); (3) several realisations can be displayed to understand the degree of variation; (4) they can be used compare the two DEMs (after Hunter et al., 1995).

5. Future work

A next step will be to incorporate multiple flow directions and add weighted thematic information to the distance and gradient information provided by the 2006 secondary DEM. The cost raster (resistance) will thus form a friction parameter that can be modified with information on the lahar rheology, bed friction, land cover etc as it varies spatially. Hydrological flow-routing aids our understanding of the valley system but doesn't predict morphological change. For this, field data from November 2007 are currently being analysed. Preliminary observations show morphology of the system has changed significantly over the year due to multiple lahar flows. Lahar modelling will be applied to a 2006 DEM to understand how the landscape has evolved. The product should be similar to a DEM derived from 2007 data (with some consideration of uncertainty). The GIS-based statistical model LAHARZ (Schilling 1998) and the more physically based Titan2D (e.g. Pitman et al 2003) are two key existing lahar models that will be evaluated. Lahars in the Belham Valley are predominantly hyperconcentrated or 'normal' flow (Barclay et al 2004) and thus are dominated by friction and turbulent behaviour. Different parameters for friction will be considered to adapt models to the more dilute flows of the Belham.

6. Summary

In active volcanic environments where quality satellite imagery is not easily obtainable, accurate representations of the terrain surface can be constructed by combining multiple in-field observations such as GPS, oblique aerial photographs and detailed field notes. A DEM with extra digitised information (such as channels and islands of vegetation) yields the best results for a simple lahar flow-routing model. The difference in the flow-routing outputs demonstrates their sensitivity to accurate surface representation. Further sensitivity analyses are in progress.

7. Acknowledgements

The first author is in receipt of an ESRC/NERC studentship PTA-036-2006-00016. She 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) and David Sharp (Secretary to the Governor of Montserrat).

References

Barclay J, Alexander J, Matthews AJ, Johnstone, C, Chivers C, Jolly A, and Norton GE (2004) Lahar activity in the Belham Valley, *Montserrat Report for the Scientific Advisory Committee to the Government of Montserrat* **36**

Darnell AR (2007) Examining the efficacy of GIS for volcanic risk analysis on Montserrat, In Probst F and Kessler C (Eds) *Proceedings of the 5th Geographic Information Days 10-12 September, Muenster, Germany pp177-181*

Hunter GJ, Caetano M and Goodchild, MF (1995) A methodology for reporting uncertainty in spatial database products, *Journal of the Urban and regional Information Systems Association* **7(2)** pp11-21

Pareschi MT, Cavarra L, Favavlli M, Giannini F and Meriggi A (2000) GIS and volcanic risk management *Natural Hazards* **21** pp361-379

Pitman EB, Patra A, Bauer A, Sheridan M, and Bursik M (2003) Computing debris flows and landslides *Physics of Fluids* **15(12)** pp3638–3646

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 efficacy of GIS for volcanic hazard assessment. She is currently in her second year of research under the supervision of Andrew Lovett, Jenni Barclay and Richard Herd.

A Statistical and GIS-Based Approach for Morphodynamic Characterisation and Modelling

A. M. Bakare, J. G. Morley, R. R. Simons

Department of Civil, Environmental and Geomatic Engineering, University College London,
Gower Street, London WC1E 6BT, United Kingdom
Telephone: +44 (0)20 7679 2741
Fax: +44 (0)20 7380 0453
Email: a.bakare@ge.ucl.ac.uk

KEYWORDS: Morphological modelling; Morphodynamic system; Empirical Orthogonal Function; Event-based; Marine GIS.

1. Introduction

The transition from static snap-shot to event-orientated approaches in Geographic Information Science (GISc) (Worboys, 2005; Goodchild *et al*, 2007 and McIntosh and Yuan, 2005) opens up opportunities for morphodynamic characterisation and modelling approaches in coastal engineering. This is due to the dynamic nature of the marine morphodynamic system that evolves at varying spatial and temporal scales in response to differing inputs (hydrodynamic, meteorological and anthropogenic) and process interactions (hydrodynamics and sediment transport). The event-orientated concepts in GISc would enable the inputs and processes that result in a particular output (morphological response) to be captured and applied to further modelling.

Current applications of geographic information science and systems in morphodynamic modelling utilise the "snap-shot" view to:

- Visualise the output of modelling procedures.
- Calculate the height differences between raster representations of the seabed depth and derive approximate seabed volumetric changes.
- Digitise vector information from the seabed raster surfaces.

There are a number of existing morphodynamic assessment approaches in coastal engineering that aim to characterise the morphodynamic system in terms of the "cause-response" relationships. These can broadly be categorised into bottom-up or top-down concepts, which differ in how they account and represent the system processes. The limitation with the bottom-up methods is the system behaviour and process interaction is inferred, so it is only feasible for short-term (hours to days) and small spatial areas (unit kilometre squared) (De Vriend, 1991a; De Vriend, 1991b and De Vriend *et al*, 1993). In the top-down case the behavioural trends are derived for each system factor, however it is not always possible to correlate the "cause-response" relationship between the input and output variables (Reeve *et al*, 2001 and Wijnberg and Terwindt, 1995). Therefore, there is a gap in morphodynamic characterisation approaches to address spatial areas of up to tens of kilometres at multi-decadal time-scales.

This paper presents a proposed morphodynamic characterisation and modelling methodology to address the described gap by utilising the event-orientated concepts from GISc with existing morphological assessment methods.

2. Methodology

The methodology presented utilises event-orientated concepts to identify the system inputs and processes that result in specific morphological response.

The functionality of the geographical information system and the morphological modelling approach is demonstrated in Figure 16. The spatial and temporal behaviour are derived for each system factor using a morphological modelling approach (1). The derived behaviour is extrapolated to calculate the morphological response at a future time step based on forcing conditions (hydrodynamic and meteorological influences) (2). The modelled response is validated against a real-world case by modifying associated weightings to the forcing conditions (3). The "cause-response" relationship is then determined by systematically varying the forcing conditions, associated weighting and the spatial and temporal averaging and aggregation of the input forcing parameters to identify characteristic morphological zones (4). The identified seabed locations that have similar responses to the varying forcing influences are spatially aggregated (5) and form the inputs into a rule-based morphological prediction procedure (6 to 8).

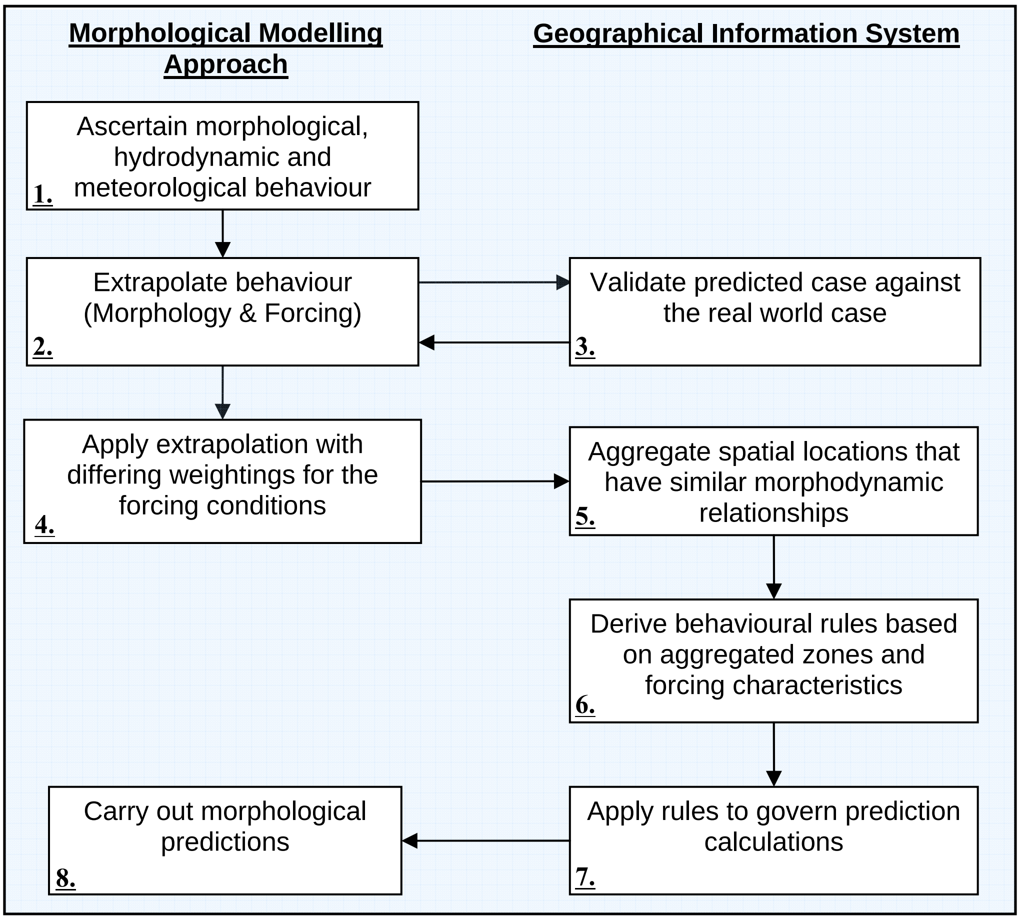


Figure 16: Proposed morphodynamic characterisation and modelling methodology.

2.1. Determining Morphodynamic Behaviour

The system behaviour is extracted from datasets of each factor using an Empirical Orthogonal Function (EOF) approach. The analysis method derives the dominant spatial and temporal eigenfunctions that account for the largest proportion of variance in the dataset (Larson *et al*, 2003 and Miller and Dean, 2007). The method has been widely used in coastal morphology analyses to identify characteristic patterns, often interpreting these in relation to the physical morphology (Wijnberg and Terwindt, 1995 and Miller and Dean, 2007). The application of the EOF method is this research involves separating the morphological data into spatial and temporal eigenvectors and associated weightings which can be linearly combined to recreate the original data based on equation 1 (Miller and Dean, 2007).

$$y(x,t) = \sum_{k=1}^{n} a_k c_k(t) e_k(x) \tag{1}$$

$$a_k = \sqrt{\lambda_k n_x n_t} \tag{2}$$

Where $e_k(x)$ is the spatial eigenvector, $c_k(t)$ is the temporal eigenvector or coefficient, a_k is determined by equation 2 and λ_k is the eigenvalue associated with the k^{th} eigenvector (where the spatial and temporal eigenvectors have the same eigenvalue), n is the lesser of n_x or n_t which are the count of spatial and temporal element respectively.

2.2. Characterising Morphodynamic Behaviour

The purpose of the validation step is to determine the appropriate morphological change function across the seabed (step 3, Figure 16). The objective of the characterisation step is to ensure that only the forcing climate that is applicable to an aggregated spatial zone is included in the change function for that zone when carrying out morphological predictions (steps 4 to 6, Figure 16).

2.2.1. Validation and Characterisation

A Geographical Information System (GIS) will be used to carry out the prediction validation (step 3, Figure 16) and enable the morphodynamic characterisation process in determining the appropriate morphological change functions for the morphological zones (steps 4 to 6, Figure 16). The system will facilitate validation by interpolating the calculated profiles into a surface and comparing the predicted output against the observed case. It will be used to highlight the distribution of the modelled difference, showing the locations where the prediction was closer to reality, due to lower differences. Future development will be to assess and determine how the calculated difference will act as a guide to modify the parameters from the forcing factor. The characterisation process would involve systematically varying the included forcing conditions and weightings to observe the locations that varied and to what degree, from which "cause-response" relationships will be established between the morphological and forcing factors.

2.2.2. Prediction

The derived behaviour and association between the system factors is used to derive a rule-based morphological model which specifies the range of conditions that influences a particular area of seabed (steps 7 and 8, Figure 16).

2.3. Study Area and Data

A study area located off the East Anglian coast between Winterton-on-Sea and Orford Ness and extends fifty kilometres offshore was selected (Figure 17). It is bounded by WGS84 geographic grid coordinates 52.763N, 1.673E and 52.005N, 2.397E for the northwest and southeast corners respectively. The area forms an important navigation route into the ports at Great Yarmouth and Lowestoft and is made up of a number of bank and channel systems that are known to be mobile (UKHO, 2002, 2003 and 2004). The bathymetric dataset comprised sixteen digitised bathymetric survey charts between 1846 and 1992, which formed the analyses set and a side-scan sonar survey output from 2002 which was used for validation.

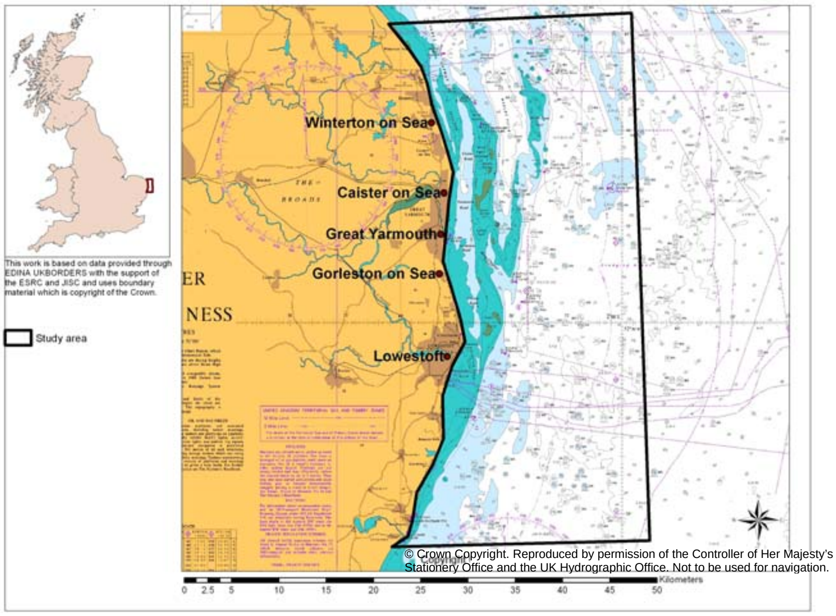


Figure 17: Study area.

3. Implementation Tests and Results

In order to investigate the applicability of the proposed methodology, initial experiments were undertaken as described in steps 1, 2 and 3 (Figure 16) for morphological data only. The purpose was to explore the capability of identifying morphological behaviour and predicting future conditions. As well as demonstrate the need to include forcing influences in morphological prediction calculations.

The temporal mean, range, minimum, maximum, standard deviation and coefficient of variance of the seabed depth were calculated for each cell in ArcGIS 9.1™, to derive an overview of the morphological behaviour within the study area (Reeve *et al*, 2001). EOF analyses undertaken in MATLAB™ were used to decompose the morphological data into spatial and temporal trends from which the seabed was reconstructed using equation 1 based on the largest eigenfunctions. To predict future morphological states the identified temporal eigenvectors were extrapolated based on a best-fit line. The coefficient derived from the extrapolated temporal trend was utilised in equation 1 to predict the bed depth at a future time step.

The results of the summary statistics for the study area support the local observation that the banks and channel systems are mobile (Figure 18 A to C). The largest change with a height range of up to thirty-three metres primarily occurs in the offshore banks and channels. The least change of approximately zero to five metres occurs closer to the shoreline (Figure 18 A). A similar distribution pattern is observed between the range and standard deviation results, which indicate that the offshore bank and channels are more variable than the nearshore features. This is possibly due to the sampling density of the original survey and the digitisation and interpolation process. Alternatively the offshore bank and channel have a differing dynamic equilibrium phase to the nearshore features, which do not coincide with the survey periodicity.

The results of the EOF analysis and extrapolation prediction for the morphological data indicate that the observed eigenfunctions (spatial and temporal) which describe the largest proportion of variance in the data identifies the cross-shore profile shape for year 2002 (Figure 19). However a root-mean-square error of 2.88 (m) indicates that there is some difference between the predicted and observed bed height, which is also demonstrated in Figure 19. This leads to the assertion that although the morphological factor has a spatial behaviour, the forcing conditions and stochastic events also contribute to morphological evolution.

4. Conclusion

The morphodynamic environment is known to evolve at varying spatial and temporal scales. The objective of the current research is to spatially and temporally characterise morphological responses to hydrodynamic and meteorological conditions in order to enable morphological prediction at multi-decadal temporal and regional spatial scales. The event-orientated approach in GISc presents opportunities to identify the forcing influences which determine a specific morphological response and can be applied to rule-based prediction calculations.

The first step of capturing morphological behaviour using morphological modelling approaches indicate that it is not sufficient to predict future morphological states based on observed morphological trends only. Instead there is a need to include the hydrodynamic influences and use the relationship between the variables to predict future states.

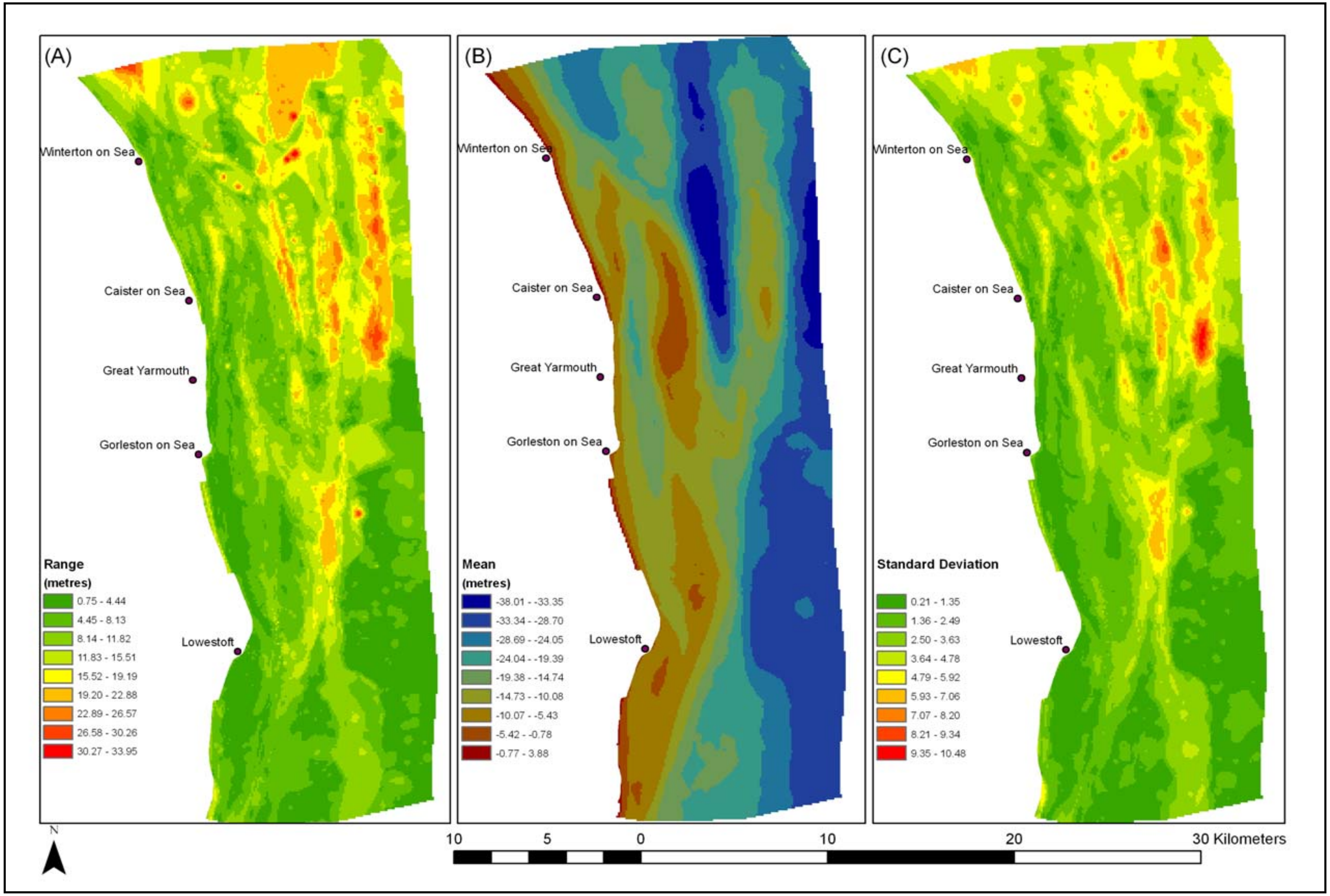


Figure 18: Results of the temporal a) range, b) mean and c) standard deviation statistics assessments for each grid cell for the 1846 – 1992 datasets.

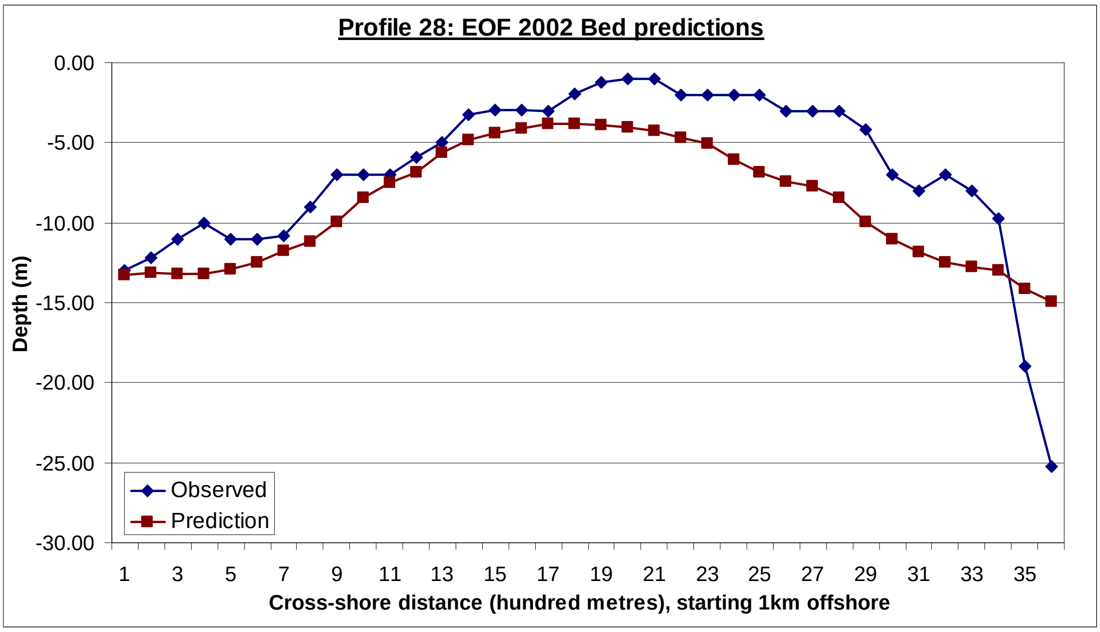


Figure 19: Comparison of the observed and predicted profile for the year 2002 derived using the EOF method for profile 28.

5. Acknowledgements

This research project is funded by the Engineering and Physical Sciences Research Council and their support is appreciated. Thanks to the Association of British Ports Marine Environmental Research for the digitised bathymetric datasets and the Maritime and Coastguard agency for the 2002 bathymetry sonar data.

6. References

De Vriend, H. J. (1991a), Mathematical Modelling and Large-Scale Coastal Behaviour: Part I. Physical Processes. *Journal of Hydraulic Research,* 29**,** 727-753.

De Vriend, H. J. (1991b), Mathematical Modelling and Large-Scale Coastal Behaviour: Part II. Predictive Models. *Journal of Hydraulic Research,* 29**,** 741-753.

De Vriend, H. J., Capobianco, M., Chesher, T., Swart, H. E. d., Latteaux, B. & Stive, M. J. F. (1993), Approaches to Long-Term Modelling of Coastal Morphology: A Review. *Coastal Engineering,* 21**,** 225-269.

Goodchild, M. M., Yuan, M. & Cova, T. J. (2007), Towards a general theory of geographic representation in GIS. *International Journal for Geographical Information Science,* 21**,** 239-260.

Larson, M., Capobianco, M., Jansen, H., Różyński, G., Southgate, H. N., Stive, M., Wijnberg, K. M. & Hulscher, S. (2003), Analysis and Modelling of Field Data on Coastal Morphological Evolution over Yearly and Decadal Time Scales. Part 1: Background and Linear Techniques. *Journal of Coastal Research,* 19**,** 760-775.

McIntosh, J. & Yuan, M. (2005), Assessing Similarity of Geographic Processes and Events. *Transactions in GIS,* 9**,** 223-245.

Miller, J. K. & Dean, R. G. (2007), Shoreline Variability via Empirical Orthogonal Function Analysis: Part I Temporal and Spatial Characteristics. *Coastal Engineering,* 54**,** 111-131.

Reeve, D., Li, B. & Thurston, N. (2001), Eigenfunction Analysis of Decadal Fluctuations in Sandbank Morphology at Gt Yarmouth. *Journal of Coastal Research,* 17**,** 371-382.

The United Kingdom Hydrographic Office (2002). *Report of Survey, Holm Channel.* Assessment EA9/2002.

The United Kingdom Hydrographic Office (2003). *Report of Survey, Holm Channel.* Assessment EA9/2003.

The United Kingdom Hydrographic Office (2004). *Report of Survey, Holm Channel.* Assessment EA9/2004.

Wijnberg, K. M. & Terwindt, J. H. J. (1995), Extracting Decadal Morphological Behaviour from High-Resolution, Long-Term Bathymetric Surveys Along the Holland Coast Using Eigenfunction Analysis. *Marine Geology,* 126**,** 30.

Worboys M (2005), Event-orientated approaches to geographic phenomena. *International Journal of Geographical Information Science,* 19 (1): 1-28.

Biography

Anna-Marie Bakare is a second year PhD in the department of Civil, Environmental and Geomatic Engineering, UCL. She holds a BSc (Hons) in Archaeology from UCL and an MSc in Geographic Information Science from UCL. Her interests are in marine GIS implementation and morphological modelling.

Jeremy Morley is a lecturer in Geographic Information Science at UCL. He obtained an MA in Natural Sciences from Cambridge University and an MSc in Remote Sensing from UCL. His interests are in the integration of remote sensing and scientific modelling with GIS.

Richard Simons is a Professor of Fluid Mechanics and Coastal Engineering at UCL. Having completed his Bachelors degree at Manchester University, he worked for Norwest Holst before joining UCL- initially as a PhD student and later as a lecturer. His particular interests are in wave-current-sediment interactions in coastal environments.

Surveying by motion tracking: modelling 3D subterranean landscape from video imagery

Anna E Mason[1], David M Mountain[2], Jonathan Raper[3]

[1,2,3] City University, Northampton Square, London, EC1V 0HB
Tel. +44 (0)20 7040 5060
Email. a.e.mason@city.ac.uk, dmm@soi.city.ac.uk, raper@soi.city.ac.uk

KEYWORDS: GIS, 2D3 Boujou, Camera Tracking, Video Imagery, Caving

1. Introduction

Computer Vision is concerned with obtaining information from image data, such as the analysis of 2D image sequences to identify and represent real world objects in 3D space. As stated by Hartley and Zisserman (2000) Computer Vision builds upon knowledge from physics, the psychology of perception and the neurosciences, but the strength of achievement rests undoubtedly upon the ability to describe complex geometric concepts and translate them into computer code. The last 15 years of substantial research on motion tracking systems has produced Augmented Reality (AR) applications ranging from robotic navigation, computer assisted surgery, augmented reality, virtual TV production and creation of special effects in the movie making industry (Welch and Foxlin, 2002; Bhanu and Paulidis, 2005). In the last seven years commercial vision-based camera tracking systems, known interchangeably as matchmoving systems, have emerged which utilise the geometry of multiple uncalibrated views to generate digital 3D representations from 2D video or film footage (Hartley and Zisserman, 2000). A synergy of this technology with GIS could provide a significant step forward in the efficiency of three dimensional data capture and modelling techniques.

One area where this could prove particularly beneficial is surveying and modelling subterranean cave and karst systems, where standard positioning systems such as Global Positioning Systems (GPS) are unavailable and surveying techniques are typically laborious or require costly laser scanner and range finder equipment. Given the challenging nature of caves, underground and confined with no access to infrastructure or view of the sky, these environments will not benefit from the investment going into the next generation of Global Navigation Satellite Systems (GNSS) such as the Galileo (European Commission, 2007) or wireless positioning systems. In this context, survey techniques that could use handheld cameras are worthy of investigation. The emphasis of this research is to present a method for using matchmoving or camera tracking software to perform 3D reconstruction of a scene and to evaluate how this technology can be applied to surveying subterranean environments.

2. Video Data Acquisition in the Field

The experiments were conducted in three UK study areas; one chalk mine and two caves formed by dissolution of limestone. Sections of each mine or cave were surveyed using conventional line survey methods and equipment (eg. measuring tape and compass) as illustrated in Figure 1.

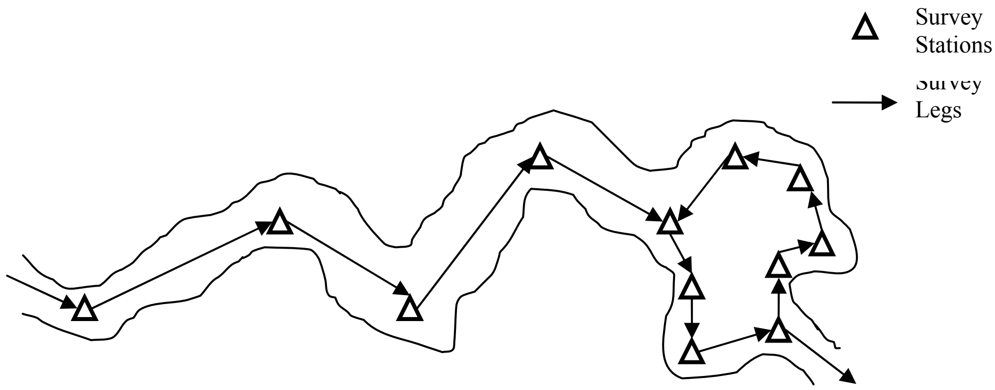


Figure 1. Line Survey. Modified from Day (2002)

The survey centreline of each passage was then filmed using portable video camera and lighting equipment. The lighting was built using a wide angle 50W halogen bulb and torch casing, connected to a lead acid battery. PYSOP markers (2d3a, 2005) were used to create recognisable features for analysis. By identifying key issues relating to technique, equipment, environment and location, each experiment contributed to systematically refine and improve the data acquisition methodology. The process is summarised in Figure 2.

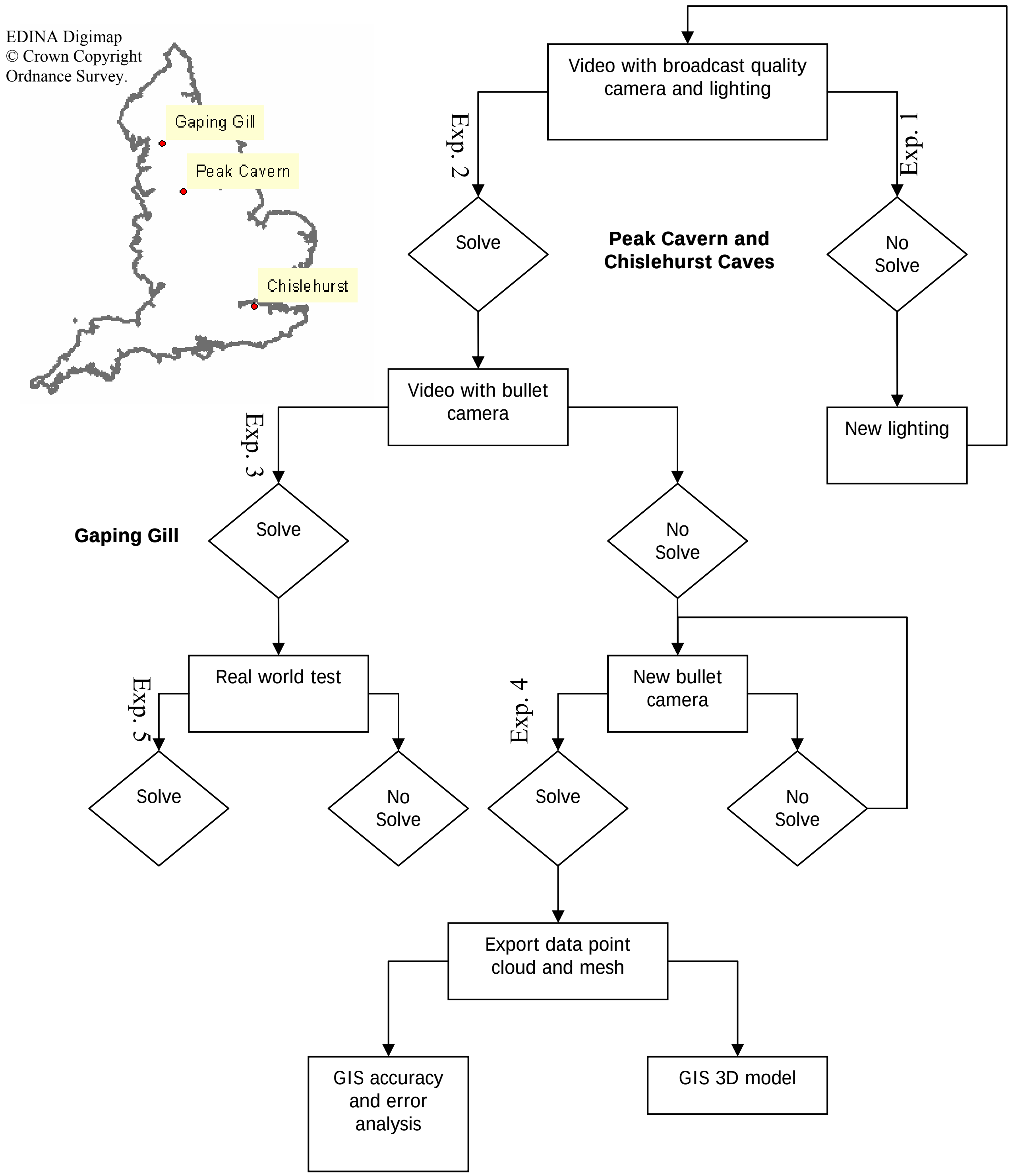
3. Three Dimensional Reconstruction and Modelling

Video image sequences were sectioned to represent survey legs of up to 40m in length. Each image sequence was processed with Boujou4 (2d3, 2007) and evaluated for quality using standard matchmoving guidelines, such as visual cues and reprojection error, as outlined by authors such as Dobbert (2005).

Matchmoving software performs analysis of raw 2D imagery to produce optimal estimates of camera movement and rotation. This is achieved by tracking 2D points of interest as they move in the frame through time (Figure 3a). Selection is based on an algorithm such as corner detection in Boujou4 (2d3, 2005b) or edge detection as in Kemp and Drummond (2004). Corner detection is likely to be a more appropriate algorithm in organic environments as there are few straight lines or edges to match that will not change radically with aspect and illumination. Also, Kemp and Drummond (2004) suggest that high density of similar features (like repeating rock texture) is not beneficial to edge detection.

2D feature points are then matched between frames using a 3D pose estimation algorithm which is often Random Sample Consensus (RANSAC) as in Nister et al. (2006), Mouragnon et al. (2006), Kemp and Drummond (2004), and Zhang et al. (2006). 3D pose estimation is described by Fisher et al. (2005) as determining the transformation of an object in one coordinate frame with respect to the same object in another coordinate frame. The software creates a virtual camera and estimates the camera parameters, movement and rotation in relation to the 2D feature tracks (Dobbert 2005; Fitzgibbon and Zisserman 2003). This process is called 3D calibration. This information is used to calculate the depth of objects in a scene, producing a 3D point coordinate cloud known as a "camera solve" (Figures 3b and 3c). In addition, a surface mesh can be generated from the coordinate point cloud using a form of delaunay triangulation (Figures 4a, 4b and 4c).

Figure 2. Summary of the data acquisition and 3D reconstruction workflow



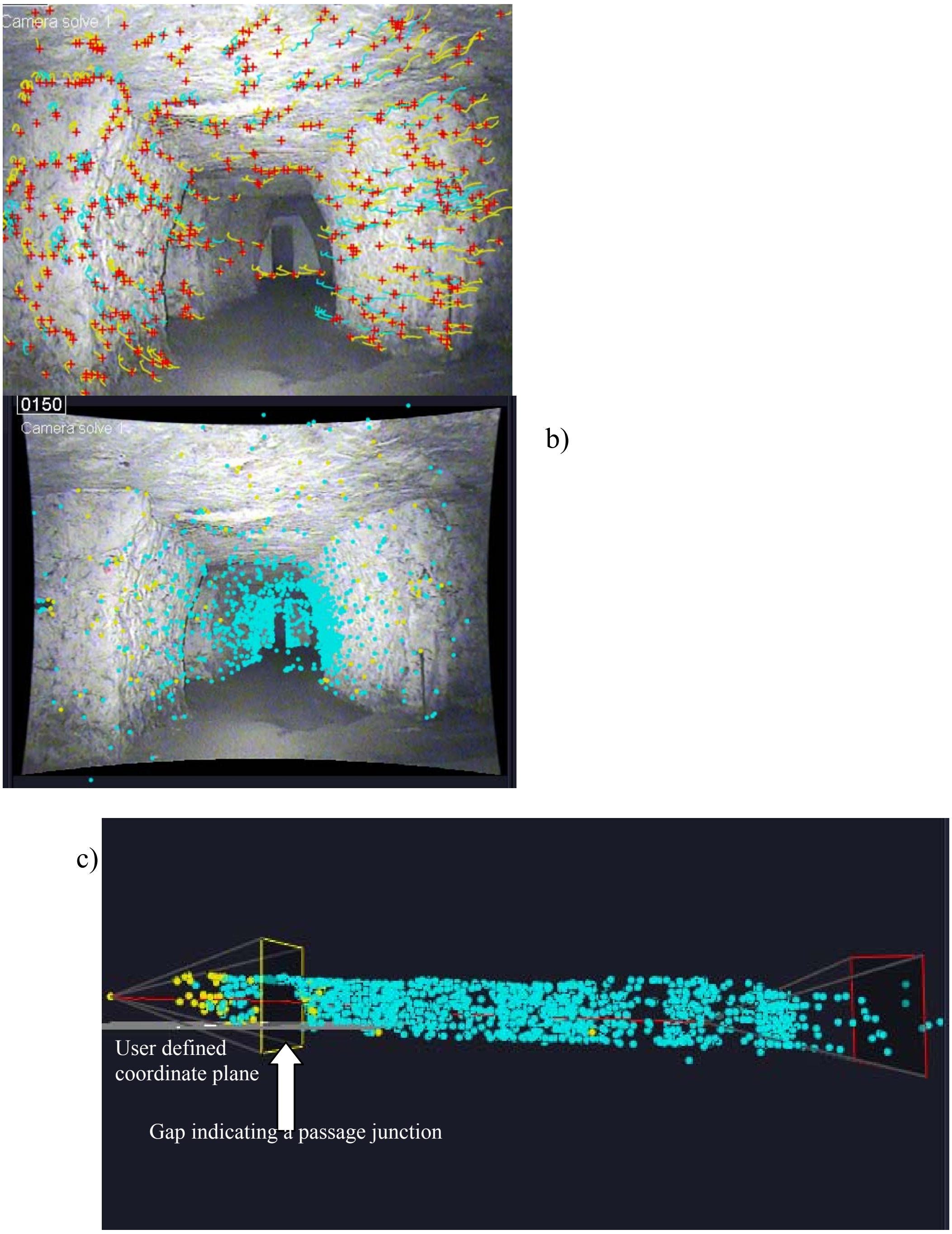


Figure. 3 a) 2D view of 2D feature tracking. The red crosses are points of interest and the yellow vectors are their tracks in previous frames. b) 3D points generated by a successful matchmoving "camera solve". c) 3D side view of 3D points generated for a survey leg of cave passage. The red line through the centre of the points is the path the camera travelled. The yellow square is the camera's projected Field of View.

In Figure 4a the cave surface appeared angular. CAD, GIS and animation tools were tested to improve appearance, including use of photorealistic lighting and materials (Figure 4c). Alternative GIS and CAD modelling approaches, such as lofting between vertical sections or creating a rectangular surface grid, were also possible using matchmove data.

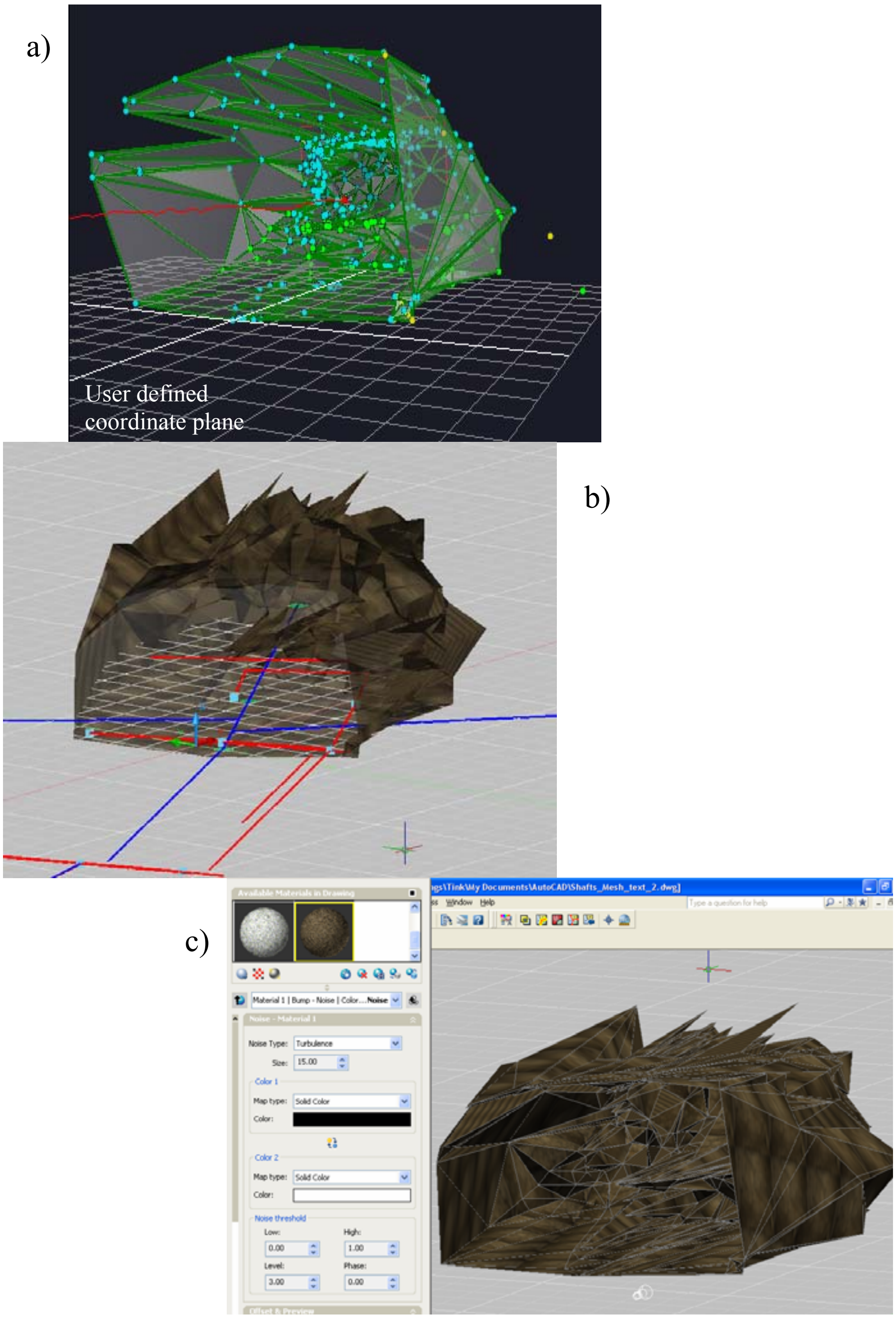


Figure 4 a) 3D view of a mesh in Boujou4. b) The same mesh mapped to the line survey results (red and blue lines) in GIS. c) Lighting and colour created from a single frame of video.

4. GIS Analysis and Modelling

GIS/CAD tools were used to assess the results visually and quantitatively. Detailed assessment of positional accuracy was precluded by insufficient control data to adequately establish a 3D model for comparative analysis. Existing LIDAR was incomparable as it mapped large main chambers (Westerman et al., 2003; Murphy et al., 2005), however, potential exists to generate 3D digital representations of entire connected systems by merging LIDAR data of main chambers with the matchmove data of interconnecting passages.

4.1 Three Dimensional Model

Survey results from Chislehurst were used to create a 3D model connecting surface and subsurface features.

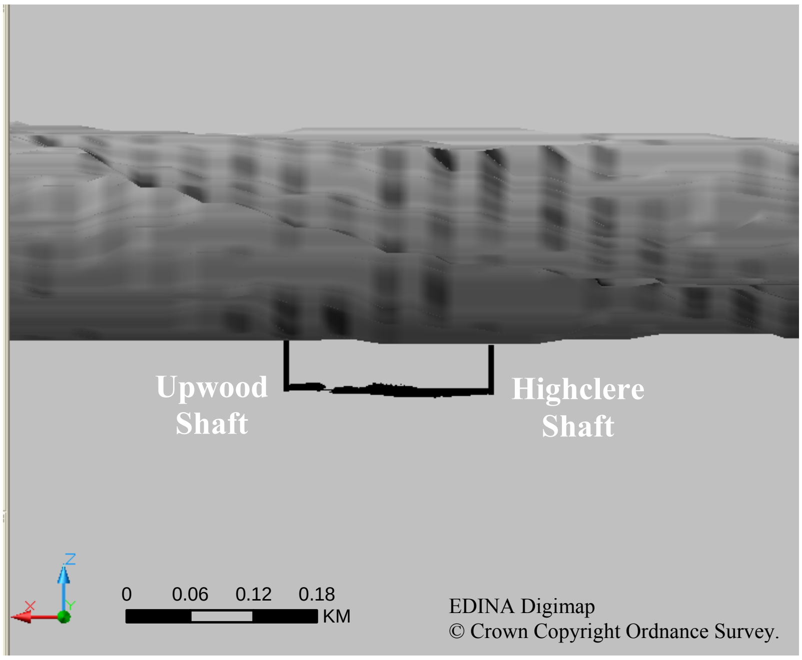


Figure 5. The mine passage connecting Upwood Shaft to Highclere Shaft at Chislehurst, Bromley

4.2 Scale

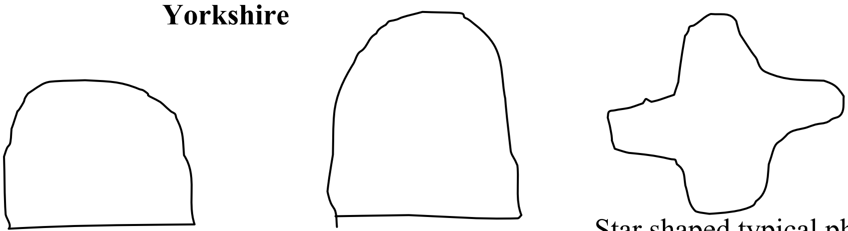
Relative scale accuracy was tested on a sample coordinate point cloud using a small sample of ground truthed measurements. The scale differences ranged from 0.2cm to approximately 20.3cm. Mean of the four measurements was 13.3cm and RMS deviation from the average was 8.1. When considered in proportion to height and width, these measurements were found to fit well within 10-15% guideline suggested by Day (2002) for sketched cave profiles.

4.3 Shape

Figure 6 shows a sample of 3D polyface meshes sectioned and analysed for accuracy of shape. The interior of each cross section is comparatively similar to the field sketches. Gaping Gill appears to be the least representative in terms of shape. This pattern is explained by the appearance of rocks inside the passage.

Club Rooms, Chislehurst Air Shafts, Chislehurst Gaping Gill, North Yorkshire

Field Sketches

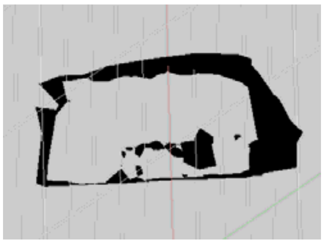


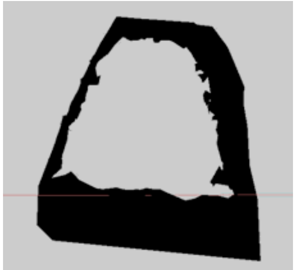
Shallow half ellipse

Arched ellipse

Star shaped typical phreatic cave morphology (Lowe and Waltham, 2002)

Vertical Cross Sections





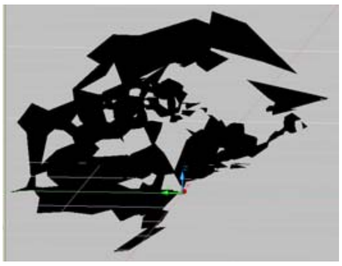


Figure 6. Cave survey sketches compared to 3D polyface meshes generated using Boujou4

5. Discussion and Conclusions

This research showed that matchmoving programs are sufficiently robust to track cave passage scenes using consumer quality portable cameras and lighting. Current research points in the direction that this technology is moving; towards faster, more accurate, real time processing, with less user input required to achieve 3D reconstruction of real scenes. The commonality of motion tracking applications is that they effectively perform digital surveying tasks of real scenes. This research demonstrates how this digital survey data can be used in GIS, thus presenting an example of a new compatible source for 3D data. The data acquisition and processing method appears particularly well suited to challenging environments where other forms of surveying may be less accessible, such as cave systems. Yet if this method was adopted as a mainstream form of surveying, it would be applicable wherever anyone has access to a video camera of reasonable resolution, any place and at any time.

Initial results and observations indicate that where good 3D calibration has been achieved, relative scale and shape are shown to be accurate, although the technique and sample sizes presented here are insufficient to give more than approximate quantitative measures. Even if the optimal pose estimation nature of monocular vision tracking systems limit the process from achieving accuracy comparable to professional surveying equipment, the process still demonstrates potential for many applications where less spatial sampling and accuracy may be appropriate to cover large areas effectively, such as cave surveying, navigation, and search and rescue. Actual comparison of such surveying techniques presents an important opportunity for further study.

References and Citations

2d3 Boujou4 (2007) http://www.2d3.com/ (accessed 20/12/07)

2d3 (2005a) What sort of tracking markers does boujou track the best? [Online] http://www.2d3.com/php/support/faqs.php (accessed 20/12/07)

2d3 (2005b) Boujou Three Help - Tutorial 1: Basic Tracking [Offline Electronic Manual]

Bhanu, B. and Pavlidis, I. (2005) *Computer Vision Beyond the Visible Spectrum: Advances in Pattern Recognition*, London, UK: Springer.

Day, A.J., (2002) *Cave Surveying – A guide to the equipment, techniques and methodology of the BCRA system*, Cave Studies Series No.11. Buxton, UK: British Caving Research Association (BCRA).

Dobbert, T. (2005) *Matchmoving: The invisible art of camera tracking*, Sybex Inc., Alameda, CA.

European Commission (2007) EU Galileo Project [Online] http://ec.europa.eu/dgs/energy_transport/galileo/index_en.htm (accessed 20/12/07)

Fisher, R. B., Dawson-Howe, K., Fitzgibbon, A., Robertson, C., and Trucco, E. (2005) *Dictionary of computer vision and image processing*, UK: Wiley

Fitzgibbon, A. and Zisserman, A. (2003) Automatic Camera Tracking. In Shah, M. and Kumar, R. (eds), *Video Registration,* UK: Kluwer, pp18-35

Hartley, R. and Zisserman, A. (2004) *Multiple View Geometry in Computer Vision (Second Edition)*, Cambridge, UK: Cambridge University Press.

Kemp, C. and Drummond, T.W. (2004) Multi-Modal Tracking using Texture Changes, *In Proceedings of the British Machine Vision Conference*, pp73–77

Mouragnon, E., Dekeyser, F., Sayd, P., Lhuillier, M., and Dhome, M. (2006) Real Time Localization and 3D Reconstruction, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 1, pp363 – 370

Murphy P., Parr A., Strange, K., Hunter, G., Allshorn S.L., Halliwell, R.A., Helm, J., Westerman, A.R. (2005) Investigating the nature and origins of Gaping Gill Main Chamber, North Yorkshire, UK, using ground penetrating radar and lidar, *Cave and Karst Science*, Volume 32, pp25-38

Nistér, D., Naroditsky, O., Bergen, J. (2006) Visual odometry for ground vehicle applications, *Journal of Field Robotics,* Volume 23(1), pp3–20

Welch, G. and Foxlin, E. (2002) Motion Tracking: No Silver Bullet, but a Respectable Arsenal, *IEEE Computer Graphics and Applications,* Volume 22(6), pp24-38

Westerman, A.R., Pringle, J.K., Hunter, G. (2003) Preliminary LIDAR survey results from Peak Cavern Vestibule, Derbyshire, UK. *Cave & Karst Science*, Volume 30(3), pp129-130

Zhang, G., An. X; Chen, W; Bao, H., Qin, X. (2006) As-consistent-As-possible compositing of virtual objects and video sequences, *Computer Animation and Virtual Worlds*, 17(3-4), pp305-314

Acknowledgements

Sincere thanks to the British Caving Research Association for funding and the resources to explore the project in such depth. I am equally very grateful to 2d3 for generously providing Boujou4. Thanks to John Farrer of Peak Cavern and Terry Hunt of Chislehurst Caves, for allowing access to the caves. Thank you to Rod LeGear, Kent Underground Research Group, for his help at Chislehurst Caves. Thanks to Shaun Puckering of Hitch n Hike, Hope Village, for his advice on building caving lamps and for lending his cave surveying gear, and to Adam Evans of Outdoor Instruction, for his guidance at Peak Cavern. Thanks to the Craven Potholing Club for making Gaping Gill such an inspiring experience. Thank you to Paul Chambers and Sarah Mason for surveying and AV assistance.

Biography

Anna E Mason is an MSc GIS graduate from City University. Her undergraduate degree is from Canterbury University, New Zealand. Anna currently works as a GIS Developer and is an operational volunteer for MapAction. Research interests include applications for graphics and animation technology in GIS.

Dr David M Mountain is a research fellow on the LOCUS project and lecturer in Geographic Information at City University. Research interests include location-based services, modelling individual accessibility, modelling time and space, virtual and augmented reality environments and mobile computing.

Prof Jonathan Raper is Head of the Information Science Department at City University. Previously he was at Birkbeck College, University of London. Jonathan has led projects and published research on location-based services, GI policy, philosophy of spatial and temporal representation, and the handling and analysis of environmental information.

A 3D Scrapbook Approach to Geospatial Analysis

S. Wang and S. Brooks
suwen@cs.dal.ca sbrooks@cs.dal.ca
Dalhousie University, Halifax, Nova Scotia, Canada
Telephone: +(1) 902-494-2512

1. Introduction

We physically relate to our environment in three dimensions. However, familiar modes of interaction reside in 2D, which mimic our interactions with traditional paper-based media. In this work, our aim is to combine the situational understanding that can be gained from 3D GIS, with familiar paper-based interaction. Our system situates 3D terrains within a metaphor of a 2D interactive scrapbook or journal. This allows the geospatial professional to maintain 3D records of their analysis over time and perform calculations directly over extracted terrain regions. In addition, the system would be useful as a teaching aid imparting an understanding of the third dimension.

2. Related Work

A number of systems have explored the third dimension including Terrafly (Rishe et al., 1999), GeoZui3D (Ware et al., 2001) and ArcGIS (ESRI, 2006), amongst others. This typically takes the form of 2.5D height fields. However, a review by Stota and Zlatanova (2003) examined the present state of 3D GIS, which came to the conclusion that 3D GIS is merely at a stage where 2D GIS was several years ago. The review noted that in some cases 3D is simply used for graphic illustrations or the rendering of computer generated fly-bys.

Recently, a hybrid approach has been developed that combines 2D and 3D elements into a unified interface (Brooks and Whalley, 2007). This system allows multiple layers of information to be continuously raised or lowered by the user, directly over the base-terrain. During elevation, the layers morph between a 3D terrain and a flattened 2D map which helps mitigate issues of terrain and data occlusion in 3D. However, we argue that further possibilities should be explored that combine 2D interaction with 3D GIS.

3. Selection and Extraction of Terrain Regions

Our prototype allows the user to interactively extract and retain portions of the 2.5D terrain, which become incorporated into an editable journal. In order to proceed, the user must specify the regions that are to be extracted from the background terrain and subsequently embedded into the scrapbook. From the user's point of view the process is quite simple: the user specifies the location of a circular or rectangular region by directly clicking on a point on the 3D surface.

This is best understood visually. In Figure 1, there are two main areas with the scrapbook on the left and the background terrain on the right. At this stage the user has already clicked on a location on the background terrain and this area has become embedded into the scrapbook.

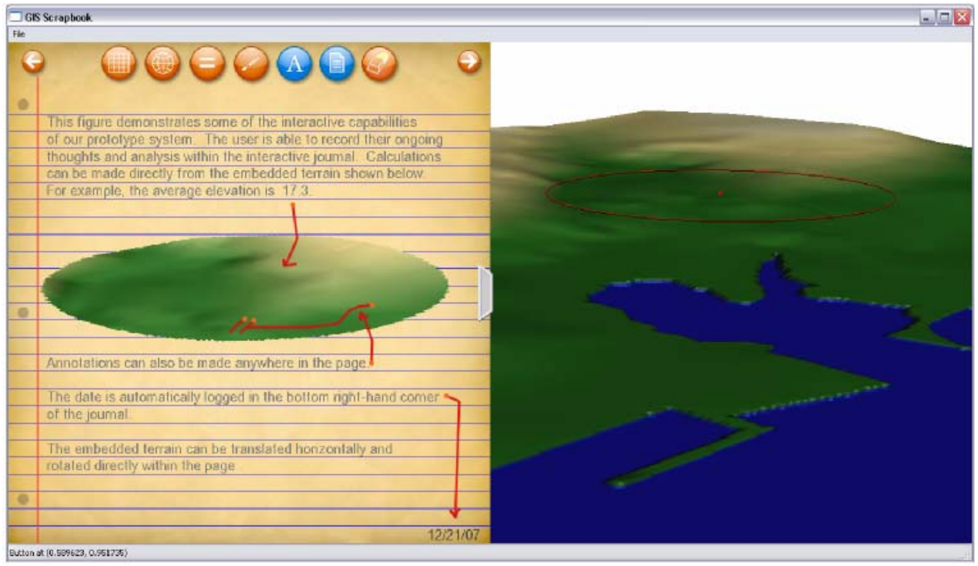


Figure 1: Journal interface with scrapbook partially occluding the 3D terrain on the left half of the screen. The scrapbook page incorporates an embedded terrain, annotations, written notes and computed equations.

However, when the user clicks on a point in the display window, this does not specify a single point in 3D space because of the scene's 3D depth. The user is in effect casting an imaginary ray that extends from the point on the display outwards into the 3D scene. From this, the system must determine which point on the ray intersects with the terrain surface, which is the location the user intended to pick. As the system is developed with the OpenGL rendering library, we utilize use the *gluUnproject* function which performs a reverse projection from the 2D display space to 3D. Using this function we can trace a mouse click from a point on the screen to the target 3D point on the terrain surface.

Moreover, the user may wish to specify any region on the terrain, and not just one that happens to be visible at present. This requires additional camera controls that allow the user to move to any location prior to specifying the selection region. These controls are typical of any virtual environment (using a combination of keyboard and mouse interaction), and so we will not discuss them in detail here.

4 Further Interactive Facilities

In addition to the primary function of extracting portions of the terrain, our system also provides a number of interactive features that support the recording of continuing analyses. These include facilities for modifying the embedded terrain region, as well as text entry and annotation.

Even after a region of the terrain has been placed in the scrapbook, the user retains some degree of interactive control over the embedded area. The user can continuously slide the center of the selected region to new locations by tracing the mouse over the surface of the 3D terrain, and the content of the embedded terrain updates in real-time to match. In addition, the selected region can be interactively rotated by dragging over the embedded terrain region itself, as shown in Figure 2.

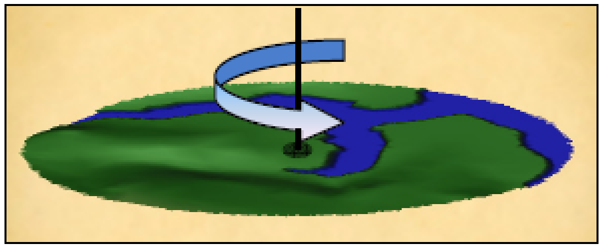


Figure 2: Interactive rotation of the embedded terrain region.

A drawback of the scrapbook interface is that it occupies half of the available screen space. To address this we have incorporated a simple hide/show control that slides the scrapbook away to the left side of the display. The control is shown in Figure 3 as a small grey handle on the right side of the scrapbook itself. The user simply clicks once on the control to hide the scrapbook, and again to reveal it.

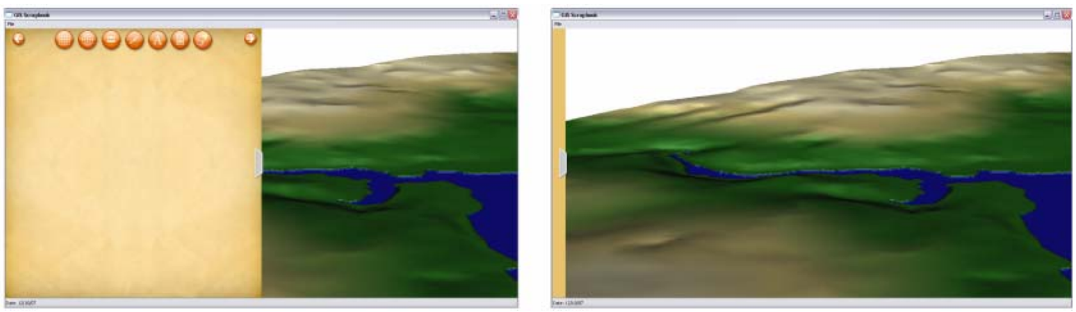


Figure 3: Scrapbook interface visible (left) and hidden away to the left (right).

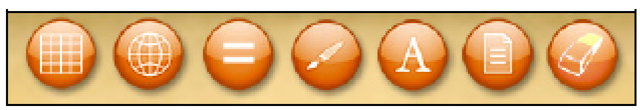


Figure 4: Detail of interaction mode buttons.

The set of accessible buttons along the top of the scrapbook represent the remainder of the available functionality (Figure 4). The first two buttons on the left set the mode of terrain selection, for choosing rectangular and square regions respectively. This is followed by the equation button which inserts equations that operate over data contained in the extracted terrain area. The equation button presents an equation dialog, as seen in Figure 5. For example, if one inserts *AVE*(elevation) it will calculate and insert the average elevation of the embedded terrain region, at the current text entry point. As a proof of concept, at this stage we have only implemented a few example equations.

The remaining buttons represent modes of interaction that support note taking and annotation. The first draws free-form 2D scribbles anywhere within the scrapbook, including directly over the embedded terrain. The second enables text-entry mode, allowing the user to make detailed notes of their ongoing analysis. Following this is a button which turns paper lines on and off, which adds more or less visual structure to the journal page. Examples of each of these note taking functions can be seen in Figure 1. Finally, the eraser button allows the removal of annotations by tracing the mouse over each of them in turn.

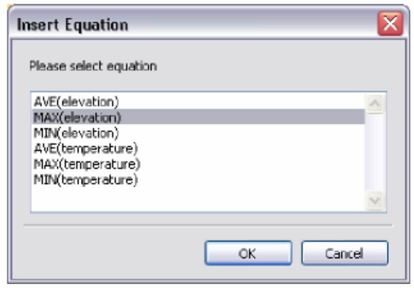


Figure 5: Equation insertion dialog.

5 Implementation Details

Both the 2D scrapbook and the 3D terrain are generated with the OpenGL rendering library using the C++ programming language. Currently, the texture for the 3D surface is algorithmically generated based on the local elevation, while the scrapbook is given a paper-like appearance by texturing a 2D polygon with a real parchment image.

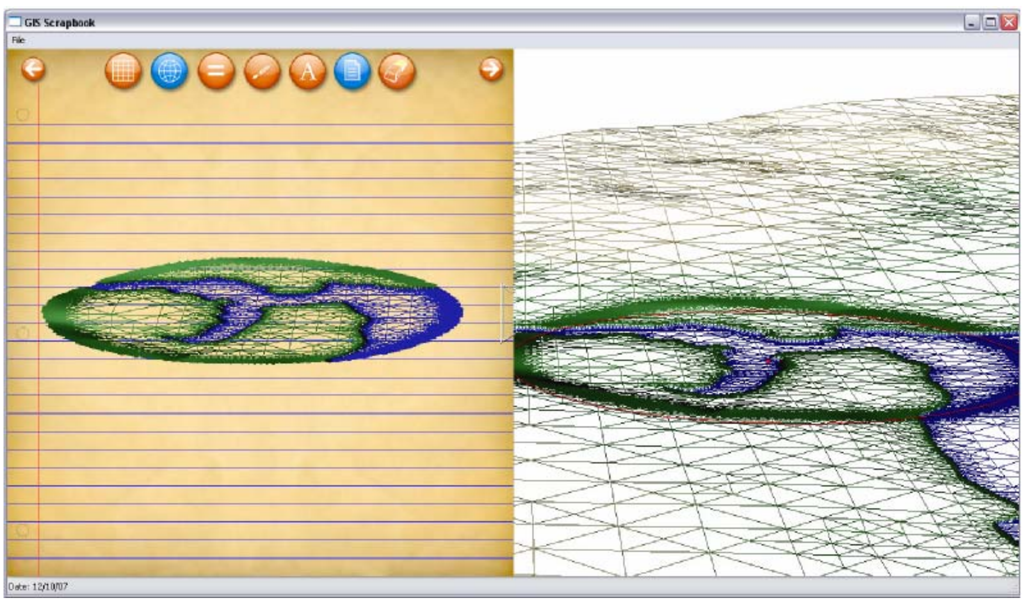


Figure 6: Wireframe rendering showing the non-uniform level-of-detail terrain.

