Using new and emerging forms of data to produce enhanced spatiotemporal population estimates

Cockings S*[1], Martin D†[1]

[1]Geography and Environment, University of Southampton

January 12, 2018

Summary

Applications in fields such as emergency planning and health require accurate estimates of population present in specific areas at specific times. New and emerging forms of data from sources such as sensors and APIs can provide signals of population activity through time for sectors such as retail, leisure and transportation, which are not well captured in official statistics. We present initial results from the ESRC-funded Pop247NRT (Population247 Near Real-Time) project, demonstrating a suite of methods to extract, process and integrate such data with more conventional sources to produce enhanced spatiotemporal population estimates for a study region in south west England.

KEYWORDS: spatiotemporal, population, dynamic, retail, new and emerging forms of data

1 Introduction

Policy-makers and practitioners frequently require information about how many people are in specific places at specific times. Despite this requirement, much population-related mapping and analysis relies on traditional census data but this does not typically provide the necessary spatiotemporal granularity. Martin et al (2015) developed an innovative modelling framework (Population247) and associated software tool (SurfaceBuilder247) for producing enhanced spatiotemporal population estimates. These have been applied to flooding and radiation in the UK (Smith et al., 2015; Alexis-Martin, 2016) and exposure assessment in Italy (Renner et al, In Press). To date, these applications have primarily involved the integration of administrative data on places of residence, work, education and health. Now, new and emerging forms of data (such as sensor data, APIs and commercially sensitive data made available via ESRC's Big Data Network investments) provide opportunities to produce space-time patterns for more dynamic, hard-to-count, population activities such as retail, leisure and transport. The ESRC-funded Pop247NRT (Population247 Near Real-Time) project aims to develop methods for harvesting, processing and combining new, emerging and existing datasets in order to produce enhanced time-specific population estimates for more informed decision-making and policy formulation in the health, emergency/crisis response and national security sectors. This paper describes the Pop247NRT methodological approach and innovations, presents an empirical example of its use in modelling population engaged in retail activity, and outlines the next steps in the project.

2 Pop247NRT: Methodological approach and innovations

The Population247 modelling framework (Martin et al., 2015, Figure 1, p.6) redistributes proportions of known population totals in demographic sub-groups (e.g. age groups) from residential locations to other locations which are known to host relevant population activities at a target time. These counts can then be recombined to produce population estimates for a given time, date and spatial resolution

* s.cockings@soton.ac.uk
† d.martin@soton.ac.uk

(typically a 100m or 200m grid). Background data on road networks, traffic densities and land are used to inform this redistribution. Other researchers have developed methods for producing enhanced spatiotemporal estimates (e.g. Ahola et al, 2007; Batista e Silva et al, 2016; Smith and Fairburn, 2008), but all are limited in some way; most often by their generic day/night or other broad time periods, or by their lack of an underlying volume-preserving model.

Population247 requires the compilation of a data library containing information about locations, magnitudes, and time profiles, for places where population activities are known to occur. The Pop247NRT project is developing an innovative suite of methods to harvest, process, calibrate and integrate data from new, emerging and existing sources in order to derive this data library. Its aim is not to compile an exhaustive data library for the area of study, but rather to develop methods for selected types of data, which can then be applied to data of the same type within different sectors or from different sources, both by this project and other users. Examples of types of data include those derived from sensors (e.g. retail footfall, traffic), APIs (retail opening hours, traffic) and data harvested from the internet (popular times, transport hub and visitor attraction statistics). Such data may be static, archived live or live. The project is particularly focusing on activities within the retail, leisure and transport sectors as these tend to contain the most dynamic and least predictable redistributions of population through time. The following section illustrates the application of these methods in an empirical example: the modelling of population engaged in retail activity for a study area in south west England.

3 Empirical example: modelling retail activity in a study area in south west England

3.1 Study area

Figure 1 shows the Pop247NRT study area, a 50km^2 area centred on Bristol and Bath in south west England (Figure 1), with a surrounding 25km buffer to allow for population travelling in and out of the area. This study area was selected to meet the requirements for collaborative case studies being developed with the project partners (Defence, Science and Technology Laboratory, Health and Safety Executive and Public Health England).

