Available online at www.sciencedirect.com
ScienceDirect
Computers, Environment and Urban Systems
31 (2007) 19–32
Computers, Environment and Urban Systems
www.elsevier.com/locate/compenvurbsys

Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps

Mitchel Langford *

School of Computing, University of Glamorgan, Pontypridd, Wales, UK

Received 30 April 2004; received in revised form 18 July 2005; accepted 18 July 2005

Abstract

Areal interpolation between one partitioning of geographical space and another remains an important topic, particular in terms of population counts and related statistics which are often required in order to compute an incidence ratio. Despite numerous recent developments in intelligent areal interpolation methods, and studies that have demonstrated their clear advantage over simple areal weighting, there is little evidence to suggest widespread usage amongst the GIS user community. It is argued that to encourage greater uptake such methods must offer simplicity and convenience. Areal interpolation based on binary dasymetric mapping is conceptually simple, but examples to date tend to use information extracted from multi-spectral satellite imagery which limits its perceived convenience. This paper examines a simple method to extract equivalent information from a raster pixel map. It is shown to offer comparable areal interpolation performance at considerably less cost in terms of both time and complexity.

Keywords: Interpolation; Dasymetric; Population; Raster

1. Introduction

GIS users often manipulate and manage thematic or attribute information either associated with, or directly generated from, a corresponding spatial zone. These zones may be defined by real natural boundaries or by abstract lines in geographical space, with

* Tel.: +44 1443 654148; fax: +44 1443 652715.
E-mail address: mlangfor@glam.ac.uk

0198-9715/$ - see front matter
doi:10.1016/j.compenvurbsys.2005.07.005

examples including government administrative units, census collection and reporting zones, zip/postcodes, travel-to-work areas, fluvial watersheds, national park limits, and the surface outcrop patterns of different geological rock types.

National census statistics are almost invariably released as aggregate counts and statistics for corresponding spatial zones. This is primarily to satisfy a legal requirement to maintain confidentiality and prevent the disclosure of information about identifiable individuals, but also aids in controlling data volume. A census is a particularly valuable dataset since it is normally the most complete source of data on population characteristics and provides a huge breadth of related socio-economic information on topics such as health, employment, ethnicity and housing stock. It has widespread uses in central and local Government activities, in the planning and management of resources related to education, health care provision, transportation and housing, and it lies at the heart of most commercial geodemographic data services.

Useful as they are, census data will always be provided for a single output geography since the release of multiple geographies presents a risk of disclosure through differencing operations. Situations therefore inevitably arise in urban studies conducted with GIS where data vital to the analysis are reported for incompatible spatial units. Such incompatibilities arise by several means: (i) different agencies report information using their own unique partitions of geographical space, (ii) administrative boundaries change over time to accommodate urban growth and redevelopment, and (iii) entirely new analytical zones may be generated by the GIS itself via operations such a spatial buffering, overlay and viewshed analysis. Even if the intent is to compare data between censuses, numerous boundary changes give rise to similar problems (Gregory, 2002). The 2001 UK Census employed entirely separate collection and dissemination geographies based on enumeration districts and output areas respectively (both sets of zones are broadly equivalent to the census block in the US Census hierarchy). The output areas that are used for data release in 2001 were engineered to satisfy minimum population thresholds, take account of postcode (zip code) geography, and achieve some optimization of shape and internal social homogeneity (Martin, 1998). Despite the obvious advantages for the analysis of 2001 data, the end result is little commonality with the 1981 Census geography in which a single geography based on Enumeration Districts was used for both collection and dissemination, thus making longitudinal studies problematic.

2. Cost-benefits of intelligent areal interpolation

The solution for incompatible spatial units is to transform all data values onto a single common set of zones using appropriate area interpolation techniques. To this end a range of techniques aimed at transferring non-spatial attributes from one partitioning of space to another have been developed (see reviews by Goodchild & Lam, 1980; Goodchild, Anselin, & Deichmann, 1993; Flowerdew & Green, 1994). The least demanding method, simple areal weighting, uses only the areas of intersection created when source zones are overlaid with target zones in a standard 'union' operation. In deriving an estimate for a target zone this method must assume uniform distribution of the mapped property within the boundaries of the source zones (Fig. 1a), a situation that is highly unlikely to occur in the real world (Xie, 1995).

One solution to this problem is to try to use additional knowledge of the local environment to better predict population distribution structure within source zones. Exactly what