We have constructed the geometry of the 3D terrain using a real-time adaptive hierarchical data structure (Wang et al, 2007), which allows us to increase frame rates without a significant loss of visual fidelity. This approach selectively adapts the number of polygons to the local topography. If the terrain's height is rapidly changing locally, more vertices are needed to accurately represent that local area. This can be seen in the wireframe rendering in Figure 6. An additional modification was required for the user selected regions. In order to produce crisp region boundaries, additional vertices are inserted around the border of a user selected region. Without this modification, the user selected regions would exhibit arbitrarily ragged boundaries, dependent on the changing precision of the local adaptive representation. The resulting selection region is shown as a wireframe rendering embedded within the scrapbook in Figure 6.

6. Conclusions and Future Work

In this paper we have presented a system that combines the situational awareness of 3D GIS, with familiar paper-based interaction. Our interactive scrapbook allows the user to save and analyse selected sub-regions of the 3D terrain, thereby maintaining 3D records of their analysis over time. However, the current system only prototypes the scrapbook concept, and there are many ways in which this work could be extended in order to increase its utility. Free form selection of embedded terrain areas (rather than circular or rectangular) would allow greater flexibility. We could also allow the user to modify the shape of the embedded selection region itself after it has been added. We would also like to significantly increase the range of equations that can be calculated over the embedded regions. In addition, it may be useful to offer facilities to add external images and diagrams into the scrapbook. Finally, thus far we have focused on the technical challenges of the system; we would also like to study the usability of the system.

7. Acknowledgements

This work was supported by NSERC and the Canadian Foundation for Innovation.

8. References

BROOKS, S. and WHALLEY, J. 2007. Towards a Comprehensive Multi-layer Hybrid Display of GIS Data. In Proceedings of GISRUK, Maynooth, pp. 215-219.

ESRI ARCGIS. (N.D.). Retrieved May 1, 2006, http://www.esri.com/software/arcgis/.

RISHE, N., SUN, Y., CHEKMASOV, M., ANDRIY, S. and GRAHAM S., 2004. System architecture for 3D terrafly online GIS. In Proceedings of Multimedia Software Engineering, 273-276.

SPRINGMEYER, R., BLATTNER, M. and MAX, N., 1992. A Characterization of the Scientific Data Analysis Process. In Proceedings of IEEE Visualization, 235-242.

STOTA, J. and ZLATANOVA, S., 2003. 3D GIS, where are we standing? In IPRS Joint Workshop on Spatial, Temporal and Multi- Multi-Dimensional Data Modeling and Analysis.

WANG, S., BEIKO R. AND BROOKS, S. 2007. Collapsible 3D Terrains for GIS Visualization. Geocomputation 2007, UK, Maynooth.

WARE, C., PLUMLEE, M., ARSENAULT, R., MAYER, L. A. and SMITH, S., 2001. GeoZui3D: data fusion for interpreting oceanographic data. Proceedings of the MTS/IEEE Conference and Exhibition, 3, 1960-1964.

Biographies

Suwen Wang is currently a Research Assistant in the Graphics and Visualization Lab at Dalhousie University in Canada. Suwen completed his B.CS. in 2007 and is scheduled to begin a Master's degree in 2008.

Stephen Brooks is an Assistant Professor at Dalhousie University in Canada. Prior to this, he acquired a Ph.D. from the University of Cambridge in 2004 and a M.Sc. from the University of British Columbia in 2000.

PDEAR Modelling of Climate Change Prediction of Rare Protea Species

Danni Guo[1], Renkuan Guo[2], Guy F. Midgley[1], A.G. Rebelo[1], and Christien Thiart[2]

[1]Kirstenbosch Research Center, South African National Biodiversity Institute, Private Bag X7, Claremont 7735, Cape Town, South Africa
[2]Department of Statistical Sciences, University of Cape Town, Private Bag, Rondebosch 7701, Cape Town, South Africa

Corresponding author: Dr Danni Guo
guo@sanbi.org

KEYWORDS: Partial Differential Equation Associated Regression, PDEAR model, climate change, Protea

1. Introduction

Global warming and climate changes are changing the environment and therefore changing the distribution and behaviour of the plant species. In this paper, we will model the Protea species in the population size of 1 to 10, in the Cape Floristic Region, in 2002, in South Africa. Since we have presence data only and these are incomplete sample data, so we will look at occurrence counts or frequency distributions of the Protea. We will use the *partial differential equation associated regression* (PDEAR) model, with consideration of precipitation and temperature, to fill in missing samples within the Cape Floristic Region, in 2002. Using future precipitation and temperature variables (Hewitson and Crane 2006, New et al. 2002, Tadross et al. 2005), we will use the PDEAR model to predict the future occurrence counts of rare Proteas population count.

2. Proteas in the Cape Floristic Region

The Cape Floristic Region is located at the southern tip of the Africa, and it covers parts of Western and Eastern Cape provinces of South Africa. It is home to some 9030 plant species, and nearly 70% of which are found nowhere else. Fynbos is the predominate ecosystem in the Cape Floristic Region, and it is under serious threat (Freeth et al., 2007). The Protea Atlas Project collected samples of Fynbos's flowering Proteas in the Cape Floristic Region, South Africa, in 2002. In this case, we are focusing on the population category of Proteas that has an estimated population size from 1 to 10 per sample site.

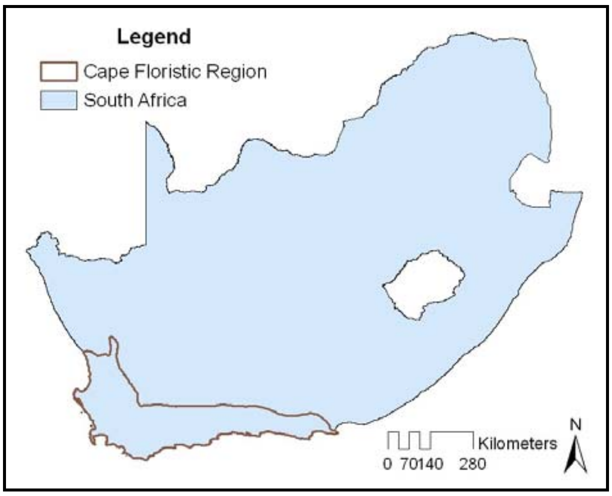


Figure 1. The Cape Floristic Region within South Africa

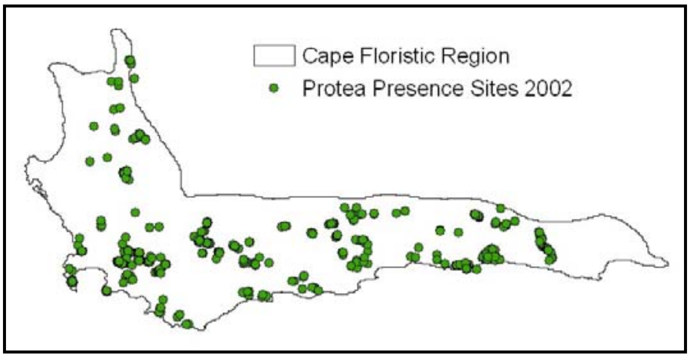


Figure 2. The Sample Locations of Proteas in the Population Size of 1-10, in the Cape Floristic Region, 2002

Figure 2 shows the locations of Proteas occurrence of the population size of 1 to 10, in the Cape Floristic Region. As one can see from the figure, the sample locations are not well spread. To solve this problem, we will use occurrence counts of the Proteas. The Cape Floristic Region is divided into grid cells, and within each cell, the presence of Protea species is counted, and the resulting value is attached to each centroid point of each cell. In Figure 3, the pink color are 0 in value, it shows the un-sampled locations. It is clear that a lot of the areas are un-sampled. The PDEAR model will be used here in order to predict the un-sampled cells.

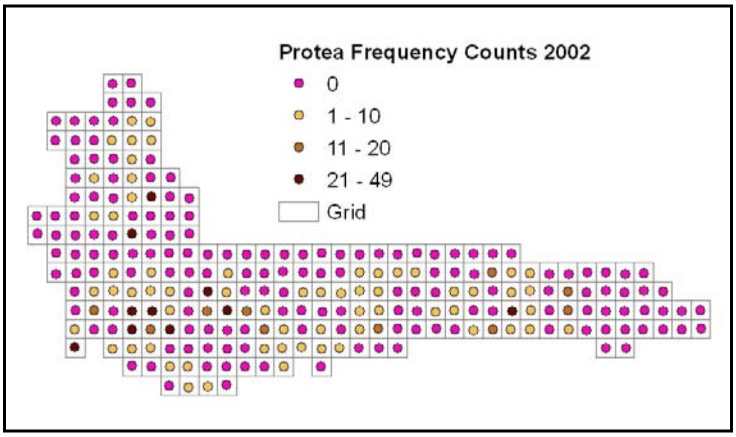


Figure 3. The Sampled Frequency Counts of Proteas in the Population Size of 1-10, in the Cape Floristic Region, 2002

3. The Theoretical Foundation for PDEAR modelling

PDEAR merges partial differential equation, multivariate regression and credibility measure theory into a new modelling family. For two-dimensional spatial modelling, the bivariate PDEAR model has following form:

$$\begin{cases} PDES \\ CREG \end{cases} \tag{1}$$

where $PDES$ denotes the first-order Pfaff linear partial differential equation system:

$$\begin{cases} \dfrac{\partial \underline{z}}{\partial x_1} = B_1 \underline{z} + \underline{a}_1 \\ \dfrac{\partial \underline{z}}{\partial x_2} = B_2 \underline{z} + \underline{a}_2 \\ \underline{z}\left(\underline{x}^0\right) = \underline{z}^0 \end{cases} \tag{2}$$

with constant coefficients:

$$B_i = \begin{bmatrix} b_{11}^{(i)} & b_{12}^{(i)} \\ b_{21}^{(i)} & b_{22}^{(i)} \end{bmatrix}, \ \underline{a}_i = \begin{bmatrix} a_1^{(i)} \\ a_2^{(i)} \end{bmatrix}, \ \ i = 1, 2 \tag{3}$$

satisfying Frobenius consistency conditions $B_1 B_2 = B_2 B_1$, and $CREG$ denotes coupled bivariate regression:

$$\begin{cases} \underline{\Delta}_{x_{1i}}^{\partial z} = B_1 \underline{z}\left(x_{1i}, x_{2i}\right) + \underline{a}_1 + \varepsilon_{1i} \\ \underline{\Delta}_{x_{2i}}^{\partial z} = B_2 \underline{z}\left(x_{1i}, x_{2i}\right) + \underline{a}_2 + \varepsilon_{2i} \\ \qquad i = 1, 2, \cdots, n \end{cases} \tag{4}$$

with partial divided difference:

$$\begin{cases} \underline{\Delta}_{\partial x_{1i}}^{\partial z} = \begin{bmatrix} \Delta_{\partial x_{1i}}^{\partial z_1} \\ \Delta_{\partial x_{1i}}^{\partial z_2} \end{bmatrix} = \begin{bmatrix} \dfrac{z_1\left(x_{1i}, x_{2i}\right) - z_1\left(x_{1(i-1)}, x_{2i}\right)}{x_{1i} - x_{1(i-1)}} \\ \dfrac{z_2\left(x_{1i}, x_{2i}\right) - z_2\left(x_{1(i-1)}, x_{2i}\right)}{x_{1i} - x_{1(i-1)}} \end{bmatrix} \\ \underline{\Delta}_{\partial x_{2i}}^{\partial z} = \begin{bmatrix} \Delta_{\partial x_{2i}}^{\partial z_1} \\ \Delta_{\partial x_{2i}}^{\partial z_2} \end{bmatrix} = \begin{bmatrix} \dfrac{z_1\left(x_{1i}, x_{2i}\right) - z_1\left(x_{1i}, x_{2(i-1)}\right)}{x_{2i} - x_{2(i-1)}} \\ \dfrac{z_2\left(x_{1i}, x_{2i}\right) - z_2\left(x_{1i}, x_{2(i-1)}\right)}{x_{2i} - x_{2(i-1)}} \end{bmatrix} \end{cases} \tag{5}$$

In Equation (1), the bivariate system is called the associated bivariate partial equation system, the bivariate regression system is called the coupled regression, and the error term is in nature of random fuzzy, in which the fuzziness comes from the replacements of the partial derivative vectors, say, $\partial \underline{z} / \partial x_1$, $\partial \underline{z} / \partial x_2$ in associated partial equation system by the divided partial differences, say, $\underline{\Delta}_{\partial x_{1i}}^{\partial z}$, $\underline{\Delta}_{\partial x_{2i}}^{\partial z}$ respectively, while the random error comes from sampling from the system under investigation. For more theoretical details, see Guo at al (2007[s], 2008).

For an intuitive explanation, we use *dear* (an abbreviation of *d*ifferential *e*quation *a*ssociated *r*egression) model, the univariate counterpart of PDEAR model to avoid the complexity in multi-dimensional spaces.

Without loss of generality, a simple linear differential equation, $dx/dt = \alpha + \beta x$ is used for illustrations. Let $\hat{x}_i^{(1)}$ denote an approximation to the primitive function $x(t)$ at t_i, and let $\Delta x_i/\Delta t_i$ be an approximation to the derivative function dx/dt at t_i, with $\Delta x_i = x(t_i) - x(t_{i-1})$, $\Delta t_i = t_i - t_{i-1}$.

Definition 2.1: If a dynamic system governed by $dx/dt = \alpha + \beta x$ is sampled at its derivative level, denoted by $X^{(0)} = \{x_1^{(0)}, x_2^{(0)}, \cdots, x_n^{(0)}\}$, the coupled equation system:

$$\begin{cases} \dfrac{dx}{dt} = \alpha + \beta x \\ x_i^{(0)} = \alpha + \beta \hat{x}_i^{(1)} + e_i, \; i = 1, 2, \cdots, n \end{cases} \tag{6}$$

is called Type I DEAR model. If it is sampled at its primitive level, denoted by $X^{(1)} = \{x(t_1), x(t_2), \cdots, x(t_n)\}$, the coupled equation system:

$$\begin{cases} \dfrac{dx}{dt} = \alpha + \beta x \\ \dfrac{\Delta x_i}{\Delta t_i} = \alpha + \beta x(t_i) + e_i, \; i = 1, 2, \cdots, n \end{cases} \tag{7}$$

is called Type II DEAR model.

The second equations in equation system (6) and (7) are called coupled regressions, while the first one is called the associated differential equation.

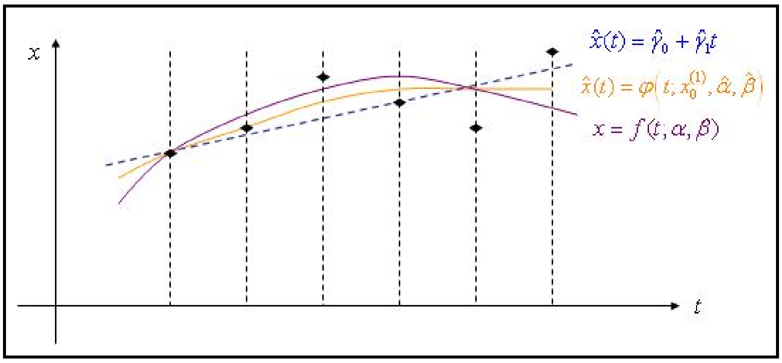


Figure 4. Two approximations to nonlinear curve $x(t) = f(t; \alpha, \beta)$ in (t, x) space

In terms of Type I DEAR, we first fit the coupled regression in the (x, x') space (Figure 5), where x' denotes the derivative of x with respect to t.

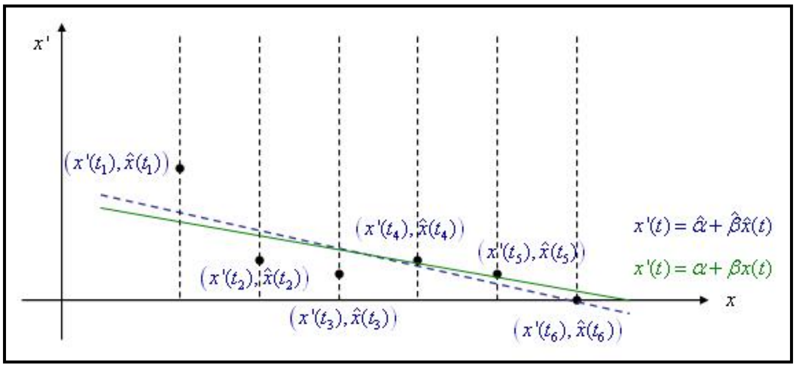


Figure 5. Type I approximation in (x, x') space

Then estimator of parameter $\underline{\theta} = (\alpha, \beta)$, denoted by $\hat{\underline{\theta}} = (\hat{a}, \hat{b})$ is obtained. Once the parameter $\underline{\theta} = (\alpha, \beta)$ is obtained, by solving the approximated linear differential equation $dx/dt = \hat{a} + \hat{b}x$, we will obtain an approximated nonlinear curve $x' = \varphi\left(t; x_0^{(1)}, \hat{a}, \hat{b}\right)$, (yellow-coloured curve in Figure 4), which is expected to approximate the primitive curve in relatively high accuracy.

If the sampling observations are collected at primitive function level, denoted as $X^{(1)} = \{x(t_1), x(t_2), \cdots, x(t_n)\}$. Then in terms of Type II DEAR Type II, the derivatives could be approximated, for example, by the divided difference, i.e., $\Delta x_i / \Delta t_i$. Just as shown in Figure 6, fitting $\hat{x}' = \Delta x / \Delta t = \hat{a} + \hat{b}x$ for approximating line $x' = a + bx$. Then, the estimated parameter $\hat{\underline{\theta}} = (\hat{\alpha}, \hat{\beta})$ will lead the nonlinear approximation $x' = \varphi\left(t; x_0^{(1)}, \hat{a}, \hat{b}\right)$ to the primitive function $x(t) = f(t; a, b)$ in (t, x) space (shown in Figure 4).

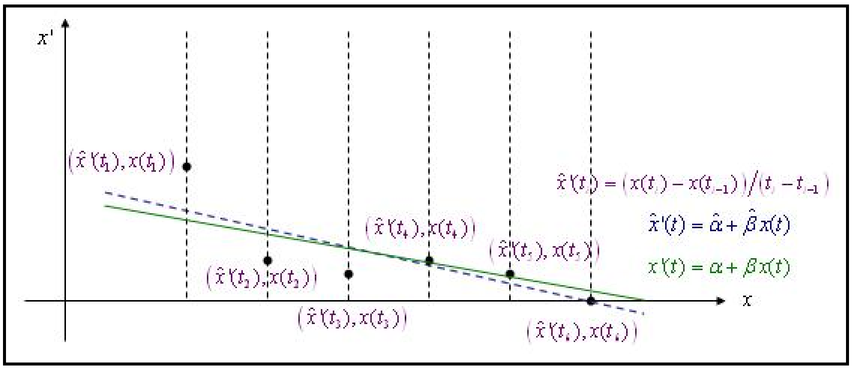


Figure 6. Type II approximation in (x, x') space

The error structure makes the coupled regression being no longer a regression in statistical linear model sense, rather in the random fuzzy variable theoretical sense, which engages Liu's credibility measure theory (2004, 2007) rather Zadeh's fuzzy mathematics (1965, 1978). More technical details can be found in 2007 publications of Guo et al listed in the reference.

4. PDEAR Predicted Protea Occurrence Counts

Using the predicted results from the PDEAR model, the un-sampled cells are predicted with frequency counts of the Protea. Figure 7 shows the PDEAR model predicted frequency counts of Proteas. The figures show a general range of predicted values, but in fact the actual predicted values are numerical, and have predicted the sample values of where there were no samples at all.

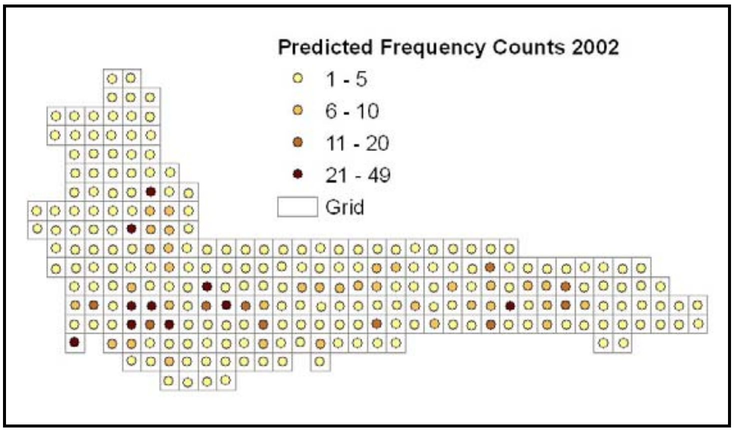


Figure 7. The PDEAR Model Predicted Frequency Counts of Proteas in the Population Size of 1-10, in the Cape Floristic Region, 2002

Now we can produce kriging prediction maps of the Proteas, using the predicted results from the PDEAR model. Figure 8 shows the distribution of occurrence counts of Proteas.

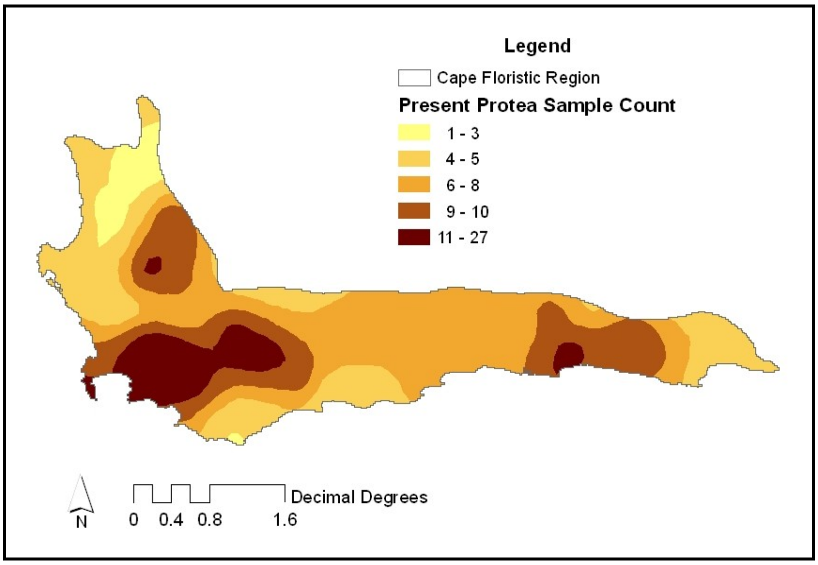


Figure 8. The Kriging Prediction Maps of Frequency Counts of Proteas in the Population Size of 1-10, in the Cape Floristic Region, 2002

Using the present precipitation and temperature data (1961-1990) and the projected future precipitation and temperature data (2070-2100) (Hewitson and Crane 2006, New et al. 2002, Tadross et al. 2005), rare Proteas counts are predicted for the future using the PDEAR model. See Figure 9.

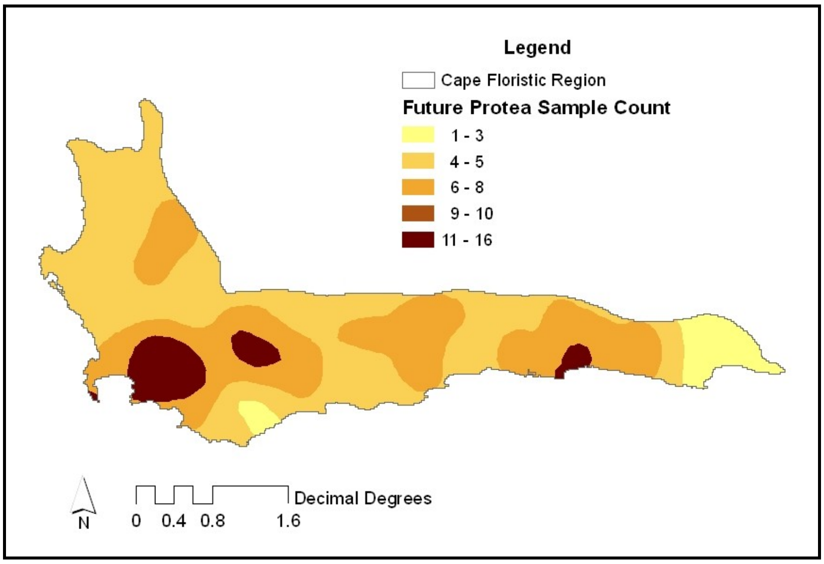


Figure 9. The Kriging Prediction Maps of Frequency Counts of Proteas in the Population Size of 1-10, in the Cape Floristic Region, in the future (2070-2100)

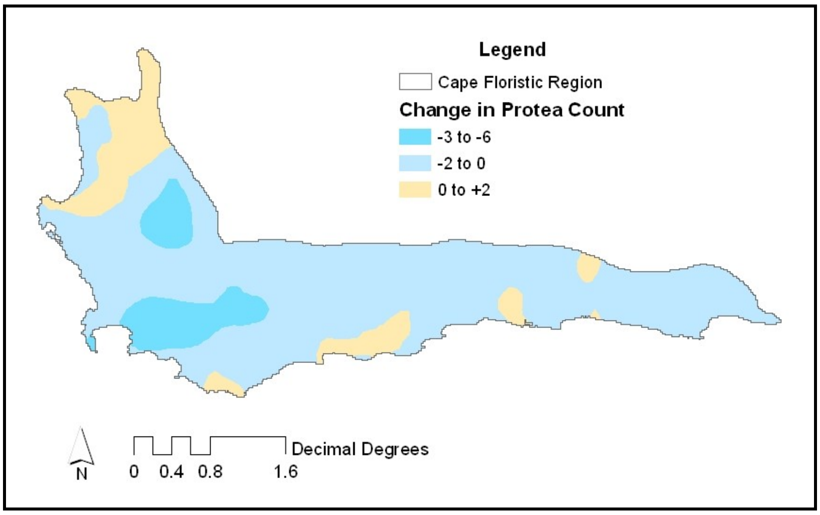


Figure 10. The Kriging Prediction Maps of Changes in Frequency Counts of Proteas between the Present and the future

Figure 10 shows the changes in Proteas counts between present and future. It is clear that in general, there is a decline in Protea counts, but in some isolated areas, a positive increase in Protea counts is shown. One must recognise that this population category of Proteas only has the population size of 1 to 10, so that it is very sensitive to environmental changes. Therefore, the Proteas between the years have changed dramatically in frequency count and spatial distribution patterns. Of particular interest is the dark blue areas, where there is a high decrease in Protea counts. It is important to recognise that under the changing climate conditions, Proteas are under threat and their distribution and numbers are changing in response to the climate, and we need to acknowledge the impacts of climate change.

5. Conclusion

The coupled regression component in a PDEAR model is in nature a special random fuzzy multivariate regression model. We developed a bivariate model for investigating the impacts from rainfall and temperature on the Protea species in the population size of 1 to 10, in the Cape Floristic

Region, in 2002, South Africa. Under same the biodiversity structure, we explore the future spatial change patterns of Protea species in the population size of 1 to 10 with future predicted rainfall and temperature. The spatial distribution and patterns are clearly will help us to explore global climate changing impacts on endangered species.

6. Acknowledgements

We would like to thank the Workshop on high-resolution climate change data for southern Africa held at the University of Witwatersrand, for providing the climate data used in this paper.
We would also like to thank the Protea Atlas Project for providing the species data for our research. This research is partially supported by South African National Research Foundation Grant FA2006042800024.

References

Freeth R, Bomhard B and Midgley GF (2007). *Adapting to Climate Change in the Cape Floristic Region: Building Resilience of People and Plants in Protected Areas*. South African National Biodiversity Institute (SANBI), Climate Change Research Group, Kirstenbosch Research Center.
Guo D, Guo R and Thiart C (2006). The Coupling of regression Modelling and Differential Equation Model in GM(1,1) Modelling and Extended GM(1,1) Models, *Journal of Grey System*, **9(2)** pp143-154.
Guo D, Guo R, Midgley GF and Ebrahim I (2007). DEAR Modelling of Brunsvigia Litoralis Plant Distribution in Western Region District Council, South Africa , *Proceedings of the 6th International Conference on Information and Management Sciences*, Lhsa, Tibet, pp770-782.
Guo, D., Guo, R., Midgley, G. F. and Rebelo, A. G. PDEMR Modelling of Protea Species in the Population Size of 1 to 10, in Cape Floristic Region from 1992 to 2002, South Africa, Journal of Geographical Information Sciences, ISSN 1082-4006, Vol. 12, No. 2, pp 67-78.
Guo R, Guo D, Midgley GF and Rebelo AG (2007) Bivariate DEAR Modelling of Range Shifts in Aulax Umbellata Shrubs From 1992 to 2002 in South Africa, *Proceedings of the 6th International Conference on Information and Management Sciences*, Lhasa, Tibet, pp783-791.
Guo, R. and Guo, D. (2007). Univariate DEMR Model – The Formation and Foundation. Proceedings of IEEE International Conference on Grey Systems and Intelligent Services, November 18-20, Nanjing, China, 2007, pp 13-18. IEEE Catalog Number: 07EX1882C, ISBN: 1-4244-1294-3, Library of Congress: 2007926987, Editor-in-Chief: Sifeng Liu
Guo, R. and Guo, D. (2007). Multivariate DEMR Model – The Formation and Foundation. Proceedings of IEEE International Conference on Grey Systems and Intelligent Services, November 18-20, Nanjing, China, 2007, pp 19-25. IEEE Catalog Number: 07EX1882C, ISBN: 1-4244-1294-3, Library of Congress: 2007926987, Editor-in-Chief: Sifeng Liu
Guo, R., Guo, D., Midgley, G. F. and Rebelo, A. G. (2007) PDEMR Model in Rare Protea Count Prediction. Journal of Grey System, Vol. 10, No. 2, September, pp 105-114, 2007. (ISSN: 1028-9488)
Guo, R., Guo, D., Midgley, G. F. and Rebelo, A. G. (2008) PDEMR Modelling of Protea Rare Species Spatial Patterns. Accepted by Journal of Uncertain Systems.
Hewitson BC and Crane RG (2006) Consensus between GCM Climate Change Projections with Empirical Downscaling: Precipitation Downscaling over South Africa, *International Journal of Climatology*, **26(10)** pp1315-1337.
Liu BD (2004) *Uncertainty Theory: An Introduction to Its Axiomatic Foundations*. Springer-Verlag Heidelberg, Berlin.
Liu BD (2007) *Uncertainty Theory*, 2nd Edition. Springer-Verlag Heidelberg, Berlin.
Midgley GF, Hughes GO, Thuiller W and Rebelo AG (2006) Migration rate limitations on climate change-induced range shifts in Cape Proteaceae, *Diversity and Distributions*, **12** pp555-562.
New M, Lister D, Hulme M and Makin I (2002). A high-resolution data set of surface climate over global land areas, *Climate Research*, **21** pp1-25.
Rebelo AG (2001). *Proteas. A field guide to the proteas of southern Africa*. Fernwood Press, Vlaeberg, Cape Town, South Africa.

Tadross M, Jack C and Hewitson B (2005). On RCM-based projections of change in southern African summer climate, *Geophysical Research Letters*, **32(23)**.
Zadeh LA (1965). Fuzzy Sets, *Information and Control*, **8** pp338-353.
Zadeh LA (1978). Fuzzy Sets as a Basis for a Theory of Possibility, *Fuzzy Sets and Systems*, **1** pp3-28.

Biography

Dr Danni Guo is a specialist scientist in the Climate Change Research Group at the South African National Biodiversity Institute (SANBI). Her current research focus is on spatial modelling and GIS in biodiversity under sparse data uncertainty.

Dr Renkuan Guo is a Professor at the University of Cape Town, South Africa. His research interest includes: spatial modelling, uncertainty modeling in quality and reliability, particularly, repairable systems, small sample inferences, stochastic process modeling, particularly, financial markets analysis, Mathematical Foundation of uncertainty theory.

Dr Guy F. Midgley has been engaged in research on climate change impacts on ecosystems and biodiversity and development of adaptation strategies since the early 1990's, and now lead SANBI's Climate Change Research Group, and contribute to the work of the National Climate Change Committee (NCCC), the South African Scientific Committee on Global Change (SASCGC), and the Intergovernmental Panel on Climate Change (IPCC).

Dr Anthony G. Rebelo is based at the Threatened Species Programme of the South African National Biodiversity Institute. His recent work includes writing up the Fynbos component of the recently completed vegetation of South Africa.

Dr Christien Thiart is a Professor at the University of Cape Town, South Africa. Her research interests includes: GIS and Graphical computing: with special interest in analysis (spatial and statistical), the Linear model with special interest in Robust estimation techniques.

Testing co-evolutionary hypotheses over geological timescales using GIS

Richard J. Butler[1], Paul M. Barrett[1], Paul Kenrick[1], Malcolm G. Penn[2]

[1]Department of Palaeontology, The Natural History Museum, Cromwell Road, London, SW7 5BD
R.Butler@nhm.ac.uk
[2] Department of Botany, The Natural History Museum

KEYWORDS: Co-evolution, Cretaceous, cycads, dinosaurs, palaeontology

1. Introduction

Ecological and behavioural interactions between species are known to exert a strong influence on their evolution (Thompson 1994). Co-evolution occurs when two (or more) species interact with each other in such as way that the evolution of one influences the evolution of the other and *vice versa*. Co-evolutionary processes are commonly divided into 'tight' and 'diffuse' co-evolution: 'tight' co-evolution occurs when there is intimate ecological coupling between pairs of species (e.g. flower/pollinator) and is thought to be important in driving speciation and adaptation, while 'diffuse' co-evolution occurs when whole groups of species interact with other groups of species, and is thought to be a key process in shaping the evolution of communities. The significance of co-evolution over ecological timescales is well established; however, it is unclear to what extent it contributes to driving large-scale evolutionary and ecological changes over geological timescales. Many instances of putative co-evolution over geological timescales have been suggested, but few have been evaluated in detail. Here we discuss proposed dinosaur/plant diffuse co-evolution during the Cretaceous period (145-65 million years ago [mya]), focusing in particular on dinosaur/cycad interactions, and the use of GIS as a tool to explore co-evolutionary hypotheses over geological timescales.

1.1 Dinosaur/plant co-evolution

Numerous hypotheses of dinosaur/plant co-evolution have been proposed, the best known of which implicates dinosaurs in the origin/radiation of angiosperms (flowering plants) in the Cretaceous (Bakker 1978; Barrett and Willis 2001). Another set of co-evolutionary hypotheses links dinosaurs and cycads (e.g. Watson and Cusack 2005; Mustoe 2007). Cycads are evergreen, insect-pollinated, seed plants with large cones and toxic foliage; in modern ecosystems cycads are of low diversity (10-12 genera) and confined largely to tropical/subtropical regions. Cycads have a long fossil record, extending at least to the Permian (250 mya), and appear to have been more abundant and diverse in the Mesozoic (Triassic, Jurassic and Cretaceous periods) than at present. Co-evolutionary interactions with dinosaurs are proposed to have driven Mesozoic cycad diversification: various groups of herbivorous dinosaurs, particularly the long-necked, large-bodied sauropodomorphs, are suggested as key dispersal agents for cycad seeds. The decline of such dinosaur groups during the Cretaceous is cited as one explanation for cycad decline (Mustoe 2007). Our ongoing work is evaluating hypotheses of dinosaur/cycad interactions using a range of approaches, including GIS.

1.2 Use of GIS in palaeontology and co-evolutionary analysis

To date, the palaeontological applications of GIS have been limited; most commonly it has been used to create geospatial databases of fossil localities, but a small number of successful analytical studies have also been published (e.g. Rode and Lieberman 2004; Rayfield et al. 2005). The advantage of using a GIS for palaeontological datasets is that it is possible to rapidly (and easily) integrate palaeobiological, palaeoenvironmental and geological data with a temporal series of maps showing former continental positions (palaeogeographical maps) – this allows identification of spatiotemporal patterns in the distributions of organisms and their environments (e.g. Rayfield et al. 2005). If two taxonomic groups are undergoing diffuse co-evolution their distributions should be closely associated both spatially and temporally; the identification and/or falsification of such spatiotemporal associations can be carried out by using GIS to compare distributions of two or more taxonomic or palaeoecological groupings.

2. Methods

Global occurrences of Cretaceous herbivorous dinosaurs were compiled as a GIS-compatible relational database in Microsoft Access; data was collected from the primary literature, based upon references cited in Weishampel et al. (2004), as well as more recent references and *The Paleobiology Database* (www.paleodb.org). Faunal information was collected for each locality and combined with data on palaeoecology (e.g. body mass, feeding behaviour), geological age, lithology and palaeoenvironment. For each geological formation yielding herbivorous dinosaur material, we collected information (systematics, physiognomy, palaeoenvironments) on macrofloras and palynofloras from the primary literature. At present, this database contains 9098 fossil occurrences at 1952 localities, representing 407 dinosaur taxa, and over 2300 macrofloral taxa.

We imported our database into ArcGIS 9.1, plotting the data onto palaeogeographical maps, shapefiles for which are available for six different Cretaceous timeslices (Scotese 2001). Modern day latitude–longitude coordinates for database localities were converted to palaeolatitude–longitude coordinates at intervals corresponding to the available palaeogeographical maps using PointTracker software (Scotese 2004). For each of the twelve Cretaceous temporal subdivisions ('stages') we used attribute selection to create new data layers (on appropriate palaeogeographical maps) displaying occurrences from that temporal stage of: a) cycads; b) 'cycadophytes' (cycads and cycad-like fossil groups); c) various groups of dinosaurs including Sauropodomorpha, Ankylosauria, Ornithopoda, and Ceratopsia. Overlayering occurrence data allows visual comparison of the spatial distributions of two groups within a time-slice. In areas in which sampling of the fossil record was good we digitised a polygon to create a minimum palaeobiogeographic range that included all occurrences of that taxonomic/paleoecological grouping – construction of minimum range polygons for two or more groups allows comparison of the degree of range overlap between two groups. The absence of range overlap between two groups proposed to be co-evolving falsifies that co-evolutionary hypothesis on the basis of available fossil data; conversely, repeated and extensive range overlap is consistent with a co-evolutionary hypothesis.

3. Results and Directions

Preliminary analyses and results highlight two major problems in assessing dinosaur/cycad co-evolutionary hypotheses: 1) the Cretaceous plant fossil record is relatively incomplete spatially and temporally, with the northern hemisphere being notably better represented than the southern hemisphere; as a result, cycad ranges can only be assessed for a relatively small number of stages and geographic areas (notably the Late Cretaceous of North America); 2) true cycad foliage is difficult to distinguish from 'cycadophytes', which include specimens probably assignable to the extinct plant group Bennettitales, thus complicating attempts to determine ranges. Cycads have a near global distribution during the Cretaceous, and are most abundant during the early part of the Cretaceous (Berriasian-Hauterivian; 145-130 mya). Co-evolution is therefore not supported for dinosaur groups which are temporally and/or geographically restricted (e.g. Pachycephalosauria, Late Cretaceous of East Asia/North America). One of the most interesting results concerns the latest Cretaceous (Campanian-Maastrictian; 83-65 mya) of North America; in this timeslice both plant and dinosaur records are well sampled geographically. Sauropodomorphs are geographically restricted in this timeslice to the SW of the United States, and do not range into the northern United States or Canada; by contrast, records of cycads from this timeslice are largely limited to the northern United States or Canada, and the ranges of sauropodomorphs and cycads do not appear to extensively overlap. This is inconsistent with the hypothesis of sauropodomorph/cycad co-evolution, whereby sauropodomorphs were the major dispersers of cycad seeds (Mustoe 2007).

Other analytical approaches also yield results inconsistent with co-evolutionary hypotheses: for example, diversity analyses indicate a major extinction of sauropodomorphs at the Jurassic/Cretaceous boundary (145 mya), but this extinction event preceded the beginnings of cycad decline by around 10 million years, and cycad decline actually begins just as sauropodomorphs (and other groups potentially capable of dispersing cycad seeds, such as primitive birds) are beginning to recover, radiate and diversify. These results suggest that although dinosaurs *may* have acted as dispersal agents for cycad seeds, there is little evidence at present that cycad diversity was reliant on dinosaur diversity and *vice versa*. Ongoing work will continue to use GIS to explore patterns of spatiotemporal diversity in cycads and dinosaurs, as well as exploring other co-evolutionary hypotheses, such as those relating to angiosperm origins. For instance, we plan to use indicator kriging to interpolate from available data, and subsequently carry out statistical tests of correlation between dinosaur/cycad distributions, using approaches that attempt to account statistically for spatial autocorrelation (e.g. Rangel et al. 2006).

4. References

Bakker RT (1978) Dinosaur feeding behaviour and the origin of flowering plants *Nature* **274** pp661-663

Barrett PM and Willis KJ (2001) Did dinosaurs invent flowers? Dinosaur-angiosperm coevolution revisited *Biological Reviews* **76** pp411-447

Mustoe GE (2007) Coevolution of cycads and dinosaurs *The Cycad Newsletter* **30** pp6-9

Rangel TFLVB, Diniz-Filho JAF and Bini LM (2006) Towards an integrated computational tool for spatial analysis in macroecology and biogeography *Global Ecology and Biogeography* **15** pp321-327

Rayfield EJ, Barrett PM, McDonnell RA and Willis KJ (2005) A Geographical Information System (GIS) study of Triassic vertebrate biochronology *Geological Magazine* **142** pp327-354

Rode AL and Lieberman BS (2004) Using GIS to unlock the interactions between biogeography, environment, and evolution in Middle and Late Devonian brachiopods and bivalves *Palaeogeography, Palaeoclimatology, Palaeoecology* **211** pp345-359

Scotese CR (2001) *Earth System History Geographic Information System v. 02b*. PALEOMAP Project, Arlington, Texas.

Scotese CR (2004) *PointTracker v. 04c*. PALEOMAP Project, Arlington, Texas.

Thompson JN (1994) *The Coevolutionary Process*. Chicago University Press, Chicago.

Watson J and Cusack HA (2005) Cycadales of the English Wealden *Palaeontographical Society Monographs* **622** pp1-189

Weishampel DB, Dodson P and Osmólska H (2004) *The Dinosauria, 2nd Edition*. University of California Press, Berkeley.

Biography

Dr Richard Butler is a postdoctoral researcher funded by the award of a NERC Standard Grant (NE/C002865/1) to PMB, PK and MP. His PhD was carried out at the University of Cambridge, and his research focuses on dinosaur evolution and the testing of macroevolutionary hypotheses using quantitative analytical techniques, including GIS.

Methods for Predicting Human Behaviour in Emergencies: An Analysis of Scientific Literature

Věra Karasová[1] and Glyn Lawson[2]

[1]Centre for Geospatial Science, School of Geography, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
Tel. +44(0) 115 846 8411 Fax. +44 (0)115 951 5249
Email: vera.karasova@nottingham.ac.uk, http://www.nottingham.ac.uk/cgs/
[2]Human Factors Research Group, School of Mechanical, Materials and Manufacturing Engineering, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
Tel. +44(0) 115 951 4040 Fax. +44 (0)115 846 6771
Email: glyn.lawson@nottingham.ac.uk, http://www.virart.nottingham.ac.uk/

KEYWORDS: agent-based modelling, human behaviour, emergencies

1. Introduction

The increased threat of terrorist event such as the explosion of a dirty bomb in densely populated urban areas necessitates modification of existing frameworks of local response. The primarily goal of first responders is to mitigate the consequence of such a disaster by reducing the number of victims, and securing resources for protection of the general public to avoid further injuries (Civil Contingencies Secretariat, 2005; Great Britain, 2005). This may be successfully accomplished if the incident commanders anticipate the reactions of affected people and the influence of the environment on human reasoning. Therefore, models representing common patterns of spatially aware human behaviour under severe, life threatening conditions, need to be incorporated into the modern emergency response plans and guidelines.

A commonly used technique for modelling human behaviour is the first principles approach (Laughery, 2005). This approach involves identification, typically from literature, of the underlying goals and principles which govern a person's performance. Several hundreds of peer-reviewed studies of behaviour have been conducted to date. These have identified a large number of principles which can be implemented as algorithms in simulation systems. The resulting simulations can subsequently be used to investigate different approaches and response scenarios for first aid and evacuation. The principles are, however, of diverse quality and much effort is required to identify which ones to use. Furthermore, they can be unspecific, un-quantified, or incomplete. For instance spatial behaviour of victims and the effects of the geographical space on human reasoning have been often neglected. This leads to difficulties with verification when used in situations other than those from which they were originally derived (Cornwell et al., 2002; Laughery, 2005; Silverman et al., 2001; Silverman et al., 2006).

Human behaviour can also be predicted through the application of other, less commonly applied, techniques. For example, sequential analysis involves the study of events and their interactions as they unfold over time. The basic method for sequential analysis is to classify behavioural patterns (either live from direct observation, or retrospectively from interview transcripts, CCTV footage, etc.) using a pre-defined taxonomy of behaviours. These data can be relevant when identifying the probability of one particular type of behaviour following another. Furthermore, these analyses enable us to extract generic behaviour patterns which can be easier transformed into simulation models (Bakeman and Gottman, 1986).

Other techniques which may prove suitable for depicting behavioural patterns include activity sampling (Kirwan and Ainsworth, 1992), expert judgements (e.g. Dombroski et al. 2006) and qualitative analysis of interview data (Kirwan and Ainsworth, 1992). These techniques will be applied and analysed in a series of future studies. This paper reports on the use of the first principles approach for modelling spatial behaviour in a CBRN (chemical, biological, radiological or nuclear) emergency.

2. Method

The first principles approach involved an extensive literature review followed by abstraction of the findings into a conceptual model of human spatial reasoning under extreme conditions. Various journal papers from psychological, sociological and human factors research related to human behaviour in emergencies were reviewed. Issues originating from these research domains have been widely studied from different perspectives, for example behavioural response to a natural disaster (Lachman et al., 1961; Perry and Lindell, 2003; Quarantelli, 1996), crowd dynamics of football supporters (Still, 2000) and the emergence of riots (El Rhalibi and Taleb-Bendiab, 2005).

However, due to the lack of empirical evidence, a thorough investigation of citizens' reactions to a CBRN terrorist incident remains a great challenge. The latest research results presented by Drury (2004) and later followed by Drury and Cocking (2007) indicate occurrence of the same kinds of human behaviour patterns and psychological processes across different scenarios, different kinds of disasters with different populations of survivors. The research presented in this paper inclines towards the belief that people are most likely unable to distinguish between an "ordinary" explosion and a CBRN incident in the first minutes after the blast. It was therefore assumed that the reaction of citizens in the first hour after the explosion will be in both cases the same. This enables the use of findings from research related to any type of sudden, impulsive and location restricted emergency.

3. Results

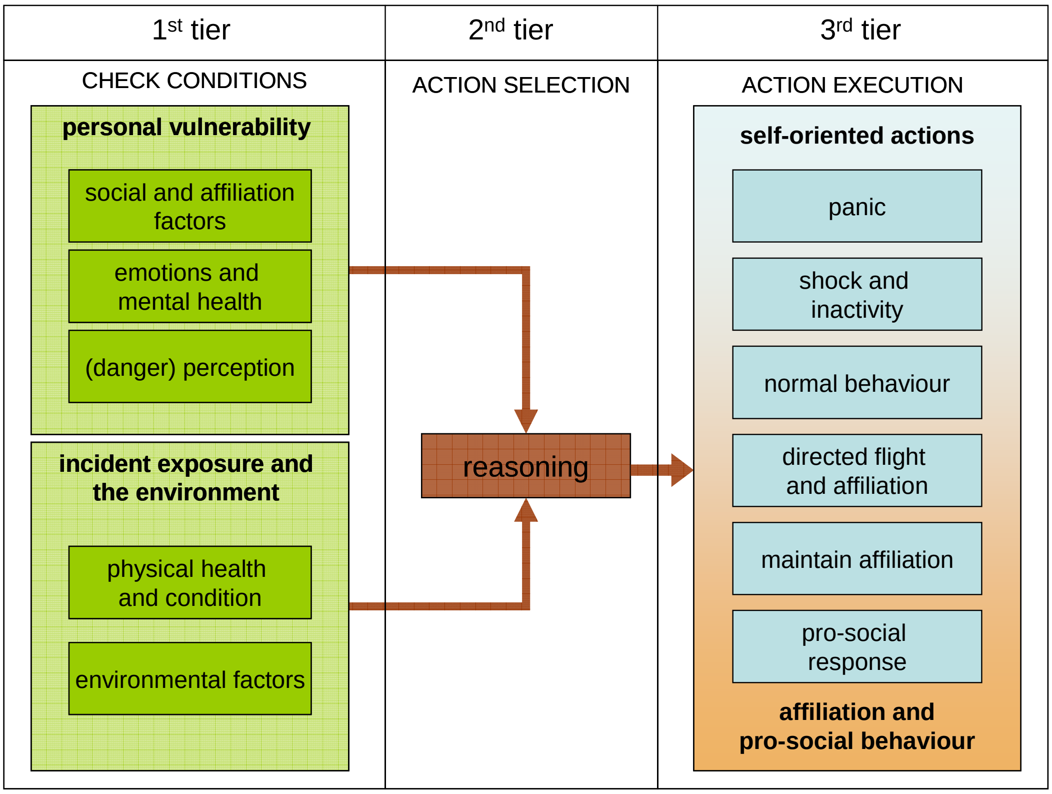


Figure 1. The three-tier agent architecture

The characteristics of human behaviour patterns obtained from the literature survey assisted with the creation of a generic conceptual model depicting the most common citizen's reaction to an urban

CBRN incident and factors contributing to their changes. It is important to mention that the whole spectrum of reviewed articles contributed to the model.

The model defines the action selection mechanism of an individual person, and forms a framework on which a software agent could be designed. The model represents a decision process based upon an assessment of the situation and selection of an appropriate action triggered by external and internal stimuli. The three-tier architecture of the agent is illustrated in Figure 1. The first tier processes the input information. This information is further decomposed according to selected factors having influence over the agent's external behaviour. These factors are either internal, representing the personal vulnerability of the agent and its personal characteristics, or external, specifying the agent's perception of the environment and incident effects. Each of the factors is affected by one or more agent attributes such as gender, energy, social affinity, mobility, etc. The result is then passed into the second tier where the reasoning determines what action to execute. This is the place where all the obtained input is analysed and evaluated against the current values of the agent's attributes. The actual action is performed in the third (output) tier. Further explanations of all the possible behaviours are provided in the Table 1.

Table 1. Characteristics of emergency behaviour

Behaviour	Specification
Panic	- irrationality - bewilderment - hysterical flight - ignorance of the environment
Shock and inactivity	- numbness - no movement
Normal behaviour	- original plan execution - no change in initial behaviour
Directed flight and affiliation	- normal to fast walk towards affiliate factors (family, home, etc.)
Maintain affiliation	- formation of ad-hoc groups - movement in groups - leader following - assisting group members - adapting to group behaviour
Pro-social response	- mutual helping based on social roles - provision of first aid and rescue - self-sacrifice for sake of helping others

4. Discussion

The literature review provided a good opportunity for the collection of diverse research results and critical points of current knowledge of human response to an explosion. There is however an apparent drawback in terms of the information accuracy. Moreover, references to the resources of the findings are often incomplete, twisted or even omitted. This generates a huge uncertainty and possible misunderstanding of the original knowledge. Utilisation of such results as a theoretical basis for the implementation of simulation models may lead to incorrect interpretation of human behaviour, which could in turn produce catastrophic consequences if consulted during the creation of evacuation plans.

The most accurate source of information appear to be direct interviews with disaster survivors or archives containing victims' testimonies, and observation reports. Example of such research can be found in Drury and Cocking (2007), Mawson (2005) and Sime (1999). In order to keep the behaviours in the conceptual model realistic, greater emphasis was put on the literature which draws upon findings from direct contact with victims or consultation of archived reports.

The literature review also revealed various models concentrating on different aspects of the human reaction. For instance Mawson (2005) focuses his attachment model and typology of response on the perceived degree of physical danger and levels of social support available in the situation. Drury and Cocking's (2007) research aim was to understand the conditions under which people start cooperating with each other, despite not knowing each other prior to the emergency. In spite of these differences, the reviewed papers have common themes such as the conclusion that panic, irrational and selfish behaviour are not the most commonly observed reaction of the public as it was believed in the past or as is widely presented in the media (Alexander and Klein, 2006; Mawson, 2005; Sime, 1999).

The spatial aspect of human behaviour has not been widely discussed in the reviewed papers. Still (2000) based his model of crowd movement through space on the assumption that people in a disaster tend to follow path of least effort. He defined such route based on the geometrical representation of the space, e.g. width of the escape routes and their congestions. Johnson Jr. (1985) argues that the initial distance of a person from the location of the incident has strong influences on the selection of egress routes and speed of evacuation. It has been reported that despite the physical nature of the area people tend to evacuate in the direction of their homes or place where their family is located by following well-known paths as oppose to searching for alternative routes (Dymon and Winter 1993; Mawson, 2005; Raphael, 2005).

5. Conclusion

The literature review was sufficient to generate a conceptual model depicting the conditions, action selection and action execution of an agent. The conceptual model has yet to be implemented for agents within a simulation; it is at this stage when other issues associated with the use of literature for modelling human behaviour in agents may become apparent (e.g. creating algorithms from un-quantified or incomplete models, verification difficulties) as well as issues related to the representation and effects of the simulation space on agent's behaviour. The process presented in this paper will be repeated using other methods for describing human behaviour than the first principles approach. The outputs of these models will be analysed for their usefulness as techniques for predicting the human response to emergency situations.

References

Alexander DA and Klein S (2006) The Challenge of Preparation for a Chemical, Biological, Radiological or Nuclear Terrorist Attack *Journal of Postgraduate Medicine* **52(2)** pp126–131

Bakeman R and Gottman JM (1986). *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge University Press, Cambridge, UK

Civil Contingencies Secretariat (2005) *Emergency Response and Recovery: Non-statutory Guidance to Complement Emergency Preparedness* [online] Available: http://www.ukresilience.info/upload/assets/www.ukresilience.info/emergresponse.pdf (accessed: 27/9/2005)

Cornwell JB, Silverman BG, O'Brien K and Johns M (2002) A Demonstration of the PMF-Extraction Approach: Modeling The Effects of Sound on Crowd Behavior *In: Proceedings of the Eleventh Conference on Computer Generated Forces and Behavioral Representation*. Orlando, Florida, 6th-9th May 2002, pp107-113

Dymon UJ and Winter NL (1993) Evacuation Mapping: the Utility of Guidelines *Disasters* **17(1)** pp12-24

Dombroski M, Fischoff B and Fischbeck P (2006) Predicting Emergency Evacuation and Sheltering Behaviour: A Structured Analytical Approach *Risk Analysis* **26(6)** pp1675-1688

Drury J (2004) No Need to Panic *The Psychologist* **17** pp118–119

Drury J and Cocking C (2007) *The Mass Psychology of Disasters and Emergency Evacuations: A Research Report and Implications for Practice*. University of Sussex, UK

El Rhalibi A and Taleb-Bendiab A (2005) Harnessing Agent-Based Games Research for Analysis of Collective Agent Behaviour in Critical Settings *In: Proceedings of the second GSFC/IEEE Workshop on Radical Agent Concepts (WRAC'05)*, NASA GSFC Visitor's Centre, Greenbelt, MD, 20th-22nd September 2005

Great Britain (2005), *Civil Contingencies Act 2005: Elizabeth II. Chapter 39*, The Stationary Office, London, UK

Johnson Jr. JH (1985) A Model of Evacuation: Decision Making in a Nuclear Reactor Emergency *Geographical Review* **75(4)** pp405-418

Kirwan B and Ainsworth LK (1992) *A guide to task analysis.* Taylor and Francis, London

Lachman R, Tatsuoka M and Bonk WJ (1961) Human Behaviour during the Tsunami of May 1960: Research on the Hawaiian disaster explores the consequences of an ambiguous warning system *Science* **133** pp1405-1409

Laughery R (2005) Simulation and modelling for analysing human performance. *In:* JR Wilson and N Corlett (Eds) *Evaluation of Human Work*. Taylor and Francis, London, pp219-240.

Mawson AR (2005) Understanding Mass Panic and Other Collective Responses to Threat and Disaster *Psychiatry* **68(2)** pp95–113

Perry RW and Lindell MK (2003) Understanding Citizen Response to Disasters with Implications for Terrorism *Journal of Contingencies and Crisis Management* **11** pp49–60

Quarantelli EL (1996) Basic Themes Derived from Survey Findings on Human Behaviour in the Mexico City Earthquake *International Sociology* **11(4)** pp481-499

Silverman BG, Might R, Dubois R, Shin H, Johns M and Weaver R (2001). Toward a Human Behaviour Models Anthology for Synthetic Agent Development *In: Proceedings of 10th Conference on Computer Generated Forces and Behavioral Representation.* Norfolk, VA, 15th -17th May 2001, SISO & IEEE, New York, pp277-285

Silverman BG, Johns M, Cornwell J and O'Brien K (2006) Human Behaviour Models for Agents in Simulators and Games: Part I - Enabling Science with PMFserv *Presence:Teleoperators and Virtual Environments* **15(2)** pp139-162

Sime JD (1999) Crowd Facilities, Management and Communications in Disasters *Facilities* **17(9/10)** pp313–324

Still GK (2000) Crowd Dynamics. *PhD thesis*, Mathematics Department, Warwick University, UK.

Biography

Věra Karasová is a PhD student at the Centre for Geospatial Science, University of Nottingham. Her research focuses on issues related to planning and management of response to large urban CBRN incidents. She is particularly interested in application of agent-based modelling into the disaster management domain. She obtained her MSc. at Helsinki University of Technology and her undergraduate degree at Czech Technical University in Prague.

Glyn Lawson in a Research Fellow and Part-Time PhD student at the Human Factors Research Group, University of Nottingham. His research concerns the analysis of methods for predicting human behaviour in emergency situations. He came to the University having previously spent several years working as an Ergonomist in industry.

Using Urban Viewsheds for Embedding Geographical Context in Photograph Databases of Urban Areas[12]

Thierry Joliveau[1] and Sanjay Rana[2]

[1]CRENAM-ISIG CNRS/UMR 5600, Université Jean Monnet - Saint-Etienne
6 rue Basse-des-Rives, 42023 Saint-Etienne Cédex 02, France
Tel. 0033477421925 Fax 0033477421924
Email thierry.joliveau@univ-st-etienne.fr

[2]CTS, University College London, Gower Street, London WC1E 6BT, UK
Tel. 00442076797873 Fax 00442076791567
Email s.rana@ucl.ac.uk Web address www.ucl.ac.uk/~ucessan/

KEYWORDS: Geo-tagging, EXIF, GPX, Urban Viewshed, Isovist

1. Introduction

Architects, geographers, historians and landscape-planners are interested in understanding the changes in the morphology of urban forms and landscape (Joliveau and Dupuis 2006). An approach to analyse and visualise these morphological changes is based on the use of photographs, which are particularly useful in documenting the rapidly changing modern urban environments. With the advent of cheap digital cameras, photographs are routinely collected for various purposes such as post-cards, public works documents, newspapers, even at regular intervals for historical records. These collections of photographs are stored in dedicated photographic-database software, which generally do not provide the geographical context of the photograph thus, it remains unknown exactly which parts (i.e. buildings, streets so on) of the city have been recorded in a photograph. The geographical context of a photograph allows the development of systematic and automated approaches to record the changes in urban form.

The geo-referencing of a photograph requires three main types of information about the photograph, namely the camera location, camera Field of View (FOV), and the camera azimuth i.e. the geographic orientation of the photograph. In the vast number of current and old photographs the location of the camera is recorded manually on a map. However some modern cameras, which have built-in GPS can usually automatically, retrieve this information except in dense built-up areas where the GPS reception can be poor. Photographs taken from the air necessarily require a GPS or some other georeferencing technique. The derivation of the camera FOV is relatively straightforward since it is directly related to the current focal length of the camera lens. The geographic orientation of the photograph is derived manually using a compass. In case of modern digital cameras, these three types of information are often stored in the Exchangeable Image File Format or EXIF[13] (only focal length) or the GPS Exchange Format or GPX[14] (except orientation) tags of digital photograph file however information on photographs taken from other kinds of cameras have to be manually recorded. Even after one has been able to record these thee types of information about the photograph, a manual identification of the

[12] A preliminary version of the paper was presented at the GeoComputation 2007 Conference as a Poster

[13] More information on EXIF is available at www.exif.org (last accessed on 24th April 2007

[14] More information on GPX is available at www.topografix.com/gpx.asp (last accessed on 24th April 2007)

urban features visible in a photograph can be a tedious and imprecise process when dealing with hundreds of photographs and the complicated urban layout.

This paper demonstrates the novel use of the concept of urban viewsheds in automatically identifying the urban features given the camera location, camera FOV and camera azimuth of the photographs.

2. Methodology and Results

An urban viewshed or isovist is the area visible from a viewpoint (Benedikt 1979). The procedure to identify the urban features visible in a photograph involves the following two steps:

Step 1 Generate the isovist at the viewpoint of the camera
Given the outlines of the buildings or building footprints, camera location (recorded by GPS etc.), camera FOV and camera orientation, the isovist of the camera viewpoint is generated using Isovist Analyst[3], an ArcView® GIS developed by Sanjay Rana. In the case of 2D open space, the isovist is naturally a 2D polygon as shown in Figure 1

Step 2 Perform Spatial Overlap Query to identify urban features
Given the isovist polygon, building footprints (with attached attributes such as names) and in fact any kind of 2D geospatial data with relevant attributes, it is trivial to perform various spatial overlap query to establish the geographical context of the photographs and objects visible from the photographs. For instance, one could find all the photographs in the database, which are supposed to represent the façade of a particular building. Fig. 1 shows the isovists of various camera viewpoints in a part of central Lyon (France). Figure 2 shows the results of an overlap query on identifying the photographs that contained the building shaded red. Specifically, the spatial query essentially involves the search of all isovist polygons that intersect the building polygon shaded red. Now since, each isovist polygon is associated or joined (in database terminology) to a photograph, it is straightforward to retrieve the photographs that contain a view of the particular buildings.

3. Future Work

In the present work, we have demonstrated the application of 2D urban viewsheds or isovists in querying the geographical context of photographs. However, photographs capture a 3D representation of the urban forms therefore, we intend to generate 3D isovists while expecting to deal with some non-trivial computation geometry in performing 3D overlap queries.

References
Benedikt ML (1979) To take hold of space: isovists and isovist fields *Environment and Planning B* **6** pp47-65
Joliveau T and Dupuis B (2006) Conception et Utilisation de Visualisations Numériques pour la Gestion Paysagère *Revue Internationale de Géomatique* (*European Journal of GIS and Spatial Analysis*) **16** pp115-134.
Biography
Professor Thierry Joliveau is Director of CRENAM at the University of Jean Monnet at Saint Etienne. His research interests include various topics in the field of physical, human and social geography.

Dr Sanjay Rana is Research Fellow at the Centre for Transport Studies. His areas of research interests include digital elevation models, visibility analyses for urban and natural environments, geovisualisation, and geospatial techniques for road traffic accident analysis.

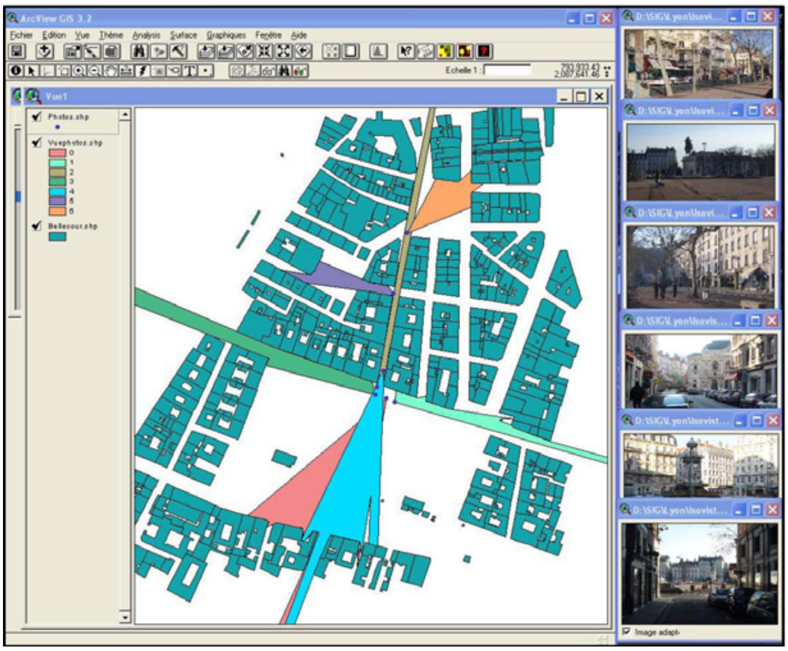


Figure 1. Photographs and their respective isovists in a part of Lyon, France. The small discs between the building blocks and the polygons extruding from these discs are respectively the location of the camera and isovists.

