A combined pycnophylactic-dasymetric method for disaggregating spatial data: the example of agricultural land use

Alexis Comber[1], Chris Proctor[2], Steve Anthony[2]

[1] Department of Geography
University of Leicester
Leicester,
LE1 7RH, UK
ajc36@le.ac.uk

[2] Environmental Systems,
ADAS,
Wergs Road, Wolverhampton,
WV6 8TQ, UK
chris.procter@adas.co.uk
steve.anthony@adas.co.uk

1. Introduction

In this paper we present a novel method for disaggregating data that for reasons of confidentiality or expediency has been summarised over spatially coarser reporting units. The method combines pycnophylactic approaches with dasymetric ones. We use the June Agricultural Census (JAC) for England and Wales by way of illustration. The JAC has been subject to a number of changes in reporting units[1], for instance:

- 1998-2003 data reported within the NUTS hierarchy:
- From 2004 data reported within ONS super output areas (SOAs).

These agricultural census reporting units are referred to as “parents” and their higher aggregations as “grandparents” in the text below.

The JAC is an annual survey of agricultural activity which collects information from carefully selected agricultural holdings in England relating to land use, crops, livestock, labour, horticulture and glasshouse. However there are a number of issues associated with using the JAC for agri-environmental modelling:

- The survey collects information about individual agricultural holdings (farms) but to preserve farm confidentiality only aggregate data are released;
- The periodic change in reporting units makes comparisons across time difficult;
- Consistent units are needed as input into models (e.g. nutrient leaching) to satisfy EU directives
- Management of the Modifiable Areal Unit Problem (MAUP) as described by Openshaw (1984), where the patterns observed at one scale change at another (this is related to the Ecological Fallacy);
- The unreasonable assumptions of internal homogeneity associated with many areal interpolation techniques;

[1]http://www.defra.gov.uk/esg/work_htm/publications/cs/farmstats_web/Publications/data_documents/data_notes.htm#nuts

The approach presented in this paper seeks to address these fundamental issues associated with aggregated holdings level data by combining dasymetric techniques with pycnophylactic techniques. *Dasymetric* techniques are those where ancillary information about area under consideration constrains the assumption of homogeneity. *Pycnophylactic* approaches are those which generate a smooth surface from polygon based data whilst preserving the mass or volume property of the original data units.

2. Dasymetric Mapping

Dasymetric mapping provides a method for refining the distribution of land use within a spatial unit parish by incorporating additional data to provide a more realistic estimate of the actual distribution of the process under investigation within the units of analysis. It does this by bringing in additional relevant information to estimate the actual distribution of aggregated data within the unit of analysis. Dasymetric mapping was first described by Wright (1936) in his analysis of population distributions in Cape Cod, Massachusetts. Wright identified unpopulated areas within towns and calculated population densities for the remaining parts of the towns. In this way dasymetric mapping provides a means to statistically represent a more realistic surface than that provided by the aggregated data. It is a form of areal interpolation that incorporates additional knowledge or data relevant to the study area (Flowerdew and Green, 1991) by breaking down the artificial structure imposed by (usually) arbitrary political boundaries. Dasymetric mapping can reveal hidden data distributions and overcomes what Langford and Unwin describe as the “spatial discontinuities” created by the imposition of an artificial boundary and shows a more realistic distribution of the data (Langford and Unwin, 1994). In this way dasymetric mapping spatially refines aggregated data and has been used as an areal interpolation tool by a number of workers (Fisher and Langford 1995; 1996; Martin, et al., 2000; Eicher and Brewer 2001).

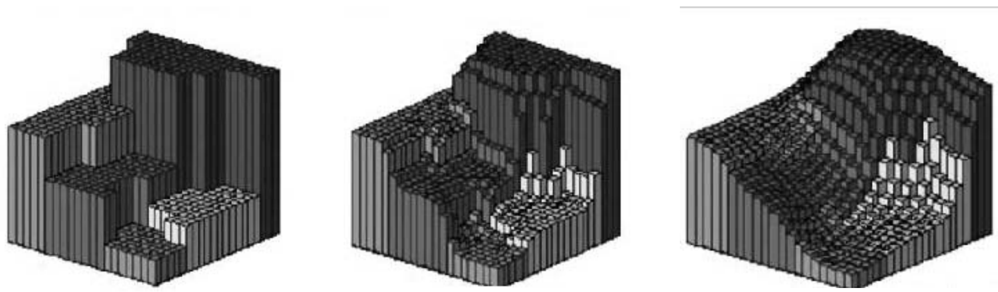
3. Pycnophylactic Interpolation

Tobler (1979) developed a procedure to generate a smooth surface from polygon based data which preserves the mass or volume property of the data called “pycnophylactic interpolation”. In pycnophylactic interpolation volumes are smoothed iteratively with the weighted average of nearest neighbours. During each iteration the total is adjusted to maintain the population count of the original (parent) polygon. The number of nearest neighbours used and of iterations determines the overall level of smoothing and is subjective. That is, the data totals for the original set of parent areal units are preserved during the transformation to a new set of areal units. Agricultural census data are only published as totals for spatial units of a certain size to ensure the privacy of individuals. Values such as the area of arable crops in each unit may be considered as a volume. Pycnophylactic interpolation computes a continuous surface from polygons-based data in 2 stages:

1) Parents are converted to a regular grid with height values (e.g. the area of agricultural land use) assigned to the grid points;
2) The height values are increased or decreased individually to make the surface smooth whilst simultaneously enforcing parent volume preservation.