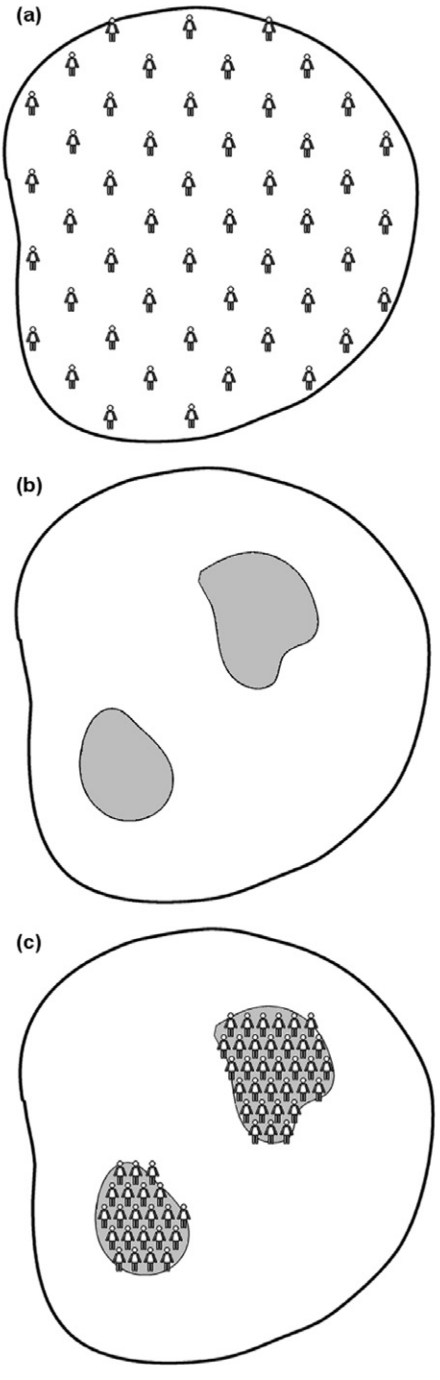

Fig. 1. An illustration of the underlying population distribution models employed by simple areal weighted and dasymetric interpolation: (a) census population distributed evenly within a source zone (simple areal weighting model), (b) ancillary information depicting populated (dark) and unpopulated (light) areas and (c) census population distributed only within populated areas of a source zone (dasymetric model).

ancillary information is used, and how it is acted upon, varies from one technique to another – but all these so-called 'intelligent' interpolation methods have a common goal of exploiting additional relevant knowledge to generate better target zone estimates (Flowerdew & Green, 1994). For example, in the binary dasymetric method (Langford & Unwin, 1994; Mennis, 2003) ancillary information on residential land use allows source zones to be internally partitioned into occupied and unoccupied sub-regions. Whilst the EM algorithm (Flowerdew & Green, 1989, 1991) provides a flexible environment whereby ancillary variables related to target zones and possessing multiple states or reported on a continuous scale can be used to enhance interpolation accuracy.

A number of recent studies have demonstrated through practical case studies that substantial improvements in estimation accuracy can be gained by using intelligent interpolation methods. For example, Reibel and Bufalino (2005) report a 20% improvement compared to simple areal weighting, while Mrozinski and Cromley (1999) demonstrated a 30–32% improvement. Despite these studies there has to date been little evidence to suggest widespread adoption amongst the broader GIS user community. To understand why greater use is not made of intelligent methods some consideration of the costs and benefits of implementing them is needed, together with associated issues such as awareness and inertia, simplicity and convenience. Firstly, the widespread use of simple areal weighting is undoubtedly encouraged by the fact that a suitable overlay tool is readily available in most GIS software. Furthermore, it does not require the user to become involved with the provision of suitable additional data resources which are needed to drive the intelligent methods. Some examples of these drawn from the research literature include: parliamentary voting patterns and census-derived car ownership statistics (Flowerdew & Green, 1994), land cover derived from classified satellite imagery (Holt, Lo, & Hodler, 2004; Langford & Unwin, 1994; Yuan, Smith, & Limp, 1997), manually digitized residential structures (Moon & Farmer, 2001), vector street networks (Mrozinski & Cromley, 1999; Reibel & Bufalino, 2005), and vector polygons depicting land use information (Eicher & Brewer, 2001; Mennis, 2003).

Most users are aware that assumptions, simplifications and errors are inherent in almost all aspects of data collection, storage, manipulation and analysis within a GIS (see Burrough & McDonnell, 1998; Chrisman, 1991; Heuvelink, Burrough, & Stein, 1989). Thus, areal interpolation error might be seen as just one more source of inaccuracy amongst a long list of potential causes, with no particular reason to be singled out for attention. General familiarity with the simple areal weighting method, the perception that it seems good enough for other co-workers, and perhaps a lack of awareness of the alternatives, all help to build inertia and prevent the uptake of more accurate techniques. The lack of functionality in common GIS packages necessitating the construction of potentially complex linkages to external statistical software packages in order to implement intelligent interpolation techniques (Flowerdew & Green, 1994) can only serve to exacerbate this situation.

It is the overall balance between positive and negative factors that will ultimately encourage or dissuade users from taking up intelligent interpolation methods. Notwithstanding the empirical evidence previously cited, and strong theoretical arguments as to their benefits (Mugglin, Carlin, Zhu, & Conlon, 1999; Sadahiro, 1999), the perceived costs are perhaps still sufficient to dampen enthusiasm and thus allow inertia to win out. Although potential users may welcome improved interpolation accuracy, the adoption of methods that involve complex statistical ideas, that can only be implemented using specialist software external to the GIS, or which require costly, esoteric or unfamiliar data resources (such as satellite remotely sensed imagery for example) are unlikely to gain widespread acceptance. The key to encouraging greater use of intelligent areal interpolation, therefore, is simplicity and convenience.