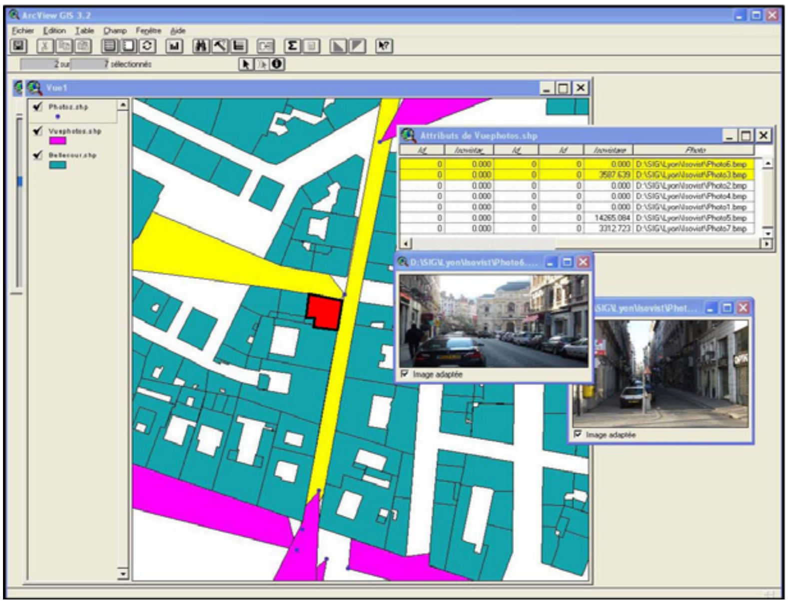


Figure 2. Automatic retrieval of the two photographs that contain a view of the building shaded red, using spatial overlap between isovists and building polygons.

Visualising Neighbourhood Well Being: GIS in an urban environment

Mike Franks
Oldham Metropolitan Borough Council
Corporate Research Team
t: 0161 770 3183
e: mike.franks@oldham.gov.uk

1. Background

The *Neighbourhood Well Being Index* is a data visualisation tool currently being developed by OMBC's Corporate Research Team for natural neighbourhoods (agreed with Area Managers and elected Members) that mostly nest within ward boundaries.[15] A partnership resource, the system will support the localism agenda and neighbourhood level work in particular. The visualisation tool will be adapted to display the Indices of Deprivation for Oldham SOAs[16] and to enable comparison between those issued in 2004 and 2007.

Compiled largely from operational and managerial information systems currently in use[17], the system provides a consistent view of population dynamics in the Borough, an ability to analyse change over time, and will complement existing focused operational and 'hot-spotting' systems. In essence, high and low rates mapped as a series of 'visual exception reports', will enable users to quickly identify issues of relevance. Data has been collated quarterly since May 2005.

Detailed analysis and examination of the data, involving Council Members, Area Managers, and Officers with local knowledge from OMBC and partner organisations will be an important part of the development process. By tapping into local knowledge and expertise, 'trigger' mechanisms can be refined. This will ensure that the intelligence provided remains relevant to Community Councils, for instance.

Whilst *well being* is a contested concept with a variety of definitions,[18] long term issues such as sustainable consumption and ecological footprints were picked up by the Audit Commission[19] and formed part of their Area Profiles. UK cities such as Sheffield, Salford and Newcastle have developed systems tracking neighbourhood change, as has the Borough of Sandwell mentioned in a study for the City of Toronto which looked at initiatives in the US, Europe and Canada.[20]

2. Objectives:

- track neighbourhood change, identify trends and sudden change within neighbourhoods;
- identify neighbourhoods with particular vulnerabilities;
- inform discussions and enable precisely targeted interventions;
- provide intelligence to Community Councils and other local initiatives.

---

[15] Apart from the Town Centre, the other 113 neighbourhoods have at least 200 households (some have near or above 1000)

[16] For explanation of acronyms see the Glossary at the end of the abstract

[17] Data sources - OMBC LLPG, CTAX/HBIS: vacancies, movers, benefit recipients; CSU/GMAC: domestic burglaries

[18] Wellbeing Concepts & Challenges, Fiona MCAllister, December 2005, SDRN

[19] Local Quality of Life Indicators, August 2005, Audit Commission

[20] Measuring Neighbourhood Vitality http://www.urbancentre.utoronto.ca/pdfs/curp/SNTF_Neigh-Vitality_RP3.pdf

The *Neighbourhood Well Being Index* collates data into four **themes**

- households in receipt of CTAX benefits (CTB)[21]
- domestic burglaries supplied by Community Safety Unit (CSU) from GMAC
- movers (closed CTAX accounts)
- vacancies (CTAX linked to LLPG)

A quarterly snapshot is taken of occupied residential properties, vacancies and households in receipt of CTB and the total of movers and domestic burglaries during the quarter is accumulated. These data yield CTB recipients as percentage of occupied households and rates per 1000 households in each neighbourhood for domestic burglaries, movers, and vacancies.

The *Neighbourhood Well Being Index* **measures**

- trends (up or down) for three or more quarters
- sudden change (for better or worse)
- severity and persistency (highest and lowest rates)

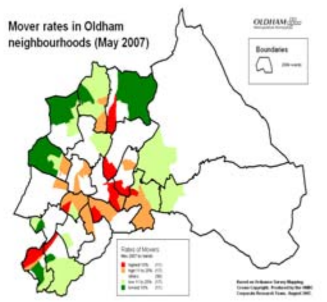
3. Headlines:

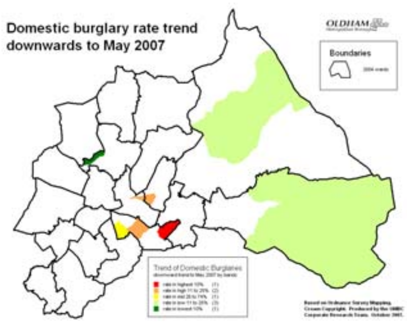
- volatility in numbers (and rates) of domestic burglaries across neighbourhoods are detected over time
- vacancies are highest in (or nearby) HMR intervention areas
- rates of movers have been falling during the last year
- rates of households in receipt of CTB are highest in Coldhurst
- an upward trend in the rates of households receiving CTB is only detected in neighbourhoods with low rates

4. Trends[22]:

CTB – rates of CTB recipients below 6.4% of all households in six neighbourhoods are ranked in the lowest three quintiles (compared to the Oldham average 7.1%) and show an upward trend over three quarters. Four have shown an upward trend for a year. Conversely thirteen neighbourhoods have shown a downward trend for three consecutive quarters. Of these, one in Coldhurst and another in an HMR intervention area in Werneth have rates of 13.2% and 11.7% respectively, ranked in the highest 10% band.

Domestic burglaries – four neighbourhoods show an upward trend including the two neighbourhoods with the highest ranked rates, located in St. James' (19.9 per 1000) and Medlock Vale (18.0 per 1000). Yet there is a downward trend in rates of domestic burglaries in eight neighbourhoods, including one in Alexandra (rate 12.7 per 1000) ranked in the highest 10% band. This is illustrated in the map opposite.[23]





[21] CTB here **excludes** households also in receipt of Income Support or Housing Benefit

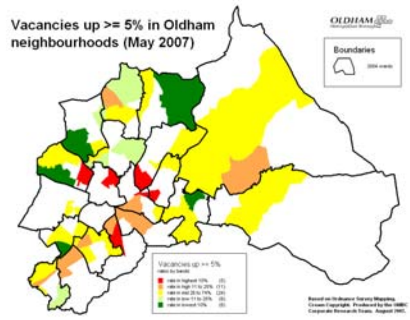
[22] Trends of rates for three or more consecutive quarters up or down to May 2007.

[23] A modified traffic light scheme is used in maps with red - indicating rates ranked in the highest 10% band, amber - the high 11-25% band, light green - the low 11-25% and dark green - indicating rates ranked in the lowest 10% . In maps displaying rates across the four themes, neighbourhoods with rates ranked in the 26-74% band are white, whereas in maps of trends and sudden changes, those neighbourhoods are shaded yellow, whilst those in white show no trend or sudden change. White is also used for neighbourhoods where no persistently high or low rates are detected.

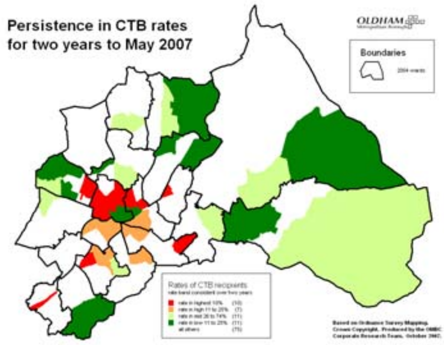
Movers – no upward trend is present over three consecutive quarters. However, a downward trend in the rates of movers over the same period can be detected in thirty five neighbourhoods.

These rates vary from the third highest with 58.8 movers per 1000 occupied households in one neighbourhood of St. Mary's (down from 100) to that ranked lowest with a rate of 6.5 (down from 55) in one neighbourhood of Shaw & Crompton. Bands of mover rates are illustrated in the map left.
Movers may be a measure of churn or of a healthy housing market, and therefore higher rates are not necessarily negative. However, the downward trend may indicate a slow down in the housing market.

5. Sudden change[24]:
CTB – households in receipt of CTB increased in twenty one neighbourhoods across Oldham. Only two had rates ranked in the highest 25% (both in Hollinwood) and nine had rates ranked in the lowest 25%.
Domestic burglaries – a sudden increase of domestic burglaries was detected in thirty three neighbourhoods, nine of which had burglary rates ranked in the highest 10% band and another eleven ranked in the highest 25%. No sudden increase was found in neighbourhoods where burglary rates were ranked in the lowest 25%.
Movers – despite the downward trend overall in rates of movers, a sudden increase in the number of movers was detected in twenty two neighbourhoods. Five had rates of movers ranked in the highest 25% and two had rates in the lowest 25%.
Vacancies – numbers of vacancies increased by over 5% in more than fifty neighbourhoods across Oldham. Sixteen of these had rates ranked in the highest 25% as well as twelve ranked in the lowest 25%. This is illustrated in the map opposite.



Persistent rates:



CTB – four of the five Coldhurst neighbourhoods outside the Town Centre have had rates of households in receipt of CTB ranked in the highest 10% of neighbourhoods across Oldham over the last two years, with two of these neighbourhoods sharing highest rankings. One Chadderton neighbourhood has a rate that is also consistently ranked in the highest 10%, and three in Hollinwood are ranked in the high 11 to 25% band. Conversely, another Hollinwood neighbourhood and four in Chadderton have CTB rates that have been consistently ranked in the lowest 25%. Broadly in keeping with the known pattern of deprivation in the Borough, nearly half the neighbourhoods in Shaw & Crompton, and in Saddleworth & Lees have CTB rates that have been consistently ranked in the lowest 25% over the two years up to May 2007. The pattern of persistent CTB rates is illustrated in the map above.
Domestic burglaries – due to the volatile incidence of domestic burglaries there are no neighbourhoods that have exhibited high or low rates consistently over the last two years. Yet one neighbourhood in the south of Waterhead had a rate of domestic burglary ranked in the highest three for over a year up to November 2006. Conversely, one neighbourhood in Royton and one in Chadderton had rates of domestic burglaries that were ranked in the lowest 25% across Oldham for three or more quarters during the two years up to May 2007.
Movers – the neighbourhood in Waterhead suffering a high rate of domestic burglaries also exhibited a rate of movers that was consistently ranked in the highest 10% over the last two years across

[24] Sudden increase in **numbers** by over 5% between February and May 2007

Oldham. Neighbourhoods nearby HMR intervention areas and the NDC area along with one in St. Mary's and another in Saddleworth & Lees also had rates of movers that were consistently in the high 11 to 25% band over the two years up to May 2007. The Chadderton neighbourhood with a low rate of domestic burglaries along with another in Royton, also had rates of movers consistently ranked in the lowest 10% for the last two years.
Vacancies – vacancy rates have been consistently ranked in the highest 10% in the HMR intervention areas and in the Waterhead neighbourhood where there is also a high rate of domestic burglaries and movers. Neighbourhoods in Alexandra, Medlock Vale, St. Mary's and one in Shaw & Crompton all have vacancy rates that have been consistently ranked in the high 11 to 25% band over the last two years. This contrasts with one neighbourhood in each of Failsworth, Hollinwood, St. James' and Saddleworth & Lees, and more in Chadderton, Royton and Shaw & Crompton, where rates of vacancies have been ranked in the lowest 25% for the last two years.

6. Other issues:
Whether links may be detected between different themes such as burglaries, movers and vacancies, and issues such as data quality and using the LLPG to link and reconcile data from independent operational systems will be explored.

7. Glossary:

CSU	Community Safety Unit
CTAX	Council Tax System
CTB	Council Tax Benefit
GMAC	Greater Manchester Against Crime
HBIS	Housing Benefit Information System
HMR	Housing Market Renewal
LLPG	Local Land and Property Gazetteer
NDC	New Deal for Communities
OMBC	Oldham Metropolitan Borough Council
SOAs	Super Output Areas

Modelling Street Safety in an Urban Environment

Claire Ellul[1], Ben Calnan[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
c.ellul, b.calnan@londonmet.ac.uk , http://www.citiesinstitute.org

KEYWORDS: transport safety urban environment isovist

1. Introduction

In the context of public transportation, *"the provision of a permeable public space contributes to an inclusive journey environment"* (Azmin-Fouladi 2007). In fact, over 11% of the public would travel more if they felt safer on the transport system (Department for Transport, 2004, cited in Azmin-Fouladi 2007). However, when planning or modelling an urban environment, architectural vision and planning principles often take precedence over the way buildings and urban features make people feel. Questions such as: *does the built environment provide a feeling of safety and security? how do such feelings contribute to an accessible transportation system? which urban features determine these perceptions?* are not generally considered.

Our research aims to redress this imbalance by providing planners and local authorities with the means to identify potential barriers to the permeability of public space. It is argued that the removal of negatively-impacting features and the resulting increase in perception of safety will increase the use of public transportation. This paper presents work in progress of which there will be two key outputs: the generation of an Index of Permeabilility (IoP) for the urban environment and the operationalisation of this through the application of an Isovist.

2. Background

This project forms part of AUNT-SUE (Accessibility and User needs in Transport for Sustainable Urban Environments). 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. (AUNT-SUE, 2007). Three key aspects are considered in the context of an inclusive journey environment – quality, safety and physical accessibility. Work carried out by Imam (2006 cited in Azmin-Fouladi, 2007) focuses on accessibility. The research presented here thus relates to quality and safety of the journey.

Although "identifying perceptual variables including elements that contribute to the fear of crime, sense of place/identify/ownership, community cohesion and place attachment requires wide-scale public consultation" (Azmin-Fouladi, 2007), this must be carried out within the context of a pre-determined range of urban elements, to focus the research in the required direction. The first step of the research thus involved an extensive review of literature to identify factors influencing a person's perception of urban space. These included both those contributing to a positive feeling about an area, such as a high number of ground floor windows being used for commercial purposes, and negative factors such as evidence of vandalism or graffiti. Additional considerations include the number of non-ground floor windows and their use, street furniture (lights, rubbish bins, bollards), the presence or absence of CCTV, buildings and garden upkeep, and the heights of boundaries such as walls, fences.

Figure 1and Figure 2 show examples of a negatively perceived area, having high boundaries and no overlooking windows. Figure 3 illustrates an area having wide-open spaces, low walls and a well-kept frontage, contributing to a positive feeling about the space. (All photos are taken in our current test area of Somerstown, London).

Figure 1 - A road with a high, long boundary one side and a building with very few windows looking down onto it. (Azmin-Fouladi 2007)

Figure 2 - A shop front with covered windows. Poor natural surveillance of the surrounding area is hypothesized as a contributing negative factor to people's perceptions of their environment (Azmin-Fouladi 2007)

Figure 20 - Some positive urban features such as well-kept frontage and promenades (Azmin-Fouladi 2007)

2. Creating an Index of Permeability (IoP)

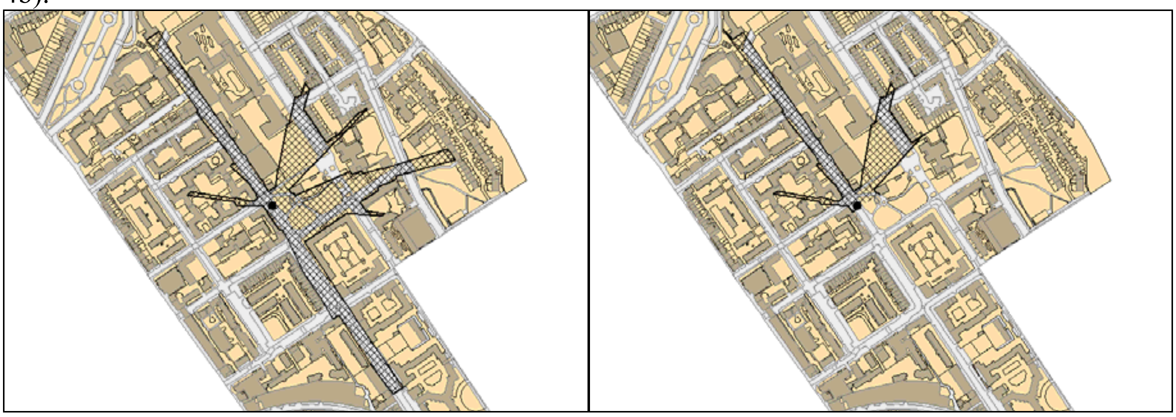
2.1. Identifying and Capturing Relevant Features

Our methodology first breaks down the urban landscape into its constituent elements, as identified by the first stage of this research including walls, fences, litter, graffiti and natural surveillance. Ordnance Survey MasterMap™ provides topographic mapping, which was then associated with local authority data for details of street furniture. This data is supplemented by data recorded in the field for the other urban elements. All data is held in a spatial database and each element type is assigned a weighting depending on its perceived importance to the permeability of an urban space.

2.2. Creating the Isovists

An isovist is a "region of space which can be described by the shapes obtained from people's vision if they rotate through 360 degrees" (Benedikt 1979, cited in Davies *et al.* 2006). They are thus particularly appropriate to model the space that can be seen from a particular point (a perspective point), and hence to identify the specific urban elements that are visible to the pedestrian.

In the case of the IoP, the boundary of the isovist is formed by tracing the boundary of the buildings that are visible from the selected point. This boundary is identified through the generation of a series of rays, which are then intersected with the building polygons forming part of the Ordnance Survey MasterMap™ Layer. Although traditionally isovists delineate the space visible through 360 degrees (Figure 4a), it is also possible to create an isovist focussed in one particular direction of travel (Figure 4b).



(a) (b)

Figure 4 (a) - Single 360 Degree Isovist (b) Single 200 Degree Isovist (created using OS MasterMap 2007 Crown Copyright)

2.3. Calculating the Index of Permeability for a single Perspective Point

The isovists are used to identify the urban elements that can be perceived by a pedestrian standing at a particular point, and the weightings for each element summed to create the Index of Permeability for that point. A high value implies good permeability. Where appropriate, the horizontal and vertical dimensions of each element are also factored into the calculation (for example, a long high wall generates a lower weighting than a long low wall). A large pavement café rates more highly than a café with no outside space. The Index also takes into account the distances of a particular element from the pedestrian – for example, streetlights further away are assigned a lower weighting than those in the immediate vicinity.

2.4. Creating a Surface of Permeability

To create a Surface of Permeability, a series of perspective points is generated in a grid across the roads and pavements of the test area. Isovists are then created for each of these points, and used to generate an Index value (Figure 5).

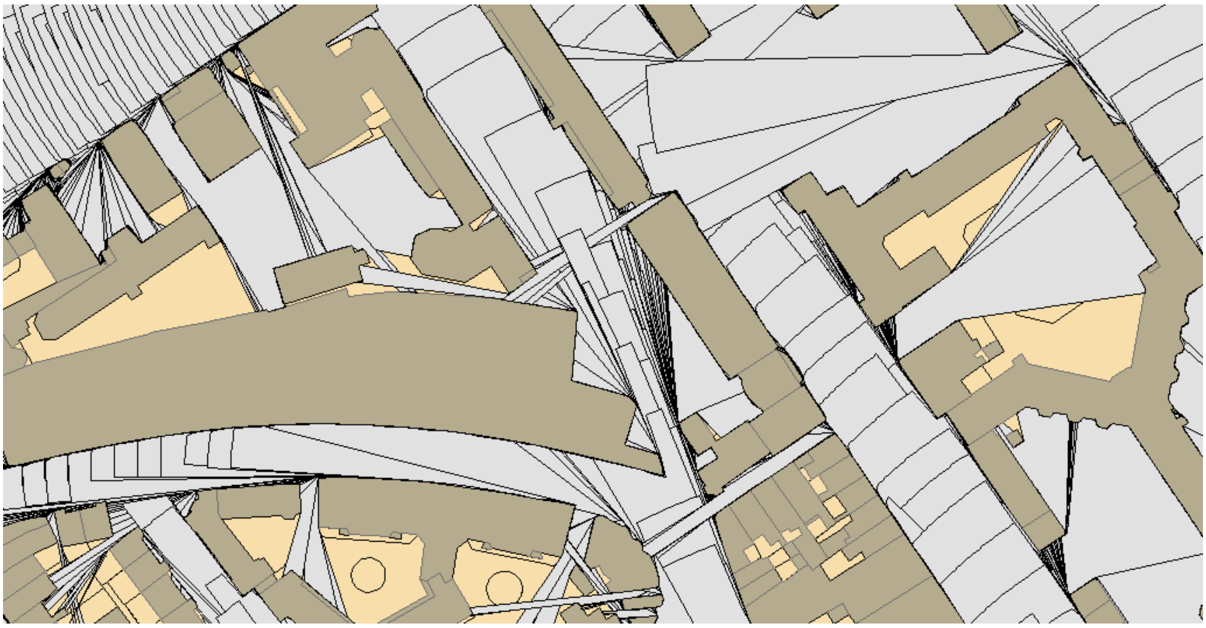


Figure 5 - Multiple Isovists from different Perspective Points on OS MasterMap base data. (Created using OS MasterMap 2007 Crown Copyright)

The Surface of Permeability is then determined by interpolating between the points. Figure 6 shows preliminary results obtained using a 5m grid of perspective points, and then using an inverse distance weighting interpolation over the nearest 5 points. As can be seen, the influence of streetlights (black diamonds) on the resulting surface is positive, whereas graffiti points (black triangles) contribute towards a more negative Index of Permeability**.**

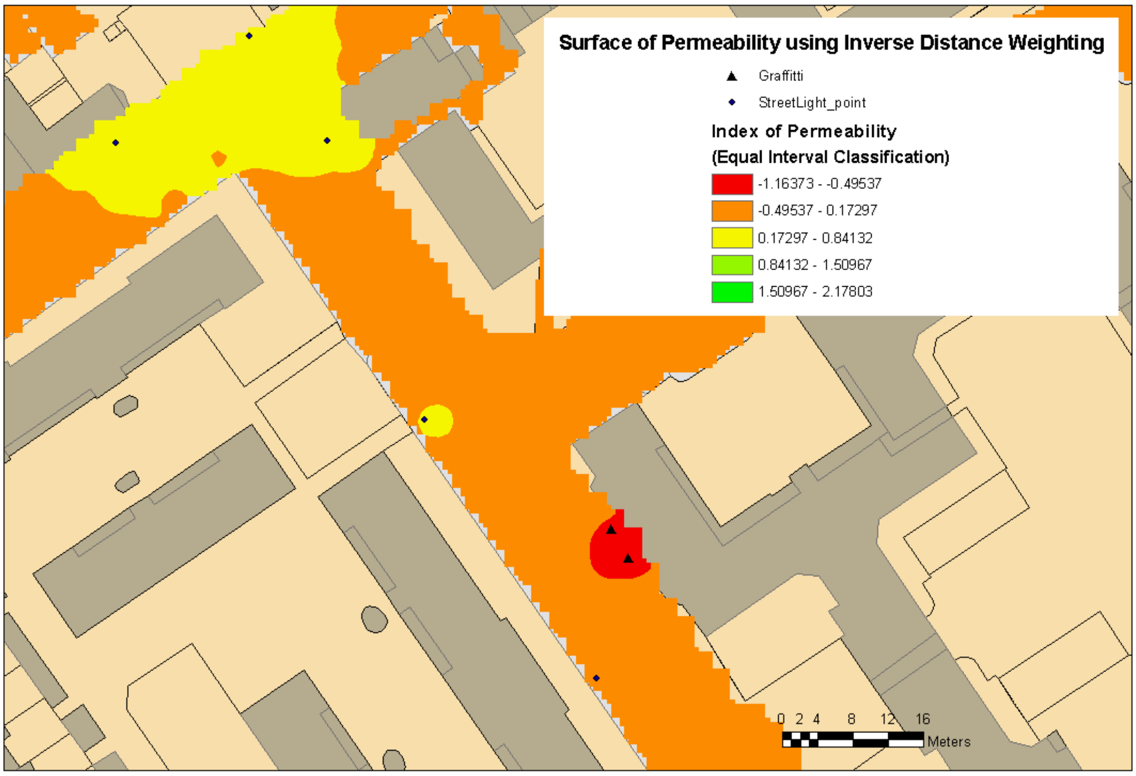


Figure 6 – IDW Surface of Permeability for part of Somerstown (Created using OS MasterMap 2007 Crown Copyright)

It is also possible to model a route or journey a pedestrian might take, for example from their house to a bus stop. Isovists can be placed at regular intervals along this route to uncover areas causing concern to the individual, and investigate the permeability of a particular route.

4. Next Steps

Modelling the effects of urban features on peoples' perceptions is a complex and potentially inaccurate and subjective task. Our implementation of the Index of Permeability innovates in its use of a commonly available desktop GIS software package (ArcGIS), providing a one-click tool for visualising perceptual data with a spatial dimension. This offers advantages over existing approaches such as *PERS* (Pedestrian Environmental Review Software), particularly in the range of urban elements. The creation of isovists entirely within the GIS contrasts with the method proposed by Rana (2005), which makes used of externally created Binary Space Partition trees to index the data.

The method used to calculate the IoP is also flexible, allowing urban elements to be added or removed and their weightings to be increased or decreased in importance. This in turn provides the ability to create a localised IoP, taking into account available datasets and local concerns. The incorporation of Isovists is of particular benefit as it provides the opportunity to model a field of vision with considerably more accuracy than a simple GIS buffer.

As mentioned, the urban elements incorporated into the Index generation process have been derived through literature review, to form the basis of a more focussed discussion with local communities. Thus, the next step in our research involves testing the method and resulting Permeability Maps with the general public by means of focus group meetings. This will allow us to confirm or refute the importance of the selected features and the accuracy of the surfaces obtained.

In the longer term, this method will form part of an AUNT-SUE toolkit to model the whole journey environment. Three potential applications can be identified in this context. Firstly, the method can be used to locate problem areas. The generation of IoP surfaces allows the hotspots (representing poorly perceived areas) and focus on these areas to see which particular features are causing the problems. It could also form the basis of a more complicated Decision Support tool whereby modelled objects could be actively placed or removed from areas, and the effects could then be compared to allow the optimum decision to be made.

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. 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

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

Rana, S. (2005) *Use of GIS for planning visual surveillance installations.* In: ESRI Homeland Security GIS Summit, 2005, Denver, USA.

7. Biography

Claire Ellul has recently completed a PhD on Topological Relationships between 3D Objects, and is currently working as a Post-Doctoral researcher at London Metropolitan University and University College London. Prior to commencing her PhD, she spent 10 years working as a GIS consultant and software engineer.

Ben Calnan is currently 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.

Measuring 'alcohol related' crime and disorder: Examining the impact of extended licensing hours in England.

Andrew Newton[1], Alex Hirschfield[2]

[1]Applied Criminology Centre (ACC), University of Huddersfield,
CSB 14, Queensgate, Huddersfield, HD1 3DH
Tel. +44 (0) 1484 47 3837 Fax +44 (0) 1484 471825
Email a.d.newton@hud.ac.uk, http://www.hud.ac.uk/hhs/dbs/acc/staff/an.htm
[2](Address as above),
Tel. +44 (0) 1484 47 3676 Fax +44 (0) 1484 471825
Email a.hirschfield@hud.ac.uk, http://www.hud.ac.uk/hhs/dbs/acc/staff/ah.htm

KEYWORDS: Licensed Premises, Crime, Spatio-temporal Analysis

1. Introduction

The recent introduction of extended licensing hours in England (Licensing Act, 2003) occurred in November 2005, amidst increasing concerns about rising levels of violence and disorder in urban centres, particularly during the late evening/night-time. This paper describes the difficulties inherent in measuring 'alcohol related' crime and disorder, and evaluating a change in policy that is county wide (as it is difficult to generate a control area). This research uses GIS to implement a multi level analysis approach to examine changes at the micro, meso and macro level, and discusses the difficulties in generating policy change data for the purposes of analysis and evaluation. As a result of this, quantitative crime analysis was supplemented with qualitative fieldwork (participant observation and semi-structured interviews) to assess the impact of the new Act.

2. Research Questions

The overall aims of the research were to provide a baseline indicator of levels of crime and disorder in and around licensed premises, and to examine the impact of the Act on patterns of crime and disorder in and around licensed premises, in each of the five case study areas. A number of specific research questions were formulated to guide this research. These were as follows:

- What patterns of crime and disorder exist in and around licensed premises?
- What other local factors may explain the prevalence of crime and disorder in and around licensed premises?
- Has the introduction of the Act, and the granting of additional opening hours to licensed premises, led to a change in violence and disorder at these licensed premises?
- Have overall levels of crime and disorder within town and city centres changed following the Act?
- Have the peaks of crime and disorder displaced to later or earlier periods?
- Has the profile of crime and disorder in and around licensed premises and associated 'hot spots' changed in relation to venues with additional opening hours?
- Has the Act resulted in unintended consequences such as geographical displacement or diffusion of benefits to surrounding areas?

3. Data and Methodology

There are inherent difficulties in measuring 'alcohol related' crime and disorder, due to the subjective nature of determining whether alcohol was a cause of crime, and this relates to both the victim(s) and or the offender(s). An additional difficulty is that traditional evaluation methodologies could not be employed, as the change in licensing represented a policy change that occurred across the whole country, thus no 'control' area could readily be generated.

The study employed both quantitative and qualitative research methods to answer these questions. Baseline conditions were identified in each of the areas using data on recorded crime, police calls for service for disorder, and data on assaults obtained from Accident and Emergency (A&E) units and from the ambulance service. Much of this covered a period of up to two years baseline and one year post implementation of the Act. The baseline conditions and changes in these offences/incidents following the implementation of the Act were analysed using a range of quantitative techniques. These included:

- Calculating monthly and yearly crime and disorder counts and rates and percentage change;
- Analysing crime and disorder by time of day and day of week;
- Analysis of domestic violence and alcohol 'flags';
- Analysing the age and gender of victims;
- The use of Geographical Information Systems (GIS) to map licensed premises and crime and disorder incidents in each area;
- The identification of clusters of licensed premises (i.e. areas of concentrated drinking) and their coalescence with crime and disorder 'hot spots';
- Analysis of crime and disorder in areas close to licensed premises (50m buffer zones);
- Proportional change analysis of crime and disorder;
- GIS analysis to test for evidence of spatial and temporal changes in crime and disorder;
- Benchmarking changes in crime and disorder in premise clusters against the remainder of the town centres;
- Constructing Resource Targeting Tables (RTTs) to identify the concentration of violence in licensed premises;
- Analysis of accident and emergency data.

One of the key difficulties presented to the research team was capturing accurate information on the capacity of licensed premises (bars and nightclubs), and the current and former opening hours of premises (to generate additional hours), and some of the approaches adopted to address this are discussed in this paper. Qualitative research was used to supplement the quantitative analysis and this involved both participant observation of licensed premises and key drinking areas and face-to-face interviews with bar, door staff and door supervisors. This was informed by the GIS crime analysis.

The data sets used in the research included:

- Police recorded crime data for violence against the person, criminal damage and sexual offences;
- Police calls for service data (disorder incidents only);
- Licensed premises data;
- A&E data;
- Ambulance data;
- Ordnance Survey AddressPoint®;
- Ordnance Survey 1:10 000 scale raster;
- UKBORDERS digital boundaries;
- Office for National Statistics (ONS) mid-2005 population estimates;
- ACORN 2006 population estimates.

5. Key Findings

A number of finding stem from this research, but some of the key findings of the changes to violence against the person are highlighted below.

- Overall, there was a 3% reduction in violence against the person post Act.
- On weekdays, the overall reduction was 4.3% but only 1.2% during weekends.
- Different patterns emerged at individual case study-level.
- The timing of violence shifted forward in time. The timing of this reflected the local areas' policy in granting extended hours.
- The evidence for temporal displacement was stronger when analysis was restricted to violence occurring within 50 metres of any licensed premises, suggesting an association with licence changes. The changes were most marked during the weekends, again suggesting a relationship with the Act.
- In all five case study areas, the concentration of violence increased among high crime licensed premises that actually used six or more additional hours per week but decreased amongst the remainder. This pattern of findings was not found when examining hours applied for.
- An examination of changes in the peaks of violence against the person suggested that post implementation there was a flattening out of peaks in two areas, no observable change in two areas, and a shift in the peak to later in the early morning in one area.
- Most licensed premises falling into the top 15 premises (i.e. the worst performing premises) for violence against the person offences within the baseline period, also occupied this position in the post implementation period.
- A new technique of synthesis mapping (where the change over time between baseline and post-implementation hot spot maps is summarised on a single map) revealed that most of the reductions and increases in violence against the person corresponded with the location of licensed premises concentrated in the key drinking areas.

6. Acknowledgements

This research was commissioned by the Home Office.

References

Hough, M., Hunter, G., Jacobson, J. and Cossalter, S. (in press[1]) *The impact of the Licensing Act 2003 on levels of crime and disorder: an evaluation*. London: Home Office.

Newton, A., Armitage, R., Hirschfield, A., Monchuk, L., Rogerson, M., Wilcox, A. and Hunter, G. (in press[1]) *The impact of the Licensing Act 2003 on levels of crime and disorder: an evaluation*. Annex A: key findings of the Birmingham case study. London: Home Office.

Newton, A., Armitage, R., Hirschfield, A., Monchuk, L., Rogerson, M., Wilcox, A. and Jacobson, J. (in press[1]) *The impact of the Licensing Act 2003 on levels of crime and disorder: an evaluation*. Annex B: key findings of the Blackpool case study. London: Home Office.

Newton, A., Armitage, R., Hirschfield, A., Monchuk, L., Rogerson, M., Wilcox, A. and Hunter, G. (in press[1]) *The impact of the Licensing Act 2003 on levels of crime and disorder: an evaluation.* Annex C: key findings of the Croydon case study. London: Home Office.

Newton, A., Armitage, R., Hirschfield, A., Monchuk, L., Rogerson, M., Wilcox, A. and Hunter, G. (in press[1]) *The impact of the Licensing Act 2003 on levels of crime and disorder: an evaluation*. Annex D: key findings of the Guildford case study. London: Home Office.

Jacobson, J. Newton, A., Armitage, R., Hirschfield, A., Monchuk, L., Rogerson, M., and Wilcox, A. (in press[1]) *The impact of the Licensing Act 2003 on levels of crime and disorder: an evaluation*. Annex E: key findings of the Nottingham case study. London: Home Office.

Hough, M, Hirschfield. A., and Newton, A. (in press[1]) *The impact of the Licensing Act 2003 on levels of crime and disorder*: Annex F. Evaluation methods. Home Office. London.

(note these publications are expected in February 2008)

[1]Note these publications are expected in February 2008

Biography

Dr. Newton is a senior research fellow at the Applied Criminology Centre (ACC), University of Huddersfield (UK). He joined the ACC in January 2005. Previously he worked for 2 years as a research associate at the Environmental Criminology Research Unit (ECRU), University of Liverpool. He received his PhD from the University of Liverpool in 2004. Prior to this attained a BSc in geography from the University of Sheffield, and an MSc in Geographical Information Science (GIS) from the University of Edinburgh. His research interests include environmental criminology, the geography of crime/place of crime, GIS, crime mapping and crime analysis, quantitative and qualitative research methods, policy analysis and evaluation, crime and disorder on public transport, crime and youths, acquisitive crime, and violence and disorder in the night-time economy.

The Importance of Workplace Quality of Life in Determining Office Location

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: workplace quality of life, indices, gravity modelling, small to medium enterprise

1. Introduction

Quality of life (QoL) is understood as "the extent to which environments, social and physical, are conducive to a state of happiness" (Wingo, 1973: 4) and for the individual, this reflects the consumption of various goods and services in the surrounding environment (Gillingham and Reece, 1979: 329). It has been implied that the well-being of employees may be an important aspect of locational choice for business (Rogerson, 1999; Lambiri *et al*., 2007; Rabianski, 2007) and that QoL offers locational advantage (Rogerson, 1999) and an economic good for business (Wingo, 1973) that in turn may shape the attractiveness of an urban location with implications for further growth and investment (Lambiri *et al*., 2007). However, to date there is limited empirical research (Grayson and Young, 1994; Rogerson, 1999; Lambiri *et al*., 2007) with much of the evidence relating to large organisations (Healey and Baker, 1993). Sufficient attention is not paid to small businesses, which are the main employers in the UK, the source of new business and innovation, and central to future aspirations for sustainable economic development through UK regions (South West Regional Assembly, 2006; South West of England Regional Development Agency, 2006) in a sub-regional and city-region context (HM Government, 2007). In order to understand and ensure future sustainable economic growth, it must be asked: What role do QoL and sustainability have in small businesses locations?

The spatial analysis of geographic information in a Geographic Information System (GIS) offers great potential to create indices that reflect the level of workplace QoL and sustainability at any given location. The wealth of data available provides opportunity for analysis (Craglia *et al,* 2004). GISs allow us to examine where businesses locate and how employees access the workplace across the transport network, building in environmental, social and monetary costs and benefits. They can be used to investigate the degree of influence that the goods and services at a location have on business decision-making. Spatial interaction and gravity modelling provide a platform for analysing the attractiveness of a location in terms of workplace QoL and accessibility in relation to the distance or time taken to reach it from the employee location. Transport mode reflects the sustainability of this access. This type of modelling has been well cited in the literature in the urban context (Ingram, 1971; Keeble *et al*, 1981; Haynes and Fotheringham, 1984; Geertmann and van Eck, 1995; Levinson, 1998; Copus, 1999).

2. Research Aims

The aim of this research is to investigate the sustainability of locational choices based on workplace quality of life indicators and accessibility measures for office-based small to medium enterprises (SMEs) in the West of England area (figure 1), through the use of geographic information analysis. A GIS has been utilized to create workplace QoL and sustainability indices at the regional and local scale and this will be used to weight potential office locations throughout the study area. The gravity modelling concept is then used to assess the sustainable economic potential of these locations and extended to understand optimum locations for business based on employee locations. The model is then compared to actual business locations in order to understand the influence of workplace QoL and accessibility on the spatial distribution.

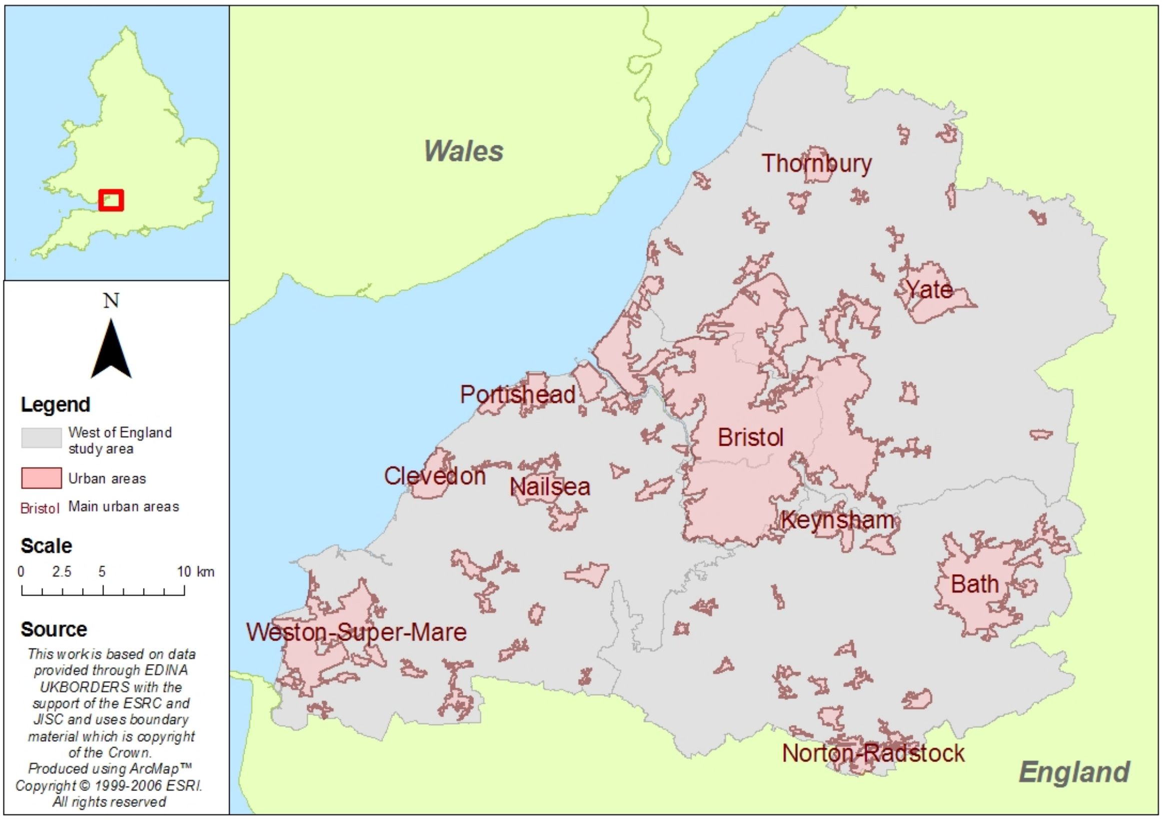


Figure 1. The West of England study area and associated urban areas

3. Research Questions

The following three research questions have been identified:

A. What constitutes a sustainable location for office-based SMEs in the West of England?
B. Which criteria are important in office-based SMEs choice of location in the West of England?
C. To what extent does workplace quality of life and sustainable accessibility influence the spatial distribution of office locations?

4. Research Design/Methodology

4.1 Definition of Terms

Exploratory face-to-face semi-structured interviews with a range of businesses carried out in summer 2007 revealed that staff well-being and ease of access to facilities is a major factor for economic success. In consultation with relevant literature, 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 Indices Creation

There are 73 urban areas in the West of England area (Urban Area Census 2001, Edina®) that represent potential office locations at the regional scale (destinations for the gravity model, section 4.5). Each potential destination is being quantified according to workplace QoL and sustainable accessibility criteria (indices) to prepare for regional analysis of company location decisions (research question A). These urban areas are then be disaggregated into local scale potential office locations, identified through the use of cluster analysis of businesses (figure 2), to be fed into the gravity model. These indices have been identified from the literature and exploratory interview research and will be adapted accordingly following a survey of businesses in the study area (section 4.3). Geographic information is being collated in order to construct a series of indices such as those listed in table 1.

The information used to construct these indices is taken from an extensive source of either publicly available data or data available under licence for academic use only. The raw data require manipulation, interpolation and formatting in a GIS to transfer vector, raster, non-spatial and non-georeferenced data (using Ordnance Survey Code Point® data from Edina®, 2007) into indicators. A weighted raster surface is created for each indicator, and then map algebra is used to combine into weighted indices covering the study area, at sub-regional and local scales.

Careful attention is given to potential sources of error, aggregation and interpolation at this stage. Inappropriate disaggregation, the arbitrary grouping of variables (Stover and Leven, 1992 cited in Lambiri *et al.*, 2007: 14) and the dictation of pre-constricted indicators to survey respondents has been carefully assessed and avoided where possible at all stages.

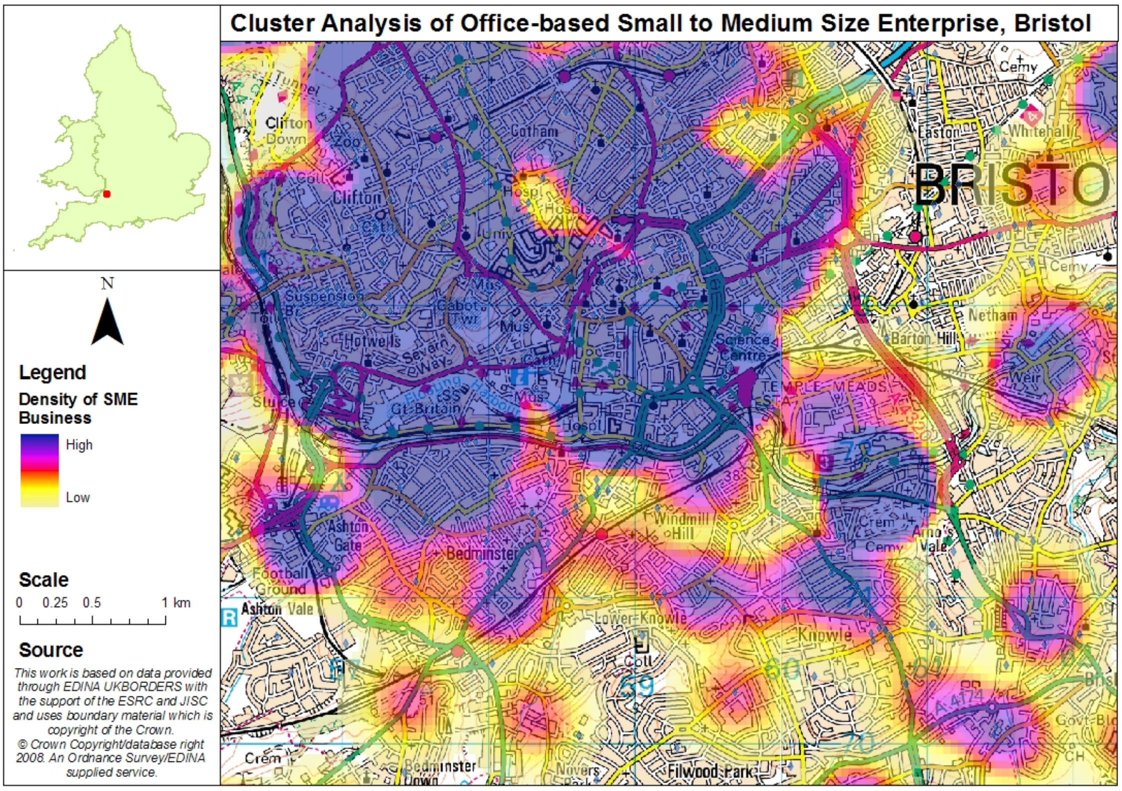


Figure 2. Identification of local scale concentrations of SME office locations through the use of cluster analysis.

Table 1. Workplace quality of life indicators

Proximity to major road network
Proximity to rail termini
Proximity to bus stops
Provision of cycle lanes/pedestrian routes
Commuting distance
Environmental quality (surroundings)
Safety
Ease of vehicular access
Parking provision
Access to services (GPs, schools, opticians, dentist, pharmacy)
Access to facilities (shops, childcare, supermarket, PO, hotels)
Access to entertainment (restaurants, public houses, theatre, cinema, museums)
Access to town centre
Access to green space/parks

4.3 Survey

An online questionnaire survey has been delivered to 1430 office-based businesses currently located in the West of England study area (research question B). These have been selected from Jordans© company information database through the Financial Analysis Made Easy (FAME) online facility (BvDEP, 2007). This stratified sample was based on the UK Standard Industry Classification (SIC) 2003 code and the European Commission definition of SME (2003/361/EC). The locations have been georeferenced using the Ordnance Survey Code Point® data. The purpose of this survey is to understand the influence of workplace QoL and sustainability on location decision-making based on the following four considerations:

a) What location criteria are considered? (criteria identification)
b) How important are these criteria? (criteria weighting – Likert scale)
c) Where are these criteria? (spatial identification)
d) How are these accessed? (network traversing)

The results from the survey will allow the validation of factors and weighting of the indices; it will identify how important these criteria are to decision-makers in relation to other locational considerations; and it will ascertain how the network is traversed by asking travel preferences. Additionally, it will identify the origin of the employees to feed in the GIS modelling.

As the answers will be pre-coded, the data can be imported and analysed directly in a statistical package (SPSS, Excel) to test for correlation and to assess weightings.

4.4 Network Analysis

The road network has been extracted from the Ordnance Survey (OS) Meridian™ 2 1:50000 data using motorways and A/B/minor roads. This has allowed the calculation of actual distances and time estimations between origin and destination to be calculated after setting correct connectivity specification through the Network Analyst OD Cost Matrix facility in ESRI's® ArcMap™ 9.2. A second 'sustainability aware' network includes railway and bus routes/stations (using the National Public Transport Access Nodes (NaPTAN) information from the Department for Transport), incorporating roads to be used for public transport, walking and cycling.

4.5 Gravity Modelling

Gravity modelling will be used to examine the relationship between office location, QoL factors and employee location (research question C). The initial modelling stage has used the Keeble *et al* (1981) gravity modelling concept (equation 1), 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. The economic potential for areas in the WoE has been assessed using the road network to model the flows between origin and destination (comprising the X and Y centroids of the 73 urban areas mentioned in 4.2). The size of the urban area was used as a proxy measure of attractiveness (figure 3) and this will be replaced with a QoL weighting at each location (initially using the literature, then compared with weightings derived from the results from the survey as stated in 4.2 and 4.3). This model is also being used with the sustainability aware network in order to assess how sustainable accessibility influences economic potential of urban areas.

The gravity model concept is used to consider actual employee locations in order to ascertain where the optimum location for an office premises would be. This is based on three scenarios: purely the locations of the employees, incorporating location QoL factors, and using a sustainable accessibility network. Correlation analysis will be carried out in order to compare rateable values of office locations with QoL values to assess the relationship between quality of life, accessibility and the value that the office commands.

The probalistic gravity model developed by Huff (1964) (equation 2) will be tested for its ability to predict flows of employees to and from origins and destinations. Locations of business and employees obtained from the survey will be used to assess the accuracy of the model at predicting the probability of employees travelling to and from destinations for employment.

$$P_i = \sum_{j=1}^{n} \frac{M_j}{D_{ij}} \qquad (1)$$

Where:

P_i	=	the economic potential for location i
Mj	=	an economic "mass" variable of location j
Dij	=	the distance between locations i and j

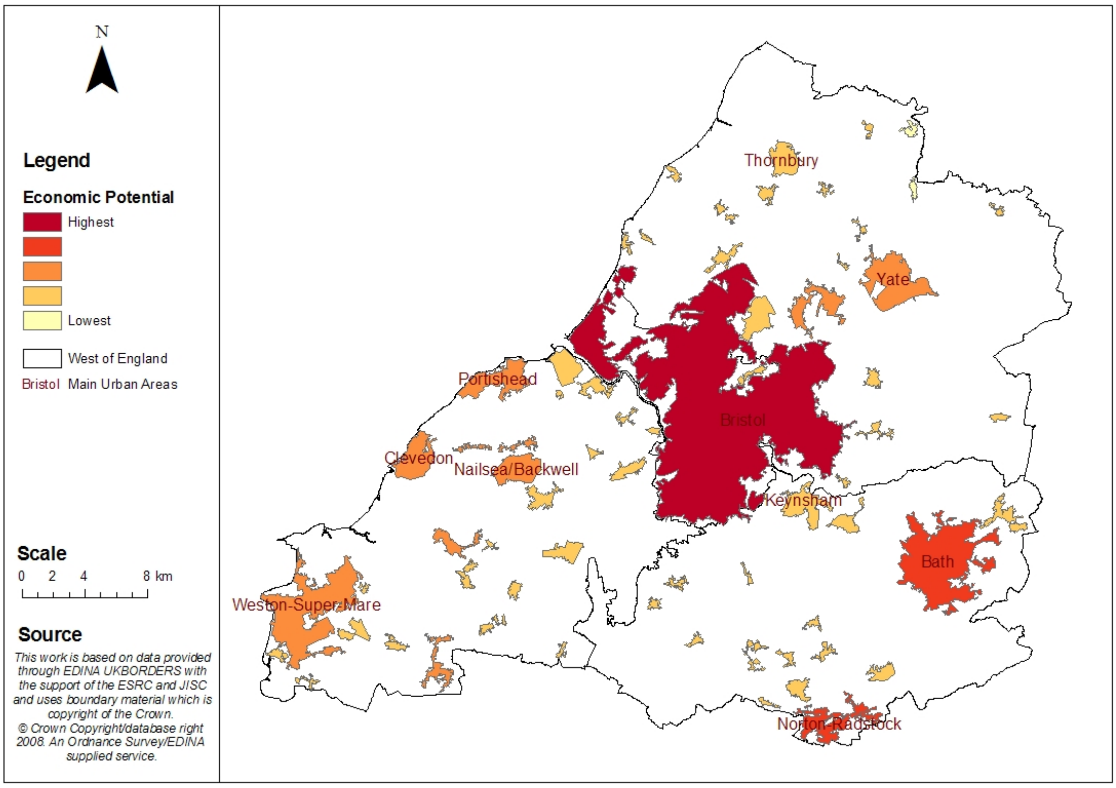


Figure 3. The economic potential of urban areas in the West of England study area based on the Keeble *et al* (1981) economic potential model.

$$P_{ij} = \frac{\dfrac{S_j}{T_{ij}^{\lambda}}}{\sum_j \dfrac{S_j}{T_{ij}^{\lambda}}} \qquad (2)$$

Where

P_{ij}	=	the probability of employees living at location i working at urban area j
S_j	=	the attractiveness of urban area j
T_{ij}	=	the travelling distance or time from residential location i to urban area j
λ	=	unknown parameter of distance decay

5. Research Outcomes

This research will provide a contribution to business location theory and to the debate regarding business and sustainability. It will critique spatial interaction theory as a method for understanding contemporary economic activity. It will identify small business needs for future policy and provision in the study region. It may be extended into a tool to aid sustainable business decision-making.

6. Acknowledgements

This project was formulated under the sustainability theme of the government-funded Great Western Research initiative, the aim of which is to further knowledge and understanding of economic activity to prepare for sustainable growth in south west England. The research is a collaboration between the University of the West of England in Bristol, the University of Bath 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

Bureau van Dijk Electronic Publishing (BvDEP) (2007) Financial Analysis Made Easy (FAME): UK and Irish company information in an instant, [Online]. Available: https://fame.bvdep.com/version-2008215/cgi/template.dll?product=1 [15/01/2008].

Copus, AK (1999) A New Peripherality Index for the NUTS III Regions of the European Union. ERDF/FEDER Study 98/00/27/130. A Report for the European Commission, Directorate General XVI.A.4 (Regional Policy and Cohesion). September 1999

Craglia M, Leontidou L, Nuvolati G and Schweikart J (2004) Toward the Development of Quality of Life indicators in the 'Digital City' *Environment and Planning B: Planning and Design* **31**(1) pp 51-64

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).

Geertmann SCM and van Eck JRR (1995) GIS and Models of Accessibility potential: An Application in Planning *International Journal of Geographical Information Systems* **9**(1) pp 67-80

Gillingham R and Reece WS (1979) A New Approach to Quality of Life Measurement *Urban Studies* **16**(3) pp 329

Grayson L and Young K (1994). Quality Of Life in Cities. London: British Library.

Haynes KE and Fotheringham AS (1984). Gravity and Spatial Interaction Models. London: Sage

Healey P and Baker (1993). European Real Estate Monitor. London: Healey and Baker.

HM Government (2007) Planning for a Sustainable Future (White Paper) © Crown Copyright 2007. Norwich, England: The Stationary Office.

Huff DL (1964) Defining and estimating a trading area *Journal of Marketing* **28** pp 34-8

Ingram DR (1971) The Concept of Accessibility: A Search For An Operational Form *Regional Studies* **5**(2) pp 101-107

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.

Lambiri D, Biagi B and Royuela V (2007) Quality Of Life in the Economic and Urban Economic Literature *Social Indicators Research* **84**(1) pp 1-25

Levinson DM (1998) Accessibility and the Journey to Work *Journal of Transport Geography* **6**(1) pp 11-21

Rabianski JS (2007) Employee Quality of Life in Corporate Location Decisions *Journal of Corporate Real Estate* **9**(1) pp 50-63

Rogerson RJ (1999) Quality Of Life and City Competitiveness *Urban Studies* **36**(5-6) pp 969-985

South West of England Regional Development Agency (SWRDA) (2006) Regional Economic Strategy For South West England 2006 – 2015.

South West Regional Assembly (2006) The Draft Regional Spatial Strategy for the South West 2006 – 2026.

Wingo L (1973) The Quality Of Life: Toward A Microeconomic Definition *Urban Studies* **10**(1) pp 3-18

Biography
Alice is currently in the second year of a PhD. Alice obtained a BSc (Hons) in Geography in 2001 (Reading) then completed an MSc in Geographical Information Science in 2005 (Birkbeck College, University of London). Alice combined her MSc studies with a position as Geographic Information Specialist at Halcrow.

OpenRouteService.org is three times "Open": Combining OpenSource, OpenLS and OpenStreetMaps

Pascal Neis[1] and Alexander Zipf[1]

[1] Chair of Cartography, Department of Geography, University of Bonn, Germany
Tel. (+49-228-73-3527) Fax (+49-228-73-5607)

<neis, zipf>@geographie.uni-bonn.de

KEYWORDS: OpenSource, OpenLS, OpenStreetMap, OGC, Routing, collaborative GI

1. Introduction

Open solutions are getting more and more momentum and support in the GI community. Nowadays we have very usable and stable open source libraries for handling, processing and visualizing spatial data. But Open Source is not the only trend, the other well known and often discussed issue within geoinformatics is the need for open standards, as specified by the *Open Geospatial Consortium* (OGC). Of course there are already open source frameworks that build on these standards such as deegree, 52°North or geoserver. Our software deals with a specific aspect of the OGCs specification: the *Open Location Services* (*OpenLS)* explained later, parts of which will be available as a new open source project at www.freeOpenLS.org in the foreseeable future.

But in addition to the already known combination of Open Source and open GI standards (OGC web services) recently with the appearance of the Web2.0, where anybody, i.e. non-specialist persons can contribute with their own data and software, also new relevant phenomena appeared concerning spatial data: People can contribute their own geodata in a collaborative attempt to build an open spatial dataset which resulted in the successfully OpenStreetMaps project (2007).

Therefore we can distinguish three aspects of openness:

- Open (GI) standards (OGC)
- Open source (GI) software
- Open (GI) data

We will introduce and discuss an example where all these three are combined in a new service that has been developed recently and is currently being extended and improved in order to meet real world requirements for a new online service. The most important application for street data and in the case of route planning we have only recently all three versions of openness combined in one new system which will soon be available at www.openrouteservice.org.

2. OpenRouteService.org

OpenRouteService.org uses free and open street data and map data from sources such as OpenStreetMaps.org and delivers route planning functionality through open standardized interfaces by the OGC. We have implemented a route service according to the OGC OpenLS Specification. It also will be made available as Open Source soon. Currently we are just setting up the website and improving the performance of the service for larger number of users and higher volumes of data. The amount of data provided by OpenStreetMap is already huge and is still growing. As first example we selected the data for Germany, and we have here already more than 400.000 streets that need to be transformed into a topological graph structure that is read and processed by our service (see figure 1).

Figure 1. OpenStreetMap for Germany used in our OGC services (WMS, WFS, OpenLS RS)

OpenStreetMaps provides free editable and usable street data (Nelson et al 2006), but for several reason this data is not provided through a standardized specification, but through a specialized API and proprietary XML schema. This has reasons which will be discussed in the full paper. While OSM has reasons for using their own formats, we think that the data provided by OSM is too important to be left out in the current trend of setting up SDIs (spatial data infrastructures) based on open OGC standards realized as web services. Therefore we also loaded parts (current scenario is focussing on Germany so far) in OGC standards based services, like WFS (Web Feature Service,), WMS (Web Map Service) and most important our OGC OpenLS Route Service (RS) and make them available as such.

3.) Web 2.0 and standard based spatial data sharing

The Web 2.0 type of free geodata (Turner 2006, Gillavry 2006, Holone et al 2007,) usually does not come in open standards, but is integrated through the Google Maps API or as *Google Earth* (GE) kml-files. While GE is now OGC principal member and KML is likely to become an OGC standard soon, that data put into GE by web2.0 users is not integrated in an open SDI, but found and visualized through a proprietary software – the GE Viewer (which is also not free for commercial users). While it includes the capabilities to search for full text entries in the data, this cannot be compared to the standardized and structured way real metadata is handled through OGC catalogue services (CS-W).

An SDI based approach (technically speaking: the use of standardized *OGC web services* (OWS)) for sharing free spatial data has been proposed by Tschirner and Zipf (2005). But due to lack of resources www.geoxchange.org remained a prototype. We consider reactivating this and developing a new portal www.openGeoXchange.org. Integrating OpenStreetMap data into WFS and WMS, as we have done for OpenRouteService.org is a first step towards such an undertaking towards the integration of Open Spatial Data and Open GI Standards.

4. Standards for Routing and Location Based Services: OpenLS

OpenLS is the acronym for OpenGIS Location Services. Since 2000 this OGC initiative has been developing implementation specifications for standardizing services that are relevant for *Location Based Services* (LB S). The OpenLS service framework consists at the moment of five core services:

- The Direcory Service,
- Gateway Service
- Location Utility Service (Geocoder/Reverse Geocoder)
- Presentation Service
- Route Service

Only recently a 6th service has been standardized by the Location Services working group: the "Tracking Service" But is not considered a "core" service. Also just since winter 2007 a new version 1.2 of the OpenLS specification has been successfully adopted by the OGC.

A further service - the "Navigation Service" - has been in discussion since 2000, but has not yet reached a stable version. Therefore we only introduce shortly the five OpenLS core services and then focus on the route service and our implementation of this together with OSM data

The Route Service offers a broad range of possibilities, among them are several parameters hat determine the result, e.g.:

1. *RouteSummary* – gives meta information about the requested route, e.g.: overall distance, units, overall needed time etc.
2. *RouteGeometry* – requests the routes geometry (line string containing all waypoints of the route). It is possible to define a maximum number of waypoints. We have realised this generalisation using the Douglas-Peucker-Algorithm.
3. *RouteInstruction* – these are "step by step" driving instructions of the calculated route. We realized this for various languages (e.g. German, English, Italian, Swedish...).
4. *RouteMaps* – The calculated route is displayed on a route map. Amongst other possibilities an overview map as well as detailed maps of the start- and destination can be requested.

5. Applying the OpenLS Specification

So far we have implemented three of the original five core services (Location Utility Service, Presentation Service und Route Service) and two more (Directory Service and Tracking Service) are under development. In particular the Route Service has already been successfully used by some of our other projects (www.ok-gis.de, www.heidelberg-3d.de, www.rewob.de).This means that several specialized or extended versions have been developed as spin-offs of the OpenLS RS:

- *Emergency Route Service* (ERS) – The ERS considers actual avoid areas from a WFS (flooded or blocked roads or areas) automatically. Neis et al (2007)
- *Accessibility Analysis Service* (AAS) – The AAS calculates a polygon representing the area that is reachable within a certain time distance based on a street network around a given location. Neis and Zipf (2007)
- *Route Service 3D* (RS3D) – The RS3D maps the route geometry onto a digital elevation model (DEM). Neis et al. (2007)
- *Route Service with Landmarks and Focus Maps.* Neis et al (2007)
- *3D Route Service with Landmarks and 3D Focus Maps.* Neis et al (2008)

6. Discussion and outlook

Our OpenLS Route Service will be made available soon as open source at a new portal we are currently working on: www.FreeOpenLS.org. The further OpenLS services shall be added also in the future, but details on this have not yet been decided. So far we have successfully tested our service based on a modified version of the geotools Dikstra library with OSM data consisting of more than 800.000 street segments. But when we cover whole countries or even Europe the response time increases to several seconds, therefore we need to use a more sophisticated approach.

Of course there are a range of other route planning algorithms and according libraries available. We will test these and then decide which one will be used in future realizations of the route service in order to speed up response times. We just are in the process of conducting more research in this topic and will select or even improve an appropriate strategy as a result. This will be presented at the conference.

References

Coast S. (2007). *OpenStreetMap.* Workshop on Volunteered Geographic Information, December 13-14, 2007 http://www.ncgia.ucsb.edu/projects/vgi/

Heidelberg – 3D: www.heidelberg-3d.de / www.gdi-3d.de

Holone, H, G Misund, H Holmstedt (2007): *Users Are Doing It For Themselves: Pedestrian Navigation With User Generated Content.* In: The 2007 International Conference on Next Generation Mobile Applications, Services and Technologies, 2007. NGMAST '07. p 9 1-99

Gillavry, E. M. (2006): *Collaberative Mapping and GIS: An Alternative Geographic Information Framework.* In: Collaborative Geographic Information Systems.

Neis, P., A. Schilling, A. Zipf (2007): *3D Emergency Route Service (3D-ERS) based on OpenLS Specifications.* GI4DM07. 3rd Int. Symp. on Geoinfo. for Disaster Management. Toronto, Canada.