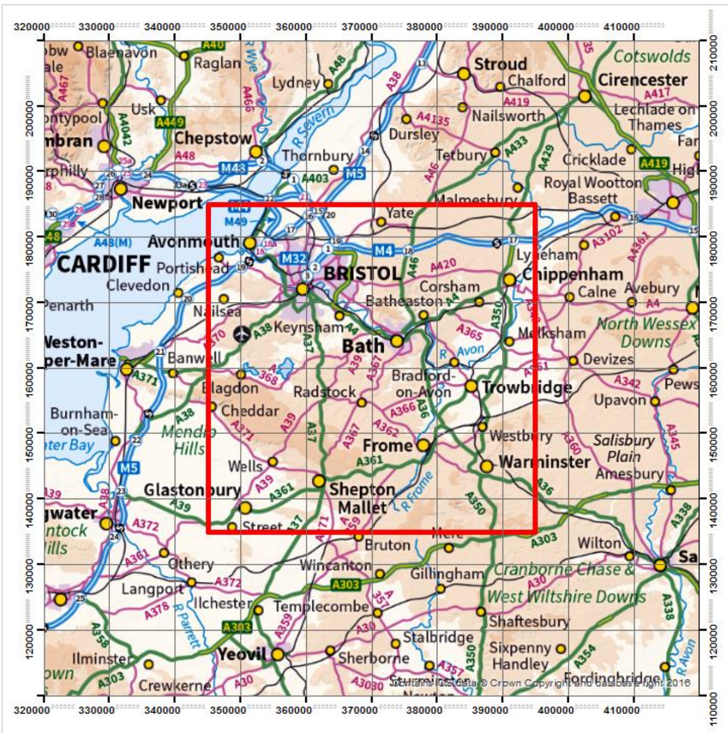


Figure 1 Pop247NRT study area

3.2 Data

Table 1 details the datasets employed in the example. As noted above, these represent different types of data, including static open data, live data harvested from the internet and a safeguarded aggregate dataset derived from sensor data.

Table 1 Datasets employed in empirical example

Dataset	Year	Data provider	Source
Classification of workplace zones (COWZ-EW)	2011	University of Southampton	http://cowz.geodata.soton.ac.uk/
Ordnance Survey (OS) Points of Interest (POI)	2016	Ordnance Survey	http://digimap.edina.ac.uk
SmartStreetSensor footfall	2016 -17	Local Data Company/ Consumer Data Research Centre	https://www.cdrc.ac.uk/
GeoLytix Retail Points	2016	GeoLytix	www.geolytix.co.uk
Google Popular Times	2017	Google	https://www.google.co.uk/maps

3.3 Methods

Figure 2 shows how the datasets are combined to model population engaged in retail activity in the study area. Two distinct types of activity are identified: (i) retail activity occurring at a stand-alone site (Figure 2a), and (ii) mixed retail activity taking place in a localised area (Figure 2b). The former represents sites identifiable as a single entity e.g. a named supermarket at a particular location, for which data specific to that site are available; whereas the latter represents a fuzzily-defined area with a mix of predominantly retail but also other activities e.g. a high street or a large shopping mall, for which data are not available for each individual retail entity and which experiences generic footfall.