3. Simplifying dasymetric areal interpolation

In the context of the discussion above areal interpolation based on dasymetric mapping has a number of attractive qualities that suggest it is well placed to overcome the inertia associated with simple areal weighting. In its most basic form an additional dataset is used to discriminate between 'populated' from 'unpopulated' areas entirely independently of the source zones (see Fig. 1b). Census population counts are then redistributed internally within each source zone such that they lie only within populated sub-zones (Fig. 1c), thereby addressing the improbable uniform distribution assumption of simple areal weighting identified earlier. Once this process is completed the remaining procedure is identical to simple areal weighting – a union overlay is followed by a tallying-up of those intersection zones needed to create the desired target zone. The benefits of dasymetric-based interpolation in terms of accuracy gain have already been demonstrated. Fisher and Langford (1995) measured interpolation errors in population counts between incompatible areal units using a Monte Carlo simulation model which allowed multiple realisations of target units (and with known true populations) to be created. The results indicated a mean reduction in interpolation error of 54% when comparing dasymetric interpolation to simple areal weighting.

Dasymetric interpolation clearly offers simplicity, both conceptually and in terms of its practical implementation. However, most reported examples of its use to date have derived the dataset depicting populated and unpopulated areas from remotely sensed imagery (e.g. Eicher & Brewer, 2001; Langford & Unwin, 1994; Mennis, 2003; Yuan et al., 1997). It may be this aspect of its implementation that has spoiled users' perceptions of its convenience, even if the robustness of the technique to the presence of error in the classified image has also been demonstrated (Fisher & Langford, 1996).

Although satellite imagery is increasingly being linked to socio-economic and urban applications (see Liverman, Moran, Rindfuss, & Stern, 1998; Mesev, 2003), it remains likely that those GIS users most often faced with areal interpolation issues will be largely unfamiliar with digital image processing techniques and the complexities of multi-spectral image classification. Without the necessary skill set, and possibly the software, to comfortably handle the manipulation and classification of multi-spectral satellite imagery, it is entirely understandable if they tend to shy away from this approach whatever its potential benefits. To ensure a perception of both simplicity and convenience it is thus necessary to seek out alternatives to remotely sensed imagery which are more familiar to typical GIS users and which are less demanding in terms of the skills required to utilise them.

The overlaid network algorithm (Reibel & Bufalino, 2005; Xie, 1995) provides one example of how such practical obstacles can be overcome. This method utilises widely available and readily accessible vector road maps to redistribute population within source zones and thus enhance the areal interpolation process. A similar solution using a buffered road network was adopted by Mrozinski and Cromley (1999). In both cases the underlying principle is that most inhabited dwellings will lie close to a road. Redistributing the source zone population such that it lies along or close to these linear networks creates a more realistic representation of the true population density surface than the even-distribution model of simple areal weighting. The research reported here follows in a broadly similar vein, but instead it employs ‘pixel maps’ in place of a classified satellite imagery to drive the dasymetric methodology. It also has a degree of similarity to the work reported by Moon and Farmer (2001) who recognise the constraints imposed by a reliance on satellite imagery, and adopted manually digitized housing distribution polygons to create their dasymetric population density surface.

4. Methodology

Pixel maps are raster scan maps delivered using the colour-palette index format, and are widely produced by most national mapping agencies. In the UK the most familiar example is the Ordnance Survey’s 1:50000 raster scan product (Ordnance Survey, 2005). These data have a broadly comparable scale to the remotely sensed images acquired by the popular Enhanced Thematic Mapper scanner (ETM+) flown on the Landsat 7 satellite (for further details see: NASA Goddard Space Flight Center, 2005; US Geological Survey, 2005). They are supplied in a format that is both familiar and easily accessible to most GIS users (i.e. GeoTIFF). Another benefit when compared to most remotely sensed imagery, is that it is unnecessary to perform any rectification tasks or to modify map projections prior to their use within the GIS.

Visually, a wealth of information is present within pixel maps but, rather like a satellite image, it is not immediately accessible without some further data processing and analysis. The thematic information is carried visually through the colour, and physically via a pixel’s colour palette index code. A particular feature of the OS data resource, and raster pixel maps in general, is that a single consistent 8-bit colour palette is utilized for all supplied data tiles. This feature is necessary to allow adjacent raster images to be spatially appended. It contrasts with the typical nature of raster data gathered by a desktop scanner, and it is highly advantageous. It means that if information can be extracted through the identification of specific colour indices, these same rules will apply equally well anywhere within the dataset. Of particular interest in the OS dataset is the identification of features shown as ‘buildings’ in the accompanying key, which appear visually in a Bisque colour. If the pixels that comprise these buildings can be identified and extracted from the raster map they would create a populated/unpopulated mask for use in dasymetric interpolation that is broadly consistent with that derived via a land cover classification of Landsat ETM+ imagery (see Fig. 2a–d). To minimize interpolation errors synchronicity between the raster map and the census data to be interpolated is desirable in order to mitigate the influence of any changes to the building stock. Of course, the same argument is true for dasymetric interpolation based on classified satellite imagery.