Neis, P., A. Zipf (2007): *A Web Accessibility Analysis Service based on the OpenLS Route Service.* AGILE 2007. Aalborg, Denmark.

Neis, P. and Zipf, A. (2008 in work, invited): *Generating 3D Focus Maps for the (mobile) Web - an interoperable approach.* In: International Journal of Location Based Services (JLBS).

Neis, P. and Zipf, A. (2007): *Realizing Focus Maps with Landmarks using OpenLS Services.* 4th International Symposium on LBS and Telecartography 2007. Hong Kong.

Nelson, A., A de Sherbinin, F Pozzi (2006): *Towards development of a high quality public domain global roads database.* Data Science Journal, V.5, November 2006- pp.223-265

OPENLS: OGC Open Location Services Version 1.1
http://www.opengeospatial.org/standards/olscore

Turner, A. (2006): Introduction to Neogeography.

www.FreeOpenLS.org

www.OpenRouteService.org

www.opengeocoder.org

Zipf, A., J. Basanow, P. Neis, S. Neubauer, A. Schilling (2007): *Towards 3D Spatial Data Infrastructures (3D-SDI) based on Open Standards - experiences, results and future issues.* In: "3D GeoInfo07". International Workshop on 3D Geo-Information. Delft, NETHERLANDS.

Zipf, A., Tschirner, S. (2005): *Finding GI-datasets that otherwise would have been lost - GeoXchange - a OGC standards-based SDI for sharing free geodata.* 2nd International Workshop on Geographic Information Retrieval (GIR 2005) Bremen, Germany. Springer.

Exploiting Volunteered Geographic Information to describe Place

Ross S. Purves and Alistair J. Edwardes

[1]Department of Geography, University of Zurich, Switzerland
Tel. (+ 41 44 635 6531) Fax.(+ 41 44 635 6848)
ross.purves¦alistair.edwardes@geo.uzh.ch, www.geo.uzh.ch/~rsp

KEYWORDS: Volunteered Geographic Information, data mining, place

1. Introduction

Traditional geographic information focuses on well-structured spatial data, generally based on either crisp objects or continuous fields, which usually reflect an institutional viewpoint and the purpose for which a dataset was collected. The attributes of such data are typically single valued at a given location or for a given object and thus, by definition, cannot reflect multiple viewpoints. Such representations are of course well suited to many tasks where we apply geographic information (for example querying the rateable value of individual houses or the elevation of a particular point in space). Equally, though it is clear that if we asked ten GISRUK participants about Manchester, we would hear ten differing perspectives on the *place* that is Manchester and that these perspectives would reflect the experiences and backgrounds of those that we asked. To date, relatively little research in GIScience has focussed on these multiple perspectives, in contrast to work in human geography where the notion of *place* is seen as central to the discipline itself (Cresswell, 2004). Fisher and Unwin (2005) recognised this gap within GIScience research:

"GI theory articulates the idea of absolute Euclidean spaces quite well, but the socially-produced and continuously changing notion of place has to date proved elusive to digital description except, perhaps, through photography and film." (p. 6).

Equally, GIScientists have recently seized upon the potential of opportunities provided by Volunteered Geographic Information (VGI). VGI is loosely defined by Goodchild (2007) as "a special case of user-generated content" whereby geographic information is created outside of a formal, official framework. Thus, for example OpenStreetMap aims to create a user-generated, editable map of the world based on data collected by volunteers (www.openstreetmap.org/). However, in this paper we are interested not in the use of VGI as a means to replace or replicate traditional structured forms of geographic information, but rather as a way of starting to describe *place* in GIScience. We contend that place lies at the opposite end of a continuum of geographic perspectives from space (Figure 1) and that VGI provides us with a real opportunity to describe place.

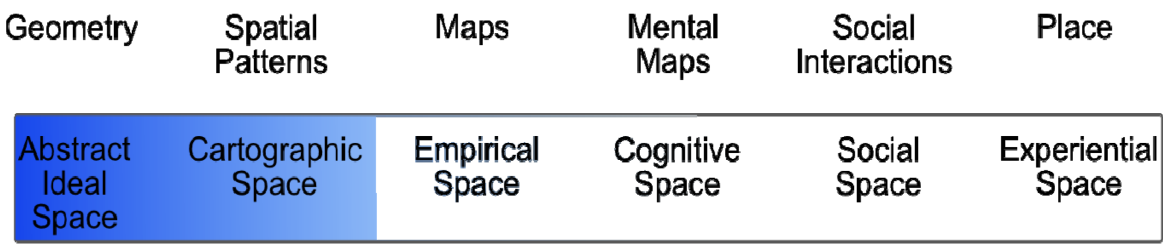


Figure 1. The space–place continuum (from Edwardes, 2007)

We set out to illustrate our contention through a case study. As part of a European project which aims to automatically add indexing terms to geo-located digital photographs we have been exploring Geograph, a classic example of VGI. The Geograph project has the aim of collecting "geographically representative photographs and information for every square kilometre of the UK and the Republic of

Ireland." Contributors submit photographs representing individual 1km grid squares for moderation and images are uploaded together with descriptive captions to a searchable web site. As of September 2007 around 5000 users had contributed more than half a millon photographs. In this paper we set out to explore two questions pertaining to the use of Geograph in exploring notions of place.

1) Can volunteered sources of geographic information provide new perspectives on describing place?

2) What limitations must we be aware of when working with such data?

In the rest of this paper we first briefly describe how place might be described, before setting out the methodology we applied in our experiments with Geograph. Finally, we describe the results of these experiments from the perspective of the questions posed above.

2. Describing place

In this paper we are primarily interested in methods for eliciting terms to describe place. Previously, much research in this domain has focussed on identifying so-called basic levels, whereby a basic level is one which is both informative and summative – thus for example *littoral zone* provides a detailed geomorphological description of a coastal feature, whereas coastline provides a very general description. Beach is both informative (in that it suggests a set of particular qualities and activities that are not offered by, for example, coastline) and summative (in that it encompasses a range of possible subclasses). Previous work to both identify basic levels and associated descriptive terms has been largely based on human subject testing. A key difficulty here is that such experiments are complex to organise, difficult to repeat and generally have relatively small numbers of subjects. We thus wished to explore the extent to which we could exploit Geograph in both replicating past research on place and forming new perspectives.

We discuss here a set of three experiments, some of which were previously described in more detail in Edwardes and Purves (2007). The first experiment set out to compare the frequencies of terms suggested as basic levels in previous empirical research with the frequency of occurrence of the same terms in the Geograph collection. In the second experiment, we explored the co-occurrence of descriptive terms with a selection of basic levels. Finally, in the third set of experiments we explored the relationships between a set of 1381 nouns identified manually within the Geograph dataset which occurred more than 100 times. To explore these relationships we analysed clusters of significant groupings amongst the nouns, using cosine similarity and hierarchical clustering techniques (Salton et al., 1975).

3. New perspectives on place?

In the first experiment, where we compared the term frequencies in Geograph with previous participant research we found that the terms identified were in many cases very similar (Edwardes and Purves, 2007). Thus, we showed that the rankings of the following terms identified by Battig and Montague (1969) as being category norms (broadly equivalent to basic levels) were significantly correlated: Mountain; Hill; Valley; River; Rock; Lake; Canyon; Cliff; Ocean; Cave. This first result suggests that Geograph can be used as a proxy for participant experiments.

In our second experiment (Table 1) we explored how people described two of the basic levels identified above (mountain and hill) and two further basic levels (beach and village) identified in previous research and in the Geograph data (Edwardes and Purves, 2007).

Table 1. Co-occurring descriptive terms with basic levels beach, village, hill and mountain

Beach		n=2824	**Village**		n=12707
Activities	**Elements**	**Qualities**	**Activities**	**Elements**	**Qualities**
Surfing	Shingle	Sandy	Conservation	Pub	Deserted
Bathing	Sand	Deserted	Reading	Shop	Pretty
Defence	Cliff	Eroded	Fishing	Inn	Green
Swimming	Headland	Soft	Playground	Church	Quiet
Tourism	Bay	Rocky	Defence	Housing	Lovely
Wading	Sea	Warm	Bowling	Edge	Pleasant
Protection	Rock	Glacial	Tourism	Cottage	Beautiful
Sport	Coast	Low	Football	Main Road	Remote
Shipping	Shore	Beautiful	Entertainment	Village green	Unusual
Golf	Island	Lovely	Sitting	Stone	Large
Hill		n=16232	**Mountain**		n=1256
Activities	**Elements**	**Qualities**	**Activities**	**Elements**	**Qualities**
Climbing	Fort	Steep	Biking	Peak	Distant
Skiing	Top	Distant	Kayaking	Summit	Black
Holidays	Summit	Wooded	Outing	Ridge	Remote
Observation	Horizon	Black	Mountaineering	Moorland	Rocky
Sitting	Ridge	Rough	Escape	Quarry	Grassy
Walking	Sheep	Grassy	Walks	Stream	Steep
Running	Valley	Round	Fun	Sheep	Natural
Cycling	Side	Big	Racing	Forest	Dark
Preservation	Trees	White	Climbing	Top	Broad
Escape	Track	Broad	Cycling	Path	Running

In the third experiment we looked at the associations between nouns through clustering techniques and used these to build sets of related terms. Table 2 shows illustrative examples of some of the associations derived using these techniques.

Table 2. Clusters and illustrative associated terms

Cluster	**Associated terms**
Road network	road, roundabout, junction
Hills and valleys	hill, valley, bridleway, walkers
River systems	river, bank, waterfalls, floodplain, water, streams, valleys, levels, aqueduct, sewage, anglers, salmon, otter
Buildings	home, wall, manor, doorway, gable, mansion, roof, glass, house, grounds, architect, foundation, columns, castle
Woodland	Forest, woodland, plantation, oak, beech, pony, heathland, commoner
Arable agriculture	Fields, pasture, crop, wheat, farmer, harvest, barley
Mountain landforms	Glen, beinn, allt, meall, loch, corrie, sgurr, garbh
Coastal features	Sea, beach, bay, peninsula, headland, sands, islands, creek, cave, foreshore, mud
Rural landscapes	Pond, fence, orchard, wire, birds, bushes, flowers, hedgerow, scenery, flock

We contend, that these experiments illustrate how we can use VGI to start to address the description of *place*, which as we have illustrated, is a more descriptive and less geometric form of geographic information. Thus for example, we identified in Table 1 typical descriptors for mountains, beaches, hills and villages. In turn, such descriptors could be extracted from more traditional sources of spatial data (e.g. identifying village greens in topographic data) to identify and locate typical and atypical examples of villages as characterised by Geograph users.

Equally, the Geograph dataset reveals notions of place with respect to the British Isles. For example, the term ‘Sea’ is used in preference to ‘Ocean’, and geographic features termed ‘Hill’ and ‘Loch’ are far more commonly encountered than those such as ‘Mountain’ and ‘Lake’. Our work on identifying descriptive terms also points to potentially stereotypical viewpoints. For instance the terms found most commonly with ‘Village’ were somewhat bucolic – “green, quiet, inn”. Whilst this is probably not due to a conscious bias by the contributors, it may relate to unconscious avoidance of taking pictures that are not in some way aesthetically pleasing.

In this paper we have ignored the spatial distribution of the terms which we discuss – that is to say, we have not considered if there are variations in the way that place is described within the British Isles – this is the subject of parallel work described in a paper by Dykes et al. at this conference.

6. Acknowledgements
This research reported in this paper is part of the project *TRIPOD* supported by the European Commission under contract 045335. We would also like to gratefully acknowledge contributors to Geograph British Isles, see http://www.geograph.org.uk/credits/2007-02-24, whose work is made available under the following Creative Commons Attribution-ShareAlike 2.5 Licence (http://creativecommons.org/licenses/by-sa/2.5/).

References

Battig W and Montague W (1969) Category norms for verbal items in 56 categories: a replication and extension of the Connecticut Norms *Journal of Experimental Psychology* **80**(2) pp1-46

Cresswell T (2004). *Place: A Short Introduction.* Blackwell, Oxford.
Edwardes AJ and Purves RS (2007) *A theoretical grounding for semantic descriptions of place* LNCS: Proceedings of 7th Intl Workshop on Web and Wireless GIS W2GIS 2007 Ware M and Taylor G (eds) Berlin: Springer pp106-120
Fisher P and Unwin D (2005) *Re-presenting Geographical Information Systems in* Peter Fisher P and Unwin DJ (eds), Re-presenting GIS, John Wiley & Sons, London. pp 1-17
Goodchild MF (2007) Citizens as sensors: the world of volunteered geography *Geojournal* **69** pp211-221
Salton G Wong A and Yang C S (1975) *A Vector Space Model for Automatic Indexing* Communications of the ACM **18**(11) pp613–620

Biography
Ross Purves is a lecturer at the Department of Geography of the University of Zurich, where Alistair Edwardes is a post-doctoral research assistant.

Exploring Volunteered Geographic Information to Describe Place: Visualization of the 'Geograph British Isles' Collection

Jason Dykes[1], Ross Purves[2], Alistair Edwardes[2] & Jo Wood[1]

[1]giCentre, School of Informatics, City University London, EC1V 0HB
Tel. (+ 44 20 7040 8906) Fax (+44 20 7040 8845)
jad7¦jwo @soi.city.ac.uk, www.soi.city.ac.uk /~ jad7¦jwo

[2]Department of Geography, University of Zurich, Switzerland
Tel. (+ 41 44 635 6531) Fax (+ 41 44 635 6848)
ross.purves¦alistair.edwardes @geo.uzh.ch, www.geo.uzh.ch/~rsp

KEYWORDS: volunteered geographic information, place, tag cloud, tree map, mashup

1. Introduction

Increasing volumes of volunteered geographic information (VGI) provide a rich but complex data source for GIScience (Goodchild, 2007). For example, Edwardes and Purves (2007) consider multiple perspectives on *place* by analysing the text and images submitted to Geograph British Isles (Geograph, 2007). We extend this work by exploring how such perspectives vary *spatially*. We do so through visualization – an informal method of data exploration that involves the development and use of interactive graphics to investigate data sets with spatial, temporal and other structure. Effective visualization results in ideation and the generation of hypotheses and the approach is increasingly popular for exploring large unknown data sets (Thomas and Cook, 2005). Visualization is particularly appropriate where exploration relies upon the synthesis of complex data that contain spatial relationships and tacit knowledge (MacEachren and Kraak, 2001). Thus we argue that visualization can assist in exploring the complexity of informal multi-authored VGI collections as we apply it to understand notions of *place*.

We support this contention through a series of case studies that use visualization techniques to reveal spatial variations and help develop initial hypotheses relating to them. Each visualizes data volunteered to Geograph as of February 2007: a collection of 340,000 photographs (or 'geographs') with titles, comments and other metadata. All geographs are georeferenced to at least 1km precision. Our interest is in use of the terms identified by Edwardes and Purves (2007) involving four 'basic levels' or 'scene types' (STs): 'beach', 'village', 'mountain' and 'hill'. We focus on the thirty most popular descriptive terms, or 'scene type descriptors' (STDs), associated with each. The STDs are categorised into ten qualities (adjectives describing the scene), ten elements (objects found in the scene) and ten activities for each for the four STs.

2. Visualization of Term Co-occurrence Hierarchy

This organisation of terms can be considered a hierarchy: *ST* - 4 nodes; *category* - 3 per ST = 12 nodes; *STD* - 10 per category = 120 nodes. Treemaps (Shneiderman, 1992, 2006) are a popular means of visualizing such hierarchies.

2.1 Static TreeMaps

A treemap of the hierarchy of terms used in geograph titles and comments reveals some structure in the use of the English language in Geograph. Figure 1 contains a leaf for each of the 50,000 cases where one of our four STs is associated with a photograph. Each is coloured using an inherited random scheme (Wood and Dykes, submitted) where root nodes are allocated a random colour that is inherited by children with a minor colour mutation. Whilst the hues have no independent meaning, the structure of the graph is reflected in the colours as well as the spatial organisation. We can see that 'hill' dominates over 'beach', 'village' and 'mountain' and that the 'none' category (no STDs associated with an ST) dominates in each case. Elements co-occur more frequently with each ST than adjectives and activities. 'Steep' and 'black' are used more frequently with 'hill' than 'mountain', etc. Our 'static' treemaps permit zooming and panning.
The 22,000 cases where STs occur in titles are shown in Figure 2 (top), again with colours that are randomised initially and so do not persist between figures. Figure 2 (bottom) removes cases where no STDs co-occur with STs in titles ('none') showing the 5,000 remaining cases. Comparison of the treemaps generated from titles and comments and focussing on the co-occurrence of STs and STDs may provide general and specific insights into how Geograph contributors use language. For example: adjectives occur less frequently in titles than comments; 'ridge' occurs less frequently with 'hill' in titles than might be expected from its use in comments (compare with 'track'); 'village green' is over-used in titles, etc.

2.2 Interactive and Spatial Treemaps

Treemaps provide a useful overview of the use of language in Geograph, but how does this vary spatially? We can add spatial symbolism to statistical graphics and information visualization (Dykes, 1995). In Figure 3 colour is used to depict location with eastings represented in red (0-700000 from the OSGB origin) and northings in green (0-1300000). This allows us to explore the spatial structure of the use of co-occurring terms. Reds represent the south-east and browns the south-west, with colours lightening as we move north. Light greens and yellows relate to locations in the north-east and greens the north-west. Consistent colouring of the leaves in any node reveals spatial autocorrelation – 'black' and 'mountains' are perhaps unsurprisingly used in the Black Mountains, but it is interesting to note that these terms do not co-occur elsewhere. 'Village' has a southern and eastern bias with little use in Scotland. 'Summit' and 'valley' have very different geographies when used in conjunction with 'hill'. The strong distinct colours associated with 'beach' (see 'shingle') relate to its peripheral geography. Consideration of the consistency, texture and dominant colours of any node may help relate geography and the language volunteered to Geograph.
Dynamic links can help us explore the relationship between the photographs and text in geographs. Our treemaps provide dynamic links to the Geograph server enabling us to download image icons as leaves are brushed and display their associated text (Figure 3).

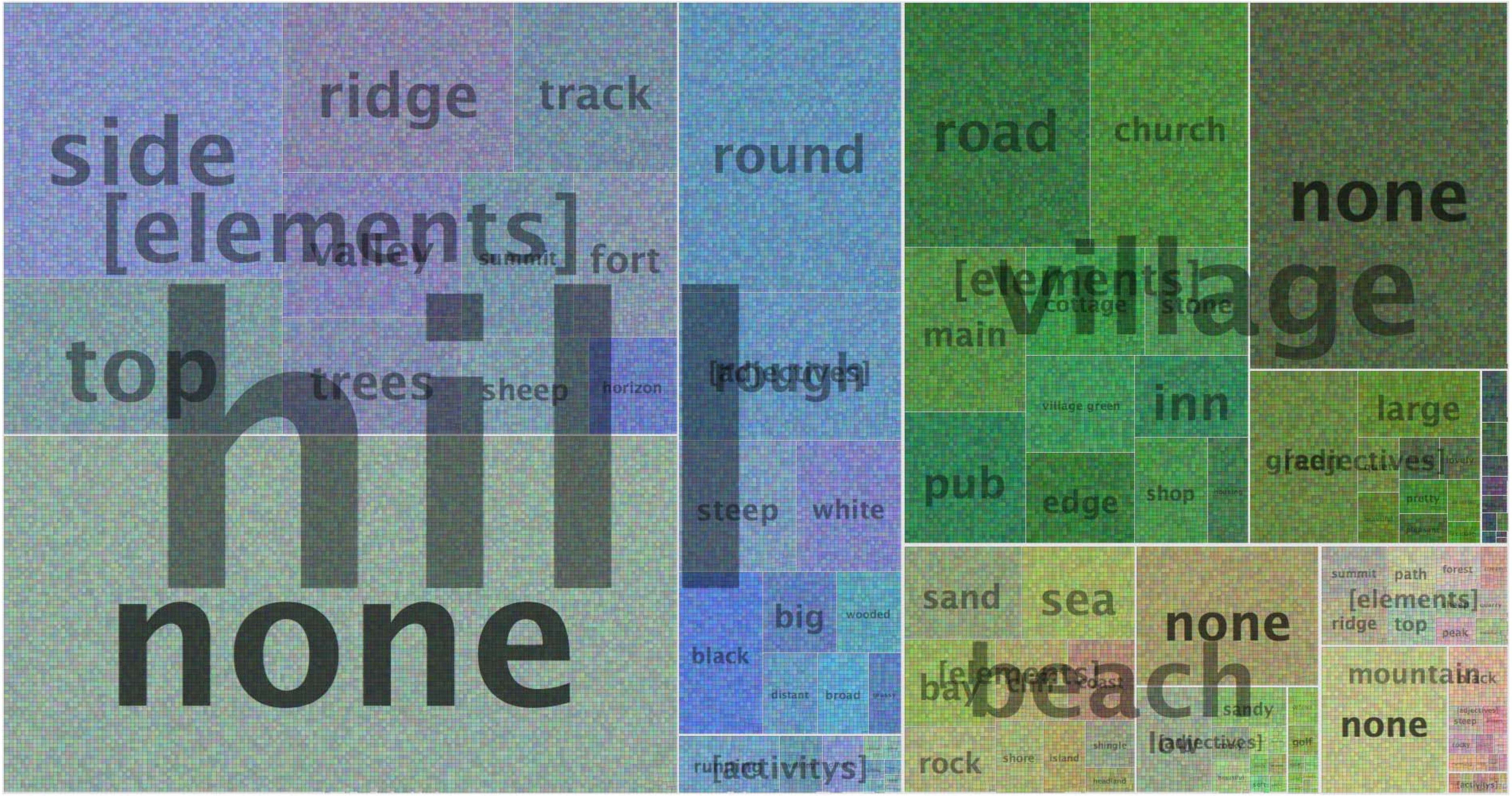


Figure 1. Treemaps of terms occurring in geograph titles and comments for STs. Node sizes represent term occurrence. Colours emphasize the ST / category / STD hierarchy with an inherited random scheme.

Figure 2. Treemaps of terms occurring in geograph titles for STs. Node sizes represent term occurrence. Colours emphasize the ST / category / STD hierarchy with an inherited random scheme. Treemaps show all occurrences of STs in titles (top) and the subset of occurrences of STs in titles where they co-occur with STDs (bottom). The second treemap is thus a focussed version of the first, with the ‘none’ category removed.

2.3 Image Treemap

Treemaps can be used in creative ways to associate hierarchical structure with complex information summaries - for example Wattenburg’s art exhibit on the colour of language (Wattenburg, 2005). Figure 4 provides a direct link between Geograph term hierarchy and photographs through an image treemap that offers considerable scope for exploring the relationships between ST, STD, text and imagery.

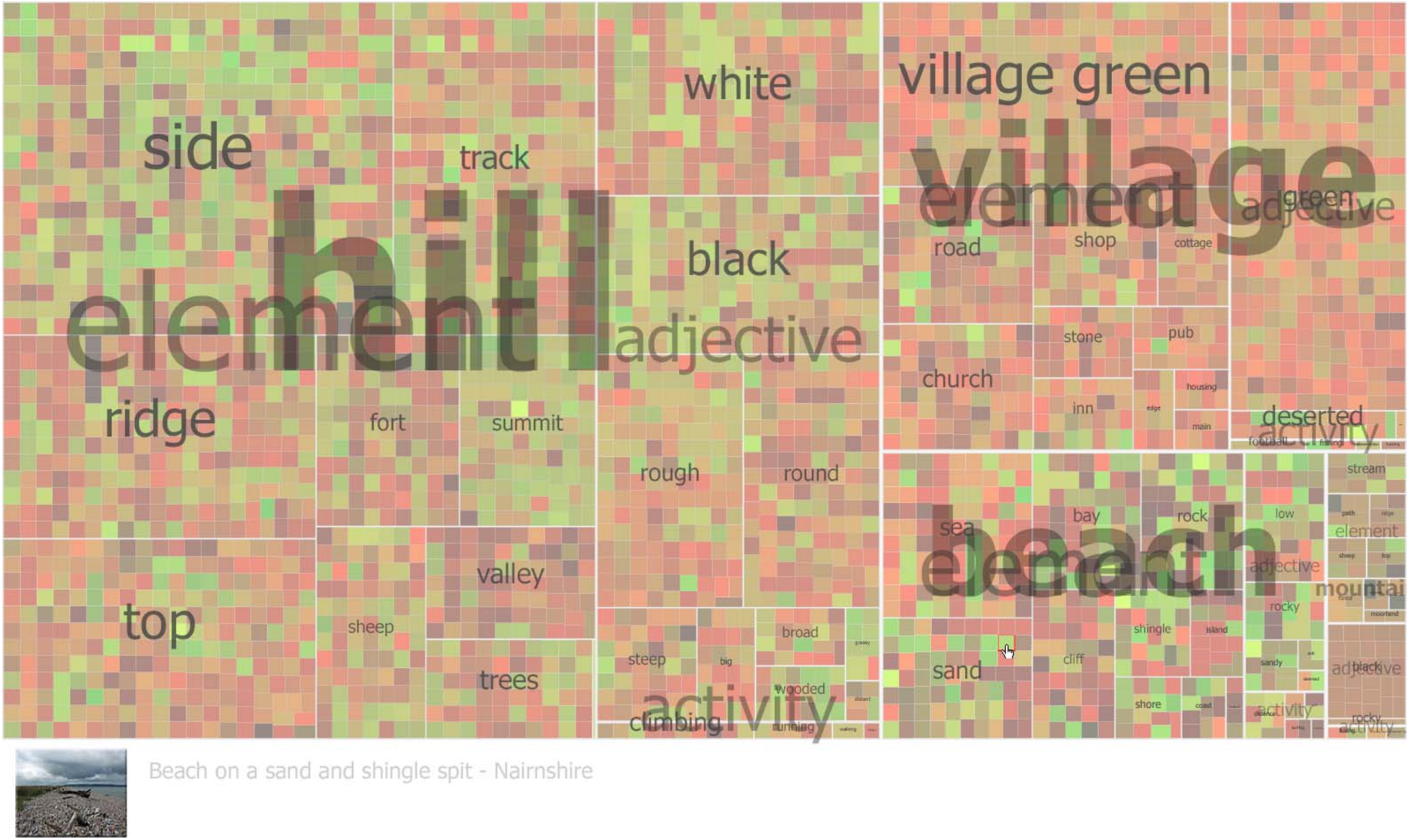


Figure 3. Treemap of terms occurring in geograph titles for STs. Node sizes represent term occurrence. Colours use geographic shading. Image © Copyright Claire Pegrum and licensed for reuse under Attribution-Share Alike 2.0 Generic Creative Commons Licence (http://creativecommons.org/licenses/by-sa/2.0/).

Figure 4. Treemap of terms occurring in geograph titles for STs (top) with detail (bottom). Node sizes represent term occurrence. Images © Copyright Geograph British Isles Photographers (http://www.geograph.org.uk/credits/2007-02-24) and licensed for reuse under Attribution-Share Alike 2.0 Generic Creative Commons Licence (http://creativecommons.org/licenses/by-sa/2.0/).

3. Exploring the Spatial Variation in Term Co-occurrence

Mapping individual terms and term co-occurrences spatially allows us to explore these relationships further. We use two techniques to do so.

3.1 Exploring National Trends

A chi statistic relates observations of phenomena with some expectation and mapping the results helps us to explore spatial variations in observations (Wood et al., 2007). The initial assumption that any STD is equally likely to occur with any ST allows us to generate chi statistics for pairs of co-occurring STs and STDs. We can do so at a number of spatial resolutions and have developed software to support the visualization of the results. Maps of the national variation in the co-occurrence of particular combinations of STs and STDs found in titles, comments or both of these can be interactively selected at three sampling resolutions. For example, Figure 5 shows that 'shingle' as a descriptor of 'beach' has a spatial bias towards the coasts of the north-west and south-east and away from the east coast and the south-west when considered at a resolution of 10km. The persistence of the pattern can be explored at other resolutions.

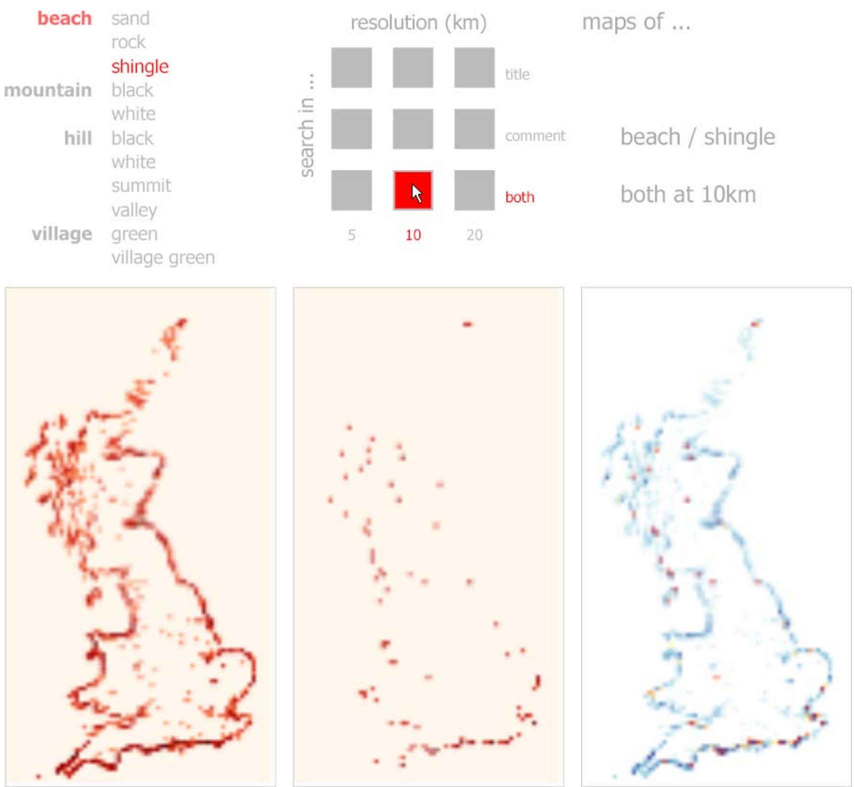


Figure 5. An exploratory interface to chi statistic maps of national trends. The three maps show: Left - ST density; Centre - STD density; Right - chi statistic for the selected ST / STD combination. The chi statistic is mapped using a diverging scheme with blues representing fewer observations than expected and reds representing more. Darker colours relate to more extreme observations.

The spatial relationship between 'black' and 'mountain' persists both at a number of scales and whether title, comment or both are considered. The maps shown in Figure 6 show that where ‘mountain’ is used to describe geographs in mountainous areas close to the Black Mountains the term ‘black’ is employed as a descriptor less than expected. This may suggest a possible geographic shielding effect, and may also reflect a difference in the predominant language used in north and mid-Wales. Our interactive tools and direct links between hierarchy, geography and text help us explore the detail. The co-occurrence of the terms ‘hill’ and ‘summit’ and ‘hill’ and ‘valley’ also have patterns that persist through the resolutions considered here. The former have a northern bias, the latter are concentrated in the south and south-west. The case of ‘hill’ and ‘summit’ is interesting as the spatial distribution varies according to whether title or comment is considered. The former case biases the distribution towards the Scottish Borders, where many toponyms include ‘Hill’, as opposed to the Highlands where hill or mountain names are typically Gaelic (Figure 7).

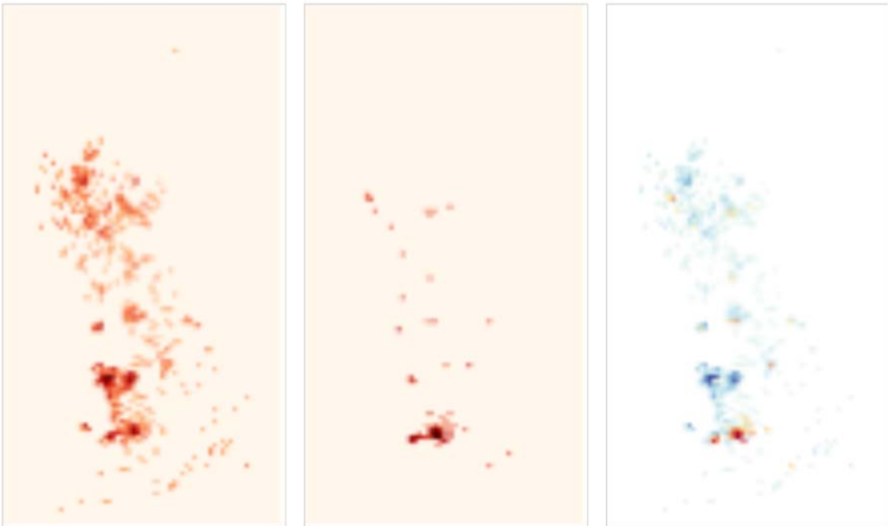


Figure 6. The geography of ‘black’ and ‘mountain’: Left - density of ‘mountain’ ST; Centre - density of ‘mountain’ ST with ‘black’ STD; Right - chi statistic of ‘mountain’ with ‘black’ STD ST. Sampling resolution 10km. Terms found in geograph titles or comments.

3.2 Exploring Local Variation

We can explore local variation in the terms used through a geovisualization mashup (Wood et al., 2007). Here we use spatial tag clouds (Slingsby et al., 2007) that are updated dynamically as we explore through a geobrowser to reveal geographic emphases. The tag cloud of all terms used in geographs located around the Manchester area shown in Figure 8 has a distinctly local flavour. Tag clouds for each of the STs show how STDs are used in the selected locality (Figure 8, top right). The tag clouds are linked to the maps so that clicking a term will reveal the associated geograph locations and selecting a location links to the geograph (Figure 8, bottom right).

The mashup allows us to relate geographs, their locations and descriptions to ancillary data including formal spatial data sets, other forms of VGI and our own derivatives such as the chi surfaces.

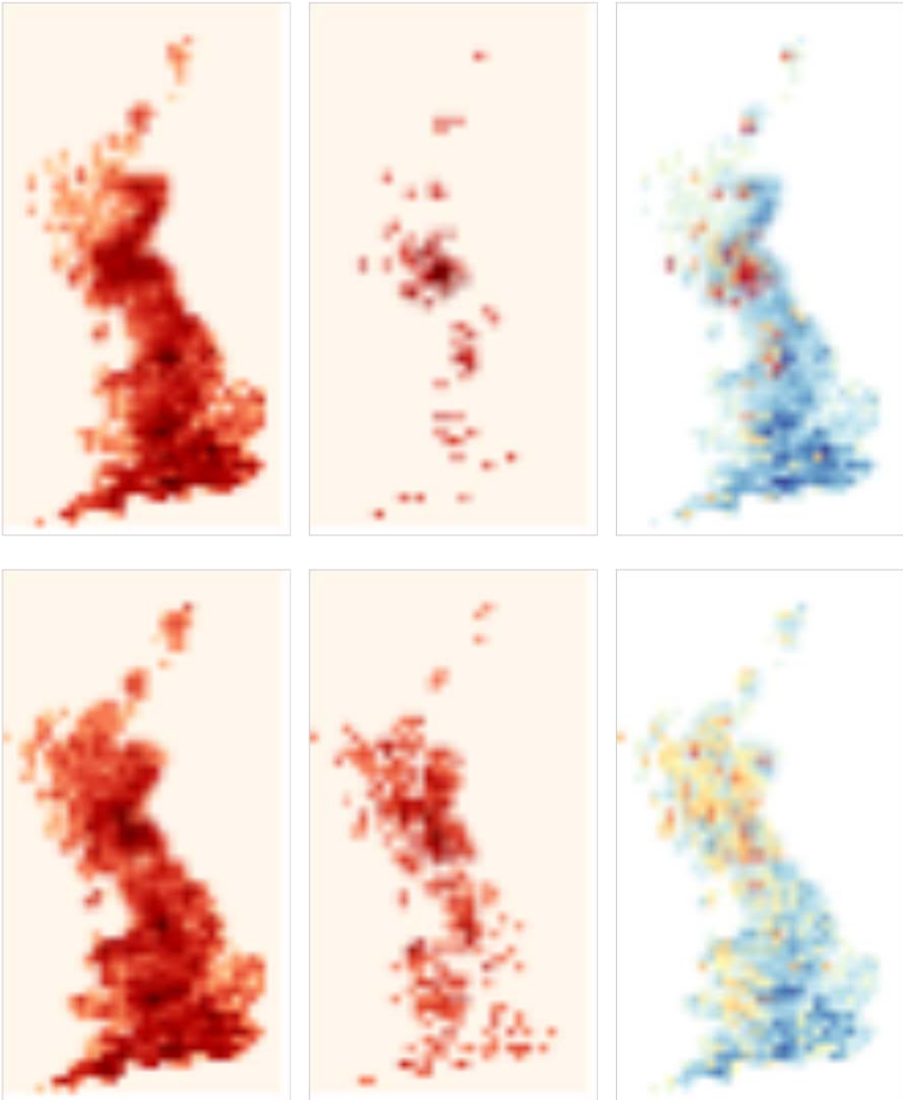


Figure 7. The geography of ‘hill’ and ‘summit’: Left - density of ‘hill’ ST; Centre - density of ‘hill’ ST with ‘summit’ STD; Right - chi statistic of ‘hill’ ST with ‘summit’ STD. Sampling resolution 20km. Compare terms found in geograph title (top) and comment (bottom) fields.

4. Conclusion

Purves and Edwardes (submitted) argue that analysis of the Geograph collection allows us to begin addressing the description of *place*. Edwardes and Purves (2007) compare the frequencies of terms derived in empirical research with those occurring in the Geograph VGI and consider the co-occurrence of terms and selected descriptors. We use visualization to explore ways in which the co-occurrence of these terms and descriptors varies geographically. Our preliminary results demonstrate ways in which visualization can contribute to this process by providing insights into the nature, structure and geography of VGI. Further developments of the treemap layout algortithm, particularly arranging nodes to reflect their geographic location (Wood and Dykes, submitted) should allow refined visual assessment of spatial pattern in co-occurrences.

Figure 8. Linked graphics showing spatial tag clouds of terms used to describe geographs in selected area. Left - all terms used; Top right - STDs used with the four STs under consideration; Bottom right - STDs used with four STs linked to geograph locations and Geograph details. Image © Copyright Chris Shaw and licensed for reuse under Attribution-Share Alike 2.0 Generic Creative Commons Licence (http://creativecommons.org/licenses/by-sa/2.0/).

There is scope for enhancing the work by exploring our assumptions and performing more sophisticated analysis of the text by, for example, reducing the rate of false positives and investigating the effects of weighting our ST / STD occurrence count towards cases where terms are used in close combination. Accounting for the effects of individuals or coordinated activity is an important generic issue relating to the use of VGI that visualization can help us detect. Work to explore Geograph further through the development of additional visualization functionality is ongoing as are efforts to compare the spatial distributions revealed in Geograph with other sources of VGI including the Flickr photographic database.

5. Acknowledgments

RSP and AJE's research for this paper is part of the project *TRIPOD* supported by the European Commission under contract 045335. We would also like to gratefully acknowledge Barry Hunter and contributors to Geograph British Isles, see http://www.geograph.org.uk/credits/2007-02-24, whose work is made available under the following Creative Commons Attribution-ShareAlike 2.5 Licence (http://creativecommons.org/licenses/by-sa/2.5/).

References

Dykes, J. A. (1997). Techniques for Adding Spatial Information to Statistical Plots of Geographical Data. EuroStat: New Techniques and Technologies for Statistics II. W. Kloesgen. Amsterdam, IOS Press**:** 261-278.

Edwardes, A. J. and R. S. Purves (2007). A theoretical grounding for semantic descriptions of place. M. Ware and G. Taylor (eds.) LNCS: Proceedings of 7th Intl. Workshop on Web and Wireless GIS, W2GIS 2007, Berlin, Springer, 106-120.

Geograph - the 3658 Geograph British Isles Photographers (2007). Geograph British Isles [available 21/12/07] http://www.geograph.org.uk/credits/2007-02-24.

Goodchild, M. F. (2007). "Citizens as sensors: the world of volunteered geography." Geojournal **69**: 211-221.

MacEachren, A. M. and M.-J. Kraak (2001). "Research challenges in geovisualization." Cartography and Geographic Information Science **28**(1): 3-12.

Purves, R. S. and A. J. Edwardes (submitted). Exploiting Volunteered Geographic Information to describe Place. GISRUK 2008 the 15th national GIS research conference for the UK, Manchester

Shneiderman, B. (1992). "Tree visualization with tree-maps: 2-d space-filling approach." ACM Transactions on Graphics **11**(1): 92-99.

Shneiderman, B. (2006). Treemaps for space-constrained visualization of hierarchies [available 20/12/07] http://www.cs.umd.edu/hcil/treemap-history/.

Slingsby, A., J. Dykes, et al. (2007). Interactive Tag Maps and Tag Clouds for the Multiscale Exploration of Large Spatio-temporal Datasets. E. Banissi (ed.) 11th International Conference on Information Visualization 07, Zurich, IEEE Computer Society, 497-504.

Thomas, J. J. and K. A. Cook, Eds. (2005). Illuminating the Path: Research and Development Agenda for Visual Analytics, IEEE Press.

Wattenburg M. (2005). The Color Project, Loop: AIGA Journal of Interactive Design Education [available 21/12/07] http://loop.aiga.org/content.cfm/the-color-project.

Wood, J. and Dykes, J. (submitted). From slice and dice to hierarchical cartograms: Spatial referencing of treemaps. GISRUK 2008 the 15th national GIS research conference for the UK, Manchester

Wood, J., J. Dykes, et al. (2007). "Interactive visual exploration of a large spatio-temporal dataset: reflections on a geovisualization mashup." IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization 2007) **13**(6): 1176-1183.

Biographies

Jason Dykes and Jo Wood are senior lecturers in geographic information at the giCentre, City University London with interests in developing and using interactive graphics to explore complex structured data sets. Jo is particularly interested in analysing surfaces and hierarchies. He wrote LandSerf and LandScript through which the surfaces were generated and implemented the Treemap algorithm. Jason is particularly interested in using developing technologies and data to augment this activity. He did the database stuff, the SVG and KML bits and wrote the LandScript.

Ross Purves and Alistair Edwardes work at the Department of Geography of the University of Zurich and have interests in geographic information retrieval and volunteered geographic information. They identified the scene types and descriptors and their research questions, analysis, ideas and insights inspired many of the techniques implemented here.

All four of the above participated in heated discussion about the nature of the data and the graphics generated from them, hypothesized at length and learned a lot about Geograph and VGI.

A Methodology for Inferring Higher Level Semantic Information from Spatial Databases

Marie-Kristina Thomson and Roderic Béra

Department of Civil, Environmental & Geomatic Engineering, University College London,
Gower Street, London, WC1E 6BT
Tel. +44 (0) 20 7679 2741
{thomson, rbera}@ge.ucl.ac.uk

KEYWORDS: ontology, knowledge representation, interpretation, land use, mapping

1. Introduction

The lack of high level semantic descriptions within geospatial data is a major critique point in regards to their handling and usage (Freksa and Barkowsky, 1995; Peuquet, 1988; Burrough and Frank, 1995; Frank, 1992). Topographic data are a striking example, as the only explicit knowledge they contain is positional. They can reveal a lot more, however, as spatial configuration are the result of functional and social processes (Lévi-Strauss, 1963). In this context the present research seeks to expose implicit information within a database to enrich its contents by generating a description that assigns functional meanings (i.e. higher level semantic information) to the map primitives. Such an approach is expected to ultimately improve the responding of data providers to customer needs and specific applications (Lüscher *et al.*, 2007).

The following sections outline how the use of knowledge representation for the explicit formalisation of knowledge about object constellations existent in a topographic map can be exploited. Section 2 discusses the pertinence of a distinction between data and semantics and how to link both aspects, section 3 proposes a model framework for the extraction of higher level knowledge and exemplifies the approach on the case of the notion of residential area, before drawing some temporary conclusions (section 4).

2. Decomposing the Mapping Problem between Knowledge and Geographic Data

Ontological approaches emerge as an alternative representation for objects stored in a geo-spatial database, thus providing new opportunities for intelligent information processing (Torres *et al.*, 2005; Hart, 2007; Hereth *et al.*, 2000; Klien and Lutz, 2005). An ontology appears as a mediating instance between the captured reality (spatial data), and higher level knowledge as perceived and conceptualised by humans. People can readily assign such knowledge to geographic data, as shown in Thomson and Béra (2007) where respondents interpreted topographic maps for land use information. The visual recognition of a 'residential' pattern on a map, for example, is characterised by the ability to interpret and categorise new observation data (Mennis *et al.,* 2000). By using existing knowledge about the objects' spatial configuration and distinctive spatial properties (such as size, shape, layout, and other criteria), a person can identify, to a certain extent, areas in a map by their primary purpose. This holds especially for the various types of residential areas (Thomson and Béra, 2007).

In order to translate this ability into mechanised ways, it is not enough to rely solely on mathematical descriptions of spatial properties and criteria since their definitions often fail to do justice to people's intuitive notions of what constitutes shape (Haggett and Chorley, 1969; Marshall, 2005). In order to capture the essence of human perception, understanding, reasoning and intuition, a machine-compatible approach needs to resemble the levels of organisation in general intelligence, as summarised by Yudkowsky (2007). This organisation can be adopted to reflect the processing of geographic data in an intelligent manner. For instance, on a set of spatial structures, such as topographic features, reasoning processes operate based on similarity judgement, proximities, shapes, sizes, etc. (Thomson and Béra, 2007), in ways similar to those investigated by Gestalt psychology. Concepts can be associated to these structures and processes to form a knowledge representation. Higher level concepts can then be built incrementally from these re-usable, primitive concepts as disposable one-time conceptual entities dependent on the information that is required by the user. Inference is achieved by classifying instances from the underlying database in terms of constrained object configurations and other concept parts as ascribed by the high level structures.

Such an approach can be termed semantic data processing since inference is based on reasoning with concepts about a data's semantics (figure 1). Similarities can be drawn to natural language processing (e.g. Katz and Fodor, 1963; Lewis, 1970; Blutner, 2002), where (a) morphological knowledge relates to the understanding of multiple forms of objects and their spatial order, (b) syntactic knowledge states structural information and relationships about the order of objects in a group signalling a particular meaning, and (c) semantic knowledge defines meaning about context and how concepts relate to objects and relations in the world. The question is how to get from raw input, i.e. spatial data, to meaning, and from meanings to customised output (ultimately new map products).

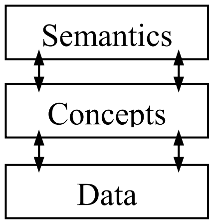


Figure 1. Semantic Data Processing

Data representation is treated as a compositional problem that consists of a hierarchy of minimal meaningful units, such as primitive entities (building, roads, etc.), that combine to form higher level meaningful composites of sets of elements (such as residential blocks). Richer knowledge is derived from meaningful object configurations, special relations, perception (Gestalt principles) and context, which needs to be grounded in the data to allow a one-to-one mapping. In a domain which manifests itself through its underlying geographic reality, such a grounding can be achieved by decomposing the rich knowledge to its finest level of detail in terms of its 'syntax', i.e. information about the structure of spatial objects (such as roads, buildings, land, and water), and how these comprise larger units that convey functional meaning, feature attributes, and spatial relations. This is exemplified in the conceptual decomposition of 'residential area' in figure 2. As identified by Barr and Barnsley (1997) land cover is organised spatially into discrete parcels whose morphological properties and the spatial relations between them convey information on land use and other higher-order 'meaning' about the scene.

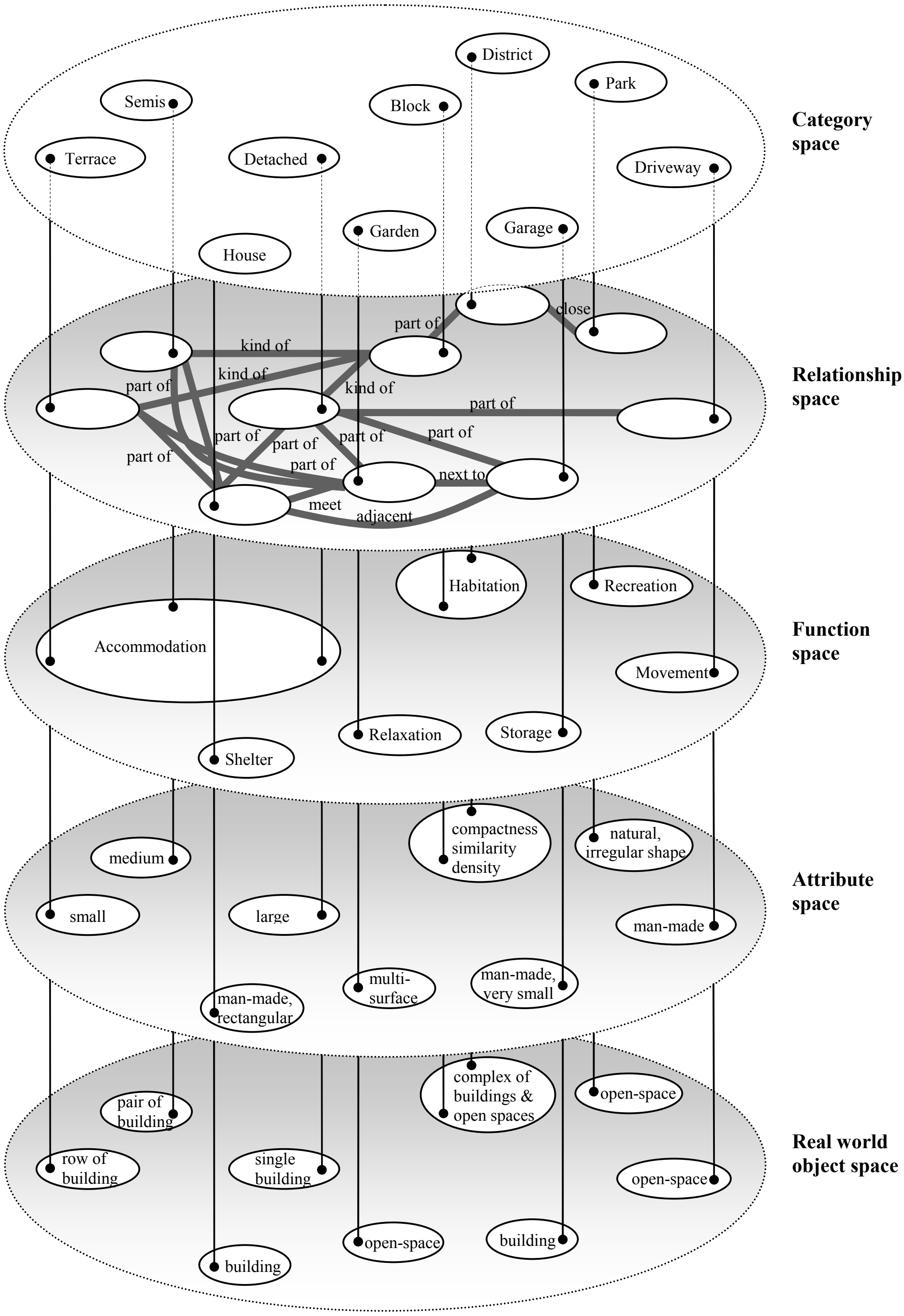


Figure 2. Decomposing the high level concept 'residential area' into its parts (derived from Thomson and Béra, 2007)

3. Model-Based Recognition of the Concept 'Residential Area' from Topographic Data

The outlined mapping in figure 2 translates into a conceptual framework that allows the processing and assignment of higher level semantic information to spatial data. Figure 3 illustrates how implicitly represented knowledge within a spatial database (such as residential area) can be inferred from knowledge that is explicitly defined through ontologies.

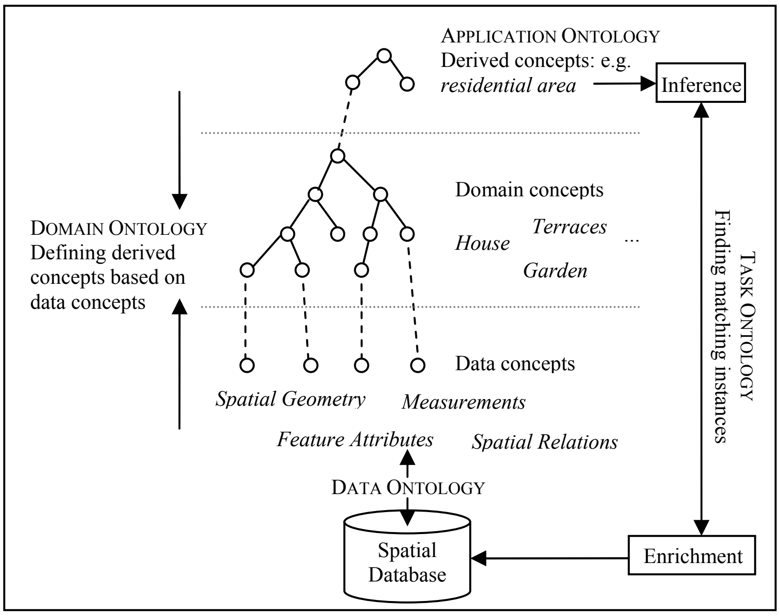


Figure 3. Relating higher level concepts to the data structure

The modelling takes a bottom-up, agglomerative approach where each primitive element, describing a separate data entity (e.g. building) becomes part of its parent aggregate concept (e.g. terraces) until the high level interpretation (i.e. residential area) of the scene is reached. Top-down modelling is adopted to define the part-whole aggregation based on certain criteria and rules which will constrain reasoning of the description logic. The incremental steps and different aggregate levels are illustrated in figure 4.

This methodology uses description logic (in a way similar to Neumann and Möller (2006) for scene interpretation), where the recognition of the whole (scene or map) arises from the recognition of its parts (aggregate concepts). It follows the tendency to continually abstract through simplification as people perceive space as a composition of simple geometries and similarities (Batty and Longley, 1994). To create a diversity of patterns and parts through a system, generic design guidance is required which specifies the elemental units or set of units to be recognised dependent on the purpose of the classification. Form plays an important role, since space is not only observed and understood in terms of its spatial pattern, but is composed of such elements. Modelling higher level meanings is difficult because of the mix of heterogeneous activities and uses which have a high complexity, threatening the classification of their geometry, as well as impeding objective and consistent categorisations (Batty and Longley, 1994). Nevertheless, through simplification a system structure can be built composed of elements and relations which decompose into further subsystems arranged into distinct hierarchies of taxonomies and partonomies. Although we have to be careful not to force the diversity of functions into a narrow range of concepts, providing a system structure enables inference of higher level information based on reasoning about the defined concepts and finding its relevant instances. Therefore, a parser working according to some configurational rule can incrementally build up the semantic interpretation of a map scene using the corresponding object semantic rules of spatial composition.

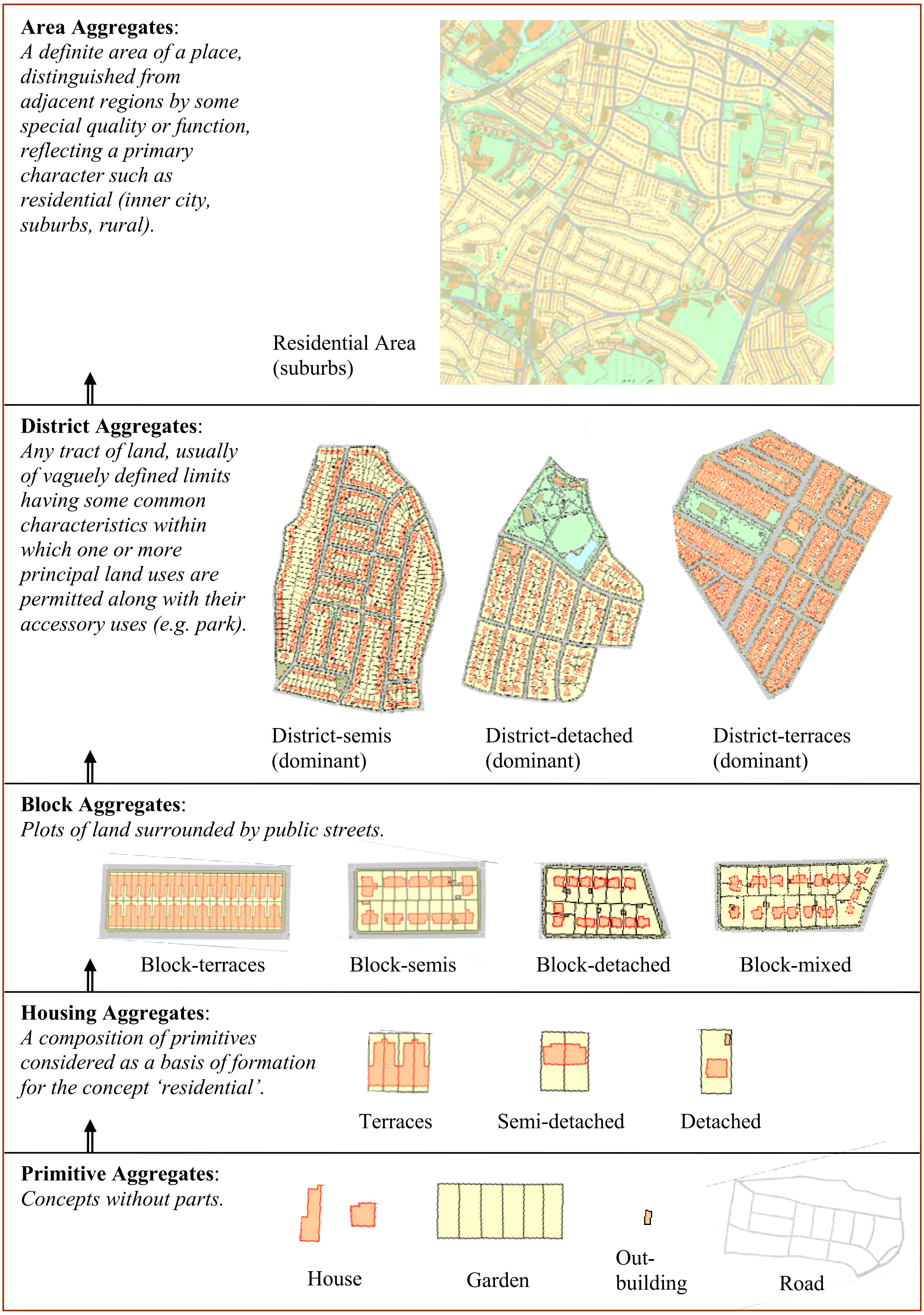


Figure 4. A system structure of aggregate concepts defining compositional elements of 'residential area'

4. Conclusion

This paper reviewed part-whole aggregation in the context of deriving higher level concepts from spatial data. By demonstrating how a hierarchy of spatial composites conveys implicit functional meaning, similar to the syntax of a natural language sentence, we can conclude that such meaning can be made explicit and derived from spatial data. The presented methodology is important for making data more definitive, intelligent and accessible. Function in particular has been identified as one of the five basic ontological relations that make geographic information more explicitly meaningful (Kuhn, 2007). The enrichment of spatial databases with such meaningful information is an essential part of the abstraction process for producing on-demand, customised, and multi-representational map products (Regnauld, 2007). Not only is there the need to reflect more the way people perceive the world (Mennis *et al.*, 2000), but to develop new frameworks that maintain a direct link to the data but can also be disassociated to ensure flexibility and allow the association with different data sources. Future work concerns the implementation of the proposed system through current technologies evolving around ontologies and description logics.

5. Acknowledgements

This work is part of a project funded by the Engineering and Physical Sciences Research Council as well as Ordnance Survey, the national mapping agency of Great Britain. Many thanks go to Nicolas Regnauld (OS) and Muki Haklay (UCL) for providing valuable comments to the writing of this extended abstract.

References

Barr, S. and Barnsley, M. (1997) A region-based, graph-theoretic data model for the inference of second-order thematic information from remotely-sensed images. *International Journal of Geographical Information Science* **11**(6): 555-576

Batty, M. and Longley, P. (1994) *Fractal Cities*. London: Academic Press Limited

Blutner, R. (2002) Lexical semantics and pragmatics. In Hamm, F. and Zimmermann, T. E. (Eds.): *Semantics, Linguistische Berichte*, Sonderheft 10, pp.27-58

Burrough, P. A. and Frank, A. U. (1995) Concepts and paradigms in spatial information: are current geographical information systems truly generic? *International Journal of Geographical Information Systems* **9**(2): 101-116

Frank, A.U. (1992) Spatial concepts, geometric data models, and geometric data structures. *Computers & Geosciences* **18**(4): 409-417

Freksa, C. and Barkowsky, T. (1995) On the Relation between Spatial Concepts and Geographic Objects. In Burrough, P. and Frank, A. (Eds) *Geographic Objects with Undetermined Boundaries*. London: Taylor & Francis

Haggett, P. and Chorley, R. J. (1969) *Network analysis in geography*. London: Edward Arnold

Hart, G. (2007) *Ontology Database Connectivity – Field Definition.* Ordnance Survey Research Labs, Internal report, Crown copyright

Hereth, J., Stumme, G., Wille, R. and Wille, U. (2000) Conceptual knowledge discovery and data analysis. In Mineau, G. and Ganter, B. (Eds): *International Conference on Conceptual Structures*, LNCS 1867, pages 421-437. Springer-Verlag

Katz, J. J. and Fodor, J. A. (1963) The structure of a semantic theory. *Language* **39**(2): 170-210

Klien, E. and Lutz, M. (2005) The role of spatial relations in automating the semantic annotation of geodata. In Cohn, A.G. and Mark, D.M. (Eds): COSIT 2005, Lecture Notes in Computer Science 3693, pp.133-148, Berlin Heidelberg, Springer-Verlag

Kuhn, W. (2007) An Image-schematic account of spatial categories. In Winter, S. *et al*. (Eds.): COSIT 2007, LNCS 4736, pp.152-168, Springer-Verlag Berlin Heidelberg
Lévi-Strauss, C. (1963) *Structural Anthropology*. New York: Basic Books
Lewis, D. (1970) General semantics. *Synthese* **22**(1-2): 18-67
Lüscher, P., Burghardt, D. and Weibel, R. (2007) Ontology-driven enrichment of spatial databases. 10th ICA Workshop on Generalisation and Multiple Representation, 2-3 August 2007, Moscow
Marshall, S. (2005) *Streets and Patterns*. London: Spon Press
Mennis, J.L., Peuquet, D.J. and Qian, L. (2000) A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science* **14**(6): 501-520
Neumann, B. and Möller, R. (2006) On Scene Interpretation with Description Logics. In Christensen, H. I. and Nagel, H.-H. (Eds.): Cognitive Vision Systems. LNCS 3948, pp.247-275 Springer-Verlag Berlin Heidelberg
Regnauld, N. (2007) A distributed system architecture to provide on-demand mapping. *Proceedings of the XXIII International Cartographic Conference*, 4-10 August 2007, Moscow, Russia
Peuquet, D.J. (1988) Representations of Geographic Space: Toward a Conceptual Synthesis. *Annals of the Association of American Geographers* **78**(3): 375-394
Thomson, M.-K. and Béra, R. (2007) Relating Land Use to the Landscape Character: Toward an Ontological Inference Tool. In Winstanley, A. C. (Ed.): *Proceedings of GIS Research UK Conference*, *GISRUK'07*. Maynooth, Ireland, 11-13 April 2007, pp. 83-87
Torres, M., Quintero, R., Moreno, M. and Fonseca, F. (2005) Ontology-driven description of spatial data for their semantic processing. In Rodríguez *et al*. (Eds.): *GeoS 2005*, LNCS 3799, pp. 242-249, Springer-Verlag Berlin Heidelberg
Yudkowsky, E. (2007) Levels of Organization in General Intelligence. In Goertzel, B. and Pennachin, C. (Eds.): Artificial General Intelligence, pp.389-501 Springer-Verlag Berlin Heidelberg

Biography

Marie-Kristina Thomson is a PhD student in the department of Civil, Environmental and Geomatic Engineering at UCL researching ways to infer higher level semantic information from spatial databases. She holds a BSc (Hons) in Remote Sensing and GIS from Bath Spa University, and she is currently student representative of the Remote Sensing and Photogrammetry Society.

Dr. Roderic Béra is lecturer in geographical information science at the UCL department of Civil, Environmental and Geomatic Engineering (CEGE). His research interests lie mainly in ontology for modelling relative proximities, proximity and space syntax, and the influence of geography in web-based social networks. He was previously a teaching and research assistant at the French Naval Academy Research Institute.

UncertML: an XML schema for exchanging uncertainty.

Matthew Williams[1], Dan Cornford[1], Lucy Bastin[1], Ben Ingram[1]

[1]School of Engineering and Applied Science, Aston University, Birmingham, UK
Tel. +44 (0) 121-204-3451
Email: williamw@aston.ac.uk

KEYWORDS: uncertainty, interoperability, WebServices, SensorML, GeoWeb

1. Introduction

Authors from Burrough (1992) to Heuvelink et al. (2007) have highlighted the importance of GIS frameworks which can handle incomplete knowledge in data inputs, in decision rules and in the geometries and attributes modelled. It is particularly important for this uncertainty to be characterised and quantified when GI data is used for spatial decision making. Despite a substantial and valuable literature on means of representing and encoding uncertainty and its propagation in GI (e.g.,Hunter and Goodchild 1993; Duckham et al. 2001; Couclelis 2003), no framework yet exists to describe and communicate uncertainty in an *interoperable* way. This limits the usability of Internet resources of geospatial data, which are ever-increasing, based on specifications that provide frameworks for the 'GeoWeb' (Botts and Robin 2007; Cox 2006).

