Kernel Density Estimation and Percent Volume Contours in General Practice Catchment Area Analysis in Urban Areas

Maurizio Gibin[1], Paul Longley[2], Phil Atkinson[3]

[1]Department of Geography – Centre for Advanced Spatial Analysis, University College London

Pearson Building, Gower Street, London WC1E 6B

Email: m.gibin@ucl.ac.uk

[2]Department of Geography – Centre for Advanced Spatial Analysis, University College London

Pearson Building, Gower Street, London WC1E 6B

Email: plongley@geog.ucl.ac.uk

[3]Southwark Primary Care Trust

Woodmill Building, Neckinger, London SE16 3QN

Email: phil.atkinson@southwarkpct.nhs.uk

KEYWORDS: GIS, Kernel Density Estimation, Percent Volume Contours, Health Geography

1. Introduction

In the National Health Service (NHS) system General Practices (GP) are providers of Primary Health Services (PHSs). PHS are managed by Primary Care Trusts (PCTs) and are generally conceived as the first point of access to public health. Good provision of PHSs can avoid patient reliance upon secondary health services like Accident and Emergency (A&E) or hospitalisation – both of which tend to be much more expensive for NHS balances. The backbone of NHS policy is set out in *Patient Choice* (Department of Health, 2006), which states the right to access treatment of uniform quality where and when a patient decides. Accessibility in health policy is defined according to two main dimensions (Luo, 2004; Higgs, 2004). Social inequalities can erect barriers to equal health treatment. The effects of non physical distance upon healthcare received are sometimes compounded by physical constraints upon access to health facilities. Physical constraints may be conceived in terms of spatial accessibility (SA) of health services. Geographic Information Systems (GIS) are useful in deriving various SA to GP services, while geodemographic and lifestyle profiling can be useful in pinpointing local social inequalities in the access to healthcare.

This paper will illustrate the deployment of spatial analysis techniques in analysing spatial accessibility and assigning "core" effective demand to General Practices, based upon the experience of a Knowledge Transfer Partnership in Southwark, a borough based in south east London, UK.

2. Kernel Density Estimation and Percent Volume Contours

Kernel Density Estimation (KDE) techniques in geospatial analysis may be applied to line or point datasets with spatially extensive attributes (de Smith et al., 2007). In

Geographic Information Systems (GIS) the result of a KDE is usually a raster dataset (Longley et al., 2005) where each cell has a density value[1] that is weighted according to distance from the starting features. The user can choose the cell size of the output raster, the attribute field to be used in the calculation, its units of measure and the search radius or *bandwidth.*

Any of a range of different functions can be used to weight density values. One of the most used is a bounded quadratic approximation to the Normal distribution called the Epanechnikov function (de Smith et al., 2007) which is defined as follows:

$$\begin{gathered} \frac{3}{4}\left(1-t^2\right) \text{ for } t=\frac{d}{h}\leq 1 \\ 0 \text{ for } t=\frac{d}{h}>1 \end{gathered} \qquad (1)$$

where *d* is the distance between the cell and the point in the dataset and *h* is the bandwidth.

Figures 1 and 2 show the kernel density estimation for a one and a two points dataset. On the left is shown a two dimensional raster (Fig. 1a and 2a), and on the right the three dimensional version where cell height values are taken from their density (Fig. 1b and 2b). Darkest colours show a higher weight to the cells closer to the points. Note how the Epanechnikov function is very similar in shape to the Normal distribution.