The most straightforward way to establish the relationship between objects and colour palette indices in the raster map data is via a cursor inquiry tool in a GIS, or something equivalent in desktop image editing software. This reveals that buildings are not associated with a single colour index, but rather with several, the most frequently occurring values being 88, 203, 210, 218, and 244. A simple recode operation is all that is needed to isolate pixels with these indices, remapping them to the value 1 whilst all others are remapped to the value 0. This operation is easily accomplished in any raster-capable GIS, or alternatively in most desktop image editing software of the type associated with the digital camera market; the outcome is illustrated in Fig. 2b. It is immediately apparent that just as build-

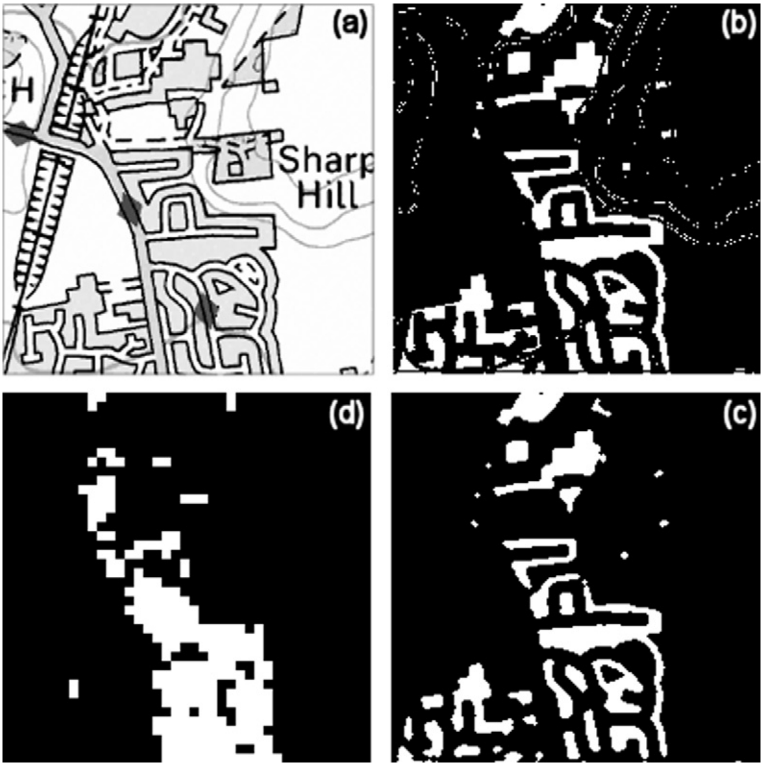

Fig. 2. The creation of populated area masks to aid dasymetric interpolation. Clockwise from upper-left: (a) sample of OS raster map data©, (b) mask obtained by a simple recode operation, (c) mask after further processing with shrink/expand operators and (d) an equivalent mask obtained from a classified Landsat ETM+ image. ©Crown Copyright/database right 2005. An Ordnance Survey/EDINA supplied service.

ings are related to more than one colour index, each colour index can itself be related to more than one map object. This confusion results from interaction between the point spread function of the scanning process and the objects themselves, effectively giving rise to a variety of 'mixed pixels'. Similar problems were encountered by Wise (1999) whilst trying to extract information from scanned paper maps. The same colours or palette indices occur in both contour lines and buildings, for example, and hence there is no way to entirely separate these features on the basis of colour palette index value alone.

Whether or not such errors will significantly affect areal interpolation performance is to be determined by experimental testing, but the recoded data are amenable to further image processing that can help to ameliorate the problem. Since the focus of this work is to simplify the task of dasymetric interpolation only modest efforts were undertaken along these lines, and the simple recoded mask was still retained for performance evaluation. In general the erroneous pixels that arise as a result of mixed pixel effects tend to occur as single isolated cells, whereas virtually all true buildings will occupy several contiguous cells, given the grid resolution of this data set (i.e. a 5 m posting). Various spatial filtering and mathematical morphology tools can therefore be employed to remove the isolated pixels and thus tidy up the urban mask. Suitable tools include, for example, the *Nibble* function within ArcGIS spatial analyst which replaces areas in a grid corresponding to a mask with the values of their nearest neighbours. A similar degree of success can be gained by using Arc/Info GRID's *MajorityFilter* which replaces grid cells by the majority of their contiguous neighbouring cells, or the broadly similar *FocalMajority* which performs the same operation but within a specified neighbourhood shape. Finally, non-linear neighbourhood operators such as *Shrink* and *Expand* are equally effective (compare Fig. 2b with Fig. 2c). Although these various functions may go by different names (e.g. *Median Filter*, *Despeckle Filter*, *Dilate* and *Erode*), they are found in most GIS software, and also in many desktop image editing suites.

For the purpose of evaluating dasymetric areal interpolation performance a simple recoded mask (Fig. 2b), and a derived mask that was tidied using shrink/expand operators (Fig. 2c) were created for the county of Leicestershire, England. An equivalent mask was also obtained from a classified and subsequently recoded Landsat ETM+ satellite image of the same study region (Fig. 2d).