In this paper we present UncertML, an XML schema which provides a framework for describing uncertainty as it propagates through many applications, including online risk management chains. This uncertainty description ranges from simple summary statistics (e.g., mean and variance) to complex representations such as parametric, multivariate distributions at each point of a regular grid. The philosophy adopted in UncertML is that all data values are inherently uncertain, (i.e., they are random variables, rather than values with defined quality metadata).

2. Use Cases

Most data contains uncertainty, arising from sources including measurement error, observation operator error, processing/modelling errors, or corruption. Processing this uncertain data, typically through models (which typically also have errors), propagates the uncertainty. The ability to optimally utilise data relies on a complete description of any uncertainty. Below are three example use cases.

2.1 Error in sensor observations

Many of the observations used in geostatistical modelling arise from sensors, which typically exhibit specific error characteristics. With increased availability of sensor observation data on the Internet, understanding these characteristics is fundamental to meaningful processing. SensorML (Botts and Robin 2007) is an Open Geospatial Consortium (OGC)-standard language that describes the lineage of an observation through the physical processes of a sensor system. Currently, it relies on simple representations of error (e.g., tolerance to two standard deviations). UncertML provides an explicit structure to describe these (potentially complex) sensor error characteristics. Harnessing the expressive capabilities of Geography Markup Language (GML), UncertML may be used directly within a SensorML document, promoting reusability; a key ingredient for interoperability (Erl 2004).

2.2 Interpolation results

The Sensor Web Enablement (SWE) (Botts et al. 2006) is an OGC-run initiative which provides XML and Web Service standards to enable the interchange of heterogeneous sensor web information on the Internet. As published sensor networks proliferate, the importance of processing chains which can consume the published observations can only increase. An example of such processing is the production of interpolated maps from discrete point observations; a context specifically addressed by the authors as part of the INTAMAP (INTeroperability and Automated MAPping) project (Williams et al. 2007). INTAMAP will provide a fully automated interpolation Web Processing Service which can be consumed by informed clients. However, for the results to have real value they must describe the uncertainties inherent in the sensor data, as well as those introduced by the interpolation process.

2.3 Processing chains

UncertML is relevant to any processing activity involving an explicit model; e.g., a simulator which maps a set of input values to output values. As simulators imperfectly describe the real systems they represent, they must account for system uncertainties. Methods for the probabilistic treatment of simulators are being developed (O'Hagan 2006) that will in future encourage more careful specification of model error and of its propagation through processing chains. We envisage such processing chains to be composed of processing applications connected in a distributed system using Web Service interfaces.

3. Requirements

Each UncertML uncertainty type can be encoded for a variety of spatial domains (Point, Line, Polygon and Grid). Over these domains, users may request summary statistics (e.g., Mean, Variance, Quantiles), or detailed descriptions such as a full parametric distribution, or a set of *n* samples from that distribution. UncertML provides hard-typed parametric distributions (e.g., GaussianDistribution) to improve usability, but also allows users to construct distributions at run-time, by defining the component variables, parameters and affordances. UncertML will also enable the use of mixture models by chaining generic distributions, effectively offering a universal density model. These approaches are consistent with our priorities of flexibility, usability, performance, portability and extensibility.

4. UncertML design

4.1 GML inheritance and common properties

GML is an established XML framework for describing the real world (Portele 2007). This framework can be extended through GML application schema to allow detailed description of real world phenomena while conforming to a standardised structure. This flexibility has led to widespread adoption of GML. UncertML utilises the benefits of GML by ensuring base types inherit from the GML *AbstractFeatureType*, via the UncertML abstract base Feature, *Uncertainty* (Figure 1). This inheritance allows UncertML types to be seamlessly integrated into any current XML language that understands the GML feature model.
Each UncertML uncertainty type has a set of common properties: *variables*, *parameters*, an *encoding block* that stores the associated values and a number of *affordances* (permitted operations for the distribution, with a description of required inputs and outputs, e.g., getMoments).

4.2 Relation to SWE Common

The Sensor Web Enablement initiative is very relevant to UncertML, thus careful attention has been paid to existing SWE technologies. For example, the SWE Common namespace (Cox 2006; Botts et al. 2006) provides a dictionary with common variable types (e.g., temperature, atmospheric pressure) which can be used with UncertML.

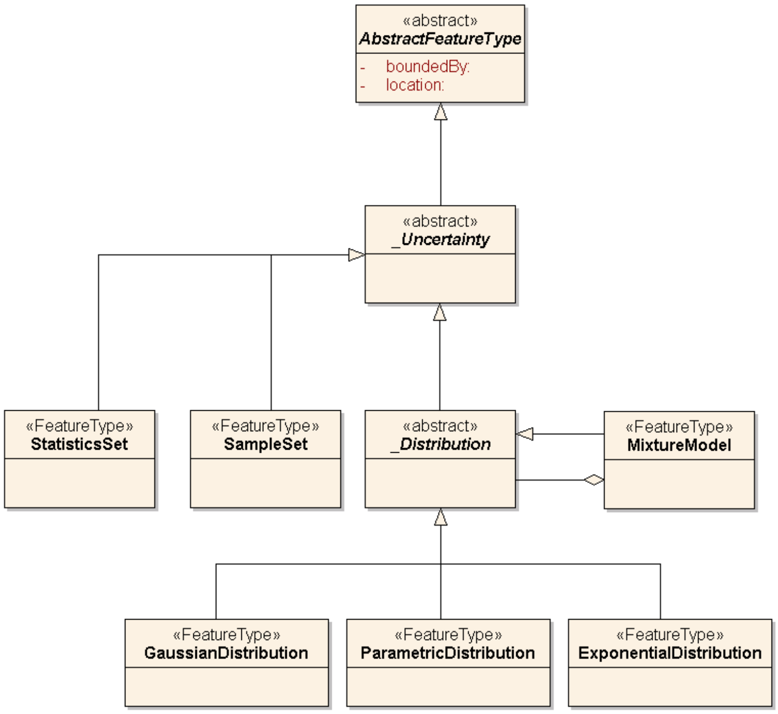


Figure 1. UML diagram displaying the inheritance hierarchy of UncertML.

5. Discussion and future directions

UncertML is currently in version 1, which has flexible support for continuous random variables. This representation is suitable for many environmental variables; however future releases will see added support for discrete and categorical random variables and possibly fuzzy representations. We note that UncertML is designed to encode *any* type of uncertainty. Whilst the above use cases highlight application to attribute uncertainty, location uncertainty is plainly an issue for geospatial data. We envisage that future iterations of GML might replace coordinates in the geometry types with *Uncertainty* types.

In the rare situation of complete knowledge the (Dirac) delta distribution can support deterministically known values. This has the advantage that one must explicitly state that the value is known with certainty (a strong statement). If one merely wishes to provide a value, then a mean, mode or median can be given without a dispersion measure, making the user aware that this is only a characterisation of the distribution’s first moment, which should be treated with due care.

An example application in which UncertML is used to inform interpolations of European raditation data, via a Web Processing Service, is presented. In further work we will develop an extensive API to support UncertML, which we hope will form the basis for many schemata dealing with uncertain objects or phenomena. We also plan to extend the work to conditional distributions and stochastic processes, providing a mechanism to communicate uncertain models as well as results.

Acknowledgements

This work is funded by the European Commission, under the Sixth Framework Programme, by Contract 033811 with DG INFSO, action Line IST-2005-2.5.12 ICT for Environmental Risk Management.

References

Botts M, Percivall G, Reed C, and Davidson J (2006) OGC Sensor Web Enablement: Overview and High Level Architecture. *Technical report, OGC.*

Botts M and Robin A (2007) OpenGIS Sensor Model Language (SensorML) Implementation Specfication.

Burrough P (1992) Development of intelligent geographical information systems. *International Journal of Geographical Information Science* **6** pp497-513.

Couclelis H (2003) The Certainty of Uncertainty: GIS and the Limits of Geographic Knowledge. *Transactions in GIS* 7(2) pp165–175.

Cox S (2006) Observations and Measurements. *Technical report 05-087, OGC.*

Duckham M, Mason K, Stell J, and Worboys M (2001) A formal approach to imperfection in geographic information. *Computers, Environment and Urban Systems* **25** pp89–103

Erl T (2004) *Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services*. Prentice Hall, USA.

Heuvelink G, Brown J and van Loon E (2007) A probabilistic framework for representing and simulating uncertain environmental variables. *International Journal of Geographical Information Science* **21(5)** pp1-11.

Hunter G and Goodchild M (1993) Managing Uncertainty in Spatial Databases: Putting Theory into Practice. *Journal of the Urban and Regional Information Systems Association* **5(2)** pp55-62.

O'Hagan A (2006) Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering and System Safety*, **91** pp1290-1300.

Portele C (2007) OpenGIS Geography Markup Language (GML) Encoding Standard. *Technical report, OGC.*

Williams M, Cornford D, Ingram B, Bastin L, Beaumont A, Pebesma E, and Dubois G (2007) Supporting interoperable interpolation: the INTAMAP approach. In *International Symposium on Environmental Software Systems*, Prague, May 2007.

Biographies

Matt Williams is a PhD student at Aston University looking at describing uncertainty using an interoperable framework (UncertML). He also maintains an interest in geospatial Web Service specifications including WFS, WMS & WPS.
Dan Cornford is a senior lecturer in Computer Science at Aston University. After a PhD in Spatial Statistics and Applied Meteorology at Birmingham University he moved to work in machine learning, focussing on application of principled probabilistic methods to complex systems, with particular emphasis on data assimilation and inference in stochastic processes.

Lucy Bastin is a Lecturer in GIS at Aston University. After a PhD on urban plant metapopulations, and research into fuzzy classification / uncertainty visualisation at Leicester University, she worked as a GIS software developer. Her research interests include spatial epidemiology and automatic interpolation via WPS.
Ben Ingram is a Research Fellow at Aston University working on the INTAMAP project. His interests lie in the development of geostatistical methods for treating large data sets, with emphasis on parallel treatments and sequential approaches.

An Agent-based Approach to Simulating Spatial Patterns of Shifting Cultivation in Vietnam

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 is a type of agricultural system where land is cleared by slashing and burning and plots are rotated for food crop production (Angelsen, 1995; Do, 1994; Gilruth *et al.*, 1995). In upland Vietnam, shifting cultivation is a popular and traditional method of agricultural production where more than 4.8 million households are engaged in shifting cultivation (Do, 1994). However, Vietnam's national government and resource managers currently view shifting cultivation as both highly destructive to the forest and the cause of serious soil erosion (Vien *et al.*, 2006). Planners, therefore, need a decision making tool to manage shifting cultivation and protect the forest, that provides visual monitoring of future impacts of policy and other external interventions. A method that has shown promise for achieving these goals is agent-based modelling (ABM) (Bonabeau, 2002). This approach allows one to effectively describe this type of agricultural practice at household level, in terms of livelihood, as well as at the landscape scale, in terms of land-cover change and natural resource management. This paper presents an agent-based model of an upland village in Vietnam where shifting cultivation is the main type of livelihood. The model framework is presented along with some initial results.

2. Previous Work

Although shifting cultivation has been a topic of research for a some time (Bui, 1990; Do, 1994; Dvorak, 1992; Le, 1993; Le *et al.*, 1990), the application of GIS and agent-based modelling to this type of cultivation system is relatively recent. For example, Gilruth *et al.*(1995) have developed a dynamic model of shifting cultivation in Guinea, Brown (2005) has built a spatiotemporal model of shifting cultivation and forest cover dynamics in Cameroon, and Becu *et al.* (2003) have developed an agent-based simulation of a small catchment for water management in Thailand. Two previous studies have been carried out in Vietnam: agent-based modelling of shifting cultivation field patterns (Jepsen *et al.*, 2006) and modelling of agrarian transition and lowland-upland interactions (Castella *et al.*, 2005). However, limitations have been identified with these studies. For example, the model by Jepson et al. (2006) is simplistic in its assumptions about the decision making context such as ignoring all economic factors, e.g. price or household potential capital that could be unreasonable to other regions. There is also no relationship between yield, which is derived from the cell age, and labour cost, which is based on the number of adjacent cultivated cells. The research in this study improves upon on both of these aspects.

3. Agent-based Model of Shifting Cultivation

The agent based model developed in this research is based on the concepts of the FEARLUS (Framework for Evaluation and Assessment of Regional Land-use Scenarios) and FLORES (Forest Land Oriented Resource Envisioning System) agent-based models (Castle *et al.*, 2006; Cioffi-Revilla and Gotts, 2003; Haggith *et al.*, 2003). However, the main difference is that the basic behaviours of the agents, which includes social decision-making (e.g. marriage, giving birth, and household partitioning) and land-use decision-making (e.g. land tenure transference, land selection for cultivation etc.) have been formulated in a much more explicit manner. This formulation is also more compatible with Vietnam's conditions than those of other existing models. For instance, FEARLUS operates at a household level and is therefore limited in its ability to simulate and monitor the dynamics of individuals within a household. Whilst FLORES is very well developed for simulating individual behaviours at a micro level such as different types of land tenure transference and labour distribution, it does not mention details of how personal perceptions and land accessibility are collectively formulated in the land-use decision making of communal lands (Haggith *et al.*, 2003).

The model of shifting cultivation proposed here is divided into a land module and a human module. The land module simulates the dynamics of the land characteristics such as land-cover and land-use, soil, fallow length, cultivation length, land ownership, etc. The human module deals with the social dynamics at both individual and household level such as marriage, birth, death, migration and livelihood seeking strategy, etc. These modules have been coded in Netlogo (Wilensky, 1999) and are integrated in one model to simulate the interaction between humans and the land within a shifting cultivation system. Human effects on the landscape through land-use activities (i.e. shifting cultivation) and the response of the land to human actions are based on land-cover change and economic return (i.e. yield). The main processes that are simulated are illustrated as in Figure 1.

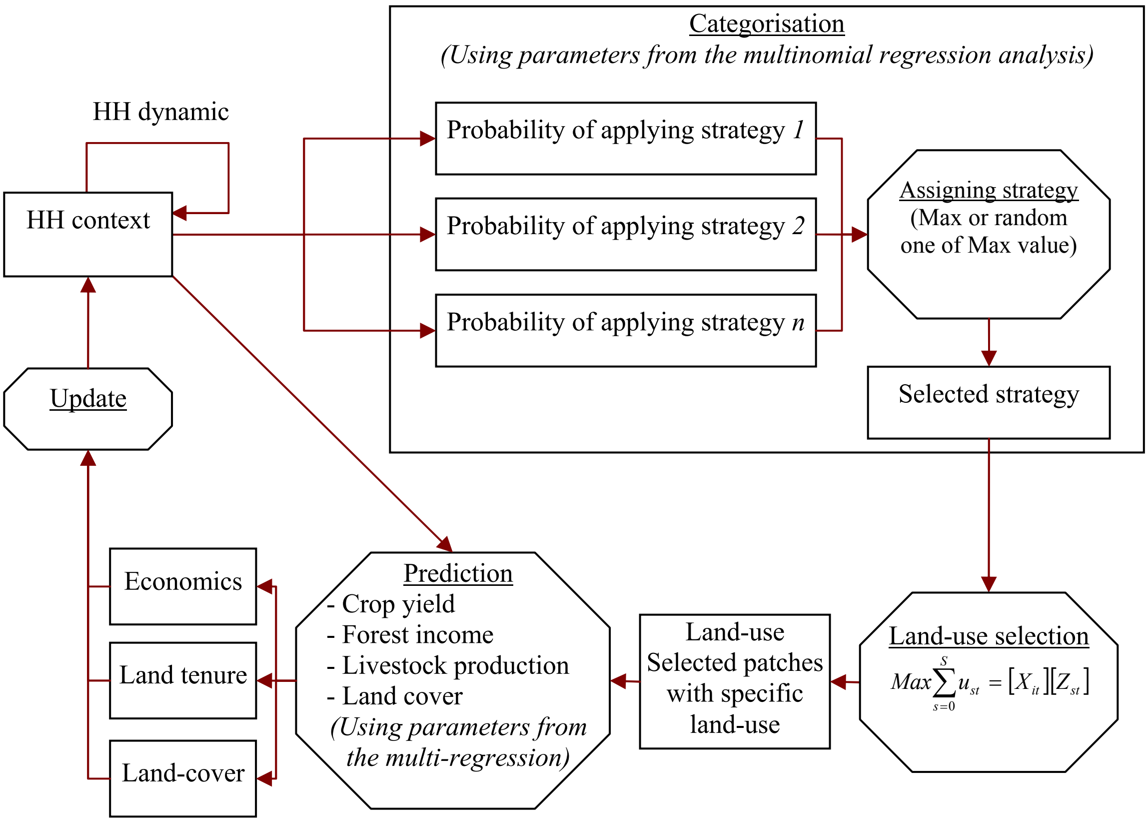


Figure 1. General process of simulation in the shifting cultivation model

The model was developed based on the assumption that the livelihood strategy of the shifting cultivator depends on their household context which includes accessibility and their personal perception of the available resources. The relationships between the household context factors and the land-use strategies were developed using statistical regression on data collected from a village in upland Vietnam. These statistical relationships have been embedded into the model to assign a land-use strategy to each household to determine the selection of fields for shifting cultivation. The selection function is adopted from Brown (2005) as follows:

$$Max\sum_{s=0}^{S} u_{st} = [X_{it}]\,[Z_{st}]$$

where u_{st} is a deterministic component of the utility function inferred by household i at patch s and time t, Z_{st} is the land specific factors and X_{it} is the land preference factors of the household, which can be predicted from a set of household specific characteristics (or context factors). A patch with the maximum value of u_{st} within the accessible land S will be subsequently selected for cultivation at time t.

The resulting land-use at each patch is then predicted by using estimated parameters from statistical analysis and assumptions from the field work data. The yield prediction function is in the form of $Y_{ist} = [H_{it}][Z_{st}]$ where Y_{st} is predicted yield for household i with specific characteristics H_{it} (e.g. education level, land per capita etc.) at patch s with properties Z_{st}. Similarly, land-cover is predicted as $C_{st} = [C_{st-1}][Z_{st-1}][\lambda_t]$ where C is the land-cover type and λ is the human intervention factors that indicate possible human interactions with the field during the fallow period.

This ABM requires information on the status of shifting cultivation at the initial stage ($t = 0$) and possible behaviours of shifting cultivators at the study site. This information has been obtained though satellite image interpretation, field observation and an initial set of interviews carried out in the summer of 2007. Time series satellite images with ground truth data were used to produce land-cover and to predict the fallow length of land patches. All the spatial information has been captured in a GIS database, and then further populated with other key land properties that are important in the farmers' decision making processes regarding shifting cultivation. The household survey data collected in the field were used to formalise the decision making and other collective rules governing swidden field selection at household level while plot interview data was used for estimating the yield prediction model.

4. Model Prototype

Figure 2 shows an example of the integrated land and human modules as a Netlogo implementation. The figure shows an example of running the model for the village of Binh Son-1 with assumptions such as: the degree of forest protection is high so farmers will not cut down forest for shifting cultivation and farmers are not yet allocated land but can move to new areas for cultivation. The model takes both individuals and households into account when modelling the human side. Figure 2 shows an example of the location of active areas of shifting cultivation (in orange) after running the model with these assumptions.

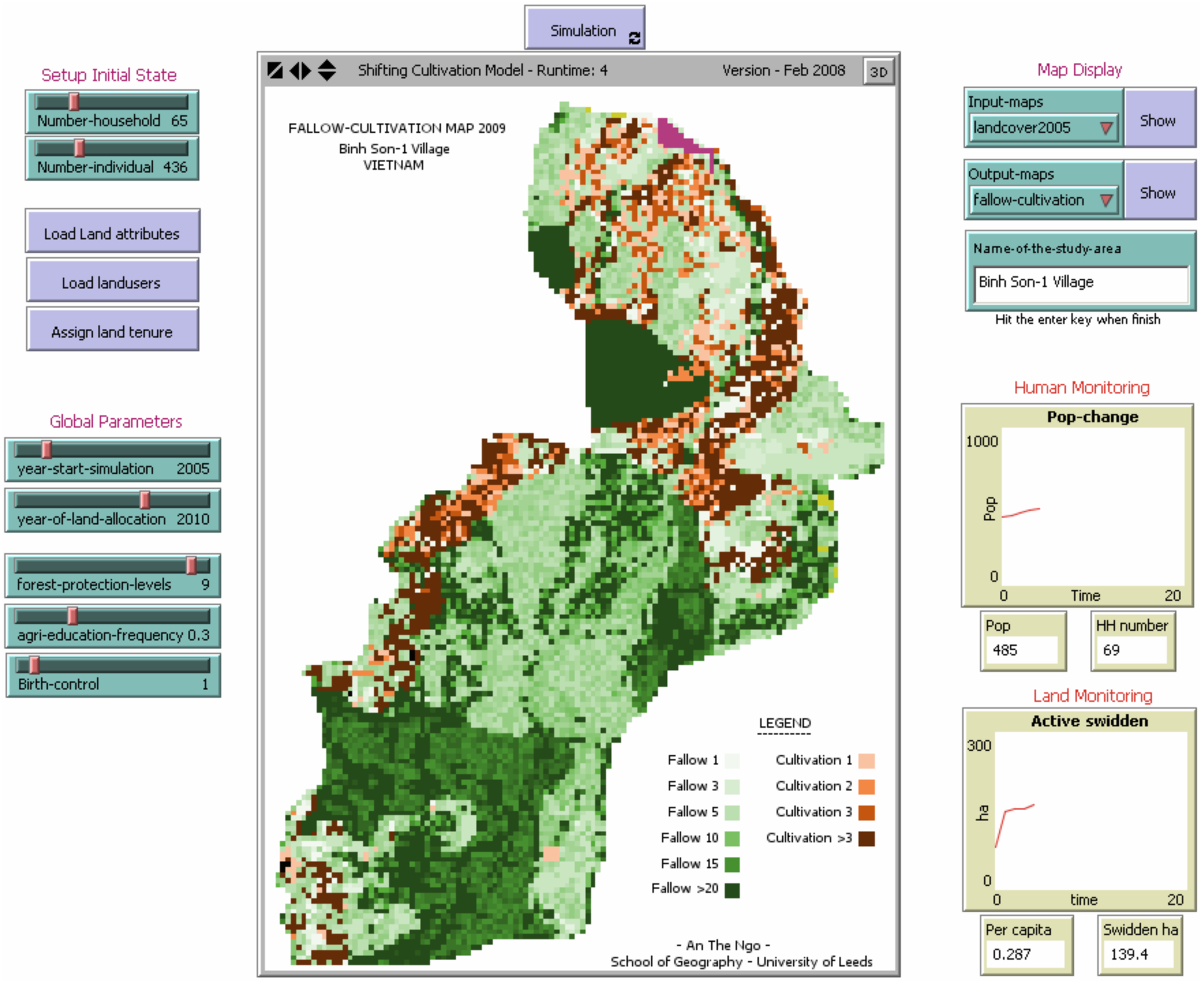


Figure 2. An agent-based model of shifting cultivation implemented in Netlogo

5. Conclusions

This paper has provided a brief overview of the agent-based model of shifting cultivation in Vietnam. Once the model is fully verified and calibrated, it will be used to consider testing different scenarios such as the effects of: (i) different levels of forest protection (ii) agricultural extension training and credit (iii) population controls and (iv) land allocation to farmers following new land laws.

6. Acknowledgements

We would like to acknowledge the support of the Vietnamese government and School of Geography, University of Leeds in funding this PhD research project.

References

Angelsen, A., 1995. Shifting Cultivation and Deforestation - a Study from Indonesia. *World Development*. 23**,** pp.1713-1729.

Becu, N., Perez, P., Walker, A., Barreteau, O. and Le Page, C., 2003. Agent based simulation of a small catchment water management in northern Thailand description of the CATCHSCAPE model. *Ecological Modelling*. 170**,** pp.319-331.

Bonabeau, E., 2002. Agent-based modelling: Methods and techniques for simulating human systems. *PNAS*. 99**,** pp.7280-7287.

Brown, D.R., 2005. A Spatiotemporal Model of Shifting Cultivation and Forest Cover Dynamics. Cornell University.

Bui, Q.T., 1990. Some problems of soil in slash and burn cultivation in the Northwestern region and the direction of its utilisation. Hanoi, Vietnam Agricultural Science Institute.

Castella, J.C., Manh, P.H., Kam, S.P., Villano, L. and Tronche, N.R., 2005. Analysis of village accessibility and its impact on land use dynamics in a mountainous province of northern Vietnam. *Applied Geography*. 25**,** pp.308-326.

Castle, C.J.E., Crooks, A.T., Longley, P.A. and Batty, M., 2006. Agent-Based Modelling and Simulation using Repast: A Gallery of GIS Applications from CASA. IN Priestnall, G. and Aplin, P. (Eds.). *The 14th Geographical Information Systems Research UK Conference.* Nottingham, UK, The University of Nottingham.

Cioffi-Revilla, C. and Gotts, N., 2003. Comparative analysis of agent-based social simulations: GeoSim and FEARLUS models. *Jasss-the Journal of Artificial Societies and Social Simulation*. 6.

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.

Dvorak, K.A., 1992. Resource-Management by West-African Farmers and the Economics of Shifting Cultivation. *American Journal of Agricultural Economics*. 74**,** pp.809-815.

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, T.C., 1993. *Swidden agriculture in the highlands of Northern Vietnam*, Baguio City: The Philippines, Cordillera Study Cerntre.

Le, T.C., Gillogly, K. and Rambo, A.T., 1990. Agroecosystems of the midlands of Northern Vietnam. Series Agroecosystems of the midlands of Northern Vietnam. Honolulu, East-West Center Environment and Policy Institute.

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.

Wilensky, U., 1999. NetLogo. [online]. [Accessed 2/11/2007]. Available from World Wide Web: http://ccl.northwestern.edu/netlogo/.

Biography

An The Ngo is a second year PhD student in the School of Geography, University of Leeds. Back in Vietnam, Mr. Ngo is a lecturer in GIS application for natural resources management at Hanoi Agricultural University.

Dr Linda See is a Senior Lecturer in GIS and Senior Developer of online learning materials. 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 particular interested in how climate change is being made amenable to government and technology choice in environmental issues.

"I'd rather be watching the telly…": Do rich media approaches offer real teaching and learning benefits for GIS software tuition of digital natives?

Nick Mount and Claire Chambers

School of Geography, University of Nottingham, Nottingham NG7 2RD
Tel. +44(0)115 951 5428 Fax. +44 (0)115 951 5249
nick.mount@nottingham.ac.uk claire.chambers@nottingham.ac.uk

KEYWORDS: software teaching, MP4, podcasting, media richness, digital natives

1. Introduction

Interest in effective approaches to teaching and learning in the spatial sciences is evident through recent funding of major initiatives such as the Spatial Literacy in Teaching (SPLINT) centre of excellence (University of Leicester, University of Nottingham and University College London). Indeed, SPLINT is concerned with the challenges and opportunities facing geographical information science educators including; (1) the increasing recognition of our students as 'digital natives' (Prensky, 2001), (2) the ease with which media-rich materials (e.g. audio and video podcasts) can be now created and distributed (Mount and Chambers, *in press*) and, (3) the pressures from universities and organisations such as the Higher Education Academy (HEA) and Quality Assurance Agency (QAA) to improve teaching and learning quality through technology adoption. To this end, it is involved in investigating the pervasive view within Higher Education that sees the adoption of the latest technologies in teaching and learning materials in any subject discipline as a positive goal, with learning benefits following such developments.

Too often it assumed that learning benefits will be inevitable and this view can be traced back to elements of the education literature in which a causal link between media richness and learning outcomes has been asserted (e.g. Kozma, 1994). Given this background, it is surprising that approaches to software tuition in GISc modules often remain locked in the production and use text-based manuals and screenshots: a very low-technology teaching solution to a very high-technology discipline. The questions of whether these traditional materials represent best practice, what alternatives exist, and what the learning benefits of these alternatives are must surely be considered.

In this paper the arguments surrounding rich media and improved learning will be reviewed and the experiences of adopting rich-media materials (video-podcasts) for

GIS software tuition in the University of Nottingham's Introduction to GIS module will be considered.

2. The media richness debate

Placing the above questions into the educational literature requires consideration of the media-richness debates of the last 25 years. These debates centre on whether learning is influenced more by content and instructional strategy than the medium of delivery or, indeed, whether the richness of the media used to deliver instructional material has any learning benefits at all (compare Kozma, (1998) and Clarke, (2001a) for very different views). Indeed, the notion of a simplistic, positive causal relationship between media richness and learning is now broadly considered to be false, with more complex questions about how the capabilities of media should be used to influence learning for particular students, tasks and situations being of crucial importance.

This is a much more helpful approach to the problem, and one which implies the importance of using media to influence cognition and the cognitive process. Several cognitive studies show, the traditional manual does not integrate information sources well and the effort required to achieve such integration can cause difficulties for many learners (Clark, 2001b). For example, Mousavi et al. (1995) show cognitive improvements in understanding integrated information for complex scientific concepts and Sweller (1999) identifies benefits in terms of data redundancy.

The debate also extends to considerations of student motivation and studies investigating the motivational benefits of media-rich, televisual instruction versus print. A positive relationship between media-richness and student motivation to learn would appear an intuitive outcome for the contemporary, digitally-native student. However, evidence from classic educational psychology studies (e.g. Salomon, 1984) highlights the fact that a student's motivation is influenced by their expectations about their chances of learning from a given media, rather than the media alone.

Whilst a simplistic link between media richness and learning does not exist, the capability of media-rich materials to improve cognition via improved information integration, reduced information redundancy and improved learner efficiency has been shown to be beneficial. Coupled with this, engagement with rich media is likely to improve student motivation, not through the media itself, but rather because of learner expectations.

3. Introduction to GIS: module strategy.

The University of Nottingham, requires all of its undergraduate students in geography programmes (between 180 and 200 students each year) to complete a compulsory level 1 module in GIS. The module, covering 10 weekly sessions, comprises a weekly lecture (approximately 1 hour) in which the emphasis is on theory, and a weekly practical (approximately 2 hours) in which emphasis is placed on linking and embedding theory through practical experience whilst gaining skills in using and applying GIS software.

Prior to 2006/07, teaching materials for the module comprised text and screenshot manuals. Student feedback indicated that practical classes were unpopular, with the text-based materials perceived as difficult to follow. Students lacked confidence about their ability to meet practical learning outcomes and had poor motivation. In response to these concerns, a strategy was developed (figure 1), centred on the replacement of text-based manuals with video podcasts, (note that the term *podcast* is used in a loose manner with no RSS delivery required).

In total 24 separate video podcasts were produced, totalling 136 minutes of viewing. These replaced 60 pages of paper-based manuals and 45 screenshot illustrations. The key benefits to learners were recognised as centring on:

1. improving student cognition through improved information integration, reduced information redundancy and an improved representation of the dynamics of software operation;
2. improving teaching material flexibility by offering rewind and recap functionality and the ability to access learning materials flexibly through web-based streaming or as MP4 downloads to PCs or mobile devices;
3. improved student motivation through engagement with the mobile devices and podcasting technologies which are an integrated component of contemporary student's lives.

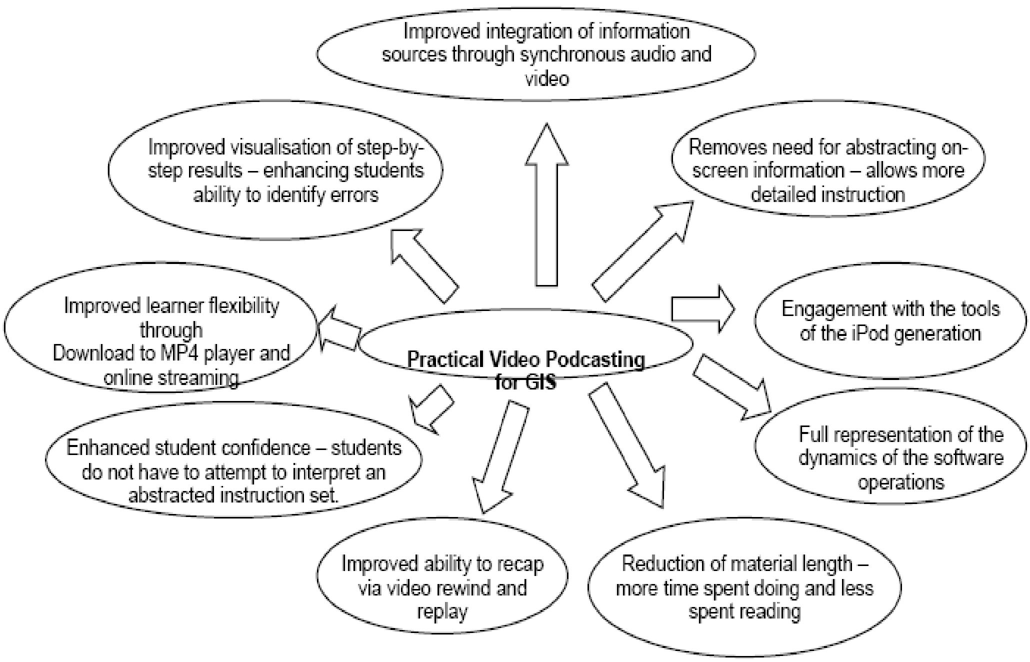


Figure 1. The perceived benefits of video podcasting in the practical elements of the Introduction to GIS module, created as part of the strategy developed during the re-design of the module.

4. Research Methodology and Process

A three-pronged evaluation process was employed, undertaken by external researchers. Quantitative data were captured through an end-of-semester questionnaire (n=96), and WebCT logs of individual student activities and marks.

Qualitative data were captured through four focus groups and nine personal interviews with students and analyzed using a grounded theory approach.

5. Results

5.1. Qualitative results

Analysis of the transcripts identified four main categories for which the use of podcasts was identified as beneficial by the groups:

1. Material flexibility and reuse
2. Promotion of independent learning
3. Cognitive benefits of rich media
4. Student engagement and motivation

The indication of significant material reuse provided by the questionnaire was confirmed in the focus groups and interviews with students commenting on their repeated use of podcasts to support coursework and allow flexible learning by undertaking practical exercises outside of the designated practical periods. Students noted the value of being able to fast forward, rewind and recap and indicated that the video podcasts, giving students control over their learning, enabling them to work independently and providing improved confidence. Several students directly highlighted the cognitive benefits of integrated rich media, over text-based instructions and recognised the benefits of dynamic materials in which a representation of *where* to go, as well as *what* to do was provided. The video podcast approach provided students with improved clarity in the instructions given and this helped to improve motivation through enhanced confidence. Students also commented on the freshness of the podcasting approach being engaging: in particular, the contrast with text-based materials requiring substantial reading was made.

5.2. Quantitative results

End-of-module questionnaires assessed the ability of students to access the podcast technology and student engagement with the podcasts. 100% of students had access to personal computing facilities over and above those provided by the university, with a high proportion of MP4 players available on personal devices including mobile telephones (21%) and video iPods (77%). The high proportion of students in university accommodation meant that 89% had broadband access for enabling podcast download. Students employed a flexible approach to using the podcasts with material reuse: 34% indicated that their use was split equally on and off campus. However, only 10% downloaded the podcasts to laptop or MP4 players. Given the high proportion of students reporting ownership of MP4 players (77%), this may indicate an unwillingness for students to place ‘work’ material onto devices they consider as ‘entertainment’.

5.3. User statistics

Individual student user statistics (n=179) were collected for total number of podcast downloads, marks for the practical assessment item and the overall mark for the module, with incomplete records removed. Podcast download numbers were regressed against practical assessment marks and overall module marks as a measure of the direction and strength of the relationships between them. Unsurprisingly, both relationships are also statistically weak with the correlation coefficients indicating that improving marks can not be explained solely on the basis of podcast downloads in either case (no coefficient was > 0.02).

7. Summary

The evaluation highlights the motivational benefits of video podcasting. Certainly, the doubling of the uptake of level 2 GIS modules following the introduction of video podcasting (an average of 17% between 2002 and 2005, and 46% in 2006) would strongly suggest improved student motivation. Video podcasting improves learner cognition through better integration of the visual and textual materials found in text-based manuals and improves learner efficiency through reduced information redundancy and a less abstract representation of the steps needed for effective software operation. It encourages a more flexible approach to learning, offers a new element of learner independence and control and improves student motivation by directly engaging with contemporary student's expectations about the benefits of mobile technologies for their learning.

Media-rich materials should be considered a realistic and achievable alternative to text-based manuals for those involved in the teaching of GISc, but only where clear evidence exists that students are equipped with the necessary technology and hardware to benefit from them. Like any other enriched media, video podcasting is not a 'quick fix' for improving student achievement within a module. Indeed, simply replacing text-based manuals with podcasts is unlikely to have any significant learner benefits unless it is accompanied by a clear teaching strategy which considers the limitations of existing media approaches, the nature of the benefits (both direct and indirect) offered by podcasting and the underlying instructional method upon which the media will be developed.

8. References

Clark, R.E. (2001a). *Learning From Media: Arguments, Analysis and Evidence*. Information Age Publishing, Charlotte, NC, USA, pp. 220.

Clarke, R.E. (2001b). What is next in the media and methods debate? In Clark, R.E. (Ed.), *Learning From Media: Arguments, Analysis and Evidence*. Information Age Publishing, Charlotte, NC, USA, Ch. 18.

Mount, N.J., Chambers, C. (*in press*). Video podcasting to support practical classes. In Salmon, G., Hawkridge, D., Nie, M. (Eds.), *Podcasting in Higher Education*, McGraw Hill.

Mousavi, S., Lowe, R., Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87, 319-334.

Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon,* 9(5), 1–2.

Salomon, G. (1984). Television is "easy" and print is "tough": the differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, 76(4), 647-658.

Sweller, J. (1999). Instructional Design in Technical Areas, Australian Council for Educational Research, Camberwell, Victoria, Australia. pp. 292.

Biographies

Nick Mount is a lecturer in GISc at the University of Nottingham, and partner on the Higher Education Academy-funded 'Informal Mobile Podcasting and Learning Adaptation' project. Claire Chambers is a learning technologist in the School of Geography, University of Nottingham.

GIS education by another name? Geographical referencing for social scientists

David Martin, Samantha Cockings and Samuel Leung

School of Geography, University of Southampton, Southampton, SO17 1BJ, UK
Tel. +44 (0) 2380 593808 Fax +44 (0) 2380 3295
D.J.Martin@soton.ac.uk, S.Cockings@soton.ac.uk; Y.Leung@soton.ac.uk
http://www.geog.soton.ac.uk/geo-refer/

KEYWORDS: georeferencing, social science, online learning, adaptive learning, GIS education

1. Introduction

This paper presents the work of an ESRC-funded project entitled *Geo-Refer: Geographical Referencing Resources for Social Scientists*. The underlying motivation for this work was to address the needs of social science researchers who have to work with geographically referenced information but who do not have any background in GIS, most of whom cannot realistically be expected to engage in formal GIS training ancillary to their primary research. The project arose from the authors' repeated experience of being approached by colleagues (and strangers!) with a range of spatial analysis needs, who were unfamiliar with the key concepts, reference datasets and software tools which are often seen as central to GIS education. We thus devised a series of online learning resources and associated workshops intended to address these learning objectives, delivered through an adaptive learning interface which permits the user to profile their own requirements and obtain a customised online tutorial. Design and implementation of this project has required a careful deconstruction of spatial concepts and skills, resulting in an extensible series of highly granular learning resources. This paper provides an overview of social scientists' needs for geographical referencing skills and describes our learning design and implementation.

2. Social science and the need for geographical referencing

Geographical location provides a key mechanism for linking data between sources, and Jones and Elias (2007) identify a need for social scientists to be better equipped with geographical linkage skills. This is apparent in the case of linkage between individual-level data such as survey responses or health records and existing secondary data such as that provided by the census of population. Similar situations arise when researchers have access to area-based data for two different sets of geographical areas which are not directly comparable – for example 1991 and 2001 census data or 2001 census data and contemporary health service areas. A third instance arises when the researcher is engaged in primary data collection, leading to questions such as 'should addresses be recorded in some standard format?' or 'Which is better: postcodes or placenames?' These questions are usually prompted by a desire to produce maps or to link data in preparation for subsequent analyses.

Public bodies have invested significantly in datasets and services which allow social science users to access geographically referenced datasets. Key data products such as census datasets include geographical codes and products such as the All Fields Postcode Directory relate many alternative geographical systems. The various UK statistical organizations have each developed neighbourhood statistics services which provide geographically referenced data, primarily in support of the neighbourhood renewal agenda (Martin, 2004).

A sound understanding of geographical referencing underpins the ability to properly conduct advanced methods such as multilevel modelling, geographically weighted regression and statistical methods for the integration of many individual and aggregate datasets. Knowledge and practice regarding these issues are frequently derived from the GIS community, but the majority of social science researchers will have neither the time nor desire to attend GIS courses. Researchers will often be experts in their own disciplines and research methods, but have highly focused and time-limited requirements for GIS use.

3. Presenting geographical referencing through adaptive online learning

We have developed a set of online learning objects (Hawley Orrill, 2001) from which customised training sequences can be assembled for individual researchers. They are designed for individual online use but can also support face-to-face courses. In addition, we are now developing alternative interfaces for specialist researcher fields (e.g. public health). Based on our conceptualization of teaching geographical referencing, four types of learning objects have been developed to meet learners' particular requirements: concepts, datasets, methods and examples, illustrated in Table 1.

Table 1. Examples of four types of Geo-Refer learning objects.

Concepts	*Datasets*	*Methods*	*Examples*
The spatial nature of social science data	National Statistics Postcode Directory	Creating a choropleth map in ArcGIS	Mapping Index of Deprivation for the Isle of Wight
Creating a choropleth map	Census Geography: England 2001	Adding data layers to ArcGIS	Relating GP QOF scores to census Super Output Areas
How to match two lists together	Census Geography: Northern Ireland 2001	How to match two lists together using Access	Mapping Devon patient data to rural/urban areas

'Concepts' objects explain fundamental georeferencing principles such as the spatial nature of social science data; map scale; how to match two lists together. Without a firm understanding of such issues, researchers may apply unsuitable techniques or derive inappropriate conclusions (Martin, 1999). Concepts objects are as far as possible context-independent and do not explicitly cite datasets or methods. 'Datasets' objects describe the characteristics and availability of specific spatial datasets (currently limited to the UK) which are used in social science research, such as census and postcode boundaries and directories. The differences between (for example) 1991 and 2001 census geographies or 2001 geographies in England and Scotland are a frequent source of confusion to research users. The independent nature of the learning objects makes it possible to incorporate new datasets within this overall resource framework. 'Methods' objects describe techniques for georeferencing, manipulating and transforming spatial data. Examples include generic and software-specific methods for accessing datasets or carrying out the tasks explained in the concepts objects, such as linking data by list-matching or spatial joins in GIS software. The focus is not only on describing the techniques themselves but also in generating an understanding of the complex issues involved and an appreciation of the potential impacts on resultant maps and analyses. The final type of object, 'Examples', provides exemplars of the use of spatial data in a range of social science applications, including some drawn from our workshop participants.

Brusilovsky and Maybury (2002) point out that many learning websites are far too static to meet the heterogeneous needs of users. The disjointed nature of web resources also means that users have to go through many iterations and undertake careful filtering before extracting the results of a search. Brusilovsky (2001) advocates the alternative use of adaptive hypermedia. Central to the Geo-Refer adaptive learning design is the principle that a user's specific geographical referencing requirements can be met by a selected subset of the resources. A profiling web form elicits user needs in terms relevant to their research and generates a unique user profile. Each of the Geo-Refer learning objects is tagged with metadata describing its content and a script searches the object database, identifying those objects most relevant to the researcher's needs and sequencing these to form a personalised tutorial. The design is summarised in Figure 1:

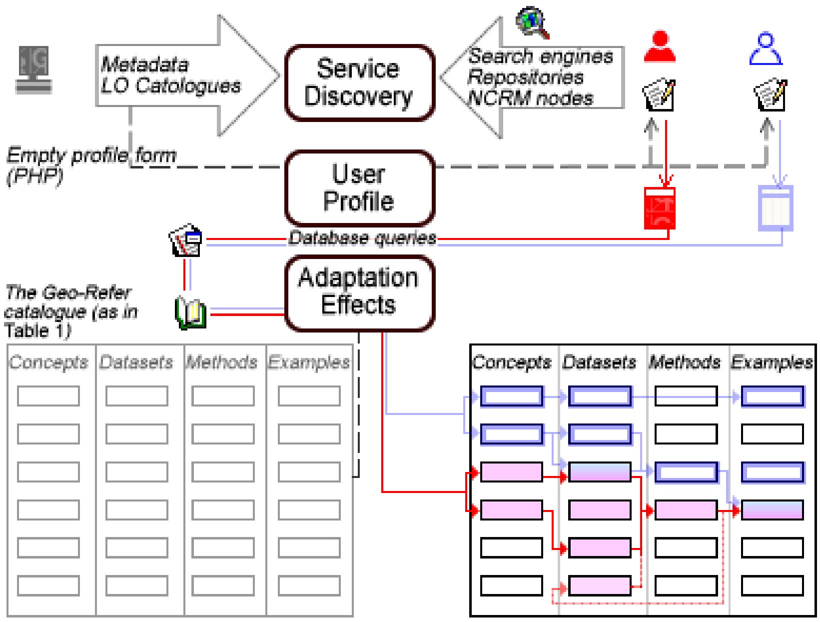


Figure 1. User profiling and Geo-Refer object selection: an example of an adaptive learning interface (after Brusilovsky and Maybury, 2002).

4. Conclusion

We have identified an important GIS learning need not met by conventional taught programmes but with significant research impacts. In response we have developed a set of extensible, adaptive learning materials for online use – either directly or through incorporation within others' teaching. We have developed a four-way classification of learning objects which allows delivery of customised sequences of essential conceptual and practical guidance in response to online profiling of learners' needs. GIS researchers approached by social science colleagues with geographical referencing questions are invited to make use of Geo-Refer at http://www.geog.soton.ac.uk/geo-refer/. We welcome feedback from GISRUK delegates.

5. Acknowledgements

We acknowledge the support of ESRC Awards PTA-035-25-0029 and ES/F018495/1.

References

Brusilovsky, P. (2001) Adaptive hypermedia. *User Modeling and user-Adapted Interaction* **11(1)** pp87-110

Brusilovsky, P. and Maybury, M. T. (2002) From adaptive hypermedia to the adaptive web. *Communications of the Association for Computing Machinery* **45(5)** pp31-33

Hawley Orrill, C. (2001) Learning objects to support enquiry-based online learning In: Wiley, D. (ed.) *The instructional use of learning objects*. Association for Instructional Technology/Association for Educational Communications and Technology http://www.reusability.org/read/

Jones, P. and Elias, P. (2007) Linking Administrative data: current issues Presented at NCRM/SRA Workshop: Data Linkage: Exploring the Potential, Royal Statistical Society, London 19 March

Martin, D. (1999) Spatial representation: the social scientist's perspective In: Longley, P. Goodchild, M. F. Maguire, D. J. and Rhind, D. W. (eds) *Geographical Information Systems: Principles, Techniques, Applications and Management* Second Edition Wiley, Chichester pp71-80

Martin, D. (2004) Neighbourhoods and area statistics in the post 2001 census era *Area* **36** pp136-145

Biography

David Martin is a Professor of Geography at the University of Southampton. His research interests are focused on social science applications of geographical information systems. He is director of the ESRC/JISC Census Programme and a co-director of the ESRC National Centre for Research Methods.

Samantha Cockings is a lecturer in Geography at the University of Southampton. She has extensive experience in teaching geographical information systems and in research particularly with geographically referenced health data.

Samuel Leung is a learning technologist with expertise in the development of online learning resources in geography. He is a surveyor by background and prior to the Geo-Refer project worked on the JISC/NSF funded DialogPLUS project.

e-Learning in GI Sciences – Approaches towards Sustainable Content

Susanne Bleisch[1], Stephan Nebiker[1], Joël Fisler[2],

[1]University of Applied Sciences Northwestern Switzerland (FHNW), CH – 4132 Muttenz
Tel. +41 61 4674 398 Fax +41 61 4674 460 Email (susanne.bleisch, stephan.nebiker)@fhnw.ch
[2]University of Zurich, CH-8057 Zurich
Tel. +41 44 63 55252 Email joel.fisler@id.uzh.ch

KEYWORDS: e-learning, sustainability, xml framework, open content, user community

1. Introduction

In the year 2000 the Swiss Virtual Campus Initiative (Edutech 2007) was initiated. During seven years over 80 projects engaged in the development of e-learning content in many different higher education subject areas were financially supported. Among them also several projects dealing with geographic information: NAHRIS 'Dealing with Natural Hazards and Risks' (NAHRIS 2007), CartouCHe 'Cartography for Swiss Higher Education' (CartouCHe 2007) or GITTA 'Geographic Information Technology Training Alliance' (GITTA 2007). Most of these projects are now or will soon be completed. However, the developed e-learning content shall remain usable and be maintained in the coming years. In the GITTA project, like in other large e-learning projects involving multiple teams and multiple languages, sustainability regarding content development has been an important issue since the project started. This paper introduces three different approaches how the future use, updating and revision of e-learning lessons developed in the GITTA project and other SVC projects can be ensured and simplified. The approaches discussed here are the XML framework eLML for structuring and storing the content (eLML 2007), the publication of the content under a Creative Commons Licence (Creative Commons 2007) and the foundation of a supportive association (GITTA Association 2007).

2. eLML – eLesson Markup Language

When in 2000 the e-learning project GITTA started there were no available formats, content standards or learning management systems able to meet our needs. We learned from the difficulties other projects had using proprietary formats and commercial learning management systems that were faced with vendors' decisions to stop support for a format or increase prices drastically. An evaluation showed that back then rarely a learning management system supported standardised import or export of content. Thus, content could not easily be transferred to another learning management system. Additionally, the GITTA project consortium was heterogeneous and multilingual. Some mechanisms, pedagogical or technical guidelines, needed to be put in place to ensure that all content authors created consistent lessons with the same look and feel.

Due to these reasons the eLesson Markup Language eLML (eLML 2007) was developed. It satisfies all of the named requirements. Its structure is defined and implemented using XML

Schema thus allowing for easy checking and validation of created content using standard tools. The content is stored separately from the layout in XML files. Layout templates are defined and updated using XSLT style sheets. With the help of different style sheets various output formats such as PDF, HTML or standard SCORM or IMS content packages (Jones 2002) for the import of the lessons into any learning management system can be generated. The creation and use of an XML based framework for e-learning content was inspired by some of the team member's positive experiences with geospatial data modelling and exchange mechanisms such as INTERLIS (INTERLIS 2007) or GML (OGC 2007). Didactically, the framework is based on the pedagogical model ECLASS (adapted from Gerson 2000) which stands for Entry, Clarify, Look, Act, Self-assessment and Summary. These are the important parts of an e-learning lesson and shall guide authors in the creation of good content. However, the authors are free in the order they wish to use some or all of these parts allowing the implementation of just about any teaching or learning structure in eLML (e.g. case studies or experimental learning).

eLML was documented for general use and published as an open source project under the General Public License (GPL) on Sourceforge.net in 2004. Since then a constantly growing number of projects and authors in Switzerland and other European countries started using eLML as their tool for creating e-learning lessons (e.g. Dykes 2007). The large user community ensures the updating and maintenance of the XML framework. A comprehensive overview of the eLesson Markup Language eLML can be found in Fisler and Bleisch (2006).

3. Open Content under a Creative Commons Licence

In the GIST (Geographic Information Systems Technology) e-learning project GITTA 43 lessons at two experience levels – basic and intermediate – in four languages were developed. The lessons cover the six topics GI Systems, Data Capture, Database Systems, Spatial Modelling, Spatial Analysis, Data Presentation and a number of case studies. Towards the end of the project the consortium had to find ways to ensure that even though funding was ending, the developed content would not become deserted and outdated. Discussing different options such as offering access via paid subscriptions or restricting the access to partner institutes the consortium decided on an open access strategy. As a consequence, in 2006 the content was made freely available under a Creative Commons licence (Creative Commons 2007). The open content GITTA e-learning lessons shall be used throughout the world and a community of users and authors interested in the maintenance of the lessons will hopefully emerge. Similarly, well-known open content initiatives include the open course ware project of MIT (MIT 2007) or the connexions project (Rice 2007).

4. The GITTA Association

However, freely available content and a possibly large user community do not yet ensure the sustainability of e-learning offerings; since they do not address the important issues of content hosting and updating. In response to these additional challenges, the GITTA Association was founded in 2006. All users of GITTA lessons and especially authors who would like to update or change existing GITTA e-learning lessons are encouraged to participate. GITTA Association members are either active individual authors or schools and institutes or sponsors. A low membership fee allows everyone to become an active member. Being an active member of the GITTA Association has the advantage of having access to different versions of a lesson and of being able to not only use the content but also to update it and to provide contributions, e.g. in the form of translations into further languages. The funds of the association are used for the hosting of the e-learning content on a content versioning server (CVS) and for special projects such as an expected cooperation with South American universities for the translation of the lessons in Spanish and Portuguese.

5. Conclusions and Outlook

e-learning has many advantages but also disadvantages. Surveys among students attending e-learning courses have shown that they like having flexible working times and being independent of a specific place such as the computer lab at the university. On the other hand, they note that it is hard to read lessons on the screen. However, with eLML we can offer them a print-version of the lessons in PDF format as a complement to the online HTML version. Looking at e-learning from the teaching side our experiences have shown us that the students achieve much better results compared to the traditional face-to-face teaching if the e-learning content is either highly interactive or teaches computer related aspects. Namely courses in the 3D modelling language X3D and the database query language SQL are highly popular among students and they achieve higher grades in tests compared to student who had not used the e-learning lessons.

Using eLML ensures that the e-learning content is independent from any proprietary format and that content and layout are handled separately. Thus, the lessons can either be stored on an independent server and made available directly through a website (see GITTA 2007) or transformed to SCORM or IMS content packages which can be imported into any modern learning management system. eLML lessons could also be made available to the students in HTML or PDF formats on a CD or DVD.

In case of the GITTA e-learning content, the publication of lessons as open content combined with the foundation of a supportive association helps ensure that the created content can be used by many different users. Some of them may also want to update or enhance the lessons they have integrated in their curriculum and thus giving back some value for the free use of the content. The GITTA Association allows structuring and guiding these cooperation processes and facilitates the integration of authors that were not GITTA consortium members in the first place.

Highly important is the sustainability of e-learning content. With the concepts described in this paper we believe to have found a way to ensure that the developed GITTA e-learning lessons will be usable also in the years to come. There are already other SVC projects who are interested in opening up their content and possibly joining the supportive association.

References

CartouCHe (2007) *CartouCHe - Cartography for Swiss Higher Education*. http://www.e-cartouche.ch/ (online December 17, 2007).

Creative Commons (2007) *Creative Commons*. http://creativecommons.org/ (online December 17, 2007).

Dykes J (2007) *InfoVis 2007 tutorial : 'Visualization with Google Earth and GIS'*. http://www.gicentre.org/infoVis/ (online February 22, 2008).

Edutech (2007) Swiss Virtual Campus. Swiss Virtual Campus, http://www.virtualcampus.ch/ (online December 17, 2007).

eLML (2007) *eLML - eLesson Markup Language*. http://www.elml.ch (online December 17, 2007).

Fisler J and Bleisch S (2006) eLML, the eLesson Markup Language: Developing Sustainable e-Learning Content Using an Open Source XML Framework. WEBIST 2006 - International Conference on Web Information Systems and Technologies, Setubal, Portugal.

Gerson S (2000) ECLASS: Creating a Guide to Online Course Development For Distance Learning Faculty. *Online Journal of Distance Learning Administration*, Vol. III, No. IV (Winter 2000). State University of Western Georgia. http://www.westga.edu/~distance/ojdla/winter34/gerson34.html (online December 17, 2007).

GITTA (2007) *GITTA - Geographic Information Technology Training Alliance*. http://www.gitta.info (online December 17, 2007).

GITTA Association (2007) *The GITTA Association*. http://www.gitta.info/association (online December 17, 2007).

INTERLIS (2007) *INTERLIS The GeoLanguage*. http://www.interlis.ch (online December 17, 2007).

Jones E R (2002) Implications of SCORM and Emerging E-learning Standards On Engineering Education. *Proceedings of the 2002 ASEE Gulf-Southwest Annual Conference*. http://falcon.tamucc.edu/~ejones/papers/ASEE02.pdf.

MIT (2007) *MIT OpenCourseWare*. http://ocw.mit.edu/OcwWeb/web/home/home/index.htm (online December 17, 2007).

NAHRIS (2007) *Dealing with natural hazards (NAHRIS)*. http://www.nahris.ch/ (online December 17, 2007).

OGC (2007) *Geography Markup Language*. http://www.opengeospatial.org/standards/gml (online December 17, 2007).

Rice (2007) Connexions - Sharing Knowledge and Building Communities. http://cnx.org/ (online December 17, 2007).

Werner M and Stern B (2003) Active and Self-controlled Web Based Education in GIS Technology and Cartography: The GITTA Project. *Proceedings of the 21st International Cartographic Conference*, Durban, p. 805-815.

Biography

Susanne Bleisch is a scientific collaborator for the e-learning projects GITTA, CartouCHe and eLML at FHNW and a PhD student at City University London. 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.

Stephan Nebiker is professor for geoinformatics, photogrammetry and remote sensing at the University of Applied Sciences Northwestern Switzerland FHNW in Muttenz, Switzerland. He is actively involved in the e-learning projects GITTA and

CartouCHe. Among his research interests are virtual globe technologies, 3D geovisualization, mobile geosensors and augmented reality.

Joël Fisler works at the IT services of the University of Zurich. His tasks cover supporting the open source learning management system OLAT and the continuing development of eLML. Both OLAT and eLML are strategic e-learning tools at the University of Zurich. He originally studied environmental sciences at the Federal Institute of Technology ETH Zurich and was the coordinator of the GIS e-learning project GITTA presented in this paper.

Geo-contextualised visualisation for teaching and learning in the field

Claire Jarvis[1], Gary Priestnall[2], Gemma Polmaer[2], Jing Li[1]

[1]Department of Geography, University of LeicesterUK
TEL: +44 (0)116 252 3808 FAX: +44 (0)116 252 3458
Email: chj2@le.ac.uk
[2]School of Geography, University of Nottingham, UK.

KEYWORDS: Augmented reality, PDAs, immersive headsets

Introduction

Teaching students themes related to geomorphological processes and landforms requires the integration of knowledge about process working at different scales, the particular processes themselves varying at different times to form unique landscapes. While the generality of individual processes may be understood, unpacking the particular history behind a specific view requires considerable spatial aptitude and imagination of the students, as the teaching vignettes of Figure 1 identify. Geo-contextualised visualisation forms one potential means of tackling this subject-specific communication issue, working alongside more traditional pedagogic approaches. This paper firstly discusses a geo-contextualised mobile PDA application used for teaching students geomorphology, and outlines further development work underway that will allow the progression of these mobile teaching and learning resources within an immersive AR system using a head-mounted display (HMD).

Teaching vignette A

Imagine that you are in the field with students, overlooking a valley that has been shaped by ice glaciers over 15,000 years ago, for example in the Lake District, Cumbria, UK. Having given the students time to explore the landscape and think about its origins, consider how helpful it would be if you could assist the students to visualise how the landscape before them has changed over time using a computer-generated virtual model of the glacier superimposed on the real landscape they are observing in the field.

Teaching vignette B

In the company of 2nd year BSc geography students, you are overlooking a set of river terraces in a landscape shaped by tectonic uplift and consequent river capture, in a zone where there has been maritime incursion twice over the past 6 million years. Unpacking the detailed sequence of processes and landform causality in the area has been the subject of many PhDs. You have 1.5 hours over which to encourage the students to explore landscape and consider how it might have been formed. Could a geo-contextualised landscape evolution model assist you in helping the students unpack the landscape in such a way that they leave the site having a good idea of the principle geographical processes shaping the view before them?

Figure 1 Examples of complex space-time communication issues in a geomorphological teaching context

Method

It has been suggested that AR seems to be an ideal candidate in almost any working context to deliver accurate, useful and up-to-date information (Regenbrecht 2007). In a geographical setting, teaching landscape process in a field context, ‘useful’ implies the provision of timely geo-contextualised view-related information in a manner that blends and supports traditional pedagogies, using methods that are practical with large student groups and cost-effective. For reasons related both to registration and graphics processing requirements in the context of visualising landscape evolution, and financial cost, we report here on a stepped approach (Figure 2) in which we propose a move from ‘light AR’ using PDAs in the field to a more immersive approach to visualisation landscape change that has realistic potential in a teaching context in the near future.

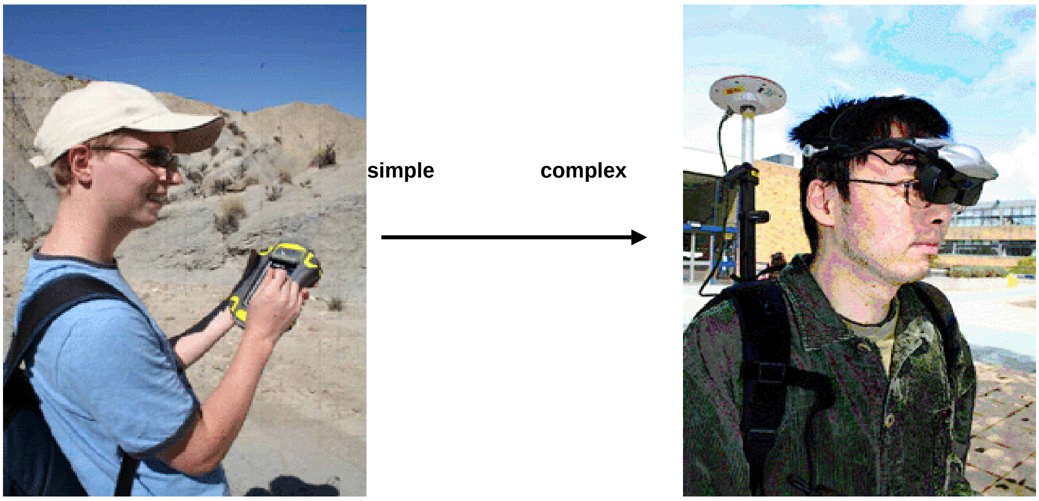


Figure2. Geo-contextualised augmented reality: From simpler GPS-enabled PDA teaching solutions to more complex immersive GPS-inertial tracker developments using laptops & headsets

Geo-contextualised PDA visualisation for learning in physical geography

Landscape visualisation has been used to enhance physical geography fieldwork based in Cumbria, North West England for a number of years. A four day residential year 1 field course included a project to engage students in the landscape history of the area they were staying in, using desktop visualisation of a terrain model to prepare 3D views in the field centre to be used to augment their view of the landscape out in the field. Digital data used included a 5m RADAR-derived Digital Surface Model, aerial photography drapes, geology drapes, and a 3D model reconstructing the retreating glacier in the valley. A general aim of the exercise was to allow students to assess the use of digital terrain modelling and landscape visualisation in recreating real landscapes they could observe at first hand and to engage them in past landscape processes such as glaciation (Figure 3).

In recent years this idea has been implemented using PDAs (Priestnall & Polmear, 2006) where several waypoint locations are chosen and 3D views prepared in the field centre before going out into the field. In the absence of mobile computing power to drive real-time landscape visualisation in the field, pre-defining viewpoints were used. Although relatively crude, this did require the students to develop a frame of reference for navigating the area to find these views and to understand their landscape context which is often acknowledged as an important aspect of fieldwork (Ishikawa & Kastens, 2005). This technique displayed few of the characteristics of a true Augmented Reality system as described by Piekarski (2006) however it did allow users to augment their view of the real scene with spatially-referenced computer generated images. The lack of automated registration between PDA screen and real scene distinguished this approach from many handheld Augmented Reality approaches such as those described by Wagner& Schmalstieg (2003) where the recognition of patterns placed in the real scene is required to generate the virtual content. Having said this, the ability to reconcile a handheld virtual scene with the real landscape is a common issue faced by recreational phone-based navigation packages such as Viewranger (Augmentra Ltd, 2006) and also studies focussing on spatial awareness in the context of location-based services (Krüger et al 2004).

Figure3. In-field visualisation
Upper Left: Photograph of real scene
Upper Right: 3D view from RADAR model
Left: RADAR scene augmented with one possible scenario for glacial retreat

Figure 4 illustrates the stages of the PDA-based field visualisation application. The first stage is to pre-generate images to be used, using the desktop 3D modelling package Bryce, immediately before going out into the field. Students select three viewpoints of interest within the study area. For each viewpoint three views are generated: the 'current' photorealistic virtual model, the 'hidden' landscape represented by the geology map draped over the surface, and the 'past' landscape including the model of retreating glacial ice. The images and details of their viewpoint locations and orientations are uploaded to a GPS-enabled PDA and form a series of waypoint symbols over a map in the navigation screen. In the second stage, out in the field, students use onscreen assistance (current location, distance and direction to the next site), to navigate to the viewpoints in order. As each viewpoint is approached, an image viewer screen is automatically triggered allowing students to compare their computer generated current, past and hidden landscape views with what they can see in the real world. The field sketch function allows them to annotate their images for use in reports that evening. Context-aware audio commentary is also delivered, providing more information about the surrounding landscape. This is based on pre-generated visibility maps indicating the landscape features visible from any given point, in a similar way to Mackaness and Bartie (2005) implemented for an urban environment.