(a) (b)

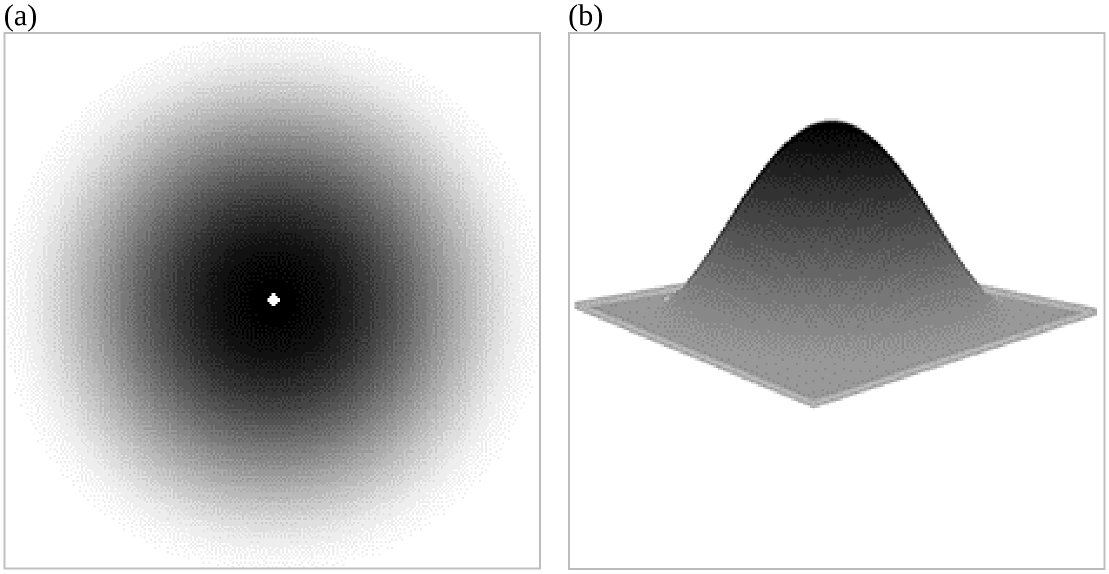


Figure 1 – Kernel Density surface and volume for a one point dataset.

When observing density values on the raster surface, it may be of interest to isolate areas that correspond to a given percentage of the total cumulative distribution. Figure 3a and 3b show the 50 *Percent Volume Contour* (PVC) for a KDE of a two points dataset. PVC is not a simple contour lines as it does not connect cells with the same value.

[1] Density values vary according to the software used (see de Smith et al. 2006). The software used in the analysis for this paper was ESRI ArcGIS 9.1© with the Spatial Analyst© extension. Percent volume contours were calculated with a free ArcGIS extension called *Hawth's Analysis Tools* developed by Hawthorne Beyer and downloadable from the website http://www.spatialecology.com.