To evaluate the relative areal interpolation performance of these dasymetric maps requires known values to be available against which the modelled estimates can be compared. This was accomplished by applying the dasymetric redistribution at the ward level in the UK Census zone hierarchy. Total ward population count is redistributed internally using the dasymetric maps, and subsequent estimates obtained for the output areas (OAs) within each ward are compared to their known true values. The results given by simple areal weighting were also computed in order to highlight the overall advantages of the dasymetric interpolation approach.

5. Results and discussion

Since a ward's population is redistributed by dasymetric mapping only within its own boundaries, the errors between estimated and true OA population values will always sum to zero within each ward. Appropriate measures of modeling error are therefore, (i) the absolute error summed by ward, and (ii) the squared error summed by ward. The 2001 census hierarchy of Leicestershire consists of 155 wards (with an average population of around 5650), which are themselves constructed from 2883 OAs (with an average population of around 300). Overall interpolation performance can be assessed by the mean of measure (i) and the root mean of measure (ii) as defined above. Finally, a normalised measure of error, the coefficient of variation computed from the root mean squared error divided by the mean ward population, is also of value.

These statistics are presented in Table 1. They indicate dasymetric areal interpolation using masks derived from raster pixel maps perform in a comparable fashion to those obtained from classified ETM+ imagery. The results from the tidied mask in particular are very close to the ETM+ mask performance, and its notable improvement over the simple recoded mask suggests that the extra processing involved is worth the effort. All dasy-

Table 1
Summary interpolation performance statistics

	Simple areal weighting	Recoded mask	Tidied mask	ETM+ mask
ABS	258	128	118	117
RMS	462	213	200	188
CoV	1.53	0.71	0.66	0.62

metric-based interpolations show very substantial improvement when compared to the simple areal weighting method.

Whilst these global statistics are useful, it is also instructive to consider the results in a more geographical context, by mapping and interpreting model performances as distributed by ward. Fig. 3 compares the performance of the tidied mask (hereafter referred to as the MAP model) and the ETM+ mask (hereafter referred to as the SAT model) on a ward-by-ward basis, identifying which model performed most strongly in each case. This highlights the fact that the MAP model was the strongest performer in the majority of wards (i.e. 91 out of 155), and is notably successful in the larger and predominantly rural wards. The most likely explanation for this is that a variety of sources of classification error tend to degrade the SAT model in these regions, providing an advantage to the relatively 'cleaner' mask of the MAP model. In particular, commission errors occur in the satellite-derived urban mask due to: (i) the presence of infrastructure features such as major roads and car parks, and (ii) from the misclassification of certain surfaces, most noticeably soils exposed through agricultural practices, and other sources of vegetation disturbance such as active building sites and quarries. There are, however, a few rural wards such as the one lying to the immediate west of Leicester city, and the southernmost ward, where MAP is outperformed by SAT. Closer inspection reveals that in all these cases the wards contained large industrial complexes in spite of their relatively rural location. This highlights the major weakness of the MAP model which lies in its inability to differentiate between buildings that are residential and those associated with commercial/industrial activity or which serve a public function. This same point also largely explains the spatial distribution of the SAT successes which are clearly associated with

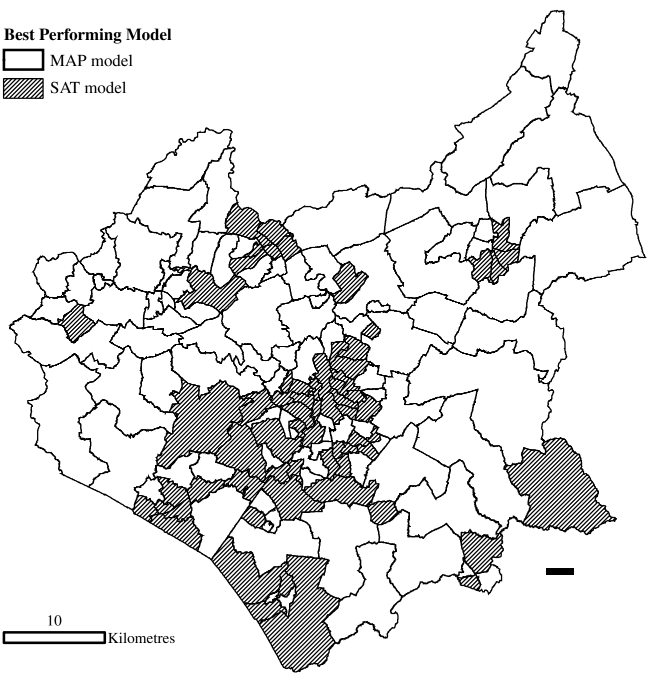

Fig. 3. Pattern of distribution of the best performing dasymetric model.

wards lying around the periphery of major urban areas. These are characterized by the presence of substantial industrial complexes and commercial buildings located within business parks, and the ability of multi-spectral satellite imagery to at least partially differentiate and exclude these land covers from the population redistribution process provides it with an advantage over the MAP model.

