Linking UK Public Geospatial Data to Build 24/7 Space-Time Specific Population Surface Models

Samuel Leung[1], David Martin[2], Samantha Cockings[3]

School of Geography, University of Southampton, Southampton, SO17 1BJ, UK
Email: [1]Y.Leung@soton.ac.uk, [2]D.J.Martin@soton.ac.uk, [3]S.Cockings@soton.ac.uk

1. Introduction

Until recently any attempt to model population distribution over space has been largely dependent on georeferencing of resident population and therefore presents an abstract representation of night-time population pattern (Bhaduri, 2008). There are however, good arguments for modelling population at different times, incorporating population movements from seasonal to diurnal timescales so as to predict, for example, vulnerable population for rapid disaster relief or potential customer numbers during a working day. This paper presents early results from a publicly-funded project to develop space-time specific population surface models of the UK. The project extends Martin's (1996) adaptive kernel density approach into a spatio-temporal kernel density estimation for building gridded surface population models. We begin by briefly reviewing relevant methods, then move on to our conceptual modelling and data linkage and conclude with some early illustrative results.

2. Spatio-Temporal Population Modelling

The need of accurate time-specific population distribution models has long been recognised (e.g. Schmitt, 1956), progress to more fully-developed models is however suppressed by the weakness of GIS software for handling spatio-temporal phenomena and the absence of detailed data on short-term population movements. Ahola et al. (2007) present a temporal model of Helsinki, using moving kernel density estimation for visualization of results on a regular grid. The LandScan project (Bhaduri et al., 2007) is currently developing "daytime" and "night-time" 90m-resolution gridded models of the US. Sleeter and Wood (2006) transfer working and school populations from home to workplace and educational locations during the daytime using small area US census data. McPherson et al. (2006) also attempt to model population in the transportation system. Smith and Fairburn (2008) provide perhaps the most relevant UK-based work through a GIS database route. Many of these methods treat daytime and night-time the only temporal dimensions. Overlaid on historical time are complex cyclical timescales including time of day, day of week, term times, public holidays and seasons. Population is further redistributed temporally and spatially on an ad hoc basis by special events. Our models thus aim to extend beyond the binary day/night split to incorporate these additional temporal complexities.

The underlying logic behind our conceptual framework is that each member of the population is engaged in one principal activity at one location at any one time. These activities and their locations can be successively partitioned based on population subgroups as in Table 1.

Table1. Population groups, principal activities and default locations considered in Pop24/7 models.

Population group	Age bands	Principal daytime activity outside home	Time profiles	Default locations
Pre-school	0-3	Child care	Child care opening hours 08:00-18:00	Nurseries and other child care premises
School-aged	Primary 4-10 Secondary 11-15	Education	School hours 09:00-15:30	Schools
Working age	16-69			
-(a) Students	College 16-17 University 18-20+ (extending to 69)	Education	College and university hours 09:00 - 17:00	Colleges and university campuses
-(b) Economically active	16-69 but excluding (a)	Employment	Work hours by standard industry classification e.g. Office workers 09:00-17:00, Retail & trading 08:00-20:00, Hotel and catering 12:00-00:00	Business premises, residential addresses (work from home, health and social visitors, Private households with employed persons, etc.), transportation network (e.g. logistics and communication)
-(c) Economically inactive	16-69 but excluding (a)	Varied	No fixed time profile	No default location
Retirement	70 and above	Varied	No fixed time profile	No default location

Our approach extends the adaptive kernel estimation model (Martin, 1989) to a spatio-temporal modelling which accepts residential and non-residential centroid locations, each of which is associated with a temporal profile describing the presence of population at different times. This combines the surface modelling approach with the spatio-temporal conceptualization of Ahola et al. (2007). Centroids are divided into “sources” and “destinations”. Each has a “region of influence” representing the area over which people may travel to that location. The temporal and spatial information interact e.g. a school will generate most travel in its area of influence at the start and end of the school day. Adjustments are made for interregional and international passenger estimates applicable to the target time. The general approach is extensible to accommodate future data series and, importantly, retains the volume-preserving characteristics of the original algorithm.