The step 2 is repeated until the remaining “roughness” – the deviation from the ideal smoothness – reaches a user-defined threshold or the maximum iterations is reached.

Pycnophylactic interpolation transforms data from one system of units (parent zones) to another. During interpolation parts of the volume are not be distributed to cells within neighbouring parents, only within the original. Tobler presents a verbal overview of the approach[2] which is illustrated in Figure 1.



Zero Iterations Five Iterations Twenty-five Iterations

Figure 1. The process of pycnophylactic interpolation from http://www.csiss.org/streaming_video/csiss/tobler_pycno.htm

4. Method

The approach taken was to combine dasymetric techniques with pycnophylactic ones:

- Dasymetric methods to identify areas of non-agricultural land use within each $1km^2$;
- Pycnophylactic ones to iteratively sum the various land use areas over groups of $1km^2$ squares and then to compare those totals with the reported totals at the higher level geographies (parishes, output areas, etc.).

The objective was to develop a database describing general classes of non-overlapping land use. Data integration was through a series of discrete stages.

Stage 1. Identification of non-agricultural areas

A $1km^2$ dataset of non-agricultural areas was produced by combining the vector maps of urban land, woodland (Strategi), common land (Rural Surveys Research Unit, University of Wales), rivers, canals, railways, and roads (Strategi) with a $1km^2$ grid. For each cell in the grid this identified the areas of each type of (non-agricultural) land use. All of the linear features used were buffered by an appropriate amount prior to integration so as to fully represent the land area that they occupy. Each buffered vector data set was intersected with the $1km^2$ grids.

Stage 2. Comparison with the PAC at the district level

An estimate of the non-agricultural land in each parent was generated from its constituent $1km^2$. A comparison (at a grandparent level) showed an under estimation of non-agricultural land area in rural areas due to incomplete vector data particularly woodland areas and small hamlets.

Stage 3. Improvement of the non-agricultural land estimates

Non-agricultural uses were divided into 5 classes for comparison with Land Cover Map of Great Britain (LCMGB) classes: Sea, Inland Water, Woodland, Urban and Rough

2 http://www.csiss.org/streaming_video/csiss/tobler_pycno.htm

Grass. A full description of LCMGB can be found in Fuller et al. (2004). For each $1km^2$ the initial land use estimates from Stage 1 were treated as being incomplete (i.e. as a lower bound). The amount of each land use was increased by *adding in* equivalent values from the aggregated LCMGB classes until they matched the total amount of non-agricultural land for each grandparent. If at the end of this constrainment the non-agricultural land was still underestimated, the additional land required was added to the Rough Grass category – the category associated with the most uncertainty in the LCMGB classification.

Stage 4. Improvements in the agricultural land estimates

After Stages 1-3 the total crop areas (Arable) for {district areas} were within +/- 5% of those reported by the parent census totals. In order to ensure the total areas for each crop matched those of the census totals completely the crop areas were adjusted in the following fashion:

- Initial crop areas were estimated from the proportions reported in the parent data;
- The total area under each crop from the census data and the $1km^2$ was compared at the grandparent level;
- The grandparental differences were used to adjust the parent crop proportions.

The proportions of crops continued to be held on a per-parent basis, but were adjusted to ensure that the totals, when calculated at a grandparent level from the disaggregated $1km^2$ data, matched those of the grandparent census totals.

5. Results

We compare the distribution of land use at the level of $1km^2$ with aggregated habitat data for Kent, an English county (Table 1 and Figure 2). The important results, Arable and Grass as these are used to model different crop types and livestock distributions, show close correspondences.

Land type	ADAS Land use	Aggregated Habitat
Arable	138175	133069
Grass	65033	85031
Rough grazing	45979	23033
Woodland	37999	60207

Table 1. Areas (in hectares) for Kent of different land types from the aggregated habitat data, the modelled land use (NB not all land types are shown).

The value and relevance of this dataset and this method is demonstrated by the number of applications which have successfully incorporated this data:

- Phosphorous and Sediment Yield, CHaracterisation In Catchments (PSYCHIC)[3];
- NEAP-N - a model developed for modelling N losses at a national scale[4];
- National Ammonia Reduction Strategy Evaluation System (NARSES)[5].

[3] http://www.psychic-project.org.uk

[4] Silgram et al. 200. Intercomparison of national and IPCC methods for estimating nitrogen loss from agricultural land. Nutr. Cycl. AgroEcosystems, 60, 189-195.

[5] http://www.defra.gov.uk/science/Project_Data/DocumentLibrary/AM0101/AM0101_3668_FRA.doc

Each of these has been developed in response to national and EU policy and has been robustly validated.