A limitation of identifying only the best performing model in each ward is the failure to communicate any idea of the magnitude of the advantage that one model has over the other. To illustrate relative modeling performance a normalized difference (ND) statistic can be computed from the total absolute ward error reported by each model:

$$\mathrm{ND} = (\mathrm{MAP\ error} - \mathrm{SAT\ error})/(\mathrm{MAP\ error} + \mathrm{SAT\ error})$$

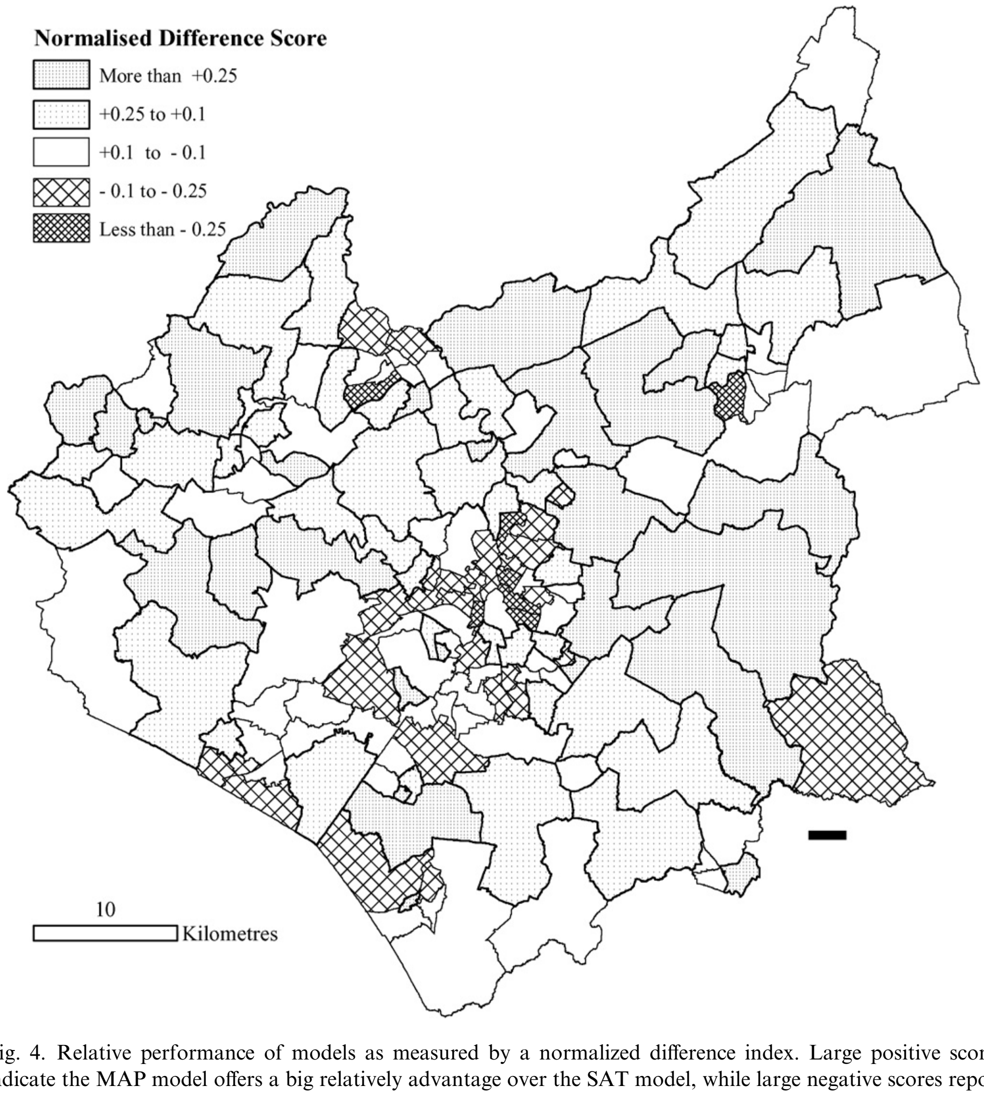

Fig. 4. Relative performance of models as measured by a normalized difference index. Large positive scores indicate the MAP model offers a big relatively advantage over the SAT model, while large negative scores report the reverse situation.

This returns a value of zero if the errors generated by each model are equal. Values tending towards +1.0 indicate that MAP provides an increasingly large advantage over SAT, while values tending towards −1.0 indicate the reverse situation. The average ND score amongst those wards where MAP outperforms SAT is +0.19, whilst in the complementary set the mean value is −0.12. This in itself implies that relative performance is stronger in wards where MAP triumphs over SAT than in the converse situation. Plotting these scores (Fig. 4) fails to reveal any particularly strong geographical pattern, but does suggest that where SAT is the best model (i.e. where negative ND scores arise) its relative advantage over MAP is large only in a few city centre wards. In contrast, strong relative performances by MAP arise mainly in large rural wards and particularly on the eastern side of the study region.