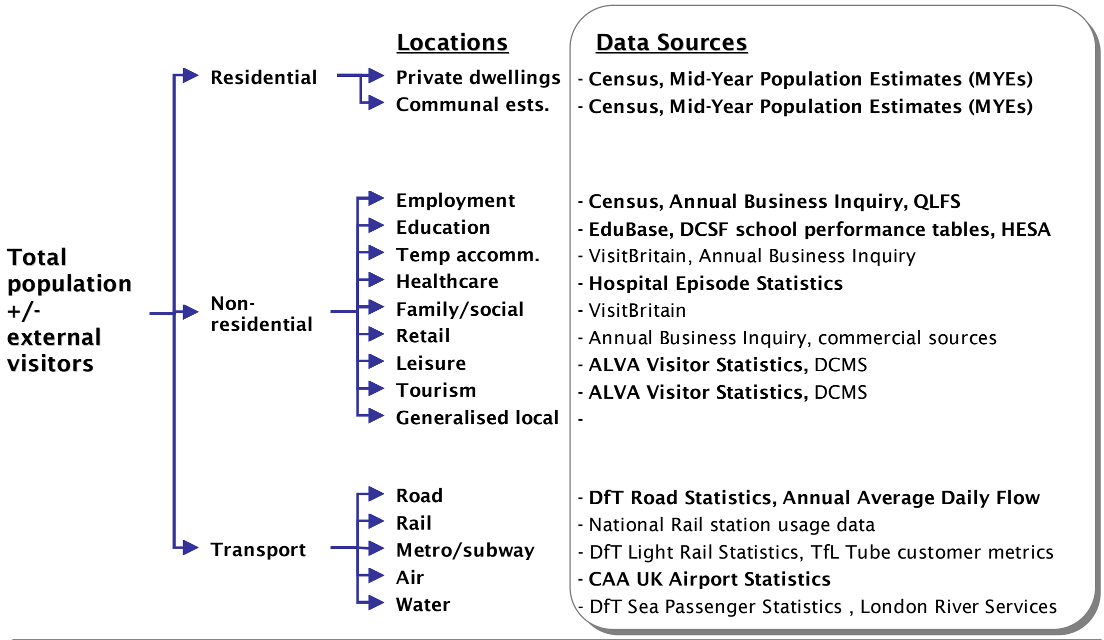
3. Data Modelling Framework

The residential population counts utilised here are based on the 2001 UK census and the annual Mid-Year Population Estimates (MYEs) released by the Office for National Statistics (ONS). By linking the census data with the residential postcodes listed on the National Statistical Postcode Directories (NSPD), the resident population can be redistributed to a finer geographical unit without compromising individual confidentiality.

Spatio-temporal data describing non-residential locations of population sub-groups tend to be patchy in coverage and/or frequency. Smith and Fairburn (2008) provide a comprehensive review of relevant UK datasets. Many are now collated within the

Neighbourhood Statistics Service (NeSS). The recent UK government's decision to free public sector data for public use and support for the linked data initiative facilitate further data access and linkage (Tennison, 2010). Many of the information required for modelling people's day-time locations including education, workplace and health care can now be interrogated from the governmental open data site at http://data.gov.uk. More specific data e.g. school censuses, annual employment counts; hospital statistics, tourist attraction visitors' figures can still be obtained directly from the responsible government departments and other public bodies.

Population in transit, unlike the two data categories above, cannot be associated with stationary point locations. The varying traffic volumes for different vehicles type at different road class add further uncertainty when assigning non-resident population to the transport network. We address this challenge of allocating the mobile population to the road network by linking the volume and distribution of traffic for each vehicle and road type spatially with the national road map data. The rail and ferry transport modes will be incorporated in the later stage. We are also aware of the potentials of real-time population monitoring data e.g. retail footfall, mobile telephone usage, vehicle tracking etc. but integration of these sources is beyond the scope of our current project. Figure 1 summaries the datasets currently used (highlighted in bold) and those to be employed later within the project under the "Residential", "Non-residential" and "Transport" activity categories.



Acronyms: **QLFS** Quarterly Labour Force; **DCSF** Department for Children, Schools and Families; **HESA** Higher Education Statistics Agency; Survey; **DCMS** Department for Culture, Media and Sport; **ALVA** Association for Leading Visitor Attractions; **DfT** Department for Transport; **TfL** Transport for London; **CAA** Civil Aviation Authority

Figure 1. Population activities and candidate datasets covering all or part of the UK.