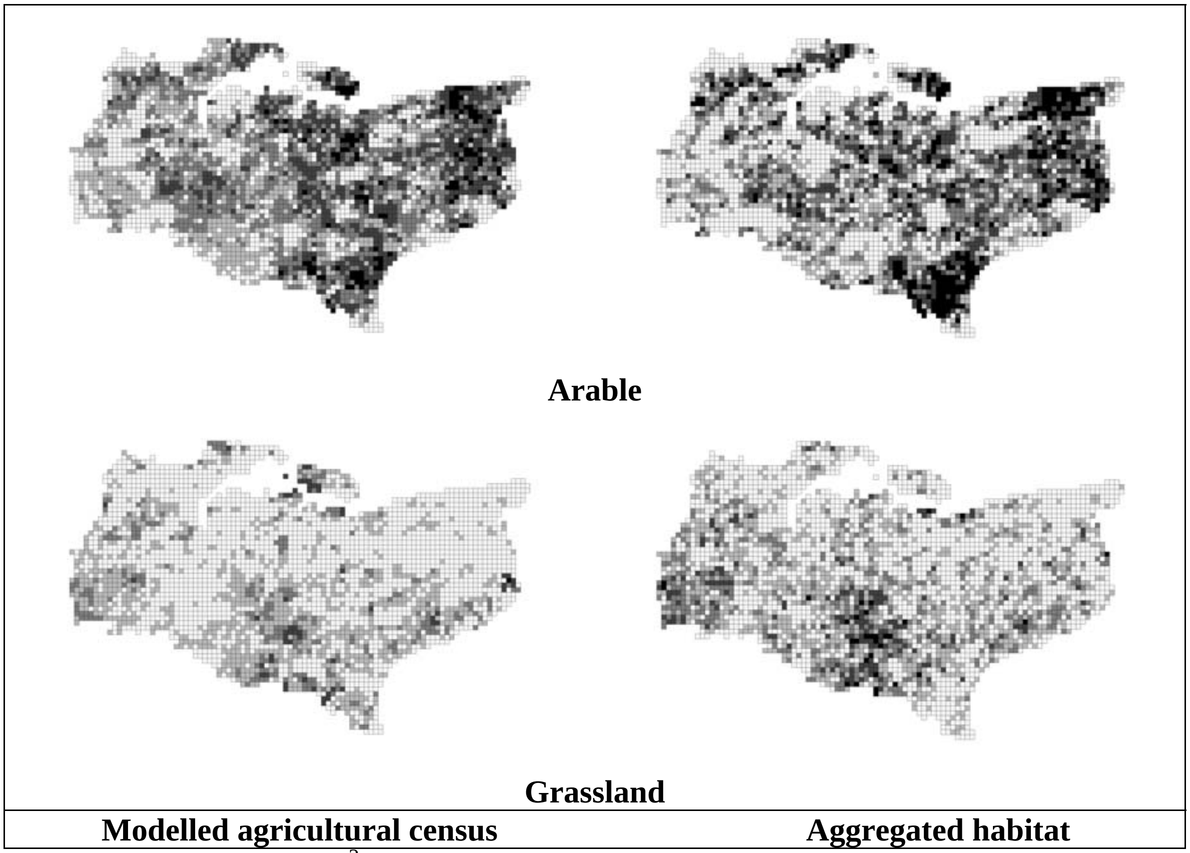


Figure 2 shows the $1km^2$ maps of arable and grassland distributions in Kent, classified into quintiles (lightest is < 20%, darkest >80%).

References

FISHER, P. F., and LANGFORD, M., 1996. Modeling sensitivity to accuracy in classified imagery: A study of areal interpolation by dasymetric mapping. *Professional Geographer*, 48: 299–309.

FISHER, P., and LANGFORD, M., 1995. Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A*:211–24.

FLOWERDEW, R., and GREEN M. 1991. Data integration: Statistical methods for transferring data between zonal systems. In Masser, I., and Blakemore, M. (eds.), *Handling Geographical Information*, pp. 38–54. Longman, New York.

LANGFORD, M. and UNWIN, D.J., 1994.Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal* 31:21–26.

MARTIN, D., TATE., N.J., and LANGFORD, M., 2000. Refining population surface models: Experiments with Northern Ireland census data. Transactions *in Geographical Information Systems* 4:342–60.

OPENSHAW, S., 1984. Ecological fallacies and the analysis of areal census data. *Environment and Planning A* 16:17–31.

TOBLER, W. R. 1979. Smooth Pycnophylactic Interpolation for Geographical Regions.' *Journal of the American Statistical Association* 74, 519–36

WRIGHT, J. K., 1936. A method of mapping densities of population with Cape Cod as an example. *Geographical Review* 26:103–10.

Biography

Lex Comber gained his PhD from the Macaulay Institute and the University of Aberdeen in 2001. Up to 2003 he worked as an RA on the EU REVIGIS project developing methods for integrating semantically discordant data. After a year in GIS consultancy with ADAS, Lex took up a lectureship at the University of Leicester where he now directs the MSc in GIS.