Whilst the ND score provides information on the relative performance of each model, another issue that deserves attention is the actual magnitude of the errors in comparison to ward population. Although an ND score may show, for example, that

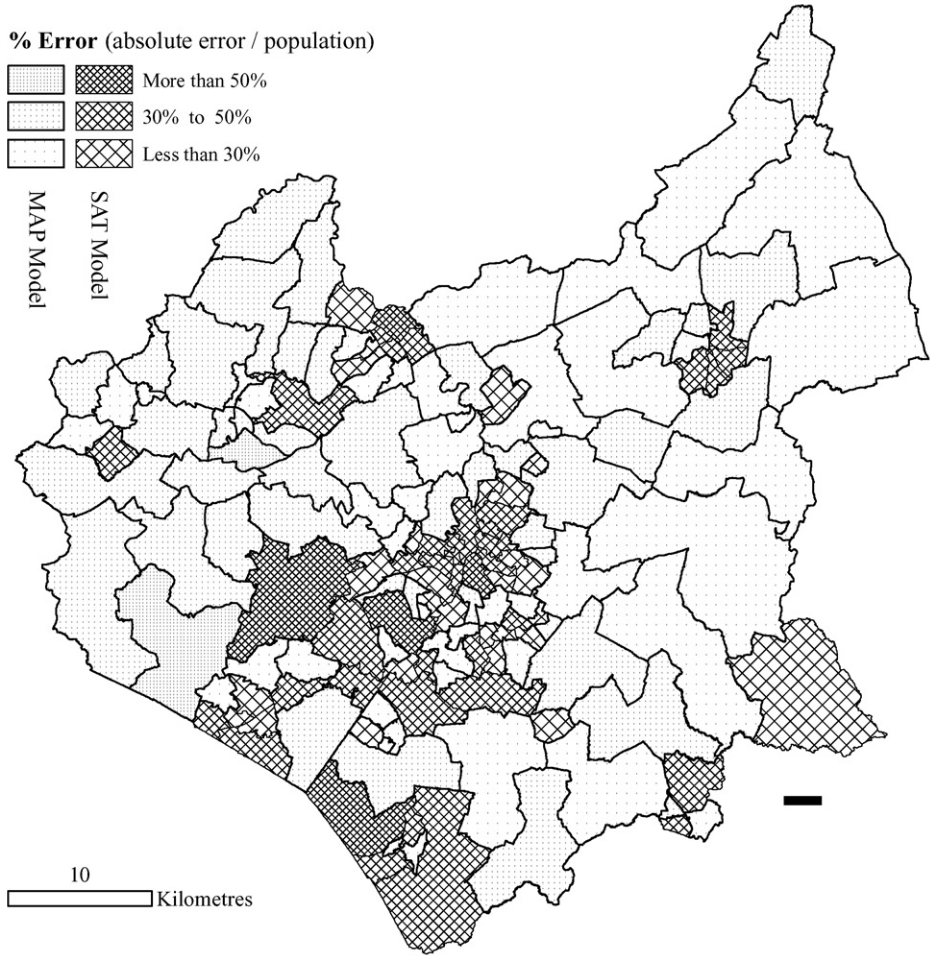

Fig. 5. Magnitude of total absolute error by ward relative to its population count.

the relative difference between models is small, the actual error could still be very large or indeed very small in respect to the ward population itself. To assess this quality it is necessary to express the total absolute ward error as a percentage of ward population. This information is mapped in Fig. 5, again classified and symbolized to indicate the best performing model in each case. Amongst wards where MAP is strongest there appears to be a general trend for errors to be higher in the west and lower in the east. The most plausible explanation is that western Leicestershire presents a somewhat more industrialised landscape than in the east, and hence errors begin to accumulate due to the presence of non-residential buildings. However, it must also be the case that it remains sufficiently rural to prevent SAT from becoming the superior model due to its ability to differentiate between these land uses, as was discussed earlier. In general, the larger errors tend to occur within or on the periphery of major urban centres. It is suggested that these arise due to a combination of factors including error in satellite image classification, the previously stated inability of the MAP model to distinguish residential from non-residential buildings, and possibly the presence of both low-rise and high-rise construction within individual ward boundaries.

6. Conclusions

Although intelligent areal interpolation methods offer clear and demonstrable advantages over simple areal weighting for estimating population in non-census reporting zones few signs exist within the research literature or elsewhere that they have been widely adopted by the GIS user community. If the proven benefits have been insufficient to encourage their widespread uptake it is perhaps the perceived costs of their implementation in terms of time, effort, and complexity that is the root of the problem. Potential users will always compare the advantages of improved target area estimates against the extra cost in terms of effort that is needed to implement the methodology, and thus the key to encouraging their wider usage is to ensure both simplicity and convenience. Areal interpolation based on the binary dasymetric method is conceptually simple, but most examples to date have used the relatively complex process of a maximum-likelihood classification of multi-spectral satellite imagery in order to obtain the redistribution mask. This paper has demonstrated that similar land cover information may be more readily accessible from raster pixel maps.

Using the Ordnance Survey 1:50000 raster map product it was possible to derive an appropriate urban mask using only simple recode and spatial filtering operations, of the sort available in any raster-capable GIS or desktop image editing software. These map data may be more familiar to those practitioners most likely to encounter the need for areal interpolation of population, and the speed and simplicity of the methodology compared to that based on satellite imagery is considerable. Furthermore, experimental results have shown that areal interpolation performance is approximately on-par with models that utilise satellite imagery. Differences in the character and nature of the satellite image-derived and map-derived dasymetric masks account for geographical patterns in the modeling performance. In particular, dasymetric interpolation based on raster map data tended to perform better in most rural areas, where classification errors can prove problematic for satellite imagery. However, the ability of multi-spectral satellite data to at least partially distinguish commercial and industrial sites from residential housing swings the balance back in its favour within urban fringe areas and in the town centres.