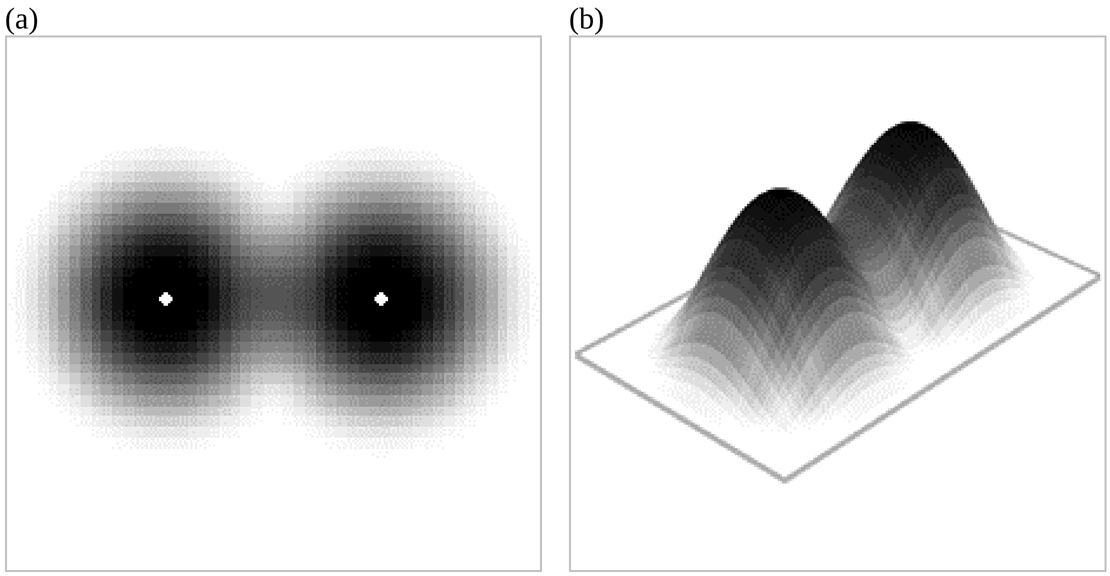


Figure 2 – Kernel Density surface and volume for a two points dataset

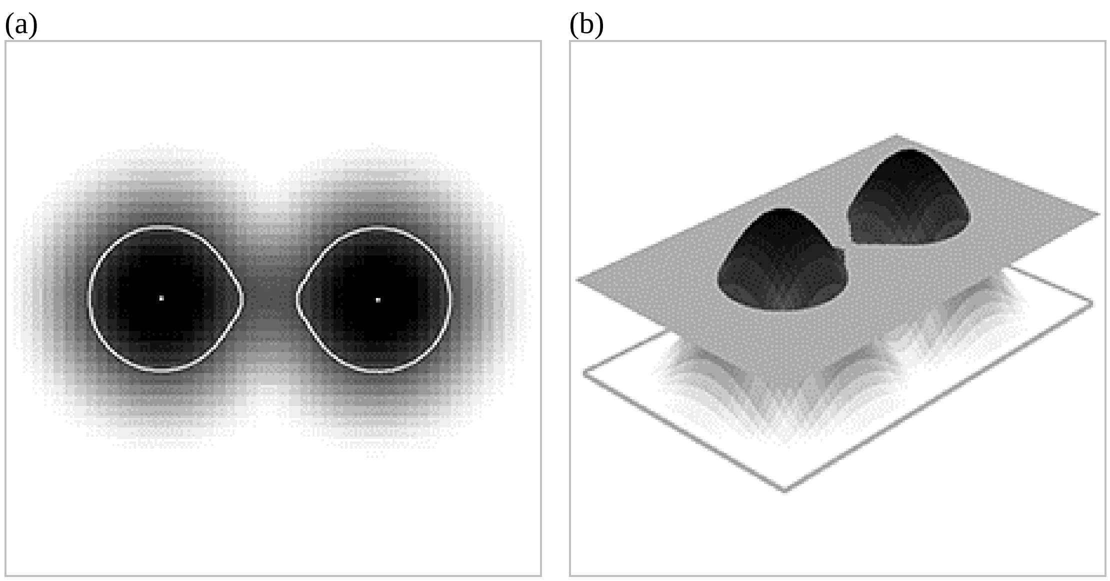


Figure 3 – 50 Percent Volume Contour for a two points dataset.

Spatial Social Sciences applications of KDE are based on spatially extensive variables (de Smith et al., 2007) such as socio-economic and health data. In the latter instance KDE is often used to measure Spatial Accessibility (SA) to health care services by comparing the densities of supply and demand in a given area (Guagliardo, 2004; Luo 2004), in order to detect uneven distributions of services as barriers of the general access and equity of treatment of healthcare mentioned above.