The usability of the application was assessed in part through the use of student video diaries which revealed some subtleties of usage which would not normally be apparent. Battery life and the audibility of commentaries did not prove to be problems and the application generally worked well to engage students in both digital terrain modelling and landscape history. There were problems however relating to screen visibility when facing the sun, the stability of the Bluetooth GPS connectivity, and occasionally the inability to access the chosen viewpoint as placed in the virtual model.

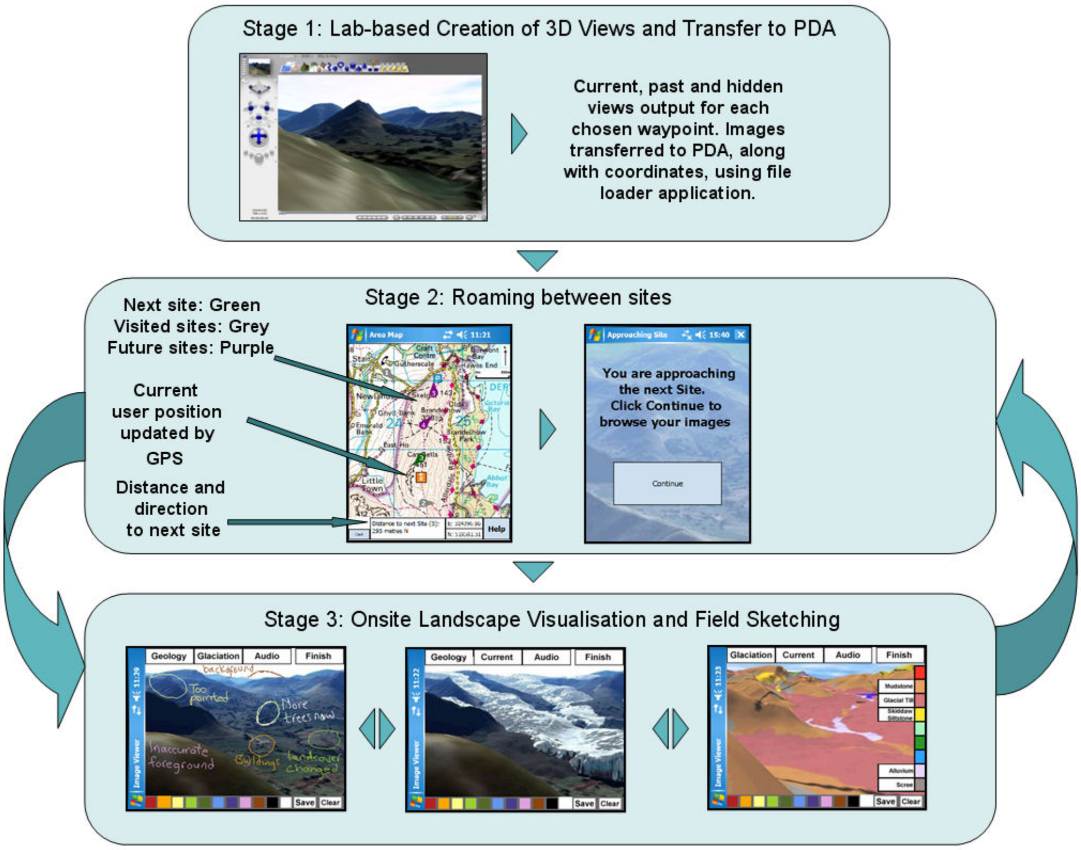


Figure 4: The PDA-based Field Visualisation Application (Kruger et al., 2004)

Geo-contextualised immersive visualisation for learning in physical geography

A logical extension to the fixed-view PDA approach of Section 2.1 is to allow similar views to be generated in the field rather then pre-specify a series of waypoints. Such a facility would allow greater flexibility when used by multiple students at one viewpoint, and offer a wider range of options for field stops. We describe here the hardware and software configuration for such a system that is currently undergoing technical testing, that is being designed firstly to work both on tablet PCs with digital compasses in an analogous manner to the PDA approach and secondly outputting views through an immersive headset linked both to GPS and motion sensors in 3 degrees of freedom as spatial inputs.

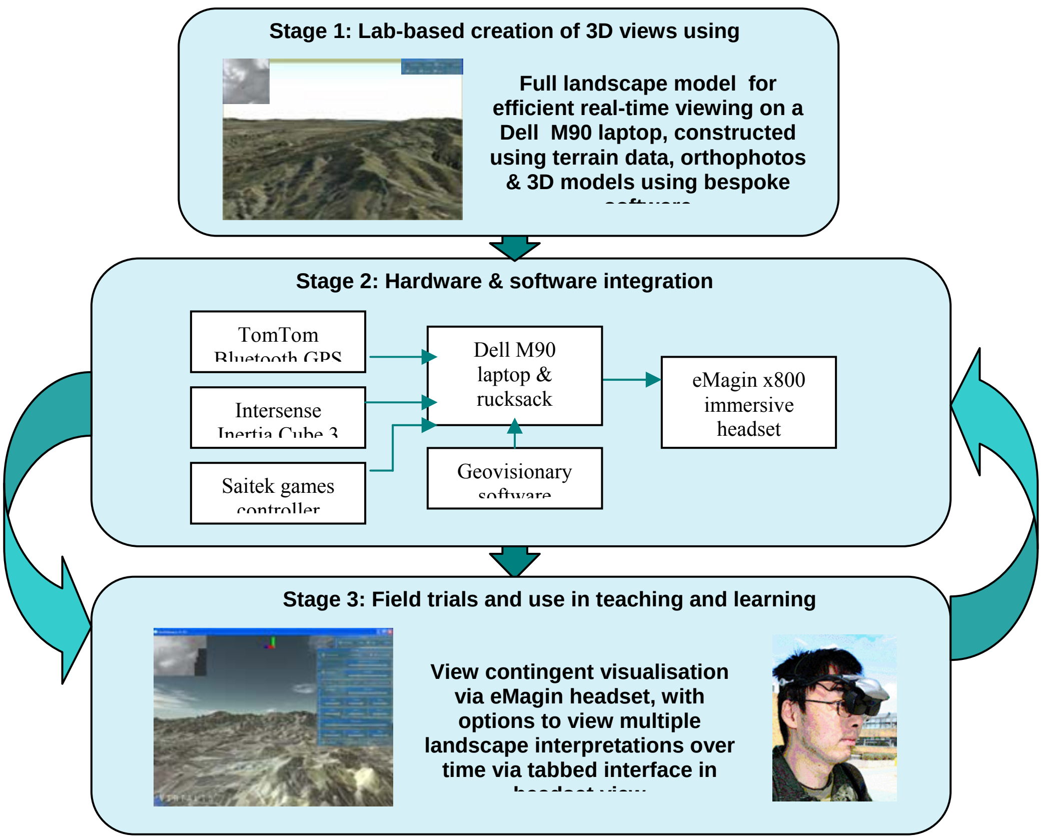
Hardware

Our immersive AR system is composed of an Intersense Cube 3 tracking system, a Bluetooth GPS, a head-mounted opaque display (HMD), a laptop computer and a user-driven input capability in the form of a games controller (Figure 5)

Scaling up to real-time surface visualisation requires mainstream and graphics processing capability, beyond that capable of current PDAs or micro-PCs. In our case, we use the dual-processor Dell M90 with inbuilt NVIDIA Quadro graphics card and 512MB dedicated graphics memory.

This system uses an eMagin x800 headset to provide a broader immersive framework for landscape scale observation than that of a PDA, albeit with a more restricted field of view that the human vision system. Wider contextualisation is important given the scale of processes operating over basin ranges, for example in relation to river capture and tectonic uplift. However, in regard to head tracking, one of the big issues in regard to augmented reality and GIS is that of registration. Typically, a GPS unit, antenna and head tracking mechanism all combine to provide such information but a number of accuracy issues surround this process and despite a considerable volume of work on the subject (e.g. Avery et al., 2005; Kealy and Scott-Young, 2006) the registration issue remains a challenging research subject. Registration accuracies matter, particularly in the context of landscape surface matching, where a student using an AR system must be able to use a system for a reasonable length of time without succumbing to disorientation or motion sickness caused by miss-registration between augmented and actual geographic realities.

Figure5. *Augmented reality system for field teaching based on Geovisionary software*



Thus, at this point in time, we consider that "true" Augmented Reality as described by Piekarski (2006) is both technically and financially prohibitive in the context of fieldwork where several groups of students often undertake the same exercise at the same time. However, a compromise scenario exists in the form of geo-contextualised real-time terrain models offered to the user visually to the user via either through a handheld device or a head-mounted display which *replace* the central portion of the user's field of vision. The latter model is similar to the idea of the 'Augurscope' (Schnädelbach *et al*, 2002) where portable computer, GPS and an Inertial Navigation System (INS) were combined to give offer a window into a virtual historical scene. Using this model a solution using a simpler and thus cheaper headset designed for the mass-market gaming, such as the eMagin x800 becomes financially feasible; slight miss-registration between images are less noticeable since actual and augmented realities are not overlain digitally but rather displayed at different times but for the same actual view.

Software

The software driving the augmented reality system is based on Virtalis's Geovisionary landscape rendering software, adapted by the University of Leicester (working with Virtalis) to take positional data from a GPS source that is then modified to orientate the view within the headset to that of the user with the assistance of the inertial orientation motion tracking device. The user is able to offset or adjust the position used for them within the renderer to give the best visual image, for example to take into account their particular height and particularly to correct for GPS height value given the use of a Sirf3 Bluetooth GPS within this system. Thus the Geovisionary software allows a user to orientate their model view as seen through the headset to that actually in front of them.

Terrain and orthophoto drapes and clusters of 3D objects are considered as "resources" within the Geovisionary software. These are displayed on the headset as a list; the user can toggle pre-specified individual objects, or groups of objects, on and off using the games controller as input device. Each of the resources has a file associated with it which contains text based information which can be displayed within the headset view as a label associated with an object. The application will allow the user to switch between 3-4 four image drapes; in a geomorphological teaching context, these could for example represent past landscapes at preset time periods. In other contexts, such resources could include historical mapping or alternative cartographic perspectives.

Conclusion

The PDA approach demonstrates that students appreciate the power of geo-contextualised visualisation to support their understanding of landscape processes, The immersive development of this core idea moves towards an intermediate AR solution that is practical and relatively cost effective, intended to support student spatial literacy. In addition to the technical aspects of this work we focus upon here, the methods presented will need to be evaluated to examine in what ways they can best be used to help students explore the spatial relationships between past and present landscape features and even the geology under their feet at that point to enhance understanding of a particular geographical system. Quite how best to blend such visualisations within the traditional fieldwork experience in a sensitive manner, such that students remain encouraged to reflect and think for themselves and other staff are comfortable adopting technologies in the field, remain the subjects of further challenge. We also acknowledge that the widespread adoption of landscape-view AR is still in its infancy, both on technical grounds and cost. The latter issue is particularly important in a teaching context, where multiple individual devices are needed. In this sense, this work can be considered "horizon watching" in teaching and learning terms; however, gaming headsets are rapidly reducing in price and high street game-boxes are converging in graphics power on professional VR installations such that the future for AR in an educational context appears inviting.

Acknowledgements

The 'NEXTMap' Digital Surface Model, derived from Intermap's IFSAR (Interferometric Synthetic Aperture Radar), and colour aerial photography were supplied by Getmapping UK. This work was funded by the UK's Higher Education Funding Council for England (HEFCE) as part of the SPLINT (Spatial Literacy in Teaching and Learning) Centre of Excellence in Teaching & Learning.

References

Augmentra Ltd (2006) *Viewranger.* Available at: www.viewranger.com/ (Accessed 15 November 2006).

Avery, B., Thomas, B.H., Velikovsky, J. and Piekarski, W., (2005), Outdoor augmented reality gaming on five dollars a day, In *6th Australasian User Interface Conference (AUIC2005)*. Newcastle.

Ishikawa, T. and Kastens, K. A., (2005) Why Some Students Have Trouble with Maps and Other Spatial Representations. *Journal of Geoscience Education* **53(2),** pp.184-197.

Kealy, A., Scott-Young, S., (2006), A Technology Fusion Approach for Augmented Reality Applications, *Transactions in GIS*, **10 (2),** pp.279-300.

Krüger, A., Aslan, I. and Zimmer, H. (2004) The effects of mobile pedestrian navigation systems on the concurrent acquisition of route and survey knowledge. *Lecture Notes in Computer Science* **3160**, 211-224.

Mackaness, W. A. and Bartie, P. (2005) A speech based augmented reality system for city tourists. *Proceedings of the GIS Research UK 13th Annual Conference*, Glasgow, 6-8th April 2005.

Piekarski, W., (2006) Simple Collaborative Indoor-Outdoor Modelling Using Mobile Augmented Reality. *Proceedings of the First International workshop on mobile geospatial augmented reality*, Banff, Alberta, Canada, 29-20th May 2006.

Priestnall, G. and Polmear, G. (2006) Landscape Visualisation: From lab to field, *Proceedings of the First International workshop on mobile geospatial augmented reality*, Banff, Alberta, Canada, 29-20th May 2006.

Regenbrecht, H., (2007), Industrial Augmented Reality Applications, *In Emerging Technologies of Augmented Reality: Interfaces and Design*, pp: 283-304.

Schnädelbach, H., Koleva, B., Flintham, M., Fraser, M., Chandler, P., Foster, M., Benford, S., Greenhalgh, C., Izadi, S. and Rodden, T. (2002) The Augurscope: A Mixed Reality Interface for Outdoors, in *Proc. ACM CHI 2002,* pp. 9-16, ACM Press.

Wagner, D., and Schmalstieg, D., (2003) First Steps Towards Handheld Augmented Reality, *7th International Conference on Wearable Computers (ISWC 2003)*, White Plains, NY, 2003. Available online at: http://www.icg.tu-graz.ac.at/Members/daniel/Publications/FirstSteps_HandheldAR.

Real-Time Landscape Visualisation: Experiences in a Teaching and Learning Context

Andy Burton[1], Gary Priestnall[2], Gemma Polmear[3], Nick Mount[4]

[1, 3]School of Geography/Institute of Engineering Surveying and Space Geodesy,
The University of Nottingham, Nottingham, NG7 2RD, UK
Tel. +44 (0) 115 846 8507
[2, 4]School of Geography, The University of Nottingham, Nottingham, NG7 2RD, UK
Tel. +44 (0) 115 9515443 Fax. +44 (0) 115 95429
Andy.Burton@nottingham.ac.uk
http://www.nottingham.ac.uk/~enzamb/

KEYWORDS: Virtual Reality, Spatial Awareness, Education, Visualisation, GPS

1. Introduction

Landscape visualisation is typically seen as the use of computer-generated 3D perspective views of predominantly natural landscapes, usually striving for some appropriate level of photorealism (Appleton & Lovett, 2003). Such forms of visual communication which mimic the public's natural view of the world, and allow past or future scenario representation, are becoming increasingly common in environmental decision making. Orland *et al* (2001) described the trend towards more interactive participatory visualisation in a planning context. Dockerty *et al* (2005) provide one example of how modern VR software can create and render realistic models of future rural landscape scenarios using GIS data. It has been recognised that 3D landscape visualisation is a powerful way to enhance the awareness of environmental phenomena and has the potential to foster action and influence policy (Sheppard, 2005). It is also important to reflect on the degree to which virtual landscapes provide representations which are true to our real world experience of that landscape (Lange, 2001; Priestnall and Hampson, 2008) or which allow us to simulate real field experience within a laboratory setting (Whitelock and Jelfs, 2005). The way in which visual imagery works for different people is complex and people are influenced in different ways by the same visual presentation (Nicholson-Cole, 2005).

In a research-led teaching and learning environment there are several ways in which landscape visualisation can be approached in a GIS curriculum. Technical procedures for data capture, processing and model building can be complemented by field-based studies (Priestnall and Polmear, 2006). The use of landscape visualisation in conjunction with an environmental decision making exercise is an additional approach and forms the focus for this paper. It has been recognised that such applications require the interface to the virtual environment to be intuitive and to support basic functions which enable users to begin to utilise some of the potential of the technology without it actually getting in the way (Furness, 1998).

This paper will focus on the case study of using semi-immersive VR as part of a Masters level student exercise to determine and communicate the location and impact of wind farms in Cumbria, UK using GIS-based analysis. Being able to assess *cumulative visual impact* of multiple wind farm proposals from many viewpoints is an important issue for ongoing wind farm planning applications such as the Berrier Hill proposal to the East of Blencathra, Cumbria (Cumbria County Council, 2006).
Specific issues addressed in this paper are:

- The practicalities of setting up a semi-immersive VR environment.
- The model building workflows for both terrain and furniture.
- An exploration of the ability of the basic VR environment to support a visual impact scenario to a small to medium sized group of people.

- Reflection on this experience and feedback to make recommendations for good practice in the use of semi-immersive VR in this context, including a description of ongoing work to address issues highlighted.

2. The Teaching and Learning Context

The activities described here are a result of Nottingham's involvement in SPLINT (SPatial Literacy in Teaching), a collaborative HEFCE-funded Centre for Excellence in Teaching and Learning (CETL) led by the University of Leicester with University College London as the third partner. The Nottingham component of SPLINT is run jointly by the School of Geography and the Institute of Engineering, Surveying and Space Geodesy (IESSG) and considers how VR and mobile computing can enhance the use of digital geographic information within and beyond geography.

As part of a Masters level module 'Fundamentals of Geographical Information Science' students from the MSc in GIS and the MSc/MA in Environmental Management undertook a group project to suggest and justify the location for a fictitious wind farm within northern Cumbria. A large amount of data was made available including RADAR-derived terrain, colour aerial photography, Ordnance Survey Mastermap vector data, wind speed data and the extent of many planning designations. The students were also given the opportunity to use a semi-immersive VR theatre as part of their final presentation to enable the visual impact of their proposals to be communicated. They were asked to supply the locations and height of three wind turbines a week before the final presentations and were given instruction in the basic operation of the VR theatre.

The following sections describe the VR theatre and the processes involved in creating a suitable virtual environment.

3. The VR Environment

At the University of Nottingham a stereo VR theatre has been installed. It features two pairs of data projectors creating a 5m wide by 2.5m high image, viewable in 3D through the use of polarised glasses. A schematic representation of the lab is given in Figure 1.

The laboratory environment can also replicate the use of location based devices in the field as it simulates the output from a Bluetooth GPS unit. Location-aware applications on mobile devices can be tested within the laboratory as previously described by Priestnall and Polmear (2007).

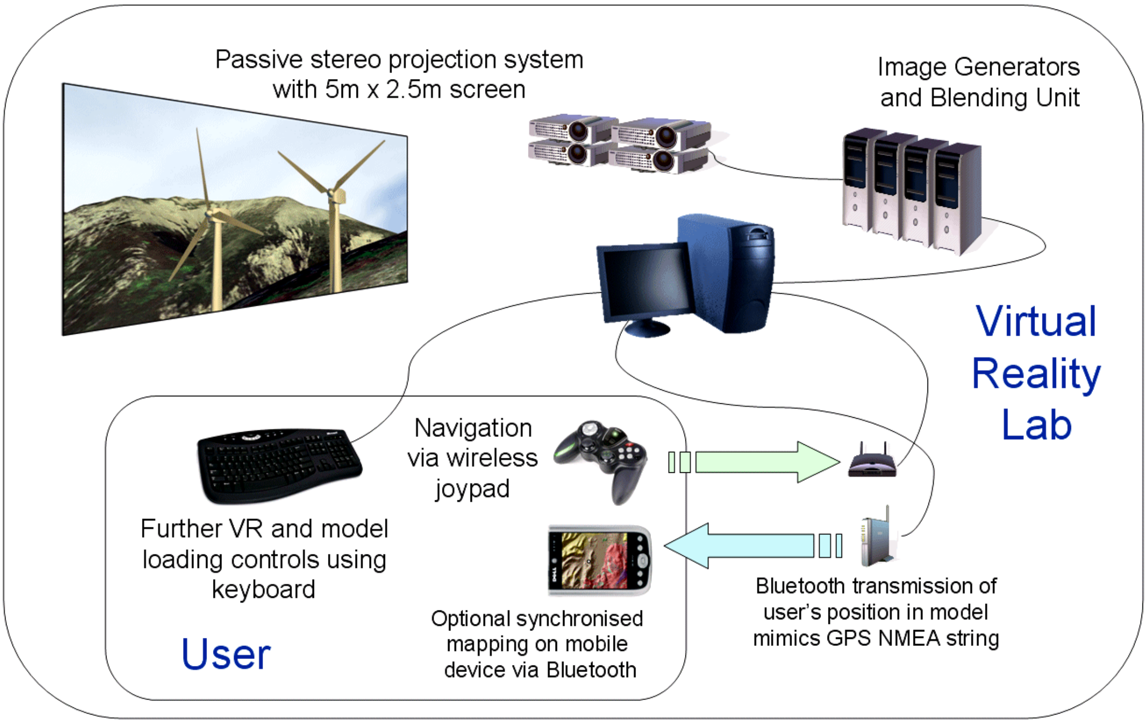


Figure 1. A Schematic Diagram of the VR Laboratory Environment.

4. Model Building

The process of the development of a model for use in the VR theatre was far from straightforward. This section will highlight the difficulties involved in the authors' method for VR world creation, using the wind-farm project as an example.

The main constraint on world development was the complexity of the software environment. The laboratory was built for use with Multigen-Paradigm's Vega Prime used in conjunction with the terrain optimisation package Blueberry (by Bionatics). These packages rely upon very specific model formats and skilled performance tuning to keep the quality high and the processing light.

The model was build up using a terrain model, overlaid with both generalised vegetation and landscape features and a collection of specific models such as buildings which added integrity and realism to the world. Firstly a georeferenced digital terrain model was tiled by Blueberry to allow rendering of a highly detailed terrain close to the viewer, and fade that down to a much less detailed terrain in the distance. Secondly a set of georeferenced aerial photographs were similarly tiled. Preprocessing of the tiles to reduce their size whilst maintaining their georeferencing was necessary to reduce processing time from days to hours.

Features such as woodland, hedge lines, power cables, and waterways were then added. A Blueberry 'plug-in' called RealNat was used to create vegetation models with suitable variable levels of detail for use in the Blueberry environment. These general features were imported to Blueberry at points, along vectors, or within areas. The level of detail settings for each model and vector were adjusted manually to optimise performance in the Blueberry Run-Time Environment.

The final part of the modelling process was to create specific models, which may be key features in a landscape, such as a fieldwork centre, to give students a known reference point in the world, or objects such as the turbines in this wind farm example. These models can be created from a variety of sources and through a variety of routes. Those currently used by SPLINT are summarised in Figure 2.

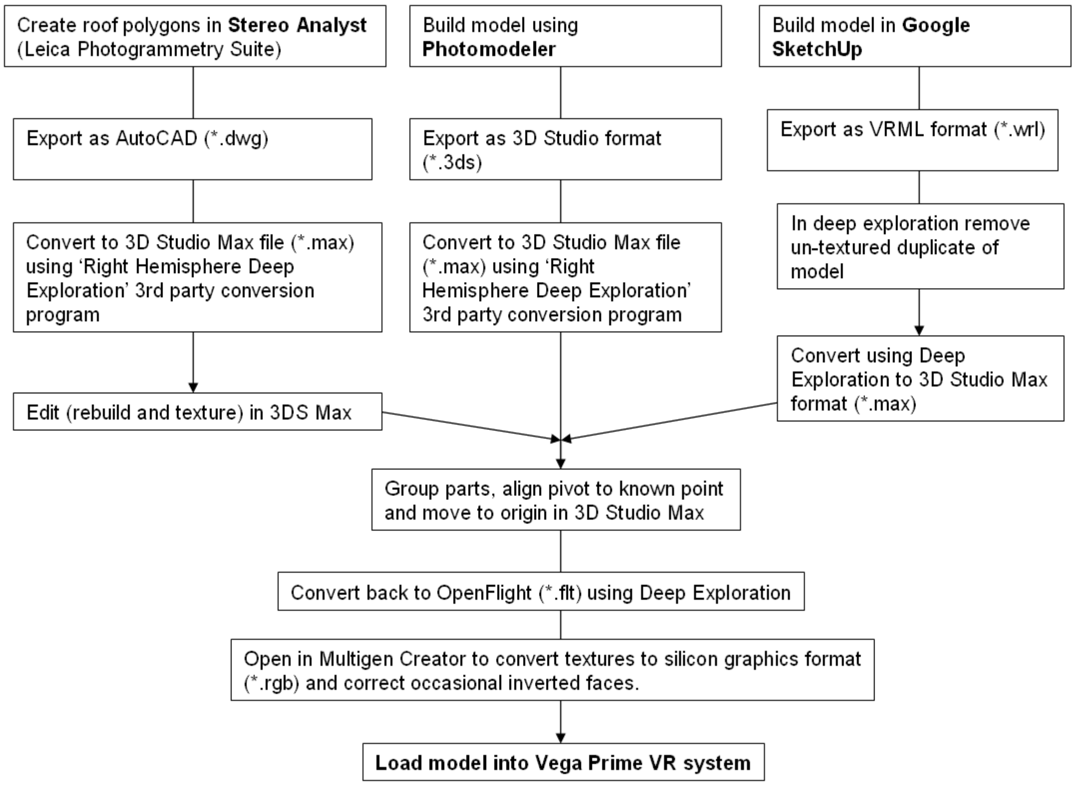


Figure 2. SPLINT's Workflows to get viable OpenFlight models from several modelling packages

5. Real-Time Visualisation in a Lecture-Room Environment

From the student perspective the emphasis of this work was not on creation of the virtual environment, but on using the semi-immersive VR theatre to supplement a presentation justifying their choice of wind farm location. The functionality available in the system was as follows:

- Representation of the region of interest as a virtual model as described in section 4.
- Production of stereo perspective views of the virtual model allowing up to 35 viewers.
- Real-time navigation via a wireless joypad (shown in Figure 3) as used with modern games consoles with two speeds of movement switchable via the joypad.
- Alternation between terrain-following and free flight movement.
- Jumping to a set of start positions (pre-defined by the students) via the keyboard
- Toggling wind turbine models on and off via the keyboard

Figure 3. Controlling the virtual environment with the joypad

The groups presented a justification for their site selection via a PowerPoint presentation to the side of the VR screen and could utilise the functionality of the VR theatre as described above in any way they wanted, as part of the presentation and in response to questions.

All groups demonstrated the location of their wind farm in free flight mode although largely focussing on the local environment and one group showed the relative proximity of major power lines to their site. Questions from the audience asking for the observer position to be relocated, for example, “Can you see the wind farm from Keswick?” were more difficult to deal with. In fact issues relating to intervisibility in the broader landscape context would seem fundamental to exploiting the benefits of a real-time VR environment, but were not well served by the default system configuration. A focus group involving the thirteen students undertaking the module was held a few days after the presentations to discuss system usability issues and to focus ongoing development work.

6. Discussion and Recommendations

Establishing a VR theatre represents a major investment of time, effort and financial expense and the creation of virtual models to support real-time landscape visualisation using such technology is non trivial. The relative virtues of large screen, stereo, real-time VR environments of this type need to be investigated: however some basic usability issues need to be addressed first. The aim of this paper has been to reflect on the practicalities of developing a VR environment which can allow landscape visualisation to play a tangible role in a student presentation context without requiring undue time and effort on the students' part.

Some issues that arose relating to the hardware, software and data required to facilitate the real-time landscape visualisation were:

- Hardware configurations should be assessed very carefully in relation to the software packages required.
- The extent of data coverages should be checked carefully to ensure an adequate landscape representation is available for both the immediate geographical area of interest but also all areas visible from that area.

In terms of the usability of the VR environment a focus group proved useful in prioritising development in order to allow some of the theoretical virtues of such environments to be realised:

- A compass or location map would allow easier orientation within the virtual model. Ad hoc changes in observer position require a high degree of spatial awareness within the virtual model.
- The ability to switch between alternative terrain drapes, such as an intervisibility map or Ordnance Survey map, was considered important.
- Of lesser importance was the ability to snap to pre-defined paths, for example a significant footpath in the case of a popular walking area such as upland Cumbria.

The priorities for ongoing development work are to develop a context map for use in conjunction with the VR display and to allow alternative map drapes to be easily defined and used by the students. The development of the context map can exploit the Bluetooth positional output from the model to allow a mapping application to be displayed a separate large monitor. A bespoke application is being written for this purpose, to provide clear display of the user's position, orientation and field of view.

7 Conclusions

This paper has reflected on the experience of using the standard interaction functionality of a semi-immersive VR theatre in the context of presentations made to medium sized audiences (10-35). The nature and magnitude of the development work necessary to build the required virtual models has been described and in consultation with students a set of development priorities have been established, including a context map remotely synchronised with the VR display.

Acknowledgements

Digital data for the development of the Cumbria virtual model was obtained from Intermap (the DSM), and Getmapping (aerial photography). Many thanks to the following students from the MSc in GIS and MSc in Environmental Management courses at Nottingham for their hard work, enthusiasm and useful feedback on the role of semi-immersive VR during this project: Koye Akinlogbin, Kazue Anan, Pantelis Arvantitis, John Bleach, Helen Clarke, Miles Glover, Abdulkadir Gulamhusein, Somwipa Lotongkum, Jesse Meredith, Alex Oulton, Luke Studden, Oana Vulcanov, Azadeh Ziaee. Thanks also to colleagues associated with the SPLINT CETL including Martin Smith, Nick Kokkas, Bill Hickin and Claire Jarvis.

References

Appleton, K. and Lovett, A. (2003) GIS-based visualisation of rural landscapes: defining 'sufficient' realism for environmental decision-making. *Landscape & Urban Planning*, Vol. 65, No. 3, pp 117-131.

Cumbria County Council (2006) Cumbria wind energy supplementary planning document: Consultation draft, August 2006. Available online at http://www.cumbria.gov.uk/elibrary/Content/Internet/538/755/2789/39007113928.pdf, Accessed 7th January 2008

Dockerty, T., Lovett, A., Sunnenberg, G., Appleton, K., Parry, M., 2005. Visualising the potential impacts of climate change on rural landscapes. *Computers Environment and Urban Systems* 29 (3), 297–320.

Furness III, T.A., Winn, W., Yu, R., 1998. The impact of three dimensional immersive virtual environments on modern pedagogy: global change, VR and learning. In: Proceedings ofWorkshops in Seattle,Washington, and Loughborough, England in May and June 1997. Human Interface Technology Lab, University of Washington, Seattle, WA. Available online at http://www.hitl.washington.edu/publications/r-97-32/, Accessed 6th January 2008

Lange, E., 2001. The limits of realism: perceptions of virtual landscapes. *Landscape and Urban Planning* 54, 163–182.

Li, C., and Longley, P. (2006), A test environment for Location-Based Services applications. *Transactions in GIS*, 10(1), pp. 43-61.

Nicholson-Cole, S.A., 2005. Representing climate change futures: a critique on the use of images for visual communication. *Computers Environment and Urban Systems* 29 (3), 255–273.

Orland, B., Budthimedhee, K., Uusitalo, J., 2001. Considering virtual worlds as representations of landscape realities and as tools for landscape planning. *Landscape and Urban Planning* 54 (1–4), 139–148.

Priestnall, G. and Hampson, D. (2008, in press) Landscape Visualisation: Science and Art. In Dodge, M, McDerby, M., and Turner, M. (eds) *Geographic Visualization: Concepts, Tools and Applications*. Wiley.

Priestnall, G. & Polmear, G. (2006) Landscape Visualisation: From lab to field, *Proceedings of the First International workshop on mobile geospatial augmented reality*, Banff, Alberta, Canada, 29-20th May 2006.

Priestnall, G. & Polmear, G. (2007) A Synchronised Virtual Environment for Developing Location-Aware Mobile Applications *Proceedings of the 15th Annual Geographical Information Science Research UK Conference (GISRUK '07)*, National Centre for Geocomputation, National University of Ireland, Maynooth, 11th – 13th April 2007, 236-240

Sheppard, S.R.J. (2005) Landscape visualisation and climate change: the potential for influencing perceptions and behaviour *Environmental Science & Policy* (8) pp. 637–654

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, Nos. 1/2, pp.121–131.

Biographies

Andy Burton is a Research fellow, currently working for the HEFCE-funded Centre for Excellence in Teaching and Learning - SPLINT (Spatial Literacy IN Teaching). His research background is in development of Virtual Reality applications for a wide spectrum of university research projects. Gary Priestnall is an Associate Professor (Senior Lecturer) whose research interests lie in geographic representation, landscape visualisation, spatially-aware mobile applications for geography fieldwork. Gemma Polmear is a Research Associate, also working on the SPLINT project, developing mobile geospatial augmented reality applications. Nick Mount is a Lecturer in Geographical Information Science and Physical Geography.

Social and environmental inequalities in rural England: spatial distributions and the effects of scale

Steve Cinderby[1], Annemarieke de Bruin[2], Meg Huby[2], Piran White[3]
[1]Stockholm Environment Institute, University of York, York, UK
[2]Department of Social Policy and Social Work, University of York, York, UK
[3]Environment Department, University of York, York, UK

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

KEYWORDS: Social and environmental Inequalities, Rural England

1. Introduction

Research on inequalities in the UK has concentrated on urban communities where the majority of the population is based. This paper describes new research from a UK Research Councils Rural Economy and Land Use (RELU) project, Social and Environmental Inequalities in Rural Areas (SEIRA), which is investigating the methods and data used to identify inequalities in rural England. Various studies, primarily in the USA, have attempted to investigate the impact that scale has on the assessment and identification of social and environmental inequalities. The paper describes the initial findings from the project on the existence and extent of inequalities in rural areas and will demonstrate the impact of changing the scale of the unit of analysis upon the identification and distribution of inequalities in an English context.

2. Deriving spatial data for assessing inequalities
2.1. Defining rural England

Rural England is defined according to the definition developed by Bibby and Shepherd (2004) that is based upon the distribution of population and consequent settlement density and morphology across the country. This allows for the identification of rural areas according to different census geographies ranging from Output Areas (OAs) upwards in scale to Local Authorities and ensures consistency with regional and national policy initiatives.

2.2. Describing rural England

The spatial dataset for describing social and environmental conditions in rural areas (SECRA) was developed under a previous RELU project. The SECRA dataset was compiled with a base unit of Lower Layer Super Output Area (LSOA). LSOAs are compiled from smaller census output areas (OAs) and have relatively consistent populations of approximately 1500 people. They were chosen as the SECRA base unit for a number of reasons. The coarsest data thought essential for inclusion in the dataset are held at LSOA level. The use of this base unit also ensures that the SECRA dataset has the potential for being updated in the future (as LSOA boundaries, unlike previous census geographies, are intended to remain constant). LSOAs are also the smallest units for which social information is readily available that avoids issues of individual confidentiality. This ensures that the SECRA dataset could be made available freely to students, policy makers and researchers to inform their thinking on rural England.

LSOAs are consistent in terms of their population size. However, for rural England their areal extent varies from $0.16km^2$ to $684km^2$. LSOA boundaries have no relation to the physical environment or people's linkages to the environment (Dorling and Fairbairn, 1997). The selection of LSOAs as the base unit for the dataset ensured that social information gathered from the census was readily available; however, the varying size of the base units meant that various geographic analysis techniques had to be employed to incorporate other social facets and specifically the environmental dimension of rural England. A variety of case-specific techniques were employed to interpolate information of different types to the LSOA base unit. These included area-weighted averaging techniques, counts and proximity measures. For example, the proximity of landfill sites to the LSOA centroid was used to assess the relative impact of these facilities on the people living within that area (Huby, Cinderby and Owen, 2005).

An overview of these techniques will be described in this paper where they relate to the identification of inequalities in the distribution of services and impacts in rural England.

3 Inequalities and scale

3.1. Identifying inequalities

The SEIRA project is utilising the SECRA dataset with a variety of interdisciplinary methodological approaches to identify inequalities. The project team is investigating geographic and statistical techniques from the social, economic and environmental fields to assess which approaches provide the most robust assessment. This paper discusses the implications for the identification of inequalities based on the varying results generated by different measures such as Lorenz curves, Gini coefficients, and inter-quartile range ratios. Examples of environmental and social inequalities identified in rural England using these complementary techniques are presented. For example, inequalities in bird species richness are associated with inequalities in socio-economic and environmental conditions at a regional scale.

3.2. The impact of scale

In the environmental justice movement of the US there has been considerable debate on the appropriate scale at which to assess environmental inequalities (Liu, 2000). Different geographies have revealed how the unit of analysis can affect the identification of inequalities. In this paper we discuss similar findings for rural England by assessing inequalities at various geographic hierarchies (built up from the LSOA base units). Preliminary results illustrate the importance of the resolution of the base units in the quantification and interpretation of inequalities.

4. Conclusions

The preliminary results of the study indicate that the holistic analysis of inequalities using the combined social, environmental and economic variables contained in the SECRA dataset is well-suited for identifying the interrelationship and interaction between spatial distributions of inequalities in rural England. The approach holds particular appeal from a policy perspective where initiatives to reduce social inequalities may be complemented by schemes designed to conserve rural ecosystems and *vice versa*. Better socio-economic conditions can improve the take-up and effectiveness of conservation and management programmes, while high quality environments and healthy ecosystems may work in turn to promote further social well-being.

The scale of analysis has an important effect on the identification of inequalities. Inequalities in the distribution of ecological, socio-economic and environmental factors are more marked when we study smaller units. This finding has particular resonance in rural England where pockets of deprivation are often hidden in areas that, at a coarser resolution, appear to be marked by affluence and high environmental quality. Improved understanding of the associations on different spatial scales between the ecological and socio-economic aspects of environmental inequality could provide an important framework for rural development policy. The findings in this paper illustrate how such a framework could be developed.

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.
Dorling, D. & Fairbairn, D. (1997) Mapping: ways of representing the world, London, Longman.
Huby, M., Cinderby, S. & Owen, A. (2005) Social and Environmental Conditions In Rural Areas. www.sei.se/relu
Lui, F. (2000) Environmental Justice Analysis: Theories, Methods and Practice. CRC Press

Biographies

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 Univeristy of York. With her background in GIS and Tropical Land Use she is responsible for the projects data gathering and spatial analysis. In addition, she will be involved in 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.

Socio-economic vulnerability units – modelling meaningful spatial units

Stefan Kienberger[1], Stefan Lang, Dirk Tiede

[1]Centre for Geoinformatics, Salzburg University, Hellbrunnerstrasse 34, 5020 Salzburg, Austria
Tel. (+43 662 8044 5210) Fax (+43 662 8044 5260)
stefan.kienberger@sbg.ac.at , http://www.uni-salzburg.at/zgis/

KEYWORDS: Vulnerability units, Spatial modelling, Visualisation, Decision Support

1. Background

Within the EU-funded research project BrahmaTWinn[25] a methodology to derive spatial vulnerability units (VulnUs) independently from administrative boundaries in the Salzach river has been developed. A major aim of the project is the integration of different indicators (social and environmental) within the DPSIR-Framework and the delineation of spatially integrated units, so-called hydrological response units (HRUs; Flügel 1996).

2. The geon concept

An increasing detail of data and complex analysis tasks opens the door for a plurality of solutions (Lang, in press). Complex real world phenomena require means for reducing and modeling the underlying complexity, for extraction the crucial information, for mapping their dynamics and the constant changes. Automated techniques of object-based image analysis (OBIA, Lang & Blaschke, 2006) making effective use of advanced analysis methods try to match the information extraction with our world view. To account for that, a flexible concept of manageable units is required. The term geon was proposed by Lang (in press) to describe generic spatial objects that are homogenous in terms of a varying spatial phenomena under the influence of, and partly controlled by, policy actions. The geon concept acts as a framework for the regionalization of continuous spatial information according to defined parameters of homogeneity. It is flexible in terms of a certain perception of a problem (specific policy realm, specific hazard domain, etc.) and using geons, we are capable of transforming singular domains of information on specific systemic components to policy-relevant, conditioned information (Tiede & Lang, 2007a). Geons are generated by transforming continuous spatial information into discrete objects by algorithms for interpolation, segmentation, regionalization, generalization; they are analyzed in terms of their arrangement, which leads to emergent spatial qualities; they are dynamic and can be monitored in terms of changes. Within the spatial extent in which a certain policy applies or a certain hazard may occur, a group of geons makes up a spatially exhaustive set (geon set). The spatial limit of the geon set, since being derived functionally, may not fully coincide with administrative boundaries. In this study, vulnerability units have been derived as a specific instance of a geon set within an area, exposed to flood risk.

3. Spatial variation of vulnerability based on spatial vulnerability units

The chosen case study covers the central area of the City of Salzburg. The urban

[25] http://www.brahmatwinn.uni-jena.de/

agglomeration of the city dominates the northern part, whereas additional rural and alpine areas (located in the central-eastern part, and to the South) characterise the chosen study area. The area is highly dynamic through its economic development. Furthermore, it is also prone to floods of various intensity (e.g. HQ-100 flooding in 2002) and therefore serves as an appropriate site to investigate vulnerability within a spatial context.

In an initial step vulnerability was conceptualised and defined. Vulnerability has been confined to the hazard type ‘floods’. The overall vulnerability to floods consists of different domains, such as susceptibility, adaptive and social capacity and resilience. For each domain, sub-domains have been identified and relevant indicators and data sources established.

Data applied originates from publicly available GIS data sources and Census data. The data is not only provided on the basis of different administration units, but additionally on a standardised grid (e.g. 250m and 1000m). This is a very unique approach and allows visualising data not only on the arbitrary administration units but shows the characteristics in a spatial distribution. Another advantage is that in subsequent analysis steps the Modifiable Areal Unit Problem (MAUP, Unwin 1996) is minimized.

To integrate the different data sets and different data sources a normalisation has to be applied. For the generation of a composite vulnerability unit each data layer or domain level has to be assigned by a weight. The methodology applied is known as Multi Criteria Analysis (MCA), Multi Criteria Evaluation or Analytical Hierarchy Process (AHP). A profound discussion on that topic and current best practices approaches is available in Carver (1991) and Malczewski (2000). Considering the fact that vulnerability is not directly measurable and due to its complex dimension and human construction an expert based approach is being chosen. The weights have been derived through a scoring exercise with stakeholders and experts in the field of disaster risk reduction and planning in the Salzach catchment. The experts have been asked to assign scores to each of the factors according to their relative importance and contribution to the vulnerability of people in the Salzach basin - with regard to floods.

To integrate the different datasets a regionalisation approach (Baatz & Schape 2000) has been chosen. This segmentation algorithm is implemented in the software Definiens Professional. As the algorithm allows weighting of the different layers for each domain delineated units can be derived. The advantage is the application of the algorithm from the beginning, but with the constraint to lose the sub-hierarchical levels. A vulnerability value is finally derived from the vector product of the different layers.

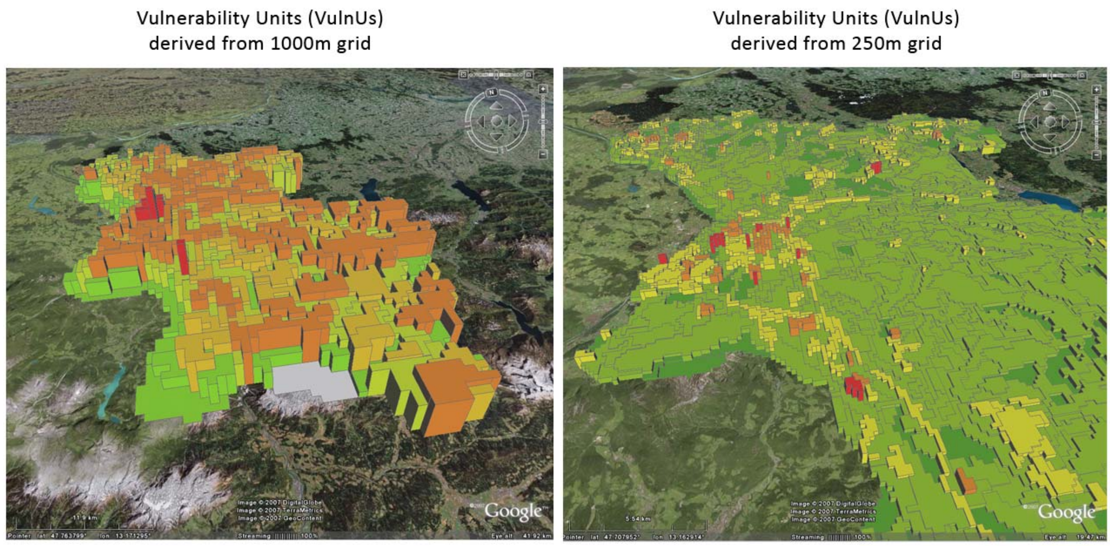


Figure 1. Geon set of vulnerability units (VulnUs) visualised in Google Earth

4. Analytical 3D Views

For visualization and information dissemination of geons (here: vulnerability units) we propose the use of virtual globes as tools for seamlessly combining data on several scales, referenced on a globe and 3D enabled (Figure 1). Freely available virtual globes (e.g. Google Earth, ArcGIS Explorer or NASA World Wind) are providing high spatial resolution contextual data enable to disseminate analytical information to a broad audience. The geons are displayed in virtual globes as Analytical 3D Views (Tiede & Lang, 2007b), which are vector information layers on different spatial aggregation (flexible aggregation units) using extrusion as a means for displaying the crucial spatial information (here: vulnerability measures).

References

Baatz M Schape A (2000) Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. AGIT 2000. In: STROBL,J. et al., Proceedings of the Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT Symposium Salzburg 2000, pp. 12-23.

Carver SJ (1991) Integrating multi-criteria evaluation with geographical information systems. Int. J. Geographical Information Systems, 1991, VOL. 5, No.3, 321-339.

Flügel W (1996) Hydrological Response Units (HRU's) as modelling entities for hydrological river basin simulation and their methodological potential for modelling complex environmental process systems – Results from the Sieg catchment. Die Erde, **127** pp43 – 62.
Jiang H Eastman JR (2000) Application of fuzzy measures in multi-criteria evaluation in GIS. Int. J. Geographical Information Systems, 2000, vol. 14, no. **2** pp173 – 184.

Lang S (in press) Object-based image analysis for remote sensing applications: modeling reality – dealing with complexity. In: Object-Based Image Analysis - Spatial concepts for knowledge-driven remote sensing applications T. Blaschke, S. Lang, G. Hay (eds.). Berlin: Springer, (pages pending)

Lang, S. & T. Blaschke (2006) Bridging remote sensing and GIS - what are the most supportive pillars? In: S: Lang & T. Blaschke (eds.): International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences vol. **XXXVI-4/C42**. CD-ROM and online at www.isprs.org.

Malczewski J (2000) On the Use of Weighted Linear Combination Method in GIS: Common and Best Practice Approaches. Transactions in GIS, 2000, **4(1)** pp5 -22.

Tiede D, S Lang (2007a) Analytical 3D views and virtual globes – putting analytical results into spatial context. ISPRS, ICA, DGfK - Joint Workshop: Visualization and Exploration of Geospatial Data, Stuttgart

Tiede D & S Lang (2007b) Pseudo-realistic and Analytical 3D views – Conditioned Information for Security Scenarios. In: G. Zeug & M. Pesaresi (eds.), Global Monitoring for Security and Stability (GMOSS). Integrated Scientific and Technological Research Supporting Security Aspects of the European Union, 365-373. JRC Scientific and Technical Reports.

Unwin DJ (1996) GIS, spatial analysis and spatial statistics', Progress in Human Geography **20(4)** pp540-441

Biography

Stefan Kienberger is presently working as a researcher and PhD candidate at the Centre for GeoInformatics (Z_GIS), Salzburg University (Austria). His research interests lie in vulnerability science, disaster risk reduction and spatial modelling. Within his PhD research he is assessing the vulnerability to hazards on the district and community level in rural Mozambique.

Stefan Lang, senior researcher and research co-ordinator at Z_GIS, completed his academic qualifications with a PhD (Dr. rer. nat.) in Geography at Salzburg University (Austria). He is a specialist in GIScience and remote sensing for urban applications, nature conservation, landscape analysis and land use/land cover mapping. His research interests include multi-scale spatial modelling and the understanding of complex environmental and societal phenomena in space. Dr Lang is responsible for RTD projects carried out at Z_GIS, and is involved in various educational activities and capacity building projects for GIS and remote sensing in Southern Africa, South Asia and South America.

Dirk Tiede is research associate at the Centre for Geoinformatics (Z_GIS) at Salzburg University, Austria. Involved in a range of EU funded research projects in the field of GMES (GMOSS, LIMES), he is specialised in GIS application development and GIS modelling as well as image analysis using VHSR / LIDAR data and object-based methodologies. His PhD research is focused on the field of bridging remote sensing and GIS – from information extraction to 3D GIS-Visualisation. He is lecturing distance learning modules within the UNIGIS network and university courses in object-based image analysis, 3D Visualization, GIS-Analysis and GIS-Application development.

A method for quantifying equity of access to local services for different demographic groups: combining network analysis with logistic regression using the example of greenspace

Alexis Comber and Chris Brunsdon

Department of Geography,
University of Leicester,
Leicester, LE1 7RH, UK
Tel +44 (0)116 252 3812 Fax +44 (0)116 252 3854
ajc36@le.ac.uk; cb179@le.ac.uk

1. Introduction

There is much concern about provision of various local amenity goods and services to different societal groups. These concerns relate to the provision of facilities and fair, equitable access to them, and range from public transport and recreation facilities, to cultural events, education and welfare provision. Local and national governments are under pressure to assess and quantify the extent of their provision of such facilities in order to demonstrate equitability or to be able to target resources to rectify any disparities in provision. Often the drivers for these concerns relate to wider social and economic development targets, initiatives and legislation. This paper presents a generic method for quantifying the provision of such goods and services based on a GIS network analysis of access to greenspaces by different socio-economic groups linked to logistic regression. The method quantifies the extent to which different sections of the population (as identified in census data in this example) are discriminated against in terms of the provision of goods and services. This work quantifies the provision of greenspace in a city in the English Midlands. The study shows how combining network analysis with logistic regression of socio-economic data can be used to strategically inform local decision-making in relation to national or international (e.g. EU) guidelines and legislation. In the example study, greenspace access for different religious and ethnic groups was analysed and contrasted. The technique of combining network analyses with regression coefficients of census or other geodemographic data allows the accessibility to those services amongst different social groups to be quantified.

2. Background

2.1 Equitable provision environmental services

The UK government is committed to "making sure that public spaces are accessible to all members of our society, able and disabled, young and old …[and to]…producing a review of social and environmental exclusion" (ODPM, 2002, p8). To this end town planning should aim to eliminate discrimination and to promote equality of opportunity. It therefore requires clear and equitable decision-making frameworks in order to reduce social and spatial inequalities (RTPI, 2001). However the concept of 'institutional discrimination' includes situations where policies and/or practices fail to provide a fair and equal service to one or more groups of people. This includes planning policies that take no account of the spatial concentrations of persons with shared characteristics (Heriot-Watt University/ ODPM, 2005). Part of the problem is that population characteristics vary enormously between different areas: in some places the cultural background of the population is changing rapidly, whilst in others the age profile is changing.

Evidence shows that in some cases issues about diversity and equity of access are not well understood in planning. For instance, the report by Sheffield Hallam University and the

ODPM (2004) noted that appraisal of the impacts of development plans and policies on different groups is weak and that planners and development control officers may find it difficult to make the connections between development proposals and diversity. UK government concerns about ensuring equitable access to local authority services are now in the mainstream of planning with planners take positive action to ensure that their practice and policies are inclusive and do not result in systematic disadvantaging of some communities or individuals (Heriot-Watt University/ ODPM, 2005). To this planners are advised to use demographic data such as census data to provide a disaggregated analysis of people and trends in their area in order to demonstrate that they deliver a fair service to everyone (Heriot-Watt University/ ODPM, 2005).

2.2 Guidelines for greenspace access

In many countries guidelines exists for the provision of greenspace. For instance in Flanders the environmental report "Milieu-en Natuurrapport Vlaanderen" (MIRA – see Van Herzele et al., 2000) specifies the different sizes, functions and proximities to which each urban resident ought to have access. Similarly in the UK English Nature (now called 'Natural England') have provided a set of standards for the provision natural places called 'Accessible Natural Greenspace Standards' (ANGSt). Such standards aim to provide benchmarks for the provision of places where people can experience and enjoy nature The ANGSt model recommend at least 2ha of accessible natural greenspace per 1000 population should be provided and specifies minimum standards for greenspace access related to greenspace area and proximity (Table 1. The standards originate from work by Harrison et al. (1995) as described in Handley *et al*. (2003) and English Nature's *A Space for Nature* (English Nature, 1996).

Table 1. ANGSt guidelines for minimum stands for greenspace

Maximum distance from home (km)	Minimum greenspace surface area (ha)
0.3	2
2	20
5	100
10	500

3. Method

The greenspace data used in the analysis was provided by the local council:

- Site of Importance for Nature Conservation
- Local Nature Reserves
- Cemeteries
- Natural Open spaces and Parks

These are areas that are accessible to the general public for everyday use. Points of access (gates, fence breaks etc) were mapped from OS 1:25,000 scale colour raster data and manually digitised. These represented the 'supply' for the network analysis. The output area polygons were provided by the Office of National Statistics and the output area census data were downloaded from Casweb (http://www.census.ac.uk/casweb/) Output area centroids represented the 'demand' for the analysis. To illustrate the equity of access amongst different sections of the community, census data on ethnicity and on religion were collated for each census output area. The road data was from Ordnance Survey Meridian 2 (1:50,000) data (©Crown Copyright/database right 2007, supplied via EDINA).

For each output area, the distance to each greenspace was calculated, allowing the shortest distance to different greenspace types, as defined by Handley *et al*. (2003) and English Nature (1996) to be quantified. The distribution of access that was analysed according to the ANGSt model: the numbers of people with and without access to different ANGSt categories in each census areas were summed for the different classes contained within the census attributes of religion and ethnicity. Generalised linear models were used to estimate likelihood of access as a function of either religion or ethnicity. A table of counts was drawn up where the rows designated whether individuals had access to green space (under rule 2) and the columns designated either the religion or the ethnicity of individuals. The count in column *i* and row *j* is denoted by c_{ij}. To test whether there is an association between the row and column effects, the Poisson regression model was applied:

$$E(c_{ij}) = \log(r + A_i + F_j) \qquad (1)$$

where c_{ij} has a Poisson distribution with mean $E(c_{ij})$, r is an intercept term, A_i is a column effect and F_j is a row effect, is compared against the model

$$E(c_{ij}) = \log(r + A_i + F_j + I_{ij}) \qquad (2)$$

where the extra term I_{ij} is an interaction effect between rows and columns. If this is significantly different from zero, this suggests some degree of association between the row and column effects. In this study, it may be used to test for association between either religion or ethnicity and access to green space. Values of I_{ij} were estimated by fitting Equation 2 to each of the tables using the R statistical software package. It is possible to relate these coefficients to a comparative index of access for each of the row categories, using the formula

$$ACCESS = 100\left(\exp\left(I_{ij}\right) - 1\right) \qquad (3)$$

Due to the way the interaction terms are calibrated, this quantity compares each column *category j* against a 'reference' category. A value of 0 suggests the likelihood of access for *category j* is the same as for the reference category. A value of 50 for *category j* suggests access is one-and-a-half times as likely as the reference category, a value of -50 that it is half as likely, and so on. The reference categories for ethnicity and religion are 'British' and 'Christian' respectively. For each of the coefficients, the *ACCESS* was calculated. An overview of the approach is presented in Figure 1.

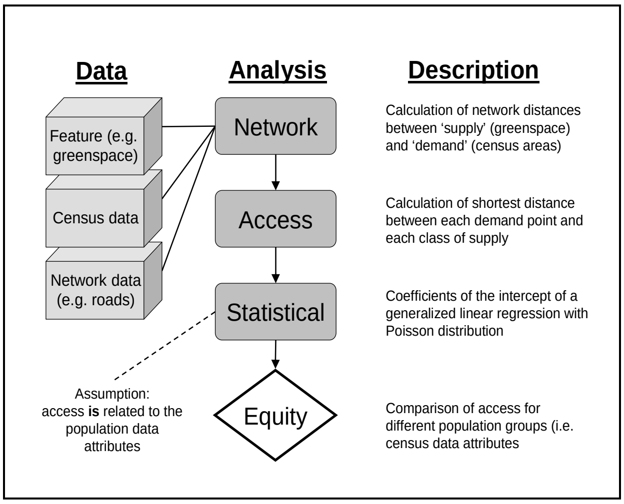


Figure 1. The analysis processing stages, with the data and analysis descriptions

4. Results and discussion

Mosaic plots are a convenient method to analyse and visualise the differences between the predicted and observed values. They plots visualise standardised residuals (often referred to as a standard normal distribution) of a loglinear model using variations in the shade, hatching and outline of the mosaic's tiles. The plots show the access to greenspace in terms of Rule 2

(20ha within 2km) for different ethnic groups (Figure 2) and religious groups (Figure 3) and the tile areas are proportional to the numbers of people affected. Negative residuals are shaded and with broken outlines and positive ones are hatched with solid outlines.

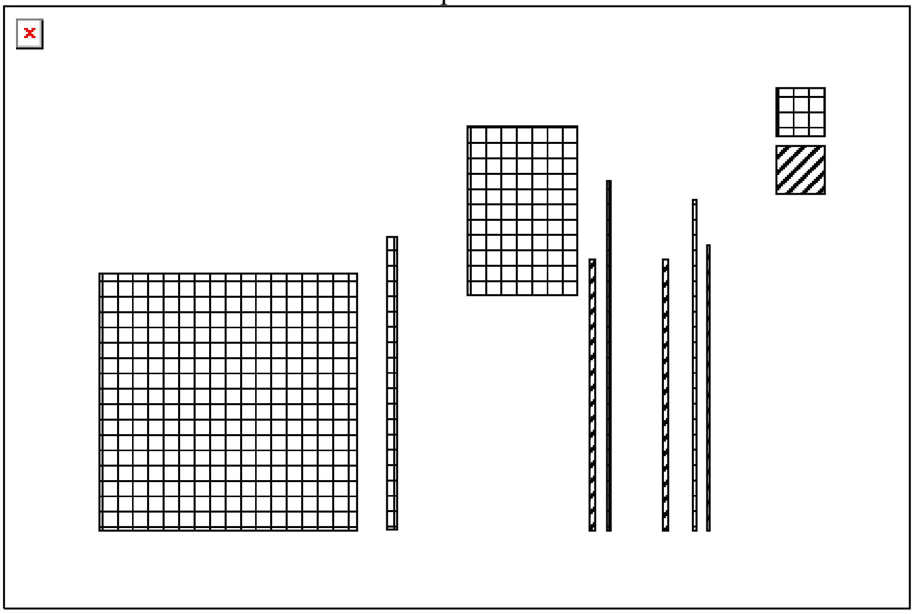


Figure 2. The mosaic plot of access ("True") to greenspace by ethnicity. Mixed1 is 'White and Black Caribbean', Mixed2 is 'White and Black African', Mixed3 is 'White and Asian' and Mixed 4 is 'Other Mixed'.

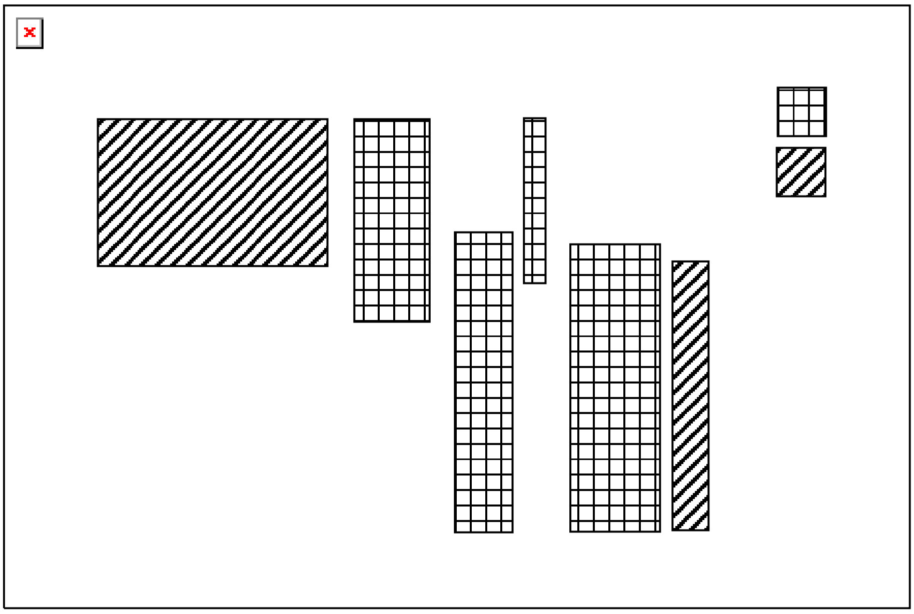


Figure 3. Mosaicplot of access ("True") to greenspace by religion

This paper presents a generic method for quantifying equity of access to some facility for different community groups. In this work we have analysed access to greenspace. But, the approach can be applied to a wide variety of demographic data, including other census variables (deprivation, disability, occupation, economic activity, household tenure and types, age and health) and other geographies such as detailed geodemographic data at household or post-code level.

References

English Nature, 1996. *A Space for Nature,* English Nature, Peterborough.

Handley, J, Pauleit, S, Slinn, P, Barber, A, Baker, M, Jones, C and Lindley, S, 2003. *Accessible Natural Green Space Standards in Towns and Cities: A Review and Toolkit for their Implementation*. English Nature Report Number 526, Peterborough.

Harrison, C, Burgess, J, Millward, A and Dawe, G, 1995. *Accessible natural greenspace in towns and cities: A review of appropriate size anddistance criteria*, English Nature Research Report No.153, English Nature, Peterborough

Heriot-Watt University / ODPM. 2005. Diversity and Equality in Planning: A good practice guide, London, ODPM

ODPM, 2002. *Living places: cleaner, safer, greener*, London, ODPM

RTPI. 2001. *Code of Professional Conduct*, London, RTPI.

Sheffield Hallam University / ODPM, 2004. *Planning and Diversity: Research into Policies and Procedures*, London, ODPM.

Van Herzele, A., Wiedemann, T., Van Overmeiren, M., 2000. Stedelijk Milieu ('Urban Environment'). In: Van Steertegem, M. (Ed.), *Milieu en Natuurrapport Vlaanderen MIRA-S 2000*. Vlaamse Milieumaatschappij & Garant Uitgevers NV, Leuven/Apeldoorn, pp. 501–515 (in Dutch – cited in Van Herzele, A and Wiedemann T, 2003. A monitoring tool for the provision of accessible and attractive urban green spaces. *Landscape and Urban Planning* 63: 109-126)

Despatialising geodemographic propensities.

Martin Callingham

School of Geography, Birkbeck College, University of London
Malet Street, London WC1E 7HX
Tel: +44 (0)20 7631 6473 +44 (0) 20 7439 4377
Email: m.callingham@geography.bbk.ac.uk

1. Summary

This paper shows that the standard way of calculating market propensities for geodemographic system can give incorrect results due to the neglecting of the spatial effect.

2. Background

Geodemographics are a way of classifying small areal units primarily on their census counts across a number of variables (Webber & Craig, 1978).

Classifications of this type found commercial use in the early 1980's when it was discovered that general behaviour, including market propensities, could be ascribed to each areal type (Bermingham, McDonald & Baker, 1979).

The standard way of attributing propensities to geodemographic classifications is by obtaining a profile of users (across a gedemographics system) and comparing it to the national profile (Harris, Sleight & Webber, 2005).

3. Purpose

This paper shows that the standard calculation of market propensities for geodemographic types using customer postcodes collected in a retail outlet is fundamentally flawed.

The problem lies in neglecting the two facts:

- geodemographic types spatially auto correlate
- distance is a powerful factor affecting propensity to visit a store.

The standard method of measuring propensities ignores the distance effect and therefore inadvertently ascribes higher propensities to geodemographic types that are on average nearer.

4. Data used

The geodemographic system used in this work is that which was produced by the Office of National Statistics and is known as the Output Area Classification or OAC for short (Vickers & Rees, 2006). The classification is a threefold hierarchical classification; this work is based on the 7 Super Groups which have been given loosely descriptive names as shown in Table 1.

Table 1. Names for OAC Super Groups

Super Group	Name
1	Blue Collar Workers
2	City Living
3	Countryside
4	Prospering Suburbs
5	Constrained by Circumstances
6	Typical Traits
7	Multicultural

The data used in this work is 48,109 customer postcodes of a restaurant chain which has 26 outlets that are nationally distributed. The customer postcodes were collected in store and verified against the Office of National Statistics Post Code Directory and the OAC code and grid reference (present on the Directory) attached.

5. Preliminary Analysis

The numbers of customers of each type were counted and percentaged on the total. These percentages were then indexed on the comparable national percentages of people of each geodemographic type. This is the standard method of calculating propensities from customer postcode files. The propensities are shown in Figure 1; an index of 1 means a propensity is equal to the national average.