Both sources have their particular, and largely complementary, strengths and weaknesses which suggest that some form of combined mask could produce even better results, although to utilise such a product would run counter to the argument that simplicity and convenience are essential. On balance it is argued that raster pixel maps offer a useful addition to the repertoire of binary dasymetric interpolation methods, facilitating their rapid deployment and hopefully encouraging wider adoption amongst the GIS user community.

References

Burrough, P., & McDonnell, R. (1998). *Principles of geographical information systems*. Oxford: Oxford University Press.

Chrisman, N. (1991). The error component in spatial data. In D. Maguire, M. Goodchild, & D. Rhind (Eds.), *Geographical information systems: principles and applications* (Vol. 1, pp. 165–174). London: Longman.

Eicher, C., & Brewer, C. (2001). Dasymetric mapping and areal interpolation: implementation and evaluation. *Cartography and Geographic Information Science, 28*, 125–138.

Fisher, P., & Langford, M. (1995). Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A, 27*, 211–224.

Fisher, P., & Langford, M. (1996). Modelling sensitivity to accuracy in classified imagery: a study of areal interpolation. *The Professional Geographer, 48*, 299–309.

Flowerdew, R., & Green, M. (1989). Statistical methods for inference between incompatible zonal systems. In M. Goodchild & S. Gopal (Eds.), *Handing geographical information: Methodology and potential applications* (pp. 239–247). New York: Longman.

Flowerdew, R., & Green, M. (1991). Data integration: statistical methods for transferring data between zonal systems. In I. Masser & M. Blakemore (Eds.), *Handling geographical information* (pp. 38–54). Harlow: Longman.

Flowerdew, R., & Green, M. (1994). Areal interpolation and types of data. In S. Fotheringham & P. Rogerson (Eds.), *Spatial analysis and GIS* (pp. 121–145). London: Taylor and Francis.

Goodchild, M., Anselin, L., & Deichmann, U. (1993). A framework for the areal interpolation of socioeconomic data. *Environment and Planning A, 25*, 383–397.

Goodchild, M., & Lam, N. (1980). Areal interpolation: a variant of the traditional spatial problem. *Geo-Processing, 1*, 297–312.

Gregory, I. (2002). The accuracy of areal interpolation techniques: standardising 19th and 20th century census data to allow long-term comparisons. *Computers, Environment and Urban Systems, 26*, 293–314.

Heuvelink, G., Burrough, P., & Stein, A. (1989). Propagation of errors in spatial modelling with GIS. *International Journal of Geographical Information Systems, 3*, 303–322.

Holt, J., Lo, C. P., & Hodler, T. W. (2004). Dasymetric estimation of population density and areal interpolation of census data. *Cartography and Geographic Information Science, 31*, 103–121.

Langford, M., & Unwin, D. (1994). Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal, 31*, 21–26.

Liverman, D., Moran, E. F., Rindfuss, R. R., & Stern, P. C. (1998). *People and pixels: linking remote sensing and social science*. Washington DC: National Academy Press.

Martin, D. (1998). Optimizing census geography: the separation of collection and output geographies. *International Journal of Geographical Information Science, 12*, 673–685.

Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer, 55*, 31–42.

Mesev, V. (Ed.) (2003). *Remotely-sensed cities*. London: Taylor and Francis.

Moon, Z., & Farmer, F. L. (2001). Population density surface: a new approach to an old problem. *Society and Natural Resources, 14*, 39–49.

Mrozinski, R., Jr., & Cromley, R. (1999). Singly- and doubly-constrained methods of areal interpolation for vector-based GIS. *Transactions in Geographical Information Systems, 3*, 285–301.

Mugglin, A., Carlin, B., Zhu, L., & Conlon, E. (1999). Bayesian areal interpolation, estimation, and smoothing: an inferential approach for geographic information systems. *Environment and Planning A, 31*, 1337–1352.

NASA Goddard Space Flight Center (2005). 'Landsat 7 Science Data Users Handbook'. Available from http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html [Retrieved June 9, 2005].

Ordnance Survey (2005). '1:50000 Scale Colour Raster'. Available from http://www.ordnancesurvey.co.uk/oswebsite/products/50kraster/ [Retrieved June 9, 2005].

Reibel, M., & Bufalino, M. (2005). Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A, 37*, 127–139.

Sadahiro, Y. (1999). Accuracy of areal interpolation: a comparison of alternative methods. *Journal of Geographical Systems, 1*, 323–346.

US Geological Survey (2005). 'Landsat Project Website'. Available from http://landsat.usgs.gov/ [Retrieved June 9, 2005].

Wise, S. (1999). Extracting raster GIS data from scanned thematic maps. *Transactions in Geographical Information Systems, 3*, 221–237.

Xie, Y. (1995). The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems, 19*, 287–306.

Yuan, Y., Smith, R., & Limp, W. (1997). Remodelling census population with spatial information from Landsat TM imagery. *Computers Environment and Urban Systems, 21*, 245–258.