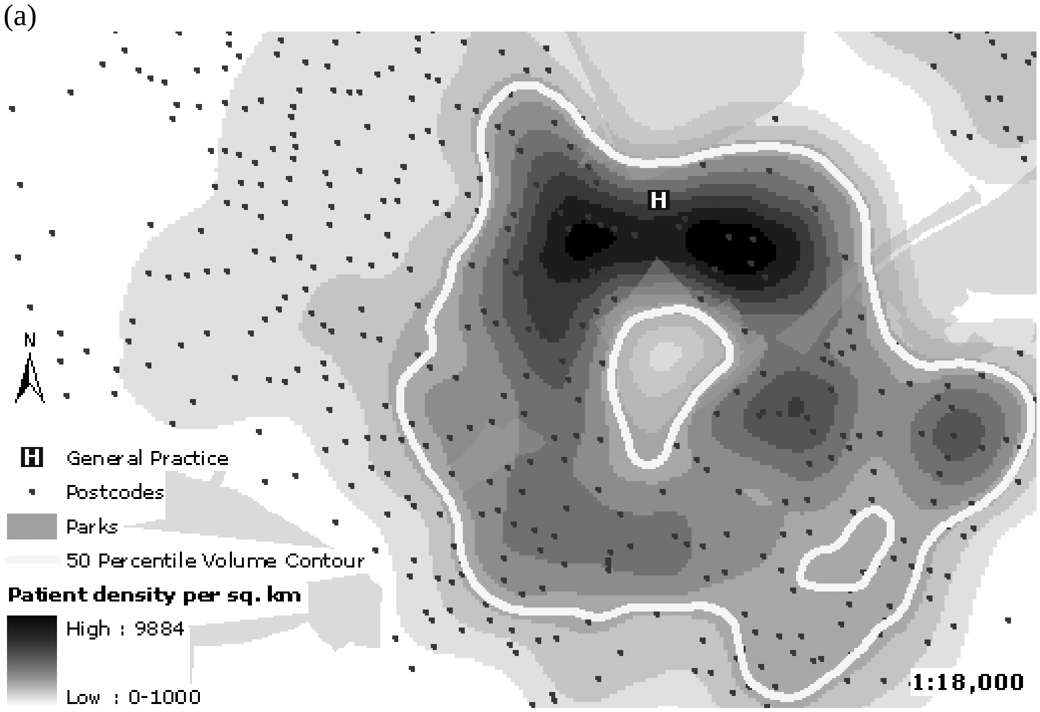
3. General practices and registered population in Southwark

Data pertaining to general practice registrations are collected at the individual level, but for dissemination purposes are often aggregated at postcode level in order to protect confidentiality. KDE techniques are particularly effective in the analysis of GP catchments areas (CAs) in urban environments because Primary Care Trusts allocate people to their closest surgeries. In the case of Southwark Primary Care Trust, the majority of patients generally live within a one mile radius of their closest surgery, although some longer distances occur when people subsequently move to another part of the Borough. Figure 4 shows the distribution of patient registrations[2] relative to a general practice in Southwark, using a point dataset and a raster with values of patient density per sq. km. The latter one is the result of a KDE applied on the postcode dataset using the attribute field containing the number of patient registered to the given GP, a bandwidth of 250 meters and a cell size of 10 meters.

If an appropriate bandwidth is chosen, the resulting surface detects the presence of physical features like roads or parks. KDE performs better than classic distance statistics and geographical dispersion measures like the standard deviational ellipse, because it takes in account the direction of distribution (Figure 4b).

A 50 percent volume contour was calculated from the KDE surface and it is shown by the thick line. Using a simple spatial query it is possible to calculate the actual number of patients registered in the 50% demand area by selecting the points falling inside the contour. These postcodes can be interpreted as the "core" CA and can be also linked with other socio-economic variables to obtain a profile of the area surrounding the GP. This is potentially very useful when targeting health campaigns

(a)



[2] To maintain confidentiality postcodes with less than 5 people registered have been masked.

(b)