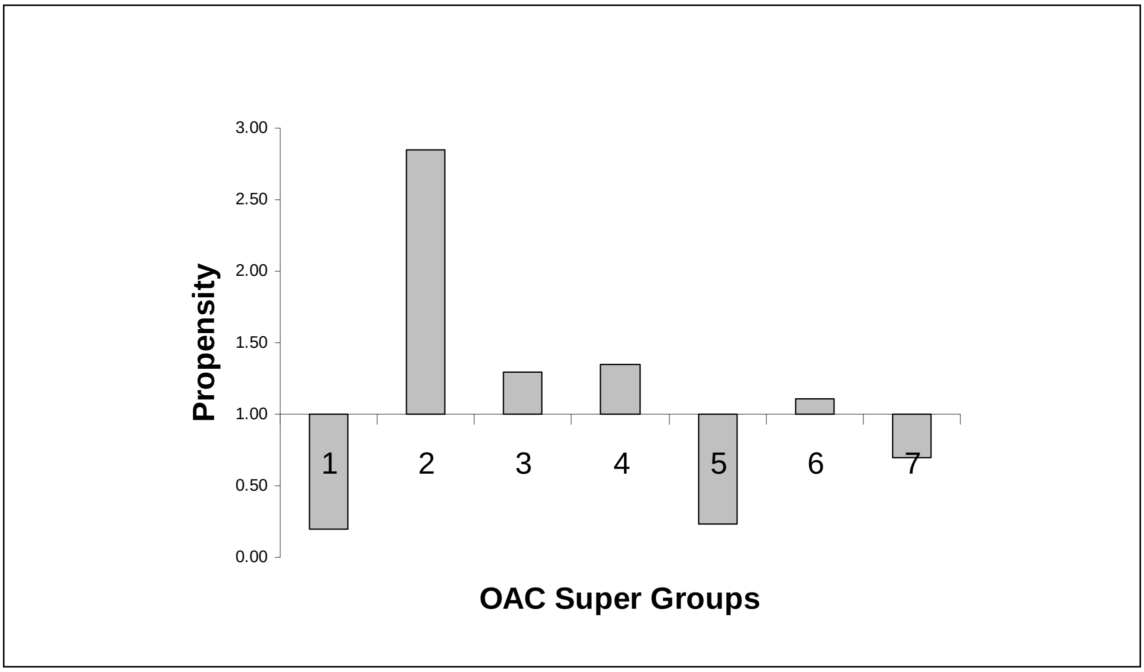


Figure 1. Chart of relative propensity to visit a restaurant by OAC Super Group

There are big differences in the propensity of people to visit this type of restaurant from different OAC Super Groups. It varies from almost 3.0 times the national average (Super Group 2 - *City Living*) to virtually 0 (Super Groups 1 - *Blue Collar Workers* and 5 - *Constrained by Circumstances*).

6. The influence of distance from the restaurant

Each customer was allocated to its nearest restaurant and its distance calculated and categorised into a distance band. This enabled a count of customers in each distance band to be made.

The catchment areas for each restaurant were similarly delineated and the numbers of residents in each distance band summed from 2001 census data.

The probability of a person living in a given distance band visiting the restaurant was calculated by dividing the customer counts by their equivalent population counts. As the customer count came from a sample of customers, the calculated set of probabilities were relative ones: for convenience these numbers were multiplied by 100. A plot of the variation of relative probability to visit is shown in Figure 2 and clearly indicates the marked effect of distance.

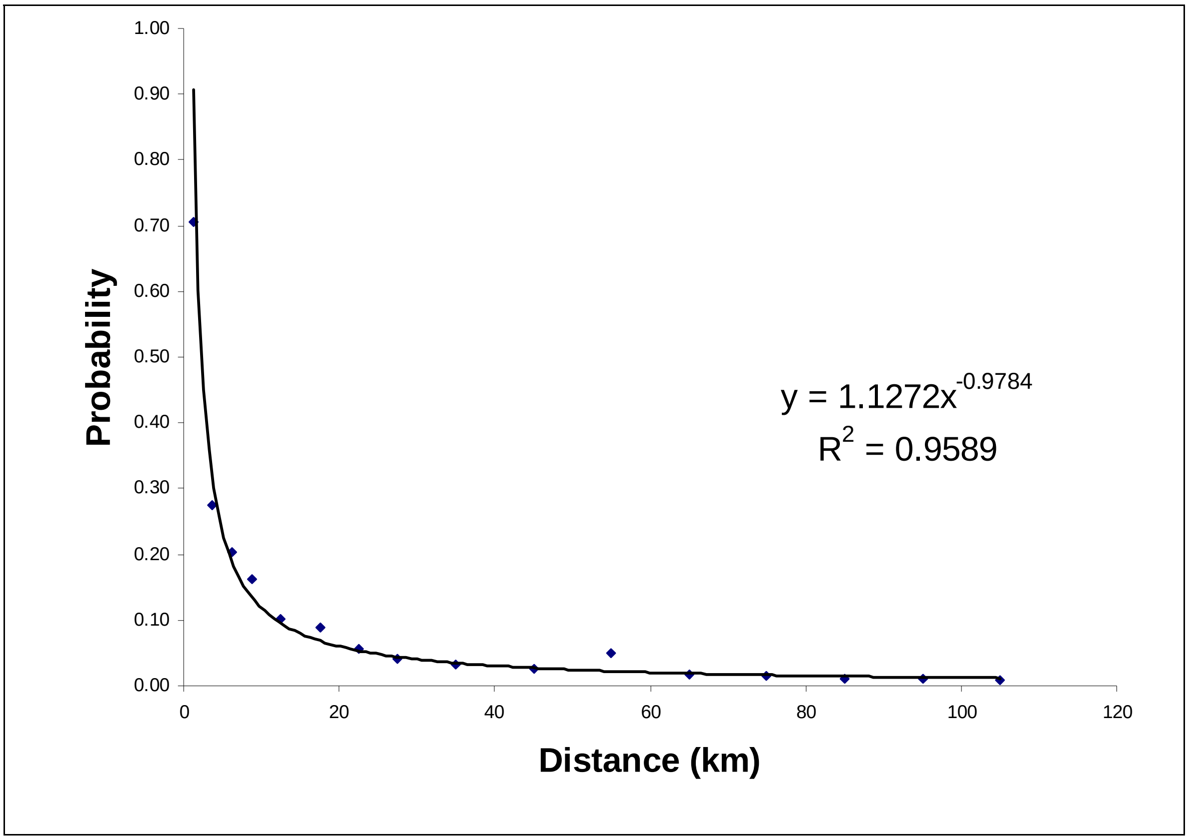


Figure 2. Probability of visiting a restaurant by distance

This indicates that the probability varies approximately inversely with distance. This has to be put in the context of the best geodemographic Super Group having a probability of about three times the average. Clearly the geodemographic and distance effects are both important.

7. The probability of visiting by broad geodemographic type at various distances

The relative probability of visiting the restaurant was recalculated as in Figure 2 but this time separately for each OAC Super Group and the results are shown in Figure 3 for each distance band. The Super Groups are shown in their general order of propensity, so that the first bar within each cluster is Super Group 3, the second is Super Group 4 and so on.

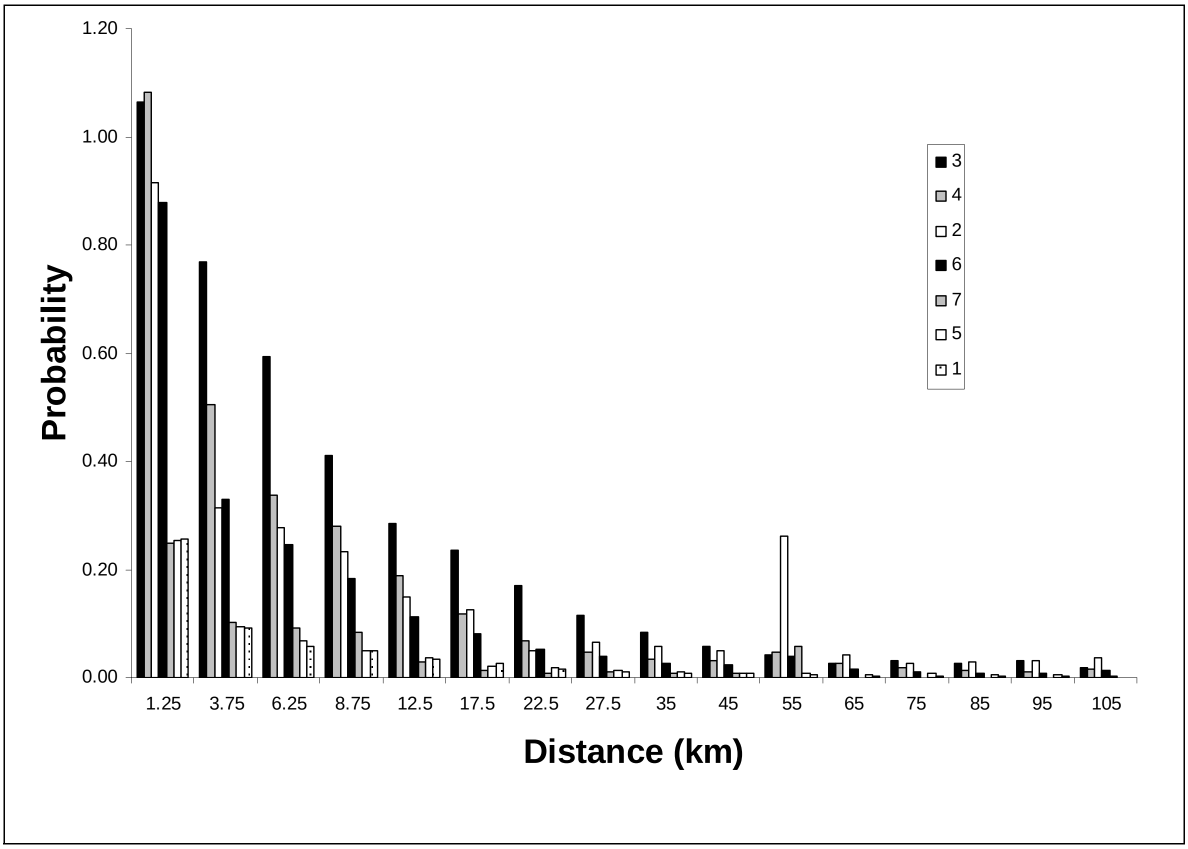


Figure 3. Probability of visiting by OAC Super Groups and distance

As would be expected, the probability for each Super Group is different and these differences are broadly replicated at each distance band.

8. Producing relative probabilities that are despatialised

An estimate of the relative propensity of people in the different OAC Super Groups to visit the restaurant can be obtained from the data shown in Figure 3 by indexing it on the distance effect shown in Figure 2 for each distance band. The best estimate of a despatialised index is the mean across the distance bands. This is shown in Figure 4 – note that the order of Super Groups is different from Figure 3, but the same as in Figure 1.

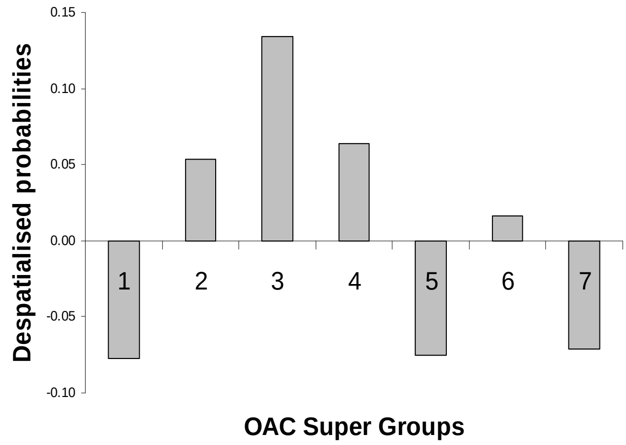


Figure 4. Despatialised indices

The comparable figures calculated using the standard method are shown in Figure 1. It is clear that there are some marked differences, the most noticeable of which is that *City Living* (OAC 2) people are shown to have the greatest probability of visiting the restaurant when estimated by the standard method whilst it is shown to be lower than *Countryside* (OAC 3) when calculated in a despatialised way. The use of the standard propensity figures would therefore lead to new restaurant locations being selected that would not be optimum.

9. Examination of the spatial effect

It has been hypothesised that the differences between the two methods arise because the mean distance of the different OAC Super Groups from the sites is not the same. In particular, the results would suggest that the *City Living* people would be closer to the sites than average whilst the *Countryside* people would be further away than average.
To test this, the mean distance of the population of OAC Super Group was calculated up to 105km. The results are shown in Table 2.

Table 2. Mean distance of OAC population within catchments up to 105 km

Super Group	Name	Dist (km)
1	Blue Collar Workers	37.92
2	City Living	19.3
3	Countryside	40.35
4	Prospering Suburbs	31.16
5	Constrained by Circumstances	34.48
6	Typical traits	31.07
7	Multicultural	14.73
All		30.61

There was a big variation of mean OAC distances and, as expected, the *City Living* people were closer than average to the restaurants and the *Countryside* people further away. It is worth remarking on the fact that the mean distance for all Super Groups was 30.61 km although the radius of calculation was 105 km. This reflects that the more distant output areas were more likely to be in countryside with a much lower population density (the restaurants were always located in urban areas)

10. Checking more formally the relationship between the two sets of probabilities

In principle it should be possible to use the distances in Table 2 to explain the differences between the indices produced by the two methods.

The standard indices were weighted by multiplying them by their mean distance and the resulting estimates of the despatialised indices adjusted to sum to the same total as the standard indices. Table 3 shows the data and the effect of applying the distance weights.

Table 3. Comparison of the indices from the two methods and conversion from one to the other by distance weighting.

Super Group	Name	Standard Index	Despatialised Index	Mean distance Km	Modelled despatialised
1	Blue Collar Workers	0.19	0.26	37.92	0.27
2	City Living	2.85	2.07	19.30	2.03
3	Countryside	1.30	2.33	40.35	1.93
4	Prospering Suburbs	1.35	1.38	31.16	1.55
5	Constrained by Circumstances	0.24	0.32	34.48	0.30
6	Typical traits	1.10	0.99	31.07	1.26
7	Multicultural	0.70	0.28	14.73	0.38
	R squared		0.64		0.96

This shows that a reasonable approximation of the despatialised (and arguably better) indices can be obtained from those calculated in the standard way simply by weighting directly by the mean distance each OAC population is away from the site.

The R-squared of the standard indices to the despatialised indices is .64 and for the modelled despatialised data and the actual despatialised data is 0.96. This suggests that the differences between the two sets of indices are at least in part due a distance effect. The relationship between the despatialised and modelled despatialised data is shown in Figure 5.

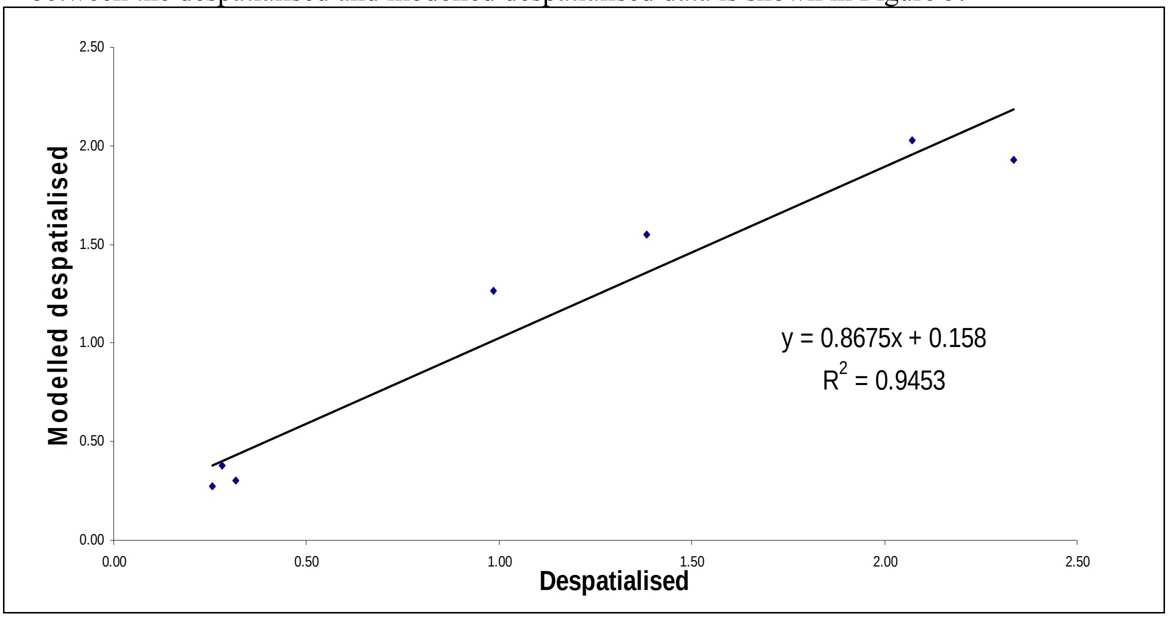


Figure 5. Relationship between the despatialised indices and those modelled from the standard method

11. Discussion

For over thirty years, market propensities have been attributed to geodemographic types using a simple index derived from comparing the percentage of customers in each type to the equivalent percentage in the population. Given a file of customer posycodes, this is a simple and easy to do which and not need specialist GIS knowledge. However, the process neglects the, until recently, unknown fact that geodemographic types spatially autocorrelate (Callingham, 2007). When this new information is coupled with the well known (and common sense) fact that propensity to visit drops off with distance, a flaw is exposed in the way the conventional index method has been calculated as it is now obvious that it favours those geodemographic types that happen to be near the site. Fortunately, it is possible to make corrections to these indices simply by weighting them by the average distance that the geodemographic type's population is from the site as explained in 10 - a figure that is easily calculated.

12. Conclusion

The market propensity indices described are regularly used to help in site selection for all types of retail outlets. Clearly, it is unsatisfactory to use metrics that have inaccuracies in them. In fact what these indices do is to falsely validate the original site selection criteria, though they may of course (if one is lucky) be correct. Therefore, in future these simple metrics should include a further stage of geodemographic distance weighting to remove the bias introduced by auto spatial correlation which appears to be a fundamental property of these systems.

References

Bermingham, J., McDonald, C. & K. Baker. 1979, *The Utility to Market Research of the Classification of Residential Neighbourhoods*. MRS 22nd Annual Conference, Brighton, March 21-23rd 1979.

Harris, R., Sleight, P. & R. Webber. 2005, *Geodemographics, GIS and Neighbourhood Targeting.* (Wiley: London).

Vickers, D. & P. Rees. 2006, *Introducing the Area Classification of Output Areas.* Population Trends 125.

Webber, R. & J. Craig. 1978, *Socio-economic classifications of local authority areas.* Studies in Medical and Population Subjects 35.Office of Population, Censuses and Surveys: London. 2520

Callingham, M. 2007, *Putting geography back into geodemographics*. GIRUK 2007 Conference Maynooth, Ireland.

Office for National Statistics 2005, *Area classification for Output Areas*. [http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/oa/default.asp].

Biography

Martin Callingham is a Visiting Professor at Birkbeck College, University of London where his research interests are areal classification and flow data. He has been involved with geodemographic systems since their inception and was a member of the ONS working party for the creation of their Output Area Classification

Inclusion and exclusion problems in geodemographic targeting

Jakob Petersen[12], Pablo Mateos[1], Paul Longley[1], Phil Atkinson[12]

[1]Centre for Advanced Spatial Analysis, Department of Geography, University College London, Gower Street, WC1E 6BT, Tel. +44 20 7525 0267, Email j.petersen@ucl.ac.uk
[2]Southwark Primary Care Trust, Public Health, Neckinger, London SE16 3QN

KEYWORDS: Targeting, Resource Allocation, Incomplete Knowledge, Hospital Episode Statistics, Geodemographics

Introduction

Most developed countries have experienced a gradual rise in health care expenditure since the 1960s (Organisation for Economic Cooperation and Development 2005). With ageing populations, this 'health care inflation' is predicted to worsen with fewer of working age to support the funding of health care systems that at the same time are expected to meet a growing demand for their services (Gray 2005). The UK National Health Service (NHS) is principally funded by the government and as such is central to the political debate not only about taxation, but also in questions of its share of the state budget in competition with other public sectors like education, policing, social security, defence, etc. (Hsiao and Heller 2007). Within the health care system itself, there are complicated trade-offs between equity and efficiency objectives with respect to improving population health, reducing risks and inequalities, and a need to ration services balanced with maintaining a certain level of user satisfaction (Musgrove 2003). There is consequently increasing pressure to reform health care systems and the NHS is undergoing reforms to make its organisation more cost-effective and to attract private enterprise (Talbot-Smith and Pollock 2006, Pollock et al. 2007). Unsurprisingly this development has resulted in a plethora of new health management solutions and planning tools. One planning tool that has garnered much interest is the use of geodemographic systems for health equity assessments and public health campaigns (Webber 2004, Farr and Evans 2005, Jones et al. 2005). Geodemographic systems are neighbourhood classifications that for three decades have been used for direct marketing and market research in the commercial sector (Webber and Craig 1978, Harris et al. 2005, Longley 2005). There are a number of competing systems and most now present extensive support material for demographic profiling and neighbourhood targeting across the public sector. The power of geodemographics in the commercial sector clearly lies in estimating demand under *incomplete* knowledge of the market, e.g. by coding the 25,000 annual responses to the Target Group Index surveys with a geodemographic system, it is possible to create maps of demand for a wide range of consumer goods and services across the whole country. Similar ideas have been proposed for the mapping of health care demand, and several systems offer health need indices based on Hospital Episode Statistics (HES: (Webber 2004, Farr and Evans 2005).

One fundamental difference between the private and public sector applications, however, may be that the NHS has already invested considerably in HES and other electronic health care databases and arguably has *complete* knowledge of the health care demand to a very fine level of geography, i.e. the user postcode. In this paper we will investigate how well a number of geodemographic systems facilitate prediction of observed demand for hospital admissions for a number of the most common chronic diseases. This work adds to the more traditional ways of evaluating geodemographic systems with the use of Lorenz curves and Gini coefficients (Novak et al. 1992). We will discuss the findings in relation to targeting of public health campaigns.

Method

To estimate the success of a geodemographic targeting strategy we set a hypothetic goal of reaching the top 20% of admissions for a given disease and compared it to reaching the same goal by a geographic targeting strategy. To do this we used a simple design used in medical diagnostics, i.e. comparing the overlap in frequencies of patients with a positive diagnosis using a high intervention test (the gold standard) with an alternative and usually cheaper low intervention test (Kirkwood and Sterne 2003). In this study we used geographic targeting as gold standard and geodemographic as the alternative test.

All admissions (spells) at hospital for a number of common chronic diseases were obtained at postcode level for Greater London, for the years 2001-2004 (Table 1).

Table 1. Number of chronic disease admissions to hospital for residents of Greater London, 2001-2004

Chronic disease indicator	*Admissions*	Chronic disease indicator	*Admissions*
Angina pectoris (I20,I25)	119,538	Asthma (J45-J46)	36,573
Breast cancer (C50,D05)	91,026	Congestive heart failure (I50)	32,759
Colorectal cancer (C17-C21)	87,988	Acute myocardial infarction (I21-I24)	31,458
All chest pain (R073-074,R101)	85,861	Cholelithiasis (K80)	29,145
Back pain (M50-M54)	55,713	Traumatic brain injury	26,333
Mental health (F20-F48)	54,180	Diabetes (E10-E14)	25,508
All arthroses (M15-M19)	51,332	Skin cancer (C43-C44)	23,663
Leukaemia (C91-C95)	49,963	Epilepsy (G40-G41)	19,087
COPD (J40-J44)	49,751	Prostate cancer (C61)	16,669
Stroke (I60-I69)	43,930	Cervical cancer (C53,D06)	10,620
Lung cancer (C33-C34)	43,774	Total	984,871

The data were aggregated at three geographical levels (unit postcode, output area and super output area) and labelled with the finest level codes of eight different geodemographic systems[26], e.g. the 50 subgroups in the Output Area Classification covering Greater London. We also included Indices of Multiple Deprivation (IMD) as a potential segmentation system. To do this we have divided Greater London into a comparable number of segments (we chose 50 quantiles) according to IMD score. The

[26] Postcode level: 1. Mosaic UK Type (60 segments in Greater London); 2. Acorn Type (58); 3. Health Acorn Type (27). Output Area level: 4. Output Area Classification Sub Groups (50); 5. London Output Area Classification Groups (49). Lower Layer Super Output Area level: 6. Index of Multiple Deprivation in 50 local quantiles according to scores (50); 7. a bespoke classification based on HES diagnostic groups (48); 8. P2 Branches (38).

performance of the systems was evaluated by: (1) Gini coefficients weighted by Census 2001 population counts. The Gini coefficient is an overall mathematical measures of inequality. By weighting target frequencies with base population frequencies across all segments in a segmentation, it is intended to give higher values to a system that minimises base population to target; and (2) ranking geographic areas according to crude rate for each disease indicator and selecting those containing the top 20% of all frequencies. These sets were treated as the diagnostic gold standard and the same procedure was repeated with area types within each geodemographic system to create the alternative diagnostic sets. For each disease indicator and geodemographic system, diagnostic sensitivity was calculated as the percentage of gold standard admissions included in the geodemographic target. The base population included in the geographic target of the top 20% of admissions (the gold standard) were divided by the number of admissions in the target for each disease to produce a numbers- to-target ratio. This was likewise repeated for the geodemographic target sets.

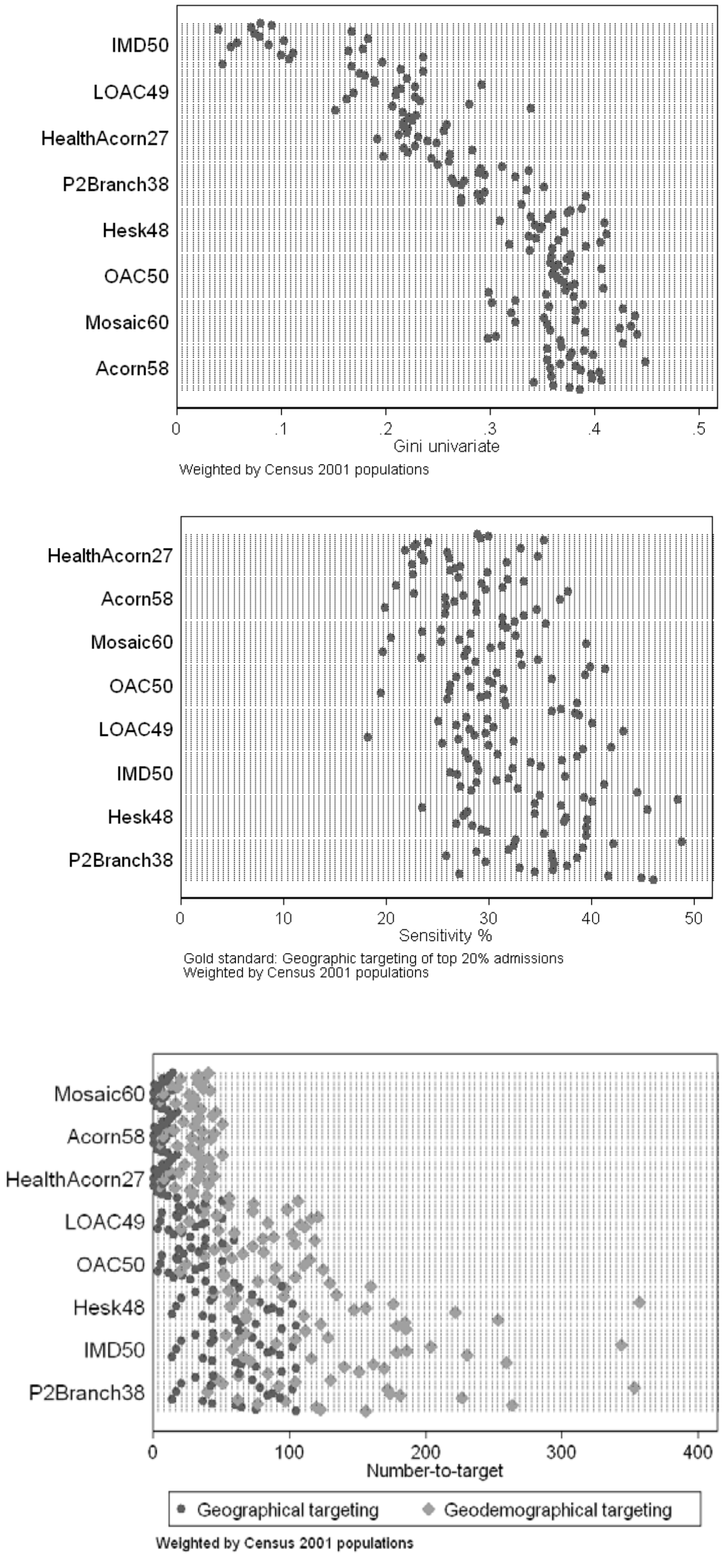


Figure 1 Top: Gini coefficients for chronic diseases across eight different geodemographic systems in horizontal panels. Middle: Sensitivity relative to geographic targeting as gold standard. Bottom: Number-to-target for same

.

Results

The Gini coefficients showed the segmentations with two general commercial systems, Acorn and Mosaic, to be best optimised relative to the base populations (Figure 1). Comparable results were obtained with OAC (output area level) and HESK[26] (super output area level); whilst the segmentation using the IMD performed the least well. Evaluating the systems relative to geographical targeting showed very low sensitivity overall. Postcode systems had the lowest sensitivity overall followed by output and super output area systems. The geodemographic strategies would in this study reach 20% of admissions, albeit not the same 20% as determined by the geographic targeting strategies. In fact the proportional overlap, i.e. the sensitivity, could be as low as 20% and never exceeded 50%. Strategies using geodemographic systems at postcode level would potentially provide a cheaper means of reaching the target population because of the relatively low base populations indicated by the lower number-to-target.

Discussion

Besley and Kanbur (1990) proposed that given scarce resources geographical targeting should favour areas in order of need until the budget is exhausted (Besley and Kanbur 1990). Any targeted strategy based on aggregated data however opens up for problems of inclusion and exclusion. A public health campaign strategy, for example, may include individuals who are not at risk for the health outcome it was designed to counter or ameliorate; i.e. problem of inclusion. Conversely there may be citizens with those exact needs that are excluded by the strategy simply by having the 'wrong' postcode; i.e. the exclusion problem. Both these problems pertain to efficiency and fairness considerations of a strategy and should be evaluated in line with other welfare policy interventions (Pellegrino and Thomasma 2004).

Our empirical analysis suggests that all of the geodemographic systems had a low sensitivity in comparison with geographic targeting. This exemplifies *exclusion* problems: geodemographic allocation strategies would still reach 20% of admissions, albeit not the same 20% displaying the highest needs as determined by geographic targeting. High numbers of base population to target (indicated by number-to-target) demonstrates the *inclusion* problems in both types of targeting; although geodemographic strategies would be more expensive to deploy, e.g. in terms of mailshots or other campaign means magnified by base population numbers.

The results of the diagnostic approach deployed here also suggest that it is the geographic order of aggregation (unit postcode, output area, or super output area), more than the geodemographic classifications themselves, that is critical for the accuracy of targeting. This also questions whether Gini coefficients, however popular, are too sensitive to the huge variability in base population size within the geodemographic classifications and applied in this way hence become a measure of this variability rather than of actual targeting 'efficiency'.

In evaluating geodemographic systems for the targeting of public health campaigns we need to consider two different situations. First, cases where we have data on actual demands and geographic targeting would thus be more accurate, more fair and potentially less costly for campaigning than a geodemographic alternative. Second, cases where we would like to predict lifestyle information. In theses cases geodemographic systems has potential value. Choosing a geodemographic system will rely on a number of factors, including budget and health data quality. This study suggests that postcode systems are not necessarily more accurate, but that they would be cheaper to deploy in campaigns, which takes us back to the central dilemma in health care delivery: to balance equity and efficiency.

Acknowledgements

We would like to thank the ESRC and DTI for support through Knowledge Transfer Partnership, no. 666. We are further indebted to Bromley Local Research Ethics Committee, Security and Confidentiality Advisory Group (SCAG) and Patient Information Advisory Group (PIAG) for permitting access to Hospital Episode Statistics data. We would also like to thank Experian Ltd., CACI and Beacon-Dodsworth for permission to use their geodemographic directories.

References

Besley, T., and R. Kanbur. 1990. The principles of targeting. Policy Research Working Paper Series **385**.

Farr, M., and A. Evans. 2005. Identifying 'unknown diabetics' using geodemographics and social marketing. Interactive Marketing **7**:47-58.

Gray, A. 2005. Population ageing and health care expenditure. Ageing Horizons **2**:15-20.

Harris, R., P. Sleight, and R. Webber. 2005. Geodemographics, GIS and neighbourhood targeting. John Wiley & Sons Ltd., Chichester.

Hsiao, W. C., and P. S. Heller. 2007. What macroeconomists should know about health care policy. International Monetary Fund, Washington DC.

Jones, C. E., P. Mateos, P. Longley, and R. Webber. 2005. Who does not eat their greens? Geodemographics, health promotion and neighbourhood health inequalities. Proceedings to 13th GISRUK Conference **13**.

Kirkwood, B. R., and J. A. C. Sterne. 2003. Essential Medical Statistics, 2nd edition. Blackwell Science, Malden.

Longley, P. A. 2005. A renaissance of geodemographics for public service delivery. Progress in Human Geography **29**:57-63.

Musgrove, P. 2003. Health economics in development. World Bank, Washington DC.

Novak, T. P., J. Leeuw, and B. MacEvoy. 1992. Richness curves for evaluating market segmentation. Journal of Marketing Research **19**:254-267.

Organisation for Economic Cooperation and Development. 2005. Health data. OECD, Paris.

Pellegrino, E. D., and D. C. Thomasma. 2004. The good of patients and the good of society: striking a moral balance. *in* M. Boylan, editor. Public health policy and ethics. Kluwer Academic Publishers, Dordrecht.

Pollock, A. M., D. Price, E. Viebrock, E. Miller, and G. Watt. 2007. The market in primary care. British Medical Journal **335**:475-477.

Talbot-Smith, A., and A. M. Pollock. 2006. The New NHS - a guide. Routledge, London.

Webber, R. 2004. Neighbourhood inequalities in the pattern of hospital admissions and their application to the targeting of health promotion campaigns. CASA working papers **90**.

Webber, R., and J. Craig. 1978. Socio-economic classifications of local authority areas. Studies in Medical and Population Subjects **35**.

Biographies

Jakob Petersen is working on a knowledge transfer partnership between Southwark Primary Care Trust and Centre for Advanced Spatial Analysis, Department of Geography, UCL. His research interests concerns public health targeting and the use of health care databases. He has an MSc in GISc and Biology and is studying for a PhD in health geography.

Dr. Pablo Mateos is lecturer in Human Geography at the Department of Geography in UCL. His research interests lie within Population and Urban Geography, and focus on investigating new ontologies of ethnicity, migration and mobility in today's rapidly changing cities and societies.

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 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.

Dr. Phil Atkinson is Public Health consultant in infectious disease epidemiology and mental health. He is head of Research & Development at Southwark Primary Care Trust and also directs the Research Development Centre for primary, community and social care research in southeast London.

Do Local Information Systems Hide the Bigger Picture? An analytical approach to measuring the strength of local boundaries.

Sarah Goodwin[1] and Jason Dykes[2]

[1]Manchester Geomatics Ltd, Beehive Mill, Jersey St, Manchester M4 6JG

sarah.m.goodwin@gmail.com

[2]giCentre, School of Informatics, City University London, EC 1V 0HB

Tel. (+ 44 20 7040 8906) Fax (+44 20 7040 8845)

jad7@soi.city.ac.uk, www.gicentre.org/jad7

KEYWORDS: Local Information Systems, polygon Wombling, boundaries, output areas, geodemographics

1 Introduction

The growth of freely available statistical data at small area geographies, and associated initiatives such as e-Government and the Freedom of Information Act 2000 have encouraged local governments and partnerships to develop web-based Local Information Systems (LIS). While the capabilities and functionality of operational LIS differ widely (Foley *et al.,* 2007), they are all constrained by scale and local boundaries. Administrative boundaries such as Local Authority (LA) boundaries are not visible on the ground and are susceptible to change, however, these are often used within LIS for decision making and analysis. Constraining data at LA boundaries in LIS could potentially restrict the information and subsequent knowledge acquired for the areas on the borders of such systems and split communities with similar characteristics. If this effect exists it may be reinforced by the growing LISfocussed approach to the consideration of geography within LAs. Here we investigate whether LA boundaries follow patterns of social change or whether they divide regions of homogenous population.

2 Methods

Two analytical techniques were used to investigate this concept in respect to the study area of Manchester and its eight surrounding LAs.

The ONS Output Area Classification (OAC) data was used to create 'homogenous areas' (HAs) within the defined study area, at the three OAC cluster levels of Super Groups, Groups and Sub Groups. These HAs were analysed to discover whether LA boundaries divide regions considered homogenous according to the three levels of classification.

A 'Geographical Boundary Analysis' technique known as 'Polygon Wombling Õ (Jacquez and Greiling, 2003) was used to analyse multiple census variables at Output Area (OA) level to identify abrupt demographic change. Boundary analysis techniques seek to define and evaluate *'edges of homogenous areas or zones of rapid change in a variable's spatialfield'* (Jacquez *et al,* 2000:225).

2.1. Polygon Wombling

Womble (1951) developed a method for discovering boundaries in a collection of continuous variables. Polygon Wombling (or areal Wombling) has been developed much more recently (Jacquez *et al,* 2000) and aims to identify difference boundaries - shared borders between two adjacent polygons that have dramatically different observed response values (Boots, 2001). There are relatively few academic articles relating to areal Wombling (Lu and Carlin, 2005), however, the method provides an effective means of identifying the important boundaries in social data. It also offers a unique representation of areal data enabling zone boundaries to be variously emphasized according to the strength of the demographic boundary that they represent.

During the Wombling process the polygon edges are analysed to distinguish the boundaries of abrupt variable change between neighbouring areas. Each polygon edge is assigned a *boundary likelihood value* (BLV) which measures the rate of variable change as well as the spatial change between two areas through a distance metric. Edges with higher BLVs represent abrupt changes in variable values and therefore are more likely to form part of the *'Wombled Boundary'* (BoundarySeer, 2007). The polygon edges with the highest BLVs are known as *Boundary Elements* (BEs). When edge BLVs exceed an established threshold , candidate BEs become part of the *Wombled Boundary.* While most analysts use the upper 5 or 10 percent relative BLV thresholds, an absolute BLV threshold that is determined by using the full range and distribution of all the edge BLVs is recommended. In order to investigate the difference between thresholds, the top 5, 10 and 20 percent thresholds were explored in this analysis along with an absolute threshold defined specifically to relate to the BLV distribution of the study area.

Multi-variable polygon Wombling allows boundaries that represent discontinuities in multivariate data to be identified. As demographic patterns are related to many social variables such as housing, employment and education, multi-variable polygon Wombling with small area geographical units, such as OAs, enables the identification of strong and weak social barriers at the local level. In order to generate a comprehensive picture of social patterns the forty-one Census 2001 variables used to create OAC were employed in multi-variable Wombling.

3.Results

The locations of the Wombled boundaries derived using an absolute threshold of 19.2% are shown in Figure 1. The LA boundaries are more likely to be identified as Wombled boundaries than other edges in the study area as a whole for each of the three relative thresholds - a third of all LA edges are within the top 20% barriers of abrupt social differences as measured through the multi -variable Wombling (Table 1). X^2 tests give us confidence that these differences are unli kely to be random in all cases (see Table 1). The OAC analysis supports this suggestion that LA edges are associated with social difference as a small percentage of LA boundaries cut across HAs. However, both analyses show that LA boundaries can divide areas that have similar social characteristics and so the consideration of LAs as distinct units may in places "*hide the bigger picture".* Importantly, there are spatial trends to this pattern – the answer to our question is geographically variable in our study area. HAs cross LA boundaries most frequently in specific areas, particularly around Manchester City Centre, indicating that the strength of the LA boundaries varies across the region. Note also that the OAC data is based on clusters generated at a national level, whereas the raw data for the Wombling is standardised within the study area to ensure that results are locally discriminating. The multi-variable Wombled boundaries identify some edges that the OAC analysis does not and these occasionally divide OAC HAs, demonstrating that even though the two techniques use the same data they detect social discontinuity with different precision.

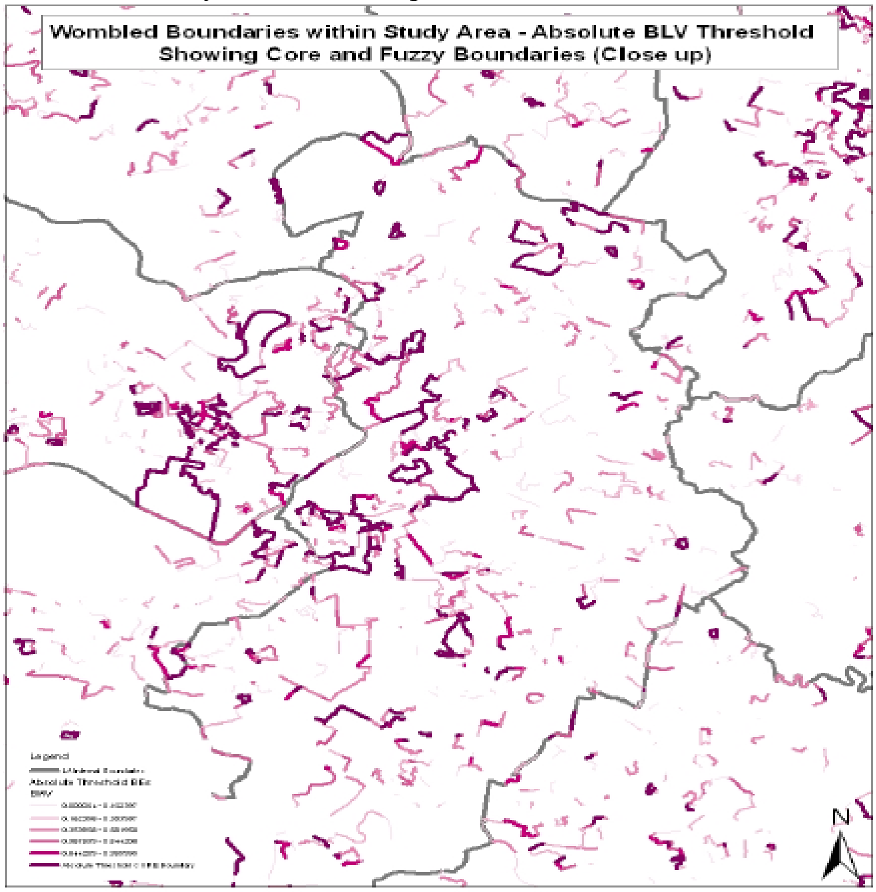


Figure 1. A close up of the Manchester City Centre area illustrating the absolute Wombled boundaries using a fuzzy representation - the darker the boundary the higher the BMV.

	threshold	expected	observed (wombled)		X^2	p	significance and trend	
All LA edges	5%	27.8	41	7.4%	6.27	**1.23%**	**at 5%**	more than expected
	10%	55.6	89	16.0%	20.07	**0.00%**	**at 1%**	more than expected
	20%	111.2	174	31.3%	35.47	**0.00%**	**at 1%**	more than expected
Manchester LA edges	5%	15.8	34	10.7%	20.79	**0.00%**	**at 1%**	more than expected
	10%	31.7	64	20.2%	32.92	**0.00%**	**at 1%**	more than expected
	20%	63.4	114	36.0%	40.39	**0.00%**	**at 1%**	more than expected
Other LA edges	5%	11.9	7	2.9%	2.05	15.22%	no	as expected (a few less)
	10%	23.9	25	10.5%	0.05	82.18%	no	as expected
	20%	47.8	60	25.1%	3.12	7.76%	no	as expected (a few more)

Table 1. Observed and expected Wombled edges for all LA edges and Manchester / non-Manchester LA edges at three thresholds. Expectations calculated under the assumption that Wombled edges will be distributed randomly across all edges. Significance evaluated with X^2 test (no Yates' correction).

Comparison of edges on the Manchester LA boundary with other LA edges helps us identify where consideration of LAs as distinct units may "*hide the bigger picture* ". Wombled edges occur much more frequently on the Manchester LA boundary than would be expected if they were randomly distributed, with 10.7% of the Manchester edges within the top 5% BLV threshold. Edges with the highest BLVs occur particularly frequently here. However this trend is not apparent when all other (non-Manchester) LA edges are considered. Indeed whilst the result is not significant, fewer Wombled edges are observed than expected in this case. The polygon Wombling provides no evidence that other LA edges constitute difference boundaries between socially distinct neighbours.

The four distributions of BLVs for a) all (OA) edges, b) all LA boundary edges, c) all Manchester LA edges and d) all non-Manchester LA edges are shown in Figure 2. The distribution of BLVs in each of the histograms shows some similarities. The outliers to the skewed distribution represent the edges of very abrupt change. This positive skewed distribution indicates that very abrupt change in social data is in fact relatively rare across the study area. Our overall findings suggest that it is more likely to occur across LA boundaries than across non-LA boundaries, but when Manchester LA boundaries and non-Manchester LA boundaries are considered separately this is only true in the former case . This trend is reflected by the less dramatic decay evident in Figure 2a than 2b. The strength of the Manchester LA boundary is apparent when comparing Figures 2c and 2d.

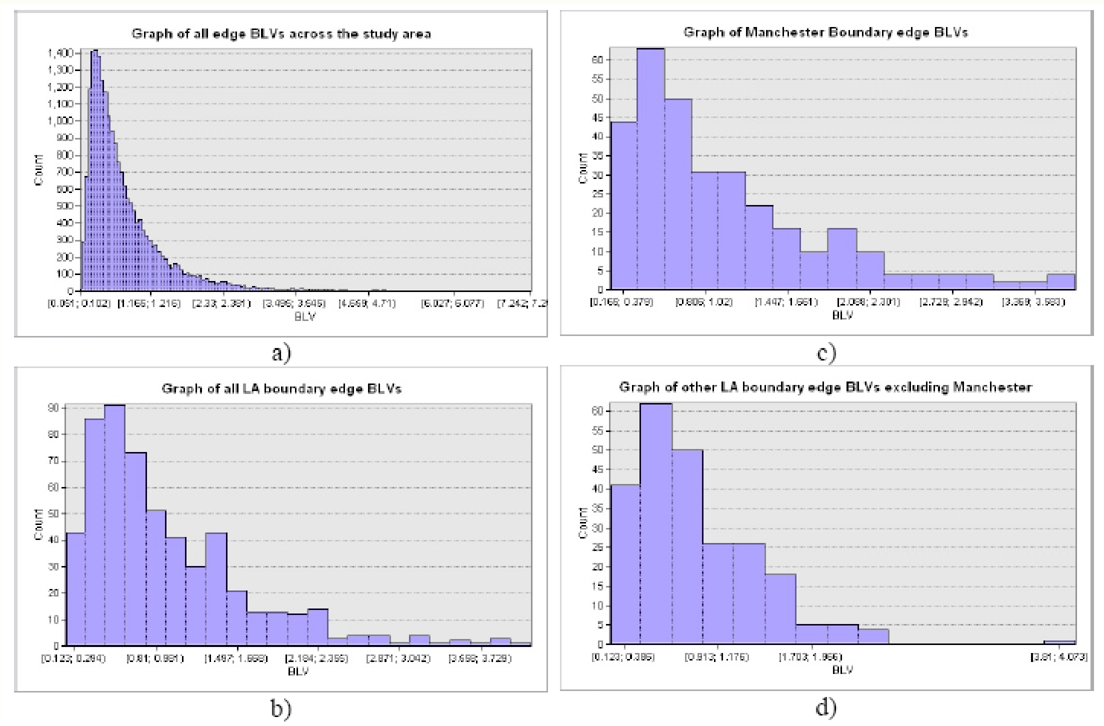


Figure 2. Histograms illustrating the distribution of BLVs for: a) all the edges in the study area, b) All LA boundary edges, c) the Manchester LA edges, d) the non-Manchester LA boundary edges

A number of specific examples of Wombled boundaries were investigated. Some occur at identifiable physical barriers, many following main street networks which are particularly well illustrated in and around the city centre of Manchester (Figure 3). *'Wombled Polygons',* whereby an area is highlighted as an island of abrupt social differences when compared to surrounding neighbours, are particularly interesting. For example, the area of Salford Quays, which was newly developed at the time of the 2001 Census through heavy investment in urban regeneration, is clearly identified as an island through Wombled boundaries.

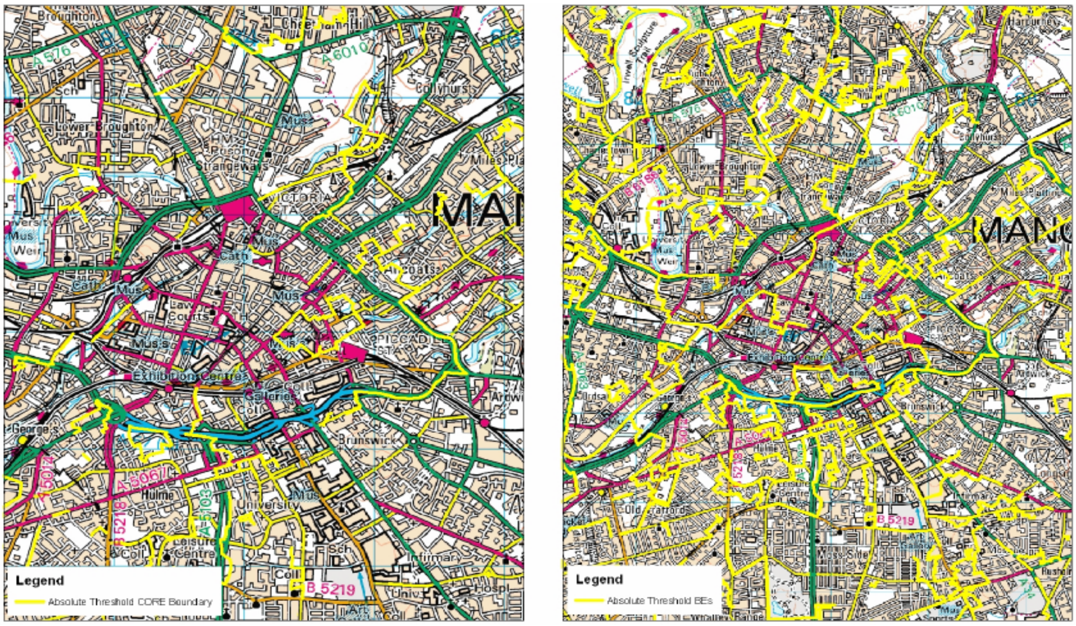


Figure 3. Absolute BLV threshold showing the core and fuzzy Wombled boundaries (in yellow) on a
1:50,000 OS raster map. © Crown Copyright/database right 2007. An Ordnance Survey/EDINA
supplied service.

4. Conclusions / Implications

Polygon Wombling allows us to measure, visualize and compare discontinuity between zones. It provides a unique perspective on area data and the opportunity to remove areal units from maps that use line thickness and colour to reflect boundary strengths (Figure 1). The degree of social difference revealed through Wombling is a particularly interesting concept for the analysis of the geography of social change in the urban environment as it enables abrupt social differences that may lead to possible tensions to be identified. The analysis of data over time could subsequently lead to a better understanding of social barriers and aid the monitoring of regeneration projects in urban environments.

In the context of LIS, we have shown that the answer to our research question is geographically variable, drawing attention to some of the potential risks associated with a discontinuous approach to data use. The high frequency of boundaries with BLVs along the Manchester LA is very unlikely to be due to random processes and the strength of these boundaries in comparison to the surrounding LA boundaries may be of interest for social analysis within the area. Whilst some of the abrupt demographic changes are due to physical barriers on the ground the frequency with which Wombled boundary edges fall on the Manchester boundary indicate that the political boundaries themselves could influence these social barriers. Where LA boundaries do not represent discontinuities in data about those who they divide, as is the case in the non-Manchester LA boundaries, then LIS may be hiding a significant 'bigger picture' by forcing administrative divisions where social divisions do not exist. Disregarding this may even contribute to social divisions in the long run. These results and the analytical techniques used in this research project are applicable to a number of areas of social research such as urban regeneration, neighbourhood renewal and local boundary definition.

5. Acknowledgements

We acknowledge Dr Robert Barr, Senior Lecturer of GIS, University of Manchester and Managing Director of Manchester Geomatics Ltd. and Dr Jo Wood, Senior Lecturer, City University London for advice and support that have contributed to this work.

References

Boots, B. (2001) Using local statistics for boundary characterisation. *GeoJournal,* 53, 339-345.

BoundarySeer (2007) Online Help - Boundary Detection - Wombling. Available online at http://www.terraseer.com/products/bsr/help/index.htm#Wombling/About_wombling.htm accessed on 2 1/12/07. TerraSeer.

Foley, P., Alfonso, X. & Wiseman, I. (2007) Local Information Systems: A review of their role, characteristics and benefits. IECRC for the Department for Communities and Local Government.

Jacquez, G. M. & Greiling, D. A. (2003) Geographic boundaries in breast, lung and colorectal cancers in relation to exposure to air toxics in Long Island, New York. *International Journal ofHealth Geographics,* 2.

Jacquez, G. M., Maruca, S. & M.-J. Fortin (2000) From fields to objects: A review of geographical boundary analysis. *Journal of Geographical Systems,* 2, 221-241.

Lu, H. & Carlin, B. P. (2005) Bayesian Areal Wombling for Geographical Boundary Analysis. *Geographical Analysis,* 37, 265-285.

Womble, W. H. (1951) Differential Systematics. Science, 114,315-322.

Biographies

Sarah Goodwin is a GeoData Consultant at Manchester Geomatics Ltd. She works closely with local government clients to ensure the best use of their spatial data. This includes the development of Local Information Systems. She recently received an MSc in Geographic Information Systems from City University London.

Jason Dykes is a Senior Lecturer in Geographic Information and member of the giCentre at City University London. He has research interests in geovisualization and is involved in work that is using and evaluating geovisualization techniques in local government.

Scenario-based Small Area Population Modelling for Social Infrastructure Planning

Yang Li[1], Allan J. Brimicombe[2]

[1,2]Centre for Geo-Information Studies,
University of East London, University Way, London E16 2RD
Tel. 0044 (0)2082232603 Fax 0044 (0)2082232918
y.li@uel.ac.uk, a.j.brimicombe@uel.ac.uk
http://www.uel.ac.uk/geo-information

KEYWORDS: small area population modelling, social infrastructure planning, multiple data sources

1. Introduction

In recent years, the geodemographic makeup of some areas in the UK has been rapidly changing. For example, immigration has put more pressure on child services, education and health care in places such as Slough, Peterborough and the Thames Gateway. Other factors affecting the Thames Gateway are housing development as part of the massive regeneration and the development and legacy of the Olympic site. This region is also experiencing high population churn, uncertainty in its demographic composition and issues in matching service delivery. There are also increasing demands for building sustainable communities that can adapt to change. A key to maintaining sustainable communities is the quality of services and opportunities afforded by the social infrastructure. Where the needs of residents rapidly change due to (im)migration, social and economic mobility and transience, there needs to be robust mechanisms for compiling and updating the evidence base on which policy and planning changes must necessarily be founded. This paper proposes scenario-base small area population modelling with multiple administrative data sources as a means of evidencing change. It is being implemented in the Thames Gateway London boroughs, with funding from UrbanBuzz (www.urbanbuzz.org) to support local social infrastructure planning.

2. Small area population modelling

The size of local population is an important determinant for the distribution of central government grants to Local Authorities, local policy-making and in the calculation of performance indicators such as crime rates. Small area population estimation and projection are crucial for local social infrastructure planning.

The social infrastructure normally includes (EDAW and Brittan, 2006):

- Primary, Mental and Acute Healthcare Services;
- Community facilities (Libraries and Adult Learning, Local Service Centres, and Youth services/facilities for young people);
- Education (early years, primary and secondary);
- Leisure, recreation and open space/green spaces;
- Emergency & Essential Services.

Social infrastructure planning aims to ensure that social services are delivered effectively and comprehensively. With small area population modelling, social infrastructure can be planned for new developments, regeneration and rationalising the efficient use of available resources.

The Office of National Statistics (ONS) provides population data which are important information sources for local policy-making at small area level (Brown and Gardiner, 2004). However ONS population data retain a degree of uncertainty and sometimes do not seem to match local events (ONS, 2003; Statistics Commission, 2003; Bates 2006). For some districts, the uncertainties in ONS population data could have arisen from sampling errors, distribution of national adjustments, and undiscovered fieldwork failures (Simpson, 2007). In addition, ONS data often have a time lag between the data collection and distribution/publication. At the time of writing, the ONS has just released lower super output area (LSOA) mid-year population estimates (MYE) for 2005 and revised all estimates for previous years (but too late for inclusion in the analyses presented here).

Scenario-based modelling has been widely used in population estimation and projection, although there are limitations for this approach (Booth, 2006). One single administrative data set may lack coverage, quality and content in order to be used directly (Bates, 2004). However, a broad range of administrative data sets and other data sources can be combined to overcome such weakness (Judson, 2007). The best way to integrate a range of administrative data sets is using small area geographies. Local administrative data sets tend to be more frequently updated. Multiple data sources can thus offer timely information for population modelling. Furthermore, with spatial analysis and statistical techniques, scenario-based population models can be constructed for small areas to form the basis for local social infrastructure planning (Alvarez and Mossay, 2006; Griffith and Wong, 2007; Oshungade, 1986).

3. Scenario-based population modelling for Thames Gateway London Boroughs

The study aims to develop demographic estimation/projection in support of social infrastructure planning for the Thames Gateway London Boroughs (Barking and Dagenham, Bexley, Hackney, Havering, Greenwich, Lewisham, Newham, Redbridge, Tower Hamlets, Waltham Forest). The geodemographic pattern in this area is complex and dynamic due to the inflow/outflow of (im)migrants, changes in household composition and residential density, and the diversity of local communities.

The ONS population data show uncertainty across the study area in estimation, projection and even in the 2001 Census figures. Figure 1 is the change between ONS 2001 and 2004 MYE which illustrates some stark contrasts between adjacent LSOA. Figure 2 shows the ONS imputation rate for the 2001 Census. Imputation rates are particularly high in the inner East London Boroughs. In Figure 3, the change of ONS MYE has been compared with the changes of dwelling stock counts and child benefit counts for the period 2001 and 2004. There are clear differences between ONS estimation and local scenarios. Figure 4 illustrates the change in the rate of school population growth from 2001 to 2006. There is a noticeably increase in the rates for the inner boroughs (except Hackney) after 2004 probably in response to immigration inflows after Eastern European countries joined the EU in 2004.

In the proposed scenario-based small area population modelling, the basic geographic unit is LSOA and the baseline is between 2001 and 2007. The datasets are from multiple sources which include a wide range of administrative datasets and other relevant datasets including electricity and water demand data. The structure of proposed modelling is shown in Figure 5. In this structure, the raw datasets of local scenarios will be firstly cleaned and then checked with each other in order to control the data quality. After that, the raw datasets are aggregated or disaggregated as necessary to the same small area geography. Latent variables that reflect the underlying true population are subsequently inferred. Comparing these latent variables with ONS population estimates, the reference indicators are created for each LSOA to show the degree of difference between the ONS estimates and local trends evidenced in administrative data sets. Such reference indicators will inform local decision makers and planners in their use of the ONS estimates. The latent variables are further modelled statistically (e.g. regression) and are used in a projection model that incorporates neighbourhood spatial dependencies. The scenario-based small area population estimation and their projection can thus be achieved.

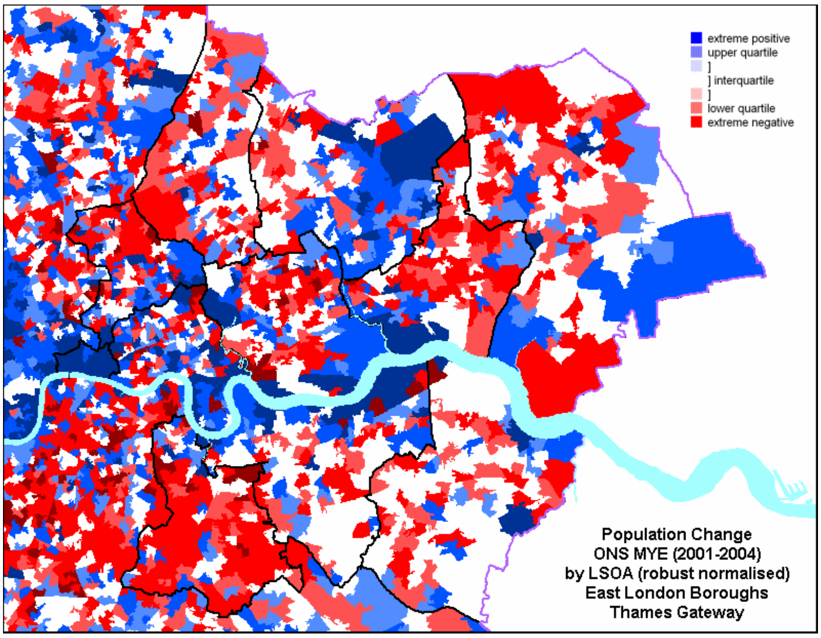


Figure1. Population change of ONS MYE (2001-2004) for Thames Gateway London Boroughs
(data and boundaries Crown copyright)

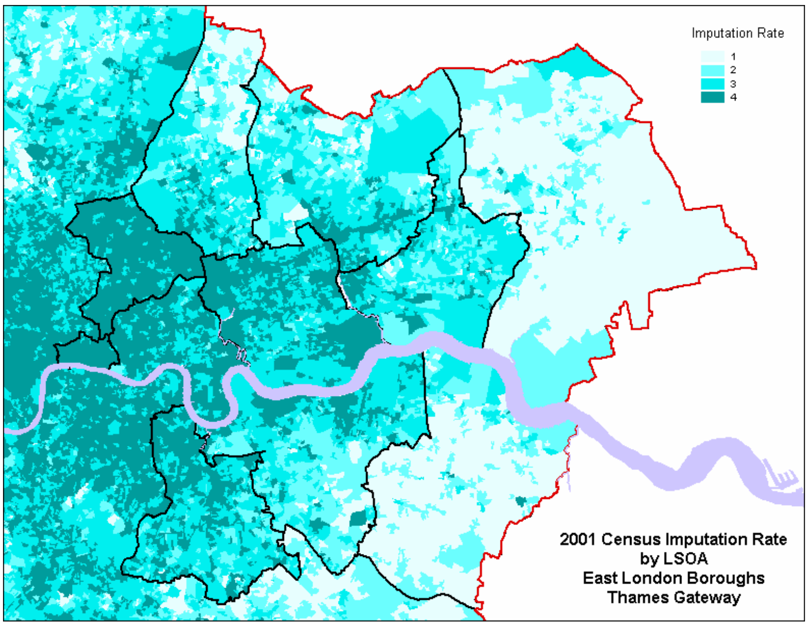
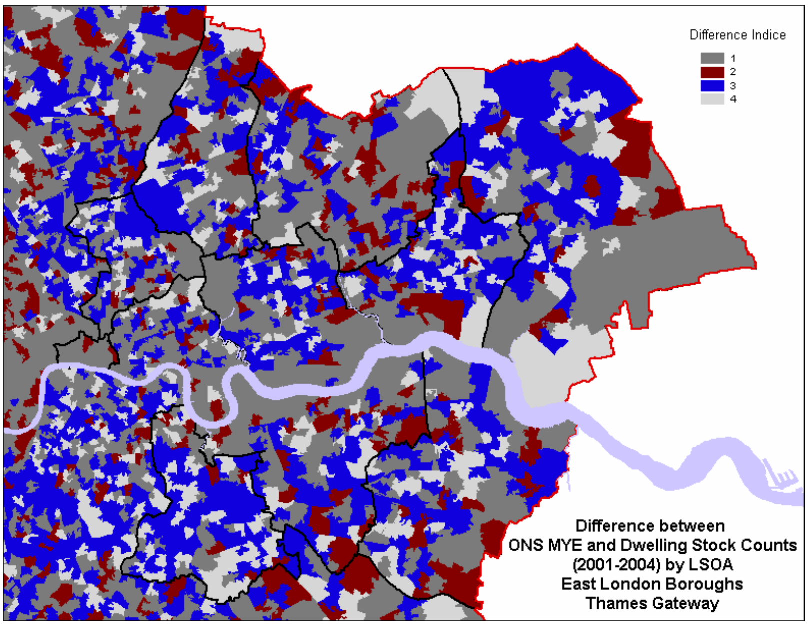
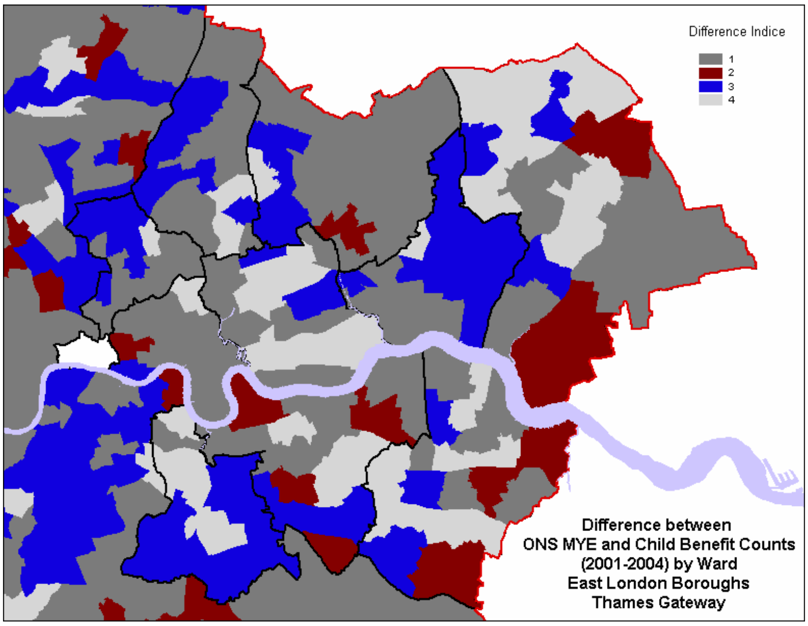


Figure2. 2001 Census imputation rate for Thames Gateway London Boroughs
(data and boundaries Crown copyright)
(ONS Imputation Rate:
1 - Less than 5%,
2 - 5% and less than 10%,
3 - 10% and less than 20%,

4 - 20% and over.)