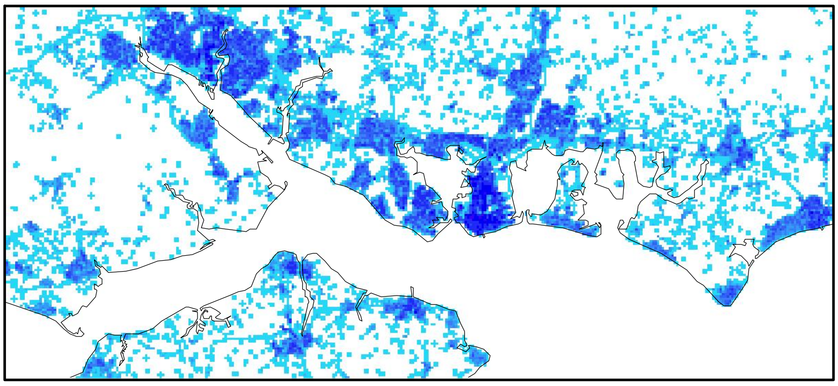
4. Empirical Example

We present here a UK study area that covers Portsmouth and Southampton and the areas between them. The M27 motorway and the Solent along the coastline provide a diverse and continuous environment for modelling a range of real and hypothetical scenarios such as intercity population movements, nature disasters like flooding and other events e.g. road accidents and oil refinery explosions.

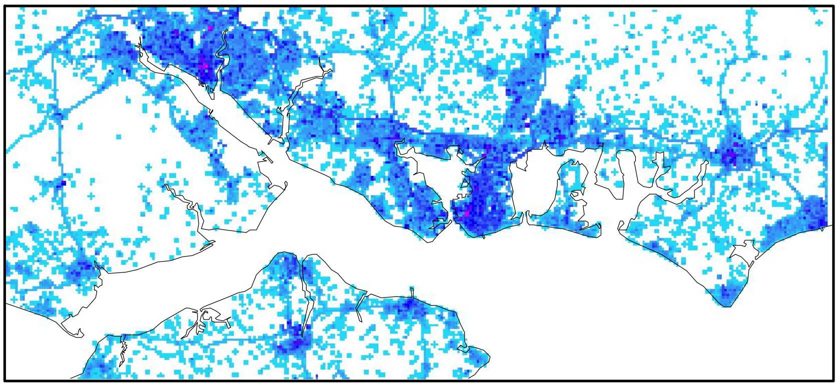
The 2001 census-based counts are first adjusted to 2006 before being divided into four age bands as in Table 1. Census baseline information on 2001 workplace populations has been reconciled with Annual Business Inquiry (ABI) datasets at the lower layer super output area (LSOA) level and reweighted to OA to obtain estimated workplace populations. Average employee numbers of large and small businesses at the local authority level are derived from the ABI workplace analysis data. Using the large user and small business count fields in NSPD, the employee number at each workplace during the working hours is given by the product of the number of businesses and their average size. Term-time population estimates for all educational establishments including state-maintained nurseries, primary and secondary schools are available from the EduBase education portal. Further and higher education institutions will be included in the portal but currently those figures can be extracted from the school and college performance tables published by the Department for Children, Schools and Families (DCSF) and the Students and Qualifiers data tables by the Higher Education Statistics Agency (HESA). Each of the aggregate datasets has been re-weighted onto postcode centroid locations. Many documentary sources are consulted to produce the most appropriate time profiles and areas of influence for each centroid type. School opening hours and catchment radii are straight forward approximations. Additional census and survey sources like the national Quarterly Labour Force Survey are used to derive the time profiles of different employment sectors and the estimated travel to work distance bands. Locally-based activities like a short trip to a neighbour, corner shop and clinic fall below the spatio-temporal resolution of our modelling but do not involve the transfer of population out of residential areas.

Similar to Smith and Fairburn's (2008) National Population Database (NPD), we incorporate the Department for Transport's (DfT) Annual Average Daily Flow (AADF) traffic data in our models. The AADF count point data are interpolated and rasterized over the Ordnance Survey Mastermap's Integrated Transport Network (ITN) map data to form a "background" layer. This background layer defines cells to which it is valid to allocate population counts, containing weights reflecting the capacity of the transportation network to contain population "in transit".