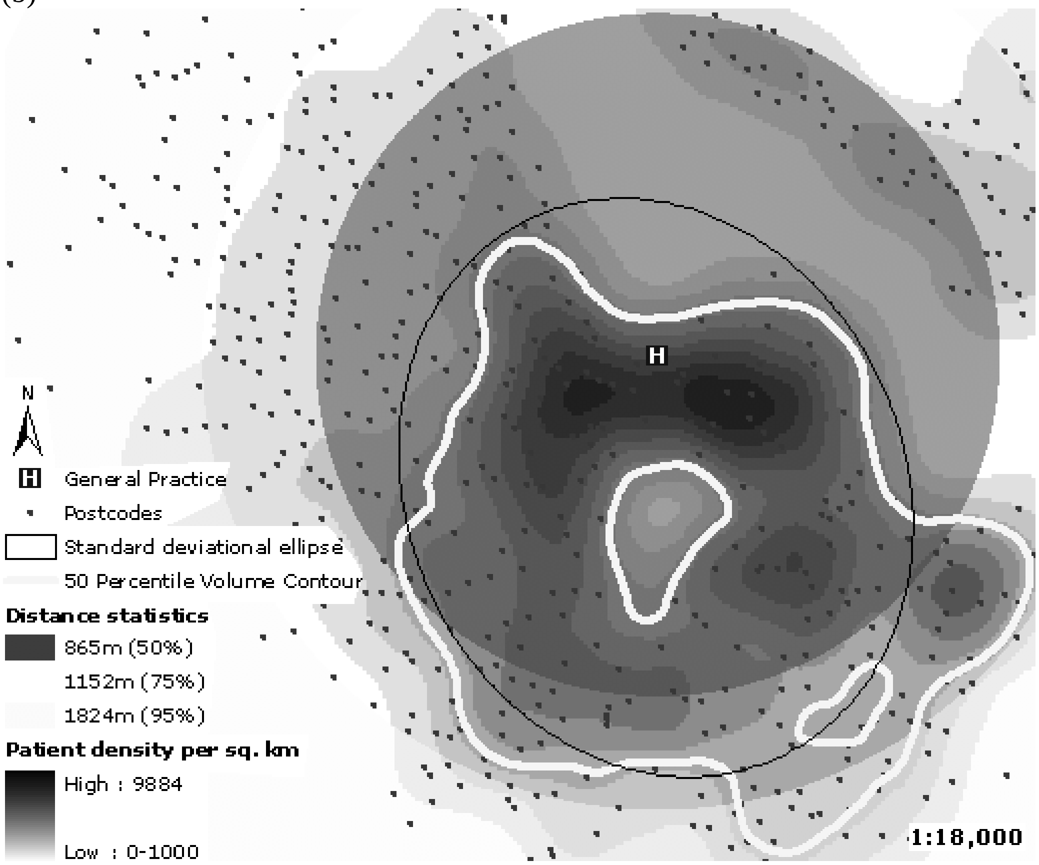
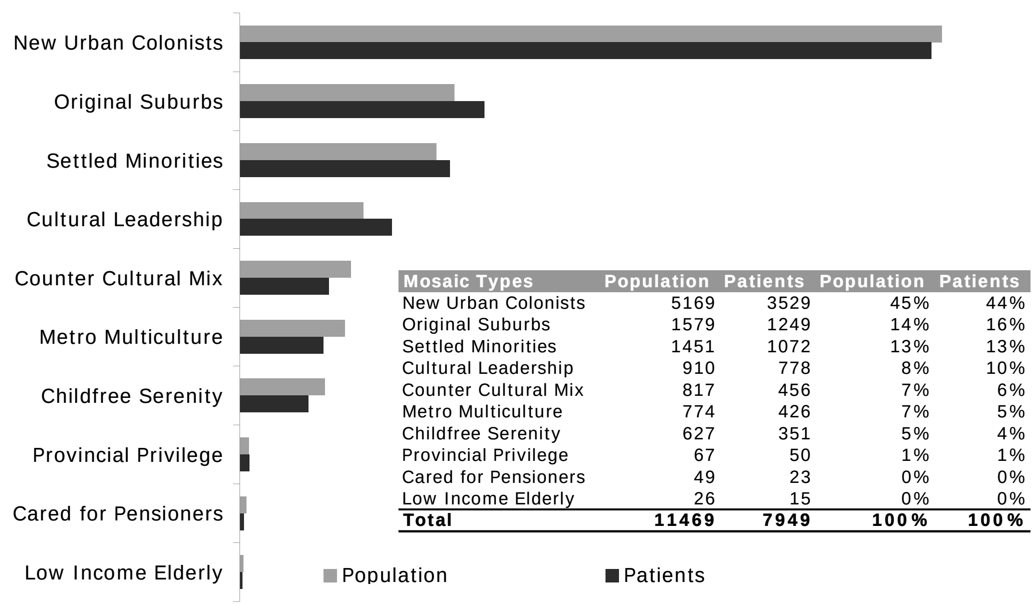