(a) Difference between the changes of ONS MYE and dwelling counts by LSOA (2001-2004)



(b) Difference between the changes of ONS MYE and child benefit counts (2001-2004)

Figure3. Differences between ONS MYE and local scenarios (2001-2004) for Thames Gateway London Boroughs by Ward (data and boundaries Crown copyright)

(Difference index:
1 – ONS population increasing / local scenario variable increasing,
2 - ONS population increasing / local scenario variable decreasing,
3 - ONS population decreasing / local scenario variable increasing,
4 - ONS population decreasing / local scenario variable decreasing.)

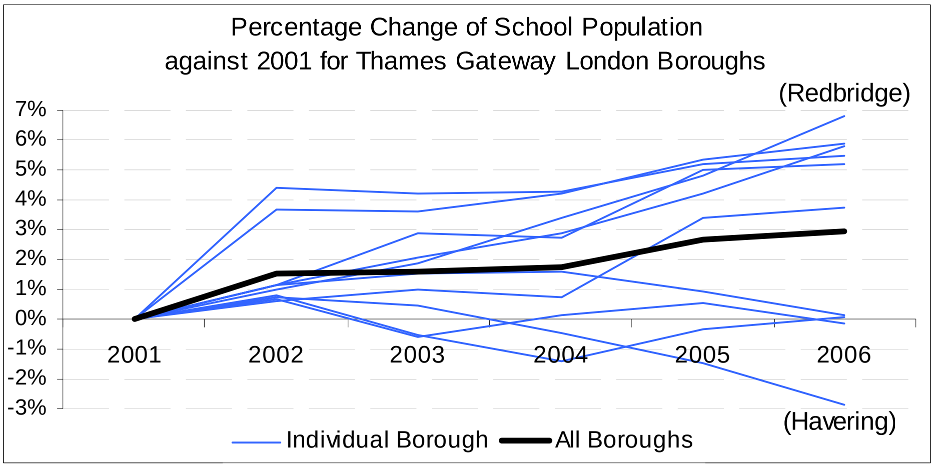


Figure 4. Percentage change of school population for Thames Gateway London Boroughs (data Crown copyright)

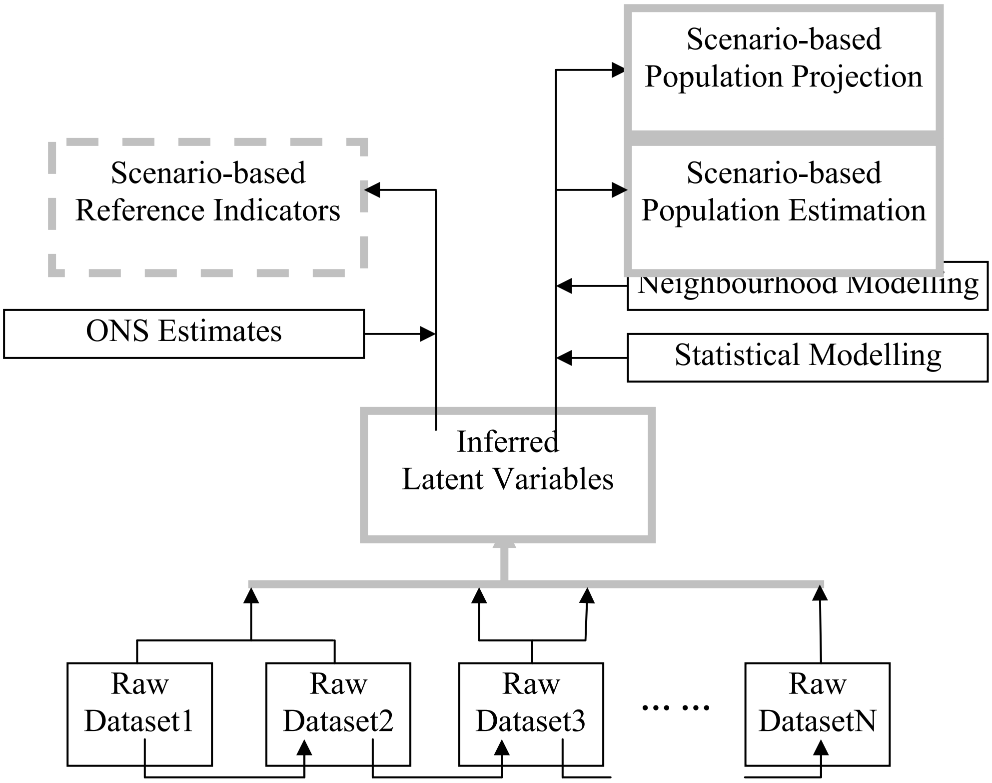


Figure 5. The structure of scenario-based small area population modelling

4. Conclusion

Population estimation and projection are crucial for local social infrastructure planning in support of sustainable communities. However, the ONS data exhibit a degree of uncertainty and sometimes do not match the locally evidenced events. The proposed scenario-based small area population modelling aims to offer an effective solution. The proposed modelling uses multiple data sources which include a wide range of administrative data sets and other relevant data sets integrated using small-area geography.

5. Acknowledgements

The authors would like to acknowledge HEFCE and the DTI's Office of Science and Innovation funding of this study through the UrbanBuzz programme.

References

Alvarez J and Mossay P (2006) Estimation of a continuous spatio-temporal population model *Journal of Geographical Systems* **8** pp307-316

Bates A (2004) Small area population estimates project: data quality of administrative datasets *Population Trends* **116** pp11-17

Bates A (2006) Methodology Used for producing ONS's Small Area Population Estimates *Population Trends* **125** pp30-36

Booth H (2006) Demographic forecasting: 1980 to 2005 in review *International Journal of Forecasting* **22** pp547-581

Brown M and Gardiner C (2004) Information policy-making at sub-local authority spatial levels: Using small area micro-data from the 2001 census *Local Government Studies* **30**(1) pp74-87

EDAW and Brittan B (2006) *London Borough Barking & Dagenham Social Infrastructure Needs Assessment, Final Report*.

Griffith DA and Wong DW (2007) Modeling population density across major US cities: a polycentric spatial regression approach *Journal of Geographical Systems* **9** pp53-75

Judson DH (2007) Information integration for constructing social statistics: history, theory and ideas towards a research programme *Journal of Royal Statistical Society A* **170(Part2)** pp483-501

ONS (2003). Census 2001 review and Evaluation: Edit and imputation report. Census Customer Service, Office for National Statistics, Titchfield, Fareham, Hants. http://www.statistics.gov.uk/census2001/pdfs/editimputevr.pdf

Oshungade IO (1986) Use of percentage change in small area statistics *The Statistician* **35** pp531-545

Simpson L (2007) Fixing the population: from census to population estimate *Environment and Planning A* **39** pp1045-1057

Statistics Commission (2003). *The 2001 Census in Westminster: Interim Report*. Statistics Commission, London.

Biography

Dr Yang Li is a Research Fellow in the Centre for Geo-Information Studies, University of East London, UK. His research interests are in Geodemographics, GIS and environmental simulation modelling, spatial data analysis, spatial data quality, agent-based modelling and location-based services.

Professor Allan Brimicombe is the Head of the Centre for Geo-Information Studies, University of East London, UK. His research interests are in spatial data quality, GIS and environmental simulation modelling, spatial data analysis and location-based services. Professor Brimicombe is a Chartered Geographer.

How to reach the ‘hard-to-reach’: the development of Participatory GIS for inclusive urban design in UK cities

Steve Cinderby[1]

[1]Stockholm Environment Institute, University of York, York, UK
Tel. (+44 1904 432994) Fax (+44 1904 432897)
Email (steve.cinderby@sei.se), web (www.sei.se)

KEYWORDS: Participatory GIS, Social Inclusion, Sustainable Urban Design

1. Introduction

Urban design is expected to play a critical role in implementing the UK Government’s urban policy agenda, in which the sustainable development of cities is seen as a key generator of national prosperity, as well as a more inclusive and equitable society. As global competition intensifies, a network of accessible, safe and attractive public spaces and walking routes oriented to leisure and tourism becomes an increasingly important feature of the ‘liveable’ city. In response to this economic imperative, there has been considerable investment in improvements to the public realm.

In post-industrial cities that are being remodelled and re-imaged to accommodate visitors there is, however, a very real possibility that the new ‘infrastructure of play’ will create isolated enclaves of affluence. These may give physical expression to urban inequalities and do little to promote social cohesion. Local Authorities and Regional Development Agencies are consciously trying to reconcile the desire to create urban environments that are attractive to high-spending consumers and public policies that prioritise social inclusion and equity.

This paper will present the methods and findings from two recent Engineering and Physical Science Research Council (EPSRC) funded projects: Inclusive and Sustainable Infrastructure for Tourism and Urban Regeneration (InSITU) (www.insitu.org.uk); and Design and Implementation Support Tools for Integrated Local LAnd use, Transport and the Environment (Distillate). The projects developed participatory GIS techniques (Cinderby, 2007; Cinderby and Forrester, 2005; Cinderby, 1999) to rapidly assess local concerns, knowledge and design ideas into the urban development process. This paper will describe these techniques, their evolution and application in three UK cities. In particular the paper will look at the application of PGIS to include the voices of so-called ‘hard-to-reach’ groups in the policy and design process.

2. The case studies

The case studies presented in this paper include: the development of a health walk in inner-city Salford; public perceptions of the design and use of streets and squares in York city-centre; and, the development of transport options and infrastructure for a suburb of Blackpool. These case studies will be used to illustrate the evolution of the PGIS techniques. The outcomes of the approach and their use (and potential application) by policy makers will also be presented.

3. Engagement of ‘hard-to-reach’ groups

Bickerstaff and Walker (2005) have highlighted the high priority given by the Government to the need to foster 'civic engagement'. Especially since 2000, there has been a considerable expression of concern to respond appropriately to 'declining public participation in political processes' and 'growing public distrust of authority and expertise (c.f. House of Lords 2000; House of Commons 2001; IPPR 2004).

On the ground the local authorities and organisations involved in the case studies presented here identified engaging with so called ‘hard-to-reach’ groups as a key area for the field work to focus upon. Defining these ‘hard-to-reach’ groups is obviously problematic, contentious and possibly divisive, however, across the local authorities they were listed as including:

- People from Black Minority Ethnic groups
- Asylum seekers
- People with disabilities
- Young people
- Older people
- People living in areas of deprivation or on a low income

Local authorities have experienced difficulties in engaging with these groups. Problems that have been identified that may be exacerbating the lack of engagement from these groups include language barriers, cultural differences, time, and ability to attend A particular remit of the research presented here was to investigate whether the use of PGIS would encourage greater participation from these target groups.

3. Rapid Appraisal – Participatory GIS (RAP-GIS)

In order to facilitate the engagement with and participation of the hard-to-reach groups a rapid appraisal form of participatory GIS was developed over the three case studies. PGIS was thought to be particularly appropriate as a way of overcoming some of the problems identified by local authorities when engaging with these communities.

Firstly, the visual nature of participatory mapping removes the barriers of literacy and to an extent language (although potentially introducing a new barrier for visually impaired groups). One of the benefits of PGIS is that the maps become the focus of participation. This removes the barriers present in public meetings where often the most vocal or confident people can dominate discussions (unless carefully facilitated (Cinderby and Potts, 2007)). Large-scale maps and detailed air photographs of urban areas do not require high degrees of literacy to interpret. Evidence from other case studies carried out during InSitu indicated that even young children (approximately six years old) can locate themselves on a map if orientated, guided and encouraged through the process by facilitators.

Secondly, the option of taking the mapping to the participants through the use of on-street events meant that people did not have to commit to attending a public meeting or event. The use of on-street PGIS mapping activities allowed people who would not (or could not) consider attending a public meeting to make a contribution whilst undertaking their everyday activities. In Salford on-street events were held at health centres, alongside a parade of shops and at a community event. This approach allowed pensioners, children, teenagers and young adults from a low income community to make their local knowledge and preferences for the proposed walking route known to the InSITU and Groundwork (Manchester, Salford and Trafford) team. In the second case study the on-street approach was developed further to include a structured set of questions to ensure consistent engagement with the various individuals who participated. The on-street approach encouraged a similar cross section of ages to participate when it was trialled in York city centre. In Blackpool this approach was taken a stage further with a combination of structured questions and individual mapping by each participant. This development allowed the different viewpoints and knowledge of young people, the elderly, men and women to be investigated individually providing a demographic framing to the participatory GIS data.

4. Advantages and drawbacks of RAP-GIS

The on-street approach had a number of advantages for encouraging the participation of particular age ranges and groups, including those who often fail to participate in conventional engagement techniques. Firstly, people did not have to make any special arrangements (childcare, transport to the venue, etc.) to participate in these events, as would be the case with a conventional meeting or special consultation event. Secondly, the time commitment for participating was less than fifteen minutes (and in many cases only two or three). This meant it was easy to fit around everyday activities. Thirdly, the one-to-one conversations between participants and facilitators meant that people did not have to justify their comments or knowledge to their peers, as would be the case at a public event. This was potentially less intimidating for those participants without the confidence or language skills to communicate effectively in group discussions. This perceived advantage of on-street individual or small group (two-teenagers for example) participation is also a potential disadvantage of the approach as it prevents any broader discussion or snowballing of conversations occurring between participants. Also unless the facilitators have some local knowledge of the area and can vet the comments, participants could potentially communicate falsehoods that would not pass muster if presented to other members of the local community.

The novel approaches to participatory mapping (stickers, flags etc.) trialled during the case studies also encouraged participation. The use of such resources and sensitive facilitation removed the necessity to draw or write thereby allowing less literate groups (children or adults) to communicate their knowledge effectively. In both the York and Salford events, this approach also appeared to encourage participation, as the flags and colourful maps drew attention. The facilitators were also available to help less able participants to add their comments and information to the maps.

The drawback of these novel-mapping approaches was that information was primarily recorded as points – even if the information given represented an area. In order to overcome this, the Blackpool case study returned to the use of full participatory mapping, but developed individual maps for all participants.

The digital nature of the PGIS database means that information obtained from specific types of participants can be extracted for assessment in isolation or used to compare specific group's perceptions with those of another. For example, during the York spaces fieldwork as well as recording the comments made about specific locations, demographic information about respondents was also recorded. In this instance the age range and gender were noted on each comment. This allowed information obtained from under 18 year olds to be extracted from the wider responses obtained from York residents. In this example the striking result from the data was how similar the opinions of the under 18s who participated was to that reported by the wider population.

5. Conclusion

Rapid Appraisal Participatory GIS shows promise as a technique for engaging with local communities in a way that encourages participation from a wider cross-section of society than may be the case for more conventional consultation exercises. The digital nature of the resulting database means the particular viewpoints of communities of concern can be assessed individually to ensure that specific issues and concerns are addressed in urban development schemes. This approach holds the potential to ensure effective and inclusive engagement with local communities over developments in their neighbourhoods.

6. Acknowledgements

The author would like to thank the EPSRC for funding both the InSitu (EP/D011671/1) and Distillate projects. The author would like to acknowledge the contribution to the findings in this paper of the researchers on the InSitu and Distillate projects who collaborated in the development of the Rapid PGIS approach. In particular, Steve Shaw, PI on the InSitu project (London Met), John Forrester (SEI), Caroline Snell (SEI), Anne Owen (SEI), Peter Schofield (University of Salford), Peter Jones (UCL), Gemma Grimes (London Met), Menna Jones (London Met), Dr Lesley Mackay (University of Salford). I would like to thank the project partners: Groundwork Manchester, Salford and Trafford; Blackpool City Council; and City of York Council. Finally, thanks must be given to the individuals from York, Salford and Blackpool who participated in the mapping.

References

Bickerstaff, K. and Walker, G. (2005) 'Shared Visions, Unholy Alliances: Power, Governance and Deliberative Processes in Local Transport Planning', Urban Studies, Volume 42, Number 12, pp. 2123-2144.

Cinderby, S. (2007) "How Communities Can Use Geographical Information Systems" in Clay, C., Madden, M. and Potts, L. (eds) Towards Understanding Community: People and Places. Palgrave.

Cinderby, S. and Potts, L. (2007) "Suspicious cartographers: Some realities of research into stakeholder understanding of the causes and possible prevention of breast cancer", Science and Public Policy, Vol. 34, No. 4.

Cinderby, S. (1999) "Geographic Information Systems for Participation: The Future of Environmental GIS?" International Journal of Environment and Pollution 11(3): 304-315.

Cinderby, S and Forrester, J. (2005) "Facilitating the Local Governance of Air Pollution Using GIS for Participation", Applied Geography, 2, 143-158.

House of Commons Public Administration Select Committee (2001) Sixth Report: Public Participation: Issues and Innovations, Session 2000-1, HC 373, London: The Stationery Office.

House of Lords Select Committee on Science and Technology (2000) Third Report: Science and Society, Session 1999-2000, HL38, London: The Stationery Office.

Institute for Public Policy Research (2004) Lonely Citizens: Report of the Working Party on Active Citizenship, London: IPPR.

Biography

Steve Cinderby is Deputy Director of the Stockholm Environment Institutes York Centre. He has been involved in the development of participatory GIS techniques for the past ten years in a wide variety of locations and contexts to facilitate local community's engagement in environmental and development decision-making.

User-centred approach for developing PPGIS.

Khitam Shraim

School of Environment & Development,
University of Manchester, England, M1 3 9PL
khitam.shraim@postgrad.manchester.ac.uk
kh shraim@hotmail.com

1. Introduction

Web-based mapping interactive tools such as public participation geographic information systems (PPGIS) have substantial promise to enhance public participation in the planning process and to serve as a communication bridge between stakeholders(Seiber, 2006). The effectiveness of such tools, however, is constrained by fundamental issues such as usability and acceptability. Thus, in order to design these 'means' to successfully achieve the desired 'ends', the development of PPGIS applications as social and technical objects should be based on a user-centred approach that focuses on "intended users and their environment and how technology can best be deployed and designed to support them" (Preece et al., 1994, p 7).

Users of web-based GIS applications notably encounter more difficulties than users of other online systems such as e-commerce (Warren and Bonaguro, 2003; Steinmann et al., 2004). It is believed that these difficulties are due to the inherent complexity of web GIS applications in terms of the specialized functionality that supports online GIS, the amount of content available and the skills required to interpret this content (ibid). This complexity can present a major impediment to public use of these applications. Generally, the call is to develop "a comprehensive user-centered design approach to geovisualization usability" (Maceachren and Kraak, 2001).

This papers aims at informing the importance of adopting user-centred approach for identifying users' needs and requirements in developing a usable and acceptable PPGIS application. So, utilizing technology to develop PPGIS is a social construct (Feenberg, 1999) that integrate the participants interest into the design process. A case study is carried out, where soft system methodology (S SM), usability engineering (UE) and acceptance of the new technology models (TA) are integrated in developing a prototype PPGIS for empowering public participation in educational planning in Palestine to meet the challenge of the Millennium Developments Goals. The proposed methodology provides opportunity for researchers to take an interdisciplinary perspective in developing PPGIS and merge different methods and techniques.

2. Methodology Framework

This research explores the possibility of using Web Information Systems Development Methodology (WISDM) developed for e-business(Vidgen et al., 2002) for generating a framework to develop the PGIS tool. This framework (Figure 1) consists of two-dimension socio-technical aspects and concentrates on organizational analysis, technical design, and HCI design. In these successive stages, issues concerning users and their environment are central to the design process (Preece et al,. 1994). The organizational analysis is based on human activity, with a seven-stage of SSM developed by Checkland (1999) with the intention of stimulating a debate about defining changes that are ‘systematically desirable’ and ‘culturally feasible’. Results from organizational analysis identify both the technical and social requirements of the new system (Avison et al., 2003). The technical design is concerned with how to integrate the stakeholders’ needs and requirements into the structure and functionality of the system (Preece et al., 1994), while the HCI design focuses on the usability and acceptability of the new application.

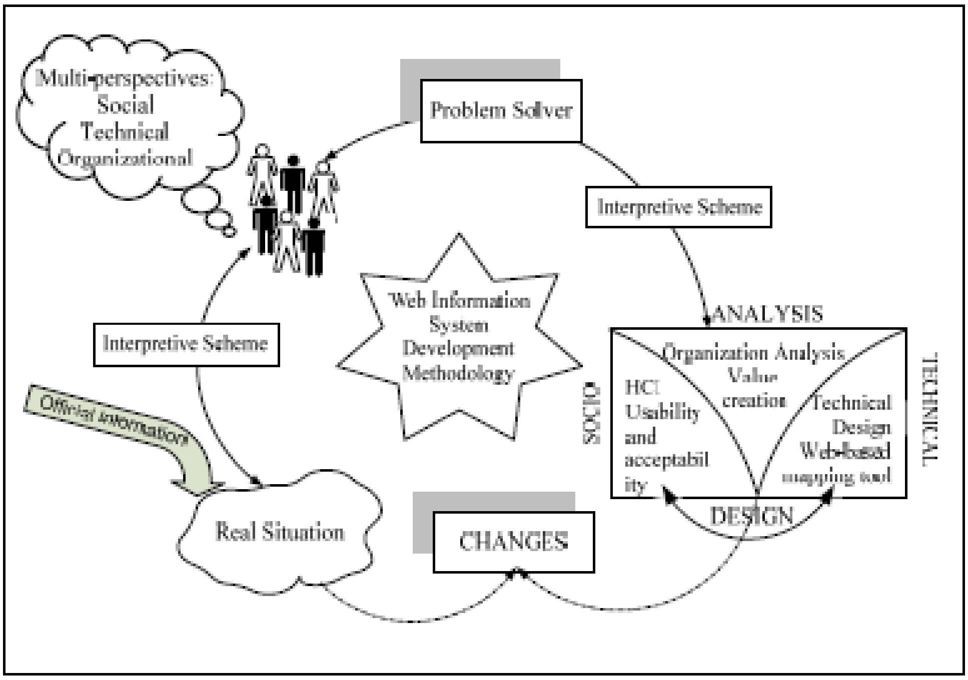


Figure 1: WISDM for EP-PPGIS (adapted from Vidgen et al., 2002)

3. Methods

Phase 1: SSM

In developing a website like PPGIS, considerable attention are given to the issue of putting people first and capture a holistic view from them. This is starting with organization analysis or human activity with the seven-stage techniques of SSM (Figure 2).

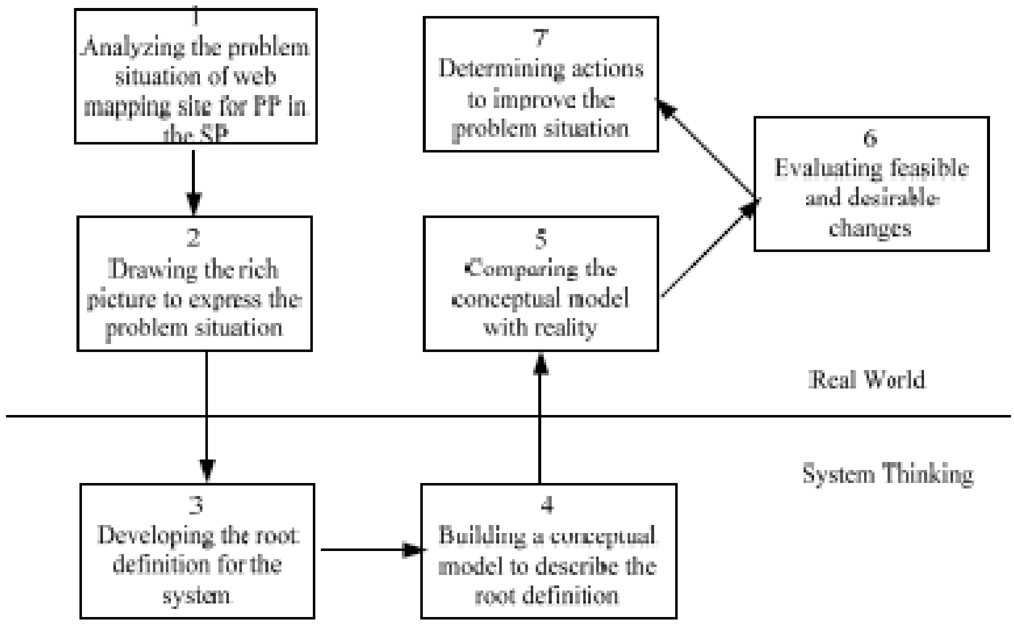


Figure 2: SSM model (adapted from Checkland and Scholes, 1990)

The techniques used to explore variations in the perspectives of stakeholders were discussion groups and in-depth semi-structured interviews with key players responsible for and interested in developing the education system which were conducted between 15th June, 2006 and 15th September 2006 in Palestine and Jordan.

Phase 2: Technical Design

Like any system, PPGIS designers are required to know the intended users and involve them early and continuously (Andrienko et al., 2002). Throughout the system development lifecycle, frequent discussions with intended users were carried out to ensure that the interface of the website designed meets their needs and is usable.

Phase3: HCI design

The recent evolution of user-centred design methods means that 'ISO-backed' usability standards are considered a crucial part of the software engineering lifecycle which must be extended to user acceptance of new IT applications (Dillon, 2000). Therefore, the process of designing the new system should take account of HCI factors and be extended to include both the usability and acceptability of the PPGIS website.

Usability framework

The main aim of usability testing is to identify difficulties and shortcomings potentially affecting an end user in using the PPGIS tool. Two methods were used to evaluate the usability of PPGIS version 1; usability testing with 10 participants and discussion with 7 individuals. The criteria used were the so-called 4 Es: Efficacy, Efficiency, Effectiveness and Elegance.

Acceptability framework

The main aim was to examine how users come to accept and use a new technology such as PPGIS. In this research, five determinants of adoption of PPGIS version 2 were chosen from different technology acceptance models (Davis, 1989; Venkatesh et al. 2003). These determinates are information quality, functionality and interactivity quality, perceived ease of use map-based interface, perceived usefulness and social influence. A 30-question online questionnaire was adapted from these models and tailored to be more specific to the PPGIS application based on literature related to the PPGIS context.

4. Results

Phase 1: SSM process

Stage 1: The starting point is to identify the 'problem situation' to be addressed by PPGIS. In the context of this research, the top-down concern of the problem owner, the Ministry of Education (MEHE), is that there is a problem in the SP process due to the ongoing crisis in Palestine and thus a need to work with local communities to improve the planning process. On the other hand, the push of bottom-up technologies from the Internet and web GIS provides an opportunity to empower the public to participate in the spatial planning process (Carig et al., 2002; Sieber, 2006).

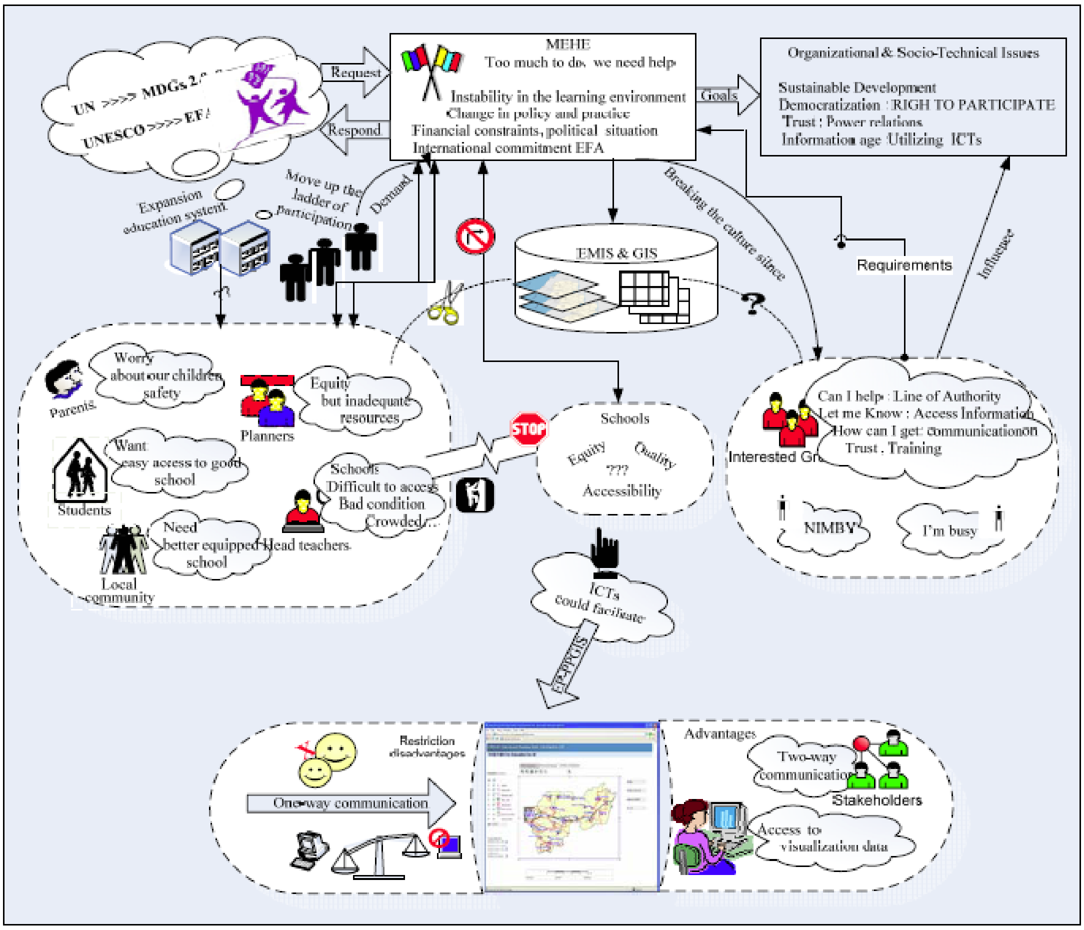


Figure 3: The RP of the SP in Palestine.

Stage 3 is the development of a Root Definition (RD), whose aim is to express some purposeful human activity system. The RD does not exist in the real world, but is concerned with the system thinking, which is more or less relevant to the problem situation. The transformation process (change), which is the core of the RD, is based on the X-Y-Z structure: *Do X by Y in order to achieve Z,* which tell us what the system will do (X), how (Y) and why it needs to be done (Z) (Checkland, ibid). Generally, various participants' perspectives assembled into an overall RD for participatory approach as follows:
The Ministry of Education and Higher Education in Palestine is responsible for the SP, looking to improve PP in the SP (X) by the introduction of web mapping tool, PPGIS, (Y), taking into consideration the technical and institutional aspects, and the willingness of the public to participate, in order to improve spatial planning issues (accessibility, quality equity) and then to achieve the EFA initiative (Z). The RD is formulated as follows **CATWOE:**

Customer	Stakeholders, interested groups, all public.
Actor	Decision makers, educational planners and developers.
Transformation	Need to improve PP in the SP------ public participation improved via PPGIS website.
Worldview	Believe that utilizing ICT could empower the public to involve in the SP and then to improve the accessibility, equity and quality of school in order to achieve the EFA initiative.
Owner	Ministry of Education and Higher Education, Palestine.
Environment constraints	The willingness of the public to participate and technical and institutional aspects such as physical accessibility to the Internet, trust, digital divide.

Table1: CATWOE for PPGIS in SP in Palestine.

Stage 4:

Figure 4 represents the overall Conceptual model (CM) which includes the minimum number of necessary activities expressed through verbs to meet the requirements of the RDs described in the above section.

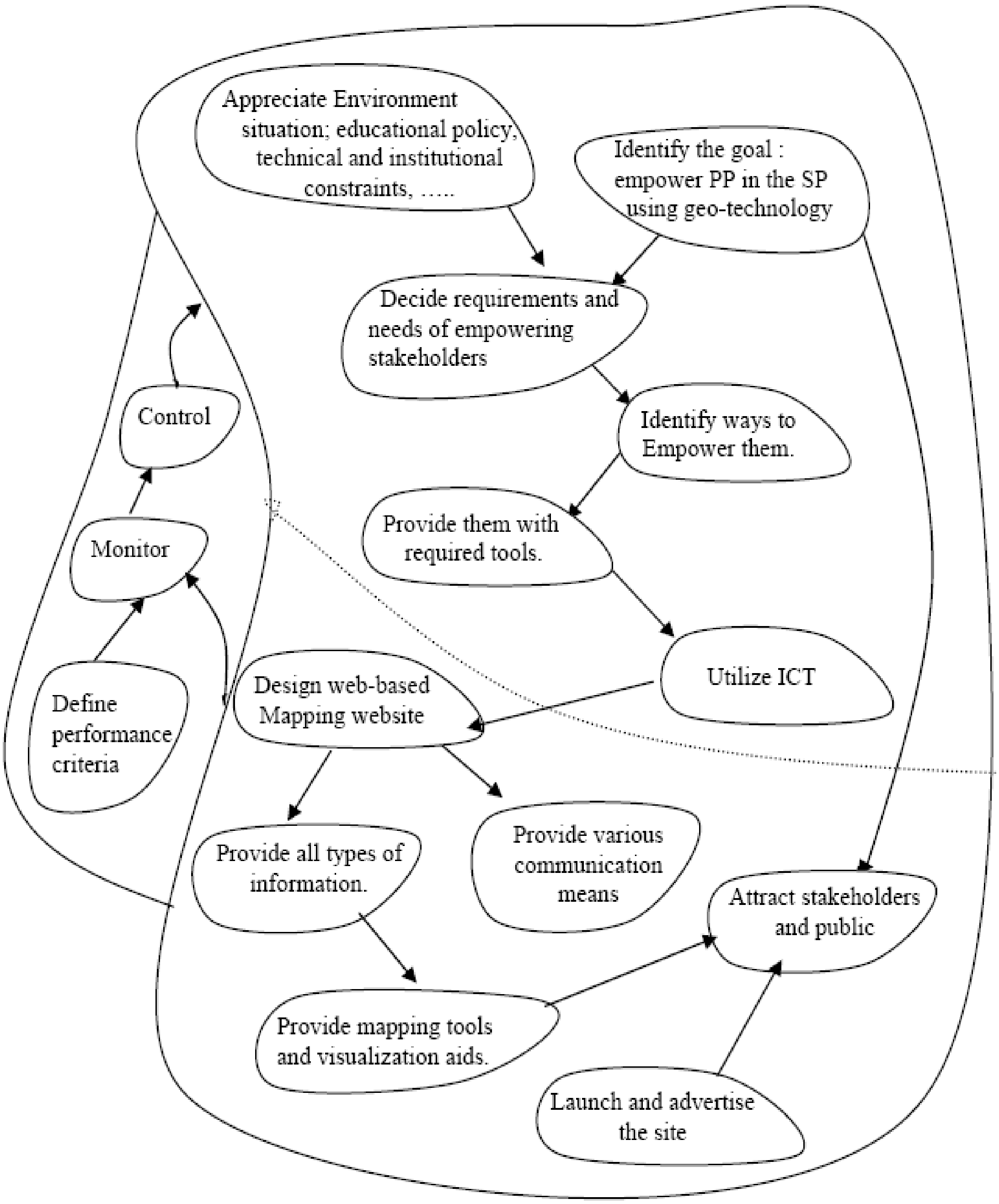


Figure 4: CM of PPGIS

Stage 5 compare the CM developed at stage four with the problem situation expressed at stage two. Given that the PPGIS tool is new to Palestine, the general comparison was based on 'what's and how's questions (Couprei, et al., 1997). 'What is the current situation of public participation in educational planning in Palestine' and 'what form should it take' were debated by decision makers at the MEHE who came to accept that ICT-based participatory approach within educational planning should beregarded as imperative and receive more attention. Changes of attitude at the national level pave the way for necessary structural and procedural modifications in order to facilitate a high degree of public involvement (Flood and Jackson, 1991).

Stage 6: During the semi-structured interviews and discussion group, the participants highlighted the essential requirements that have to be taken into consideration to achieve the structural and procedural changes desired.

Stage 7: The final stage is to make recommendations to address the problem situation, not actually to implement specific solutions (Checkland, ibid). The MEHE has already recommended setting up the required procedures such as improving technological infrastructure, raising awareness, advertising the websites and controlling the process and building trust.

Phase 2: Technical Design

The PPGIS website Figure 5 is currently being developed as a pilot application with a technical platform based on Open Source Software (OSS) using Open Geospatial Consortium (OGC) compliant Web Map Server (WMS) technology, mySQL for the database and PHP scripting language.

Figure 5 The EPPGIS website interface

Phase 3 : HCI Design

Usability

The result of usability testing indicates that the website partially met its objectives by providing sufficient information. Although, it provides several communication channels that meet the needs of different groups, yet does not take into consideration the geo-referencing and restrictions of users contributions seriously. Participants were generally satisfied with the map-based interface and found that visual representation is more effective to attract them. Feedback was used immediately to improve the functionality and interface design of the prototype.

Acceptability

Since PPGIS application has been introduced recently and individuals have no or limited experience in the early stage, several workshops was conducted to the participants prior filling the questionnaire. The survey was conducted from 5th of November 2007 through 10th January 2008. A total of 163 different cases responded to the survey. Seven of the received responses were dropped because they were incomplete. This was an extremely complex task considering the political situation in Palestine.

The regression analysis revealed that functionality perceived usefulness and social influence would be influential in the Palestinian decision to adopt the PPGIS in the future while perceived information and perceived ease of use interface would not.

The open comments of respondents were also analyzed alongside the quantitative analysis. The assessment confirmed that there were a number of issues that would need to be addressed, the most significant being related to institutional (organizational) and procedural (logistics) considerations.

5. Conclusion

In this paper three methods; SSM, UE and AT are integrated to identify needs and requirements for and responses to ICT–based development activities for effective design and to determine a set of variables that have a significant influence on Palestinian intention to adopt new technology such as PPGIS. This outlined framework to support user-centred design approach for PPGIS applications and offered researchers and practitioners a valuable framework upon which a more complete understanding of user needs and response to new technology can be built.

6. References

Checkland, P. (1999). Soft *Systems Methodology. a 30-years retrospective.* Chichester, UK : Wiley.

Craig, W., Harris, T. and Weiner, D. (2002). *Community Participation and Geographic Information Systems.* London: Taylor & Francis.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly,* 13(3), 3 19-340.

Feenberg, A. (1999) *Questioning Technology,* Routledge, New York.

Haklay, M. and Tobon, C. (2003). Usability evaluation and PPGIS : towards a user-centred design approach, *International Journal of Geographical Information Science,* 17(6), 577-592.

Peng, Z.R. (2001). Internet GIS for Public Participation, *Environment and Planning B. Planning and Design,* 28(6), 889 - 905.

Preece, J. Rogers, Y. Sharp, H. Benyon, D. Holland, S. and Carey, T. (1994). *Human Computer Interaction.* Harlow : Addison-Wesley.

Rattray, N. (2006). A User-Centred Model for Community-based Web-GIS, *URISA Journal,* 18 (2), 25-34.

Seiber, R.(2006). Public Participation Geographic Information Systems: A Literature Review and Framework. *Annals of the Association of American Geographers,* 96(3), 491-507.

Steinmann, R., et al. (2004). Analysis of Online Public Participatory GIS Application Applications With respect To the Differences Between The US and Europe. http://www.iapad.org/publications/ppgis/udms%20paper 2004 last.pdf

Vidgen, R., Avison, D., Wood, B.and Wood-Harper, T. (2002). *Developing Web Information Systems.* Elsevier.

Warren, D. and J. Bonaguro. (2003). "Usability Testing of Community Data and Mapping Systems. http://www.gnocdc.org/usability/usabilitytesting.html

Participatory Strategic Planning Using a Virtual Reality Environment

David Miller, Paula Horne, Jane Morrice, Jonathon Ball and Pernette Messager

Macaulay Institute, Craigiebuckler, Aberdeen, AB15 8QH
Tel. (+44) (0)1224 498200 Fax (+44) (0) 1224 311556
d.miller@macaulay.ac.uk

KEYWORDS: Virtual Reality, Public Participation, Awareness Raising, Planning

1. Introduction

The Brundtland Commission (WCED, 1987) identified the importance of information and participation in issues relating to the management and planning of change. Subsequent policies on sustainable development (Scottish Executive, 2006; United Nations, 2005) emphasise needs for education and capacity building to increase levels of public and professional engagement in environmental decision-making. This paper describes the use of virtual reality tools, in a mobile theatre, to facilitate information dissemination and consultation with 'hard to reach' public audiences, such as children, on issues of landscape change.

2. Background

2.1 Public Participation

Arnstein (1969) defined a multi-level participatory classification within three main levels: (i) educational (primarily information dissemination), (ii) active participation (public opinions sought and considered in expert decision-making), (iii) citizen power (direct influence on the decision-making process). Sheppard (2005) argues that the impact of participation is linked to the level of involvement, with greater impact from more citizen power, which is one aim of public policy (Scottish Executive, 2007).

Visually representing the real world, and potential alterations, is essential for landscape planners to communicate their thinking to the wider public (Ball, 2008; Bishop, 2007; Lange and Bishop, 2005). The Scottish Executive (2007), in reforming the planning process, also noted that 3D modelling has potential for "engaging communities and assisting planners and Councillors to visualise and assess the visual impact of development proposals". One tool being used to support such engagement is a virtual reality theatre, with associated software tools, for use in public venues.

2.2 Participation in Local Plan Development

The Loch Lomond and The Trossachs National Park Authority is preparing a new local plan which will inform decisions pertaining to land use and development. In spring 2007, the Authority ran consultation events throughout the Park to elicit public aspirations and fears over future land use. Over 900 people attended 26 events, but few younger or disadvantaged people were attracted (The Loch Lomond and the Trossachs National Park Authority, 2007).

With the Park Authority, the Macaulay Institute developed a programme of events, the aims of which were to engage younger people in the planning process, eliciting opinions on preferences for, or concerns over, future land uses. The outputs would contribute to the Park Authority's 'issues report', for use in a subsequent phase of the consultation on plan development.

2. Methods

2.1 Audience

In June 2007, events run at Gartocharn and Killin, in central Scotland, attracted 147 children, plus teachers, Park Rangers, the Park Board and Community Councils. To attract participation, the events were designed to align with the school Geography curriculum, addressing themes of conflict in the countryside, landform development, National Parks, and citizenship.

2.2 Virtual Reality Theatre Facility

A mobile virtual reality theatre (the 'Virtual Landscape Theatre', www.macaulay.ac.uk/planning) was central to these events, in which 3D models of western Scotland and the Park were projected. The theatre comprises a curved screen (8 m x 6 m x 3 m), with three projectors through which rendering of landscape models was distributed. The theatre was designed for use in public venues, such as halls and schools, with audiences of 16 at a time.

2.3 Virtual Reality Models

Two categories of model were used (Table 1), created in ESRI ArcScene, but rendered using packages best suited to model purpose. Vega Prime (MultiGen-Paridigm, 2007), was used for a regional model, for landscape interpretation and feature recognition. An interface developed in Octaga Professional (Octaga, 2007) provided interactive movement of features in models, including hotkeys to: (i) switching between images (*e.g.* 1:50,000 map and aerial images); (ii) introducing new features (*e.g.* houses, wind turbines, trees); (iii) 'drag and drop' features, guided by the audience. Feature icons were colour coded as green (*i.e.* more/good) or red (*i.e.* fewer/bad).

Table 1. Summary of 3D models used in the Virtual Landscape Theatre

Description	**Model(s)**	
	Regional	**Local Models**
Purpose	Land use interpretation, features identification, Park delimitation	Identifying options for future land use and discussion of local issues
Extent	100 km x 100 km	20 km x 30 km
Image source	Landsat Thematic Mapper, June 2001	Aerial photography
Data resolution	25 m x 25 m	2 m x 2 m
Elevation source	Ordnance Survey 1:50 000 DEM	Ordnance Survey 1:10 000 DEM
Distributed rendering software	VegaPrime, with 'clickable' attributes	Octaga/Octaga Professional, with feature interactions

2.4 Event operation

Presentations were tailored to suit age-range and Geography class, with 'ice-breakers' and introduction to the event, Park familiarisation, and tests of feature identification using the regional model. The local models supported detailed consideration of options for future land uses (Figure 1), as set-out below, with an electronic voting system used to record participant answers, and prompt group discussion.

(i) Identify features in the VR model (to familiarise with the area and voting equipment).
(ii) Vote on changes likely to impact on the local landscape, and priorities for the event.
(iii) Select and place icons of features of specific types of change.
(iv) Group voting on choice of change and placement, with reconsideration of choices in (iii) above if the group votes accordingly.
(v) Evaluation and feedback questions.

Figure 1. Overview of south Loch Lomondside, showing a merging of aerial and Ordnance Survey 1:50 000 map imagery during a period of audience voting.

The locations of the features were recorded for conversion into Shapefiles for further reporting and analysis. In addition, the group voting on each feature was recorded, as were images of the three projection screens, for supporting documentary information.

3. Results
3.1 Group Recommendations

Voting from each audience group was recorded individually, and pooled for the two venues, with reporting on factors such as age, organisation and residence in or out of the Park. Broadly, the results suggest:

(i) A preference for limited wind turbine development in and around the Park to cater for local needs.
(ii) Limited expansion of housing supply, and building design, consistent with the character of villages in the Park.
(iii) Disagreement was apparent with respect to provision of shops in the south of the Park, between those living in the Park (reporting a lack of clothes, sports and convenience food shops), and those outside (arguing that sufficient shops exist close-by). In the north-east, a similar lack of shops was reported, plus no dentist or ambulance provision, but too many public houses.

Figure 2 shows the siting of icons of features around Killin during one session. The green trees indicate an extension of woodland sought upslope of the village, and both red and green wind turbines indicates some audience disagreement as to renewable energy provision.

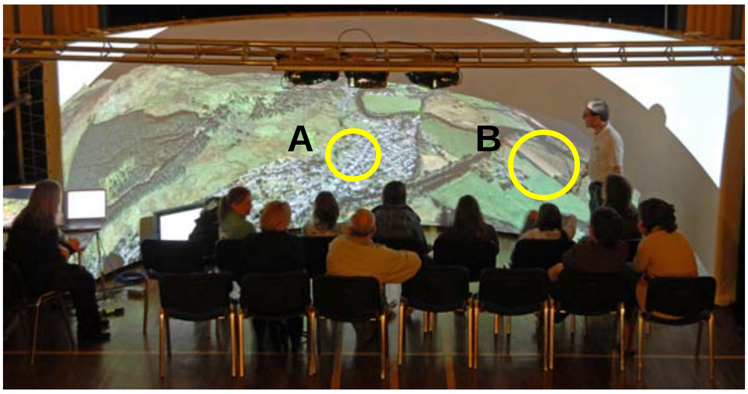


Figure 2. Audience discussion on future changes in land use around Killin. 'A' icons of green trees indicating a preference for further woodlands upslope of village; 'B' red and green wind turbines.

Electronic voting was also used to provide pupils and teacher feedback on the event, such as relevance to work in the Geography class, awareness of selected issues, limitations of outputs, and reactions to technology.

4. Discussion

Designing event structure and materials was possible by linking the event themes to specific items in the primary and secondary school Geography curriculum. Effectiveness and impact of theatre and tools is hard to assess without a directly comparable type of event using different tools. However, feedback from teachers, Community Councils and Park Rangers suggests a high level of effectiveness of the approach, as an extension of the conventional consultations, with some new issues identified.

The nature of audience interaction with the models appears to have been appropriate to satisfy the aims of the participation. Based on voting results from feature recognition (e.g. lochs, islands, mountain peaks, villages, woodlands), the virtual environment provided materials with levels of familiarity suitable for credible suggestions for consideration of existing and new features. Audience surveys suggest that the package (*i.e.* the evidence of recording views, relevant models, the facility and its interactivity) supported material participation, beyond that of information dissemination. The level of influence on final decisions remains to be assessed after completion of the process of plan development.

5. Conclusions

Attendance numbers and reported levels of relevance and interest across ages, suggest this mechanism provided a means of enabling young people to input to the planning process, supporting information dissemination (e.g. land use interpretation), consultation (eliciting opinions), and collaboration (public and private stakeholding bodies). The impact will be assessed through the process of plan development, including follow-up activities by teachers on the issues raised, in the context of the Geography curriculum. The final feedback is being compiled for the Park Authority in the form of information sheets as per their main format of reporting, with further such events planned for public consideration of the issues raised during the wider consultations.

6. Acknowledgements

The research was funded by the Scottish Government Rural Environment and Research Analysis Directorate. Thanks to the Loch Lomond and The Trossachs National Park Authority for their collaborative support, and the population, staff and pupils of participating schools and community groups.

References

Arnstein, S. (1969) A ladder of citizen participation. *Journal of the American Institute of Planners*, **35 (4)**, pp216-224

Ball, J., Capanni, N. and Watt, S. (2008), Virtual Reality for Mutual Understanding in Landscape Planning, *International Journal of Social Sciences*, **2(2)** (http://www.waset.org/ijss/v2/v2-2-12.pdf).

Bishop I.D. (2007), Developments in public participation and collaborative environmental decision-making, In: *GIS for Environmental Decision-Making, Innovations in GIS* (Eds. A Lovett and K Appleton), CRC Press, pp181 – 191.

The Loch Lomond and the Trossachs National Park Authority (2007), http://www.lochlomond trossachs.org/park/default.asp?p=320

Multigen-Paradigm (2007), Vega Prime, http://www.presagis.com/products/multigen_paradigm/ details/vegaprime/

Octaga (2007), Octaga Professional, www.octaga.com

Scottish Executive (2006) Learning for our Future: Scotland's first action plan for the UN Decade of Education for Sustainable Development, Scottish Executive, pp. 35.

Scottish Executive (2007) Planning Advice Note 81: *Community Engagement "Planning with People"*. March 2007. Scottish Executive, pp. 30.

Sheppard, S.R.J. (2005) Participatory decision support for sustainable forest management: a framework for planning with local communities at the landscape level. *Canadian Journal of Forest Research*, **33(7)** pp1515-1526.

United Nations (2005) United Nations Decade of Education for Sustainable Development http://portal.unesco.org/education/

WCED (1987), *Our Common Future*. World Commission on Environment and Development, Report of United Nations Commission.

Biography

David Miller is Head of the Integrated Land Use Systems Group at the Macaulay Institute. He leads research into landscape change, public preferences, spatial modelling and the representation of land use change. He is responsible for developing knowledge transfer strategies for target audiences, including elected representatives and the general public.

Using a touch table to support interactive land use planning

Gustavo Arciniegas, Ron Janssen, Nancy Omtzigt, Corinne Cornelisse

Institute for Environmental Studies, Vrije Universiteit, De Boelelaan 1085 1081 HV Amsterdam
Tel. +31-20-5989 555 Fax +31-20-5989 553
Gustavo.Arciniegas@ivm.vu.nl, http://www.ivm.falw.vu.nl

KEYWORDS: Participatory planning, multi-user, touch table, land use planning

1. Introduction

Dutch fen meadow areas will face major changes in the near future. With current water management ground levels will continue to drop and the peat layer will slowly oxidize. Agriculture will face increasing difficulties to produce in a cost-effective way. The Gouwe Wiericke area is located in the west of the Netherlands and consists predominantly of fen meadow. This type of landscape has high natural, cultural and historical values. Multiple groundwater levels are present and have been managed for centuries to make them suitable for functions such as agriculture, recreation and nature conservation. At present, several issues are critical in Gouwe Wiericke namely, ground subsidence, the physical preservation of the fen meadow landscape, fragmented water management, water quality below the standards of the Water Framework Directive (EU-WFD), and the changing economical position of dairy farming. These issues suggest that in the area multiple stakeholders with different, and often conflicting, objectives are present. Defining both short- and long-term land use policies for the area is thus a complex task. This paper presents progress on a method to interactively develop land use plans that combine and integrate the objectives of the various stakeholders involved.

2. The approach

In recent years, complex decision problems have increasingly been supported by workshop-based decision sessions, often called decision conferences. The same interactive approach (design-assess-evaluate-design) carried out by a team of experts or decision makers is extended to spatial negotiation. For spatial negotiation, plans are assessed in real-time by GIS-based tools and interactive instruments. Results are discussed and fed to design sessions to produce alternative options in real-time. This facilitates appraisal of existing options, and the creative exploration of alternatives. The team set-up ensures participation, but also efficient exploration of the major decision perspectives. Two major objectives are defined for this case study:

- To develop and implement a method for supporting interactive land use planning processes in fen meadow areas
- To integrate and communicate knowledge generated in the process to public and private stakeholders through several workshops, with the help of interactive instruments

Spatial information plays a central role in our approach. Maps are used to communicate and exchange knowledge among stakeholders. This is done through workshops, where interaction between stakeholders is prompted through the use of printed and digital maps. For the case of digital maps, a device called the ‘touch table’ is used. The touch table is an instrument to visualize and handle spatial information in a multi-user setting. It is used in this study as a tool for stakeholder input via a touch-enabled screen. This device allows multiple users to both input their views simultaneously and to draw map features using different backgrounds, keeping a record of which user has drawn what. It is also an objective of this research to measure the added value of such devices for interactive land use planning processes and its complementary role next to that of printed maps.

3. The workshops

Our approach involves six workshops. For Gouwe Wiericke, the sequence of the workshops is indicated in Table 1. The nature of each workshop is defined according to one of the three frames in the framework for map use in policy making (Carton, 2007):

- Analysis — Map use as science; Map as a research model
- Design — Map use as art; Map as a design language
- Negotiation — Map use as politics; Map as a decision agenda

Table 1. Plan of workshops for Gouwe Wiericke

	Workshop	*Frame*
1	Exploratory policy workshop	Design
2	Scientific workshop	Analysis
3	Exploratory stakeholder workshop	Design
4	Scientific workshop	Analysis
5	Stakeholder workshop	Negotiation

The first three workshops were exploratory and focused mostly on the exchange of knowledge. These workshops took place in October, December 2007 and February 2008. The remaining workshops will be more oriented to negotiation and are planned for June, October and December 2008.

3.1 Workshop 1: design

Policy goals of the first workshop were to communicate spatial information, exchange knowledge and explore/develop possible land use alternatives for the area. This workshop was organized for civil servants from province, municipalities involved and the water management authority. We used digital maps presented on the touch table and printed maps. Research goal was to compare how participants use different types of maps to support the various design tasks. Three initial land use alternatives were developed previously for the area, based on land use functions that are related to groundwater levels. Participants were divided into two groups and then asked to provide feedback (by indicating strengths and weaknesses) on the alternatives and make changes if needed (Figure 1). Group 1 was asked to provide feedback on printed maps and Group 2 on digital maps on the touch table.

Figure 1. Participants of the first workshop providing input on the touch table (left) and on printed maps (right)

Participants could use different background maps to support making changes to a plan. These maps were available both on paper and digitally on the touch table. At the end of the workshop, criteria were defined to evaluate each plan. The criteria are derived from economical, social, ecological, spatial and political objectives concerning the area. A multi-criteria analysis (MCA) was conducted to derive scores of each alternative with respect to each criterion and to rank these alternatives.

3.2 Workshop 2: analysis

The second workshop was held with members of the various research groups in charge of conducting studies for the area. Each group was asked to present their results as a map. These maps were displayed on the touch table. Objectives of the workshop were to group, combine and integrate this knowledge. It was also a goal to measure the robustness and suitability of each land use alternative and, with this as a basis, to choose a plan for further adjustment. To do this, participants were asked to try different land use restrictions based on groundwater levels (See Table 2), dynamically seeing its feasibility displayed on the touch table. A combination of the touch table and GIS-based decision support tools was used. Participants were asked to experiment with the plans, adjust land uses where they considered them needed and make changes to the thresholds to test the robustness of the plans.

Table 2. Land use restrictions for 'Extensive agriculture', based on groundwater levels

Interval	*Suitability*
Open water - groundwater level ≤ 0	Unsuitable - below groundwater level
0 - 10 cm below groundwater level	Unsuitable - below groundwater level
10 - 20 cm below groundwater level	Unsuitable - below groundwater level
20 - 30 cm below groundwater level	Suitable
30 - 40 cm below groundwater level	Suitable
40 - 50 cm below groundwater level	Suitable
50 - 60 cm below groundwater level	Unsuitable - above groudwater level
60 - 70 cm below groundwater level	Unsuitable - above groundwater level
70 - 80 cm below groundwater level	Unsuitable - above groundwater level
80 - 90 cm below groundwater level	Unsuitable - above groundwater level

3.3 Workshop 3: design

Policy goal of the third workshop was to evaluate and adjust three initial land use plans for the area. Research goal was to explore the possibilities offered by the touch table as support instrument for exchange of both knowledge and insights concerning the planning of the area. Participants were public servants, project leaders and nature organizations. They were asked to make changes to a previously chosen land use plan and also to design a new plan. To create a new plan, three options were available namely, changing existing land uses from one of the three alternatives and the current situation, allocating land use classes on a parcel-based blank map and drawing polygons in free-form mode allocating land use classes. For each user's input, suitability, which is based on ground water levels, was dynamically calculated and mapped (See Figure 2). These tasks were performed on the touch table in a multi-user setting. Results were presented at the end of the workshop. With this as a basis, participants were asked to evaluate each plan by filling scores for each criterion in an effects table in order to rank all three plans and the current situation. Figure 3 shows a flowchart with the major steps followed during the third workshop.

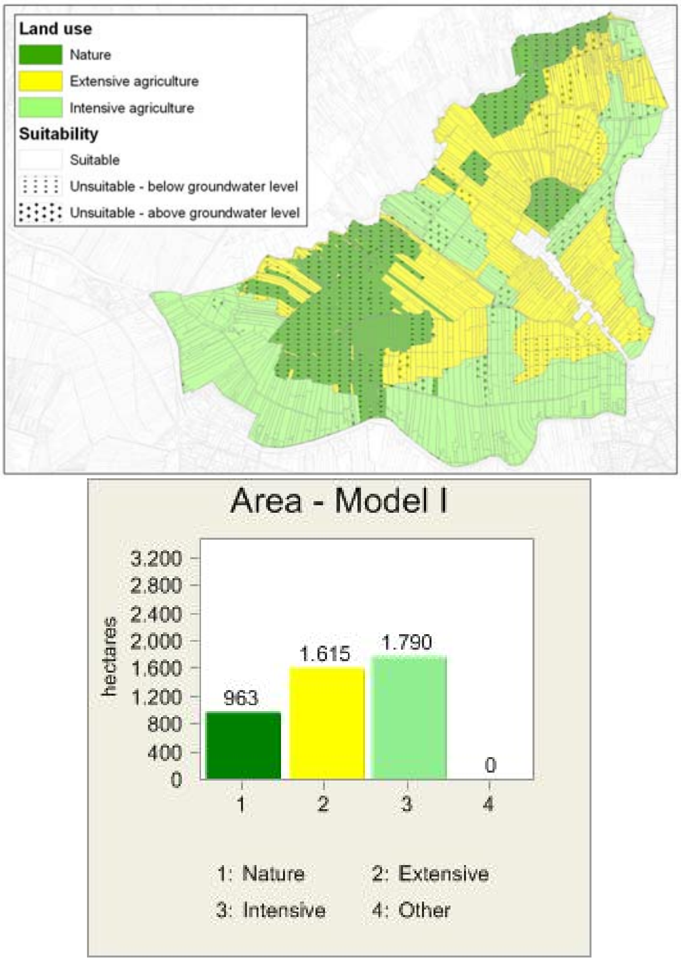


Figure 2. Example of a land use plan designed by the participants on the touch table with dynamic suitability and areas per land use class

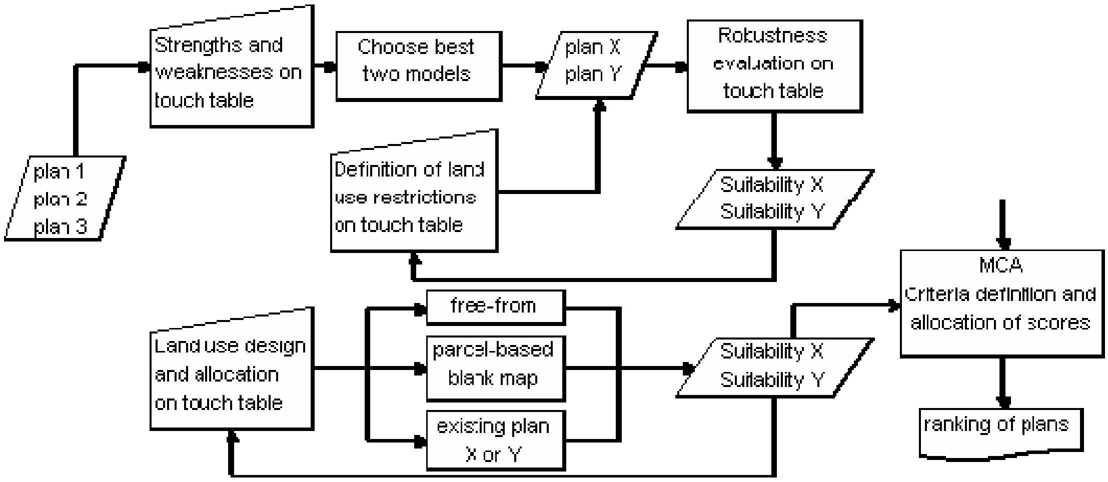


Figure 3. Major steps of the process followed during workshop 3

4. Feedback from the participants

Surveys were conducted before and after the workshops to measure the participants' expectations as well as their reflections on the process and the impact of the touch table in comparison to traditionally used printed maps. We were particularly interested in exploring three aspects namely, the main differences between the usage of printed and digital maps, the added value of the touch table (in the participant's view) and its major functionalities/advantages. Results of the surveys showed a preference (65% of 35 participants) for the touch table over printed maps. Their reasons included the possibility to choose different background maps, to overlay and combine different maps using transparencies, to zoom and pan throughout the area, and to see separately what other participants have drawn. Another important reason was the possibility to both quantitatively and visually evaluate land use alternatives in a team-setup. The surveys also indicated that the use of the touch table prompted significant discussions and knowledge exchange on the area, crucial for the reaching of a compromise plan.

For the remaining workshops, it is intended to add more evaluation and visualization tools. Functionality will be added to help participants suggest changes in land use and support a negotiated result that is the best possible for all parties involved. So far we did not use the table to support negotiation. It will be interesting to see whether, when real interests are at stake, participants are still willing to participate within the framework set by the table.

5. Acknowledgements

This project is funded by the Dutch BSIK research programs: 'Climate changes spatial planning' and 'Living with water'.

References

Carton L (2007) *Map making and map use in a multi-actor context* PhD thesis Delft University of Technology.

Herwijnen M and Janssen R (2002) The use of multi-criteria analysis in a spatial context *Spatial Information and the Environment* pp259-272

Herwijnen M (1999) *Spatial decision support for environmental management* PhD thesis Vrije Universiteit, Amsterdam.

Janssen R (1992) *Multiobjective decision support for environmental management* Kluwer Academic Publishers, Dordrecht.

Janssen R (2001) On the use of multi-criteria analysis in environmental impact assessment in the Netherlands *Journal of multi-criteria decision analysis* **10(2)** pp101-109

Janssen R and Uran O (2003) Presentation of information for spatial decision support: A survey on the use of maps by participants in quantitative water management in the IJsselmeer region, the Netherlands *Journal of physics and chemistry of the earth* **28** pp611-620

Uran O and Janssen R (2002) Why are spatial decision support systems not used? Some experiences from the Netherlands *Computers, environment and urban systems* **27(5)** pp511-526

Biography

Gustavo Arciniegas *holds B.Sc. and M.Sc. degrees in civil engineering and geoinformatics. His research interest is the use of spatial information technologies for land use planning, hazard monitoring and environmental problem solving. Currently he works as Ph.D. researcher in spatial decision support tools at the Institute for Environmental Studies, Amsterdam.*

Ron Janssen *specializes in decision support for environmental management. His research topics are decision analysis and spatial analysis/evaluation, focusing on the effective use of information and the development of spatial decision support systems. Currently he is head of the department 'Spatial analysis and decision support' within the Institute for Environmental Studies, Amsterdam.*

Nancy Omtzigt *has been employed at the Institute for Environmental Studies in Amsterdam since 1999. Her work deals with the design and implementation of geographic information systems in environmental research and education. Research activities include conceptual design and implementation of spatial decision support systems and visualization techniques for spatial data.*

Corinne Cornelisse *holds M.Sc. degree in environment and resource management. Her research topics are water management and energy transitions with a focus on the participation of relevant stakeholders. She is at present employed as junior researcher at the Institute for Environmental Studies of the Vrije Universiteit Amsterdam.*