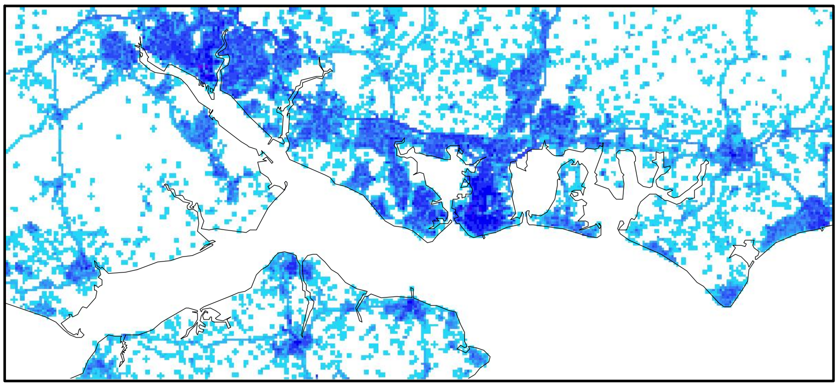
Working with the available datasets shown as bold text in Figure 1, we have used *SurfaceBuilder* to statically model the study area population at several key time intervals to prove the concept. Figure 2 illustrates early results, representing population densities for three reference times on a typical term-time weekday in 2006 at 200m resolution. Note that the figures do not yet contain correctly calibrated population in transit between locations. The maps below display the daily population redistributions of an entirely residential pattern at 02:00; daytime many-at-work/school distribution at 09:00 and mixed rush hours and evening night-out model at 18:00.



(a) Night-time resident population, dark shading indicating a high population density.



(b) Workday model at 09:00 when children and workers arrive in schools and workplaces.



(c) Evening model showing rush hour traffic on roads and residual business and retail activities in city centres.

Figure 2. (a)-(c). Gridded population models for south Hampshire study region on a term-time weekday in 2006. Coastline and road network data © Crown Copyright/database right 2010. An Ordnance Survey/EDINA supplied service.

5. Next Steps

The current stage of the project involves intensive data integration, software programming the development of standard model runs for specified timescales. The model software and the standard outputs will be available for download and use via the Internet. To facilitate a wider reuse of the tool, we are closely monitoring the open data initiatives with an expectation to incorporate comparable and free-of-charge alternative datasets. We are also actively exploring the use of 3-dimensional block models and shaded polygon maps, overlaid in Google Earth (Figure 3) which, combined with time slider tools, can provide a powerful and multiscale visualization option for exploring time-space population distributions together with recognisable geographical features and placenames.

(a) Night-time resident population of Portsmouth. The column height of each 200m grid cell represents the estimated population count at 02:00.

(b) Day-time population distribution corresponding to Figure 2(b), high columns can be found in city centres or major out-of-town industrial estates and retail parks.

(c) Evening model showing a less concentrated population distribution. Population counts at the high volume cells are gradually redistributed to home locations via the transport network.

Figure 3 (a) – (c). Population densities over three different time periods in Portsmouth, rendered as a 3-dimensional KML layer using the Google Earth™ mapping service. © 2010 Tele Atlas © 2010 Infoterra Ltd & Bluesky © Europa Technologies Image © 2010 Getmapping plc.

Acknowledgements

Economic and Social Research Council (ESRC) award RES-062-23-0081. AADF: DfT. ABI: NOMIS, licence NTC/ABI07-P3020. EduBase: DCSF. ONS 2001 Census Standard Area Statistics: ESRC Census Programme, Census Dissemination Unit, Mimas (University of Manchester). NSPD: ONS, ESRC Census Programme, EDINA (University of Edinburgh). QLFS: Economic and Social Data Service, usage number 40023.

References

Ahola T, Virrantaus K, Krisp JM and Hunter GJ 2007, A spatio-temporal population model to support risk assessment and damage analysis for decision-making. *International Journal of GIS*, 21:935-953.

Bhaduri B 2008, Population distribution during the day. In *Encyclopedia of GIS* Shekhar S and Xiong H (eds) 880-885, New York: Springer.

Bhaduri N, Bright E, Coleman P and Urban ML 2007, LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal*, 69:103–117.

McPherson TN, Rush JF, Khalsa H, Ivey A and Brown MJ 2006, A day-night population exchange model for better exposure and consequence management assessments. Sixth Symposium on the Urban Environment, American Meteorological Society, Atlanta, USA, 28 January – 3 February.

Martin D 1989, Mapping population data from zone centroid locations. *Transactions of the Institute of British Geographers*, 14(1):90-97.

Martin, D 1996, An assessment of surface and zonal models of population. *International Journal of GIS*, 10(8):973-989.

Schmitt RC 1956, Estimating Daytime Populations. *Journal of the American Planning Association*, 22:83-85.

Sleeter R and Wood N 2006, Estimating daytime and nighttime population density for coastal communities in Oregon. In proceedings of the Annual Conference of the Urban and Regional Information Systems Association, Vancouver, Canada, 26-29 September.

Smith G and Fairburn J 2008, Updating and improving the National Population Database to National Population Database 2. *Research Report 678*, London: Health and Safety Executive.

Tennison J 2010, Why linked data for data.gov.uk? White Paper, Norwich: The Stationery Office.