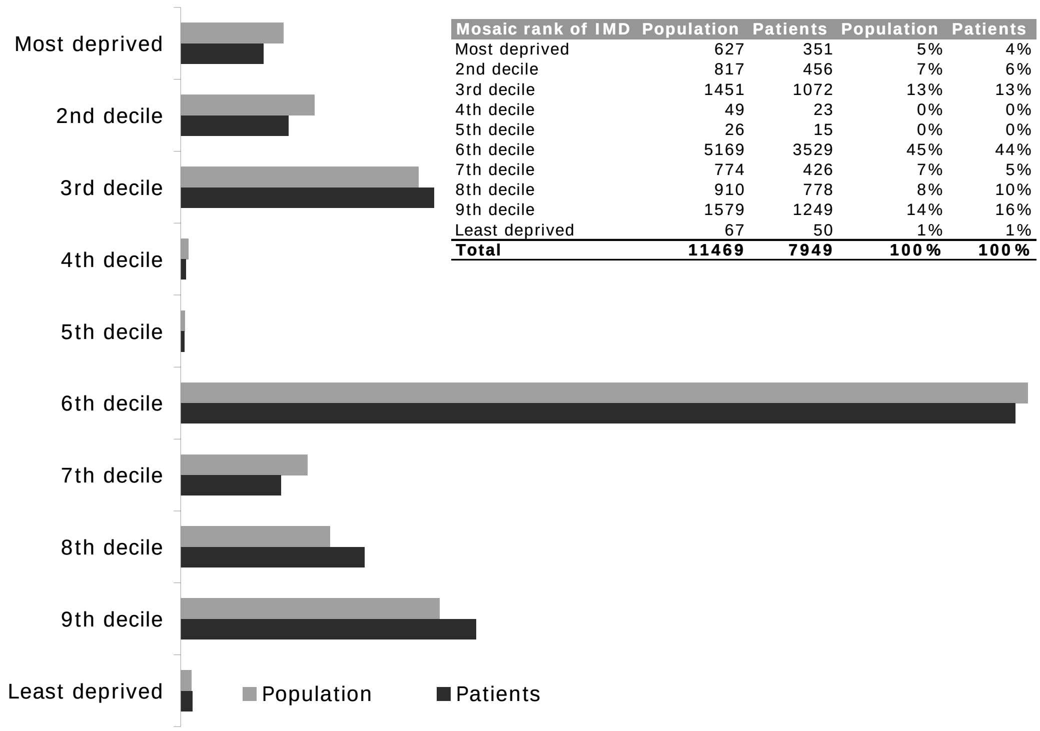
Figure 4 – Patient distribution at postcode level for a General Practice in Southwark.

(a)



Mosaic Types	Population	Patients	Population	Patients
New Urban Colonists	5169	3529	45%	44%
Original Suburbs	1579	1249	14%	16%
Settled Minorities	1451	1072	13%	13%
Cultural Leadership	910	778	8%	10%
Counter Cultural Mix	817	456	7%	6%
Metro Multiculture	774	426	7%	5%
Childfree Serenity	627	351	5%	4%
Provincial Privilege	67	50	1%	1%
Cared for Pensioners	49	23	0%	0%
Low Income Elderly	26	15	0%	0%
Total	**11469**	**7949**	**100%**	**100%**

(b)



Mosaic rank of IMD	Population	Patients	Population	Patients
Most deprived	627	351	5%	4%
2nd decile	817	456	7%	6%
3rd decile	1451	1072	13%	13%
4th decile	49	23	0%	0%
5th decile	26	15	0%	0%
6th decile	5169	3529	45%	44%
7th decile	774	426	7%	5%
8th decile	910	778	8%	10%
9th decile	1579	1249	14%	16%
Least deprived	67	50	1%	1%
Total	**11469**	**7949**	**100%**	**100%**

Figure 5 - Profile of the postcodes using the 50 percent volume contour.

Figure 5 shows the geodemographic profile of the 50% of the CA using Mosaic, a commercial geodemographic system developed by Experian. Mosaic types of registered patients have similar proportions to the underlying population distribution (Figure 5a). The predominant type is *New Urban Colonists*, characterised by young professionals living in high-density gentrified areas with easy access to the city.

In the health targeting decision-making process area deprivation plays a fundamental rule. Figure 5b shows the rank of the Index of Multiple Deprivation coded at postcode level for the 50% CA. Greatest part of the registered patients falls in the 6^{th} decile, showing that the area surrounding the practice is not particularly deprived.

4. Conclusions

Use of surfaces and contours can improve exploratory spatial data analysis and help in communicating geographic information (GI) to the general public. Geodemographic information is particularly important for the health sector because of its correlations with health outcomes and policy performances.

PVC and KDE can be very useful in the analysis of GP registered population and can help in selecting the closest postcodes to the surgery. Postcodes profiles are a good source of information when targeting lifestyles associated with particular health conditions.
Selecting only the areas with high densities of patients can avoid issues of confidentiality in communication and data sharing. Although useful in preserving data

confidentiality, KDE surfaces can lead users to believe that attributes are continuously distributed in the study area. In a cartographic representation this problem can be avoided by overlaying other sources of information in a sort of dasymetric mapping (in Fig. 4a, for example, the distribution of parks was overlaid on top of the density surface).

Optimising bandwidth remains one of biggest irresolvable problems (Brunsdon, 1995). Mathematical algorithms (Kao et al. 2002 and Silverman, 1986) can improve the selection process but are no complete substitute for personal experience and knowledge of the study area and of the attribute upon which density analysis is performed.

5. References

BRUNSDON, C. (1995). Estimating probability surfaces for geographical point data: an adaptive kernel algorithm, *Computer & Geosciences*, 1995 21.

DEPARTMENT OF HEALTH. (2005). *Departmental report 2005*. Department of Health website. http://www.dh.gov.uk/PublicationsAndStatistics/Publications/AnnualReports/DHAnnualReportsArticle/fs/en?CONTENT_ID=4113725&chk=1krOlR

DE SMITH, M. J., GOODCHILD, F. M., LONGLEY, P. A. (2007). *Geospatial Analysis*. Leicester: The Winchelsea Press.

GUAGLIARDO, M. F. (2004). Spatial accessibility of primary care: concepts, methods and challenges, *International Journal of Health Geographics,* 2004 3:3.

HIGGS, G. (2004). A literature review of the use of GIS-based measures of access to health care services, *Health Services & Outcomes Research Methodology,* 2004, 5.

KAO, D., LUO, A., DUNGAN, J. L., PANG, A. (2002). Visualizing spatially varying distributions data, *Proceedings of the Sixth International Conference on Information Visualisation*, IEEE Computer Society 2002.

LONGLEY, P. A., GOODCHILD, F. M., MAGUIRE, D. J., RHIND, D. W. (2005). *Geographic Information Systems and Science*. West Sussex: John Wiley & Sons.

LUO, W. (2004). Using a GIS-based floating catchment method to assess areas with shortage of physicians, *Health & Place*, 2004 10.

SILVERMAN, B.W. (1986). *Density Estimation*. London: Chapman and Hall.

6. Author's Biography

Maurizio Gibin is working with the Department of Geography – Centre for Advanced Spatial Analysis, University College London, on a Knowledge Transfer Partnership project with Southwark Primary Care Trust. He obtained a degree in Economics from University of Eastern Piedmont and a PhD in Geography from University of Trieste. He is currently a PhD student under the supervision of Paul Longley and Dave Ashby from Department of Geography, University College London and Phil Atkinson from Southwark Primary Care Trust.