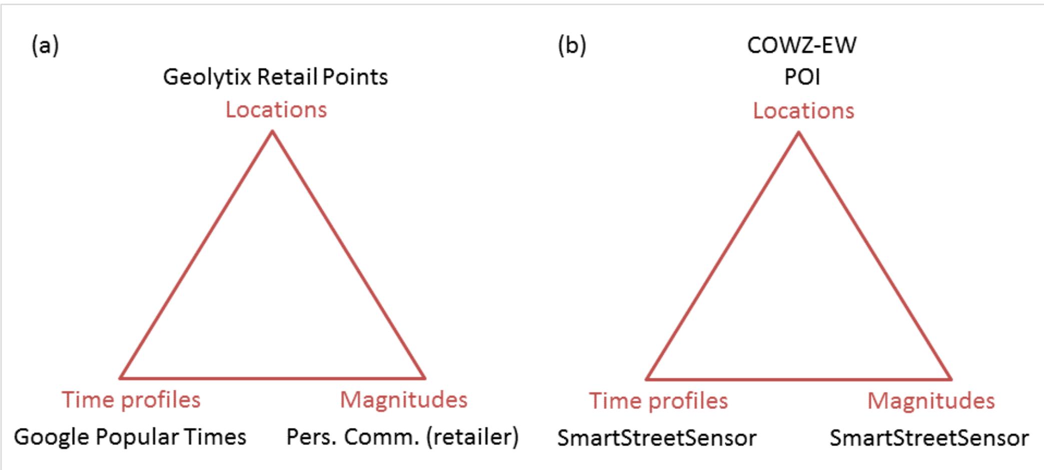


Figure 2 Datasets employed to model population engaged in retail activity for (a) supermarkets, and (b) areas of mixed retail activity

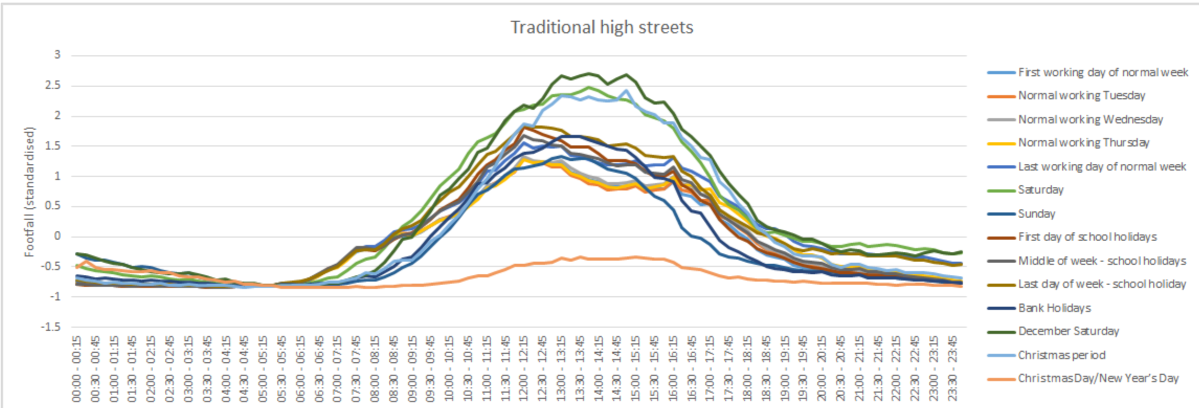
For (a), point locations of supermarkets are obtained from GeoLytix Retail Points data and rasterised to a 100m grid: only size bands C and D are included as the other (smaller) categories do not generate sufficient hourly/daily footfall. Site-specific time profiles are obtained from Google Popular Times data, enabling typical hourly percentage of capacity estimates by day of week to be derived for size bands C and D. Magnitudes are not openly available from any known source; indicative footfall estimates for stores in the study area in the GeoLytix size bands were therefore obtained by personal communication with a retail analyst.

For the mixed retail areas in (b), workplace zones (WZs) containing a specified number of retail-related OS POI sites are identified. The selected WZs are members of the Retail, Servants of society and Top jobs COWZ-EW supergroups, as might be anticipated. The spatial extent of retail areas within these WZs is then further refined by retaining only those 100m grid squares containing at least three retail-related POIs within the selected WZs. Time profiles are derived from analysis of SmartStreetSensor aggregate footfall counts from 176 sensors in 12 selected cities. 5 minute counts are aggregated to 15 minutes and classified by Traffic England day type (with the addition of a December Saturday category) and retail group (i.e. COWZ-EW group/supergroup, GeoLytix band C and D). Typical time profiles, giving the percentage present at a specific time, are then generated across the series for these retail groups and day types. Magnitudes are similarly derived from the SmartStreetSensor data, providing an average total capacity for each retail group/day type grouping.

3.4 Preliminary results

Figure 3 shows illustrative standardised footfall per 15 minute interval by day type for sensors located in WZs classified as (a) Traditional high streets, and (b) Top jobs. Clear differences can be seen between the working week, weekend and holiday periods. The Top jobs profile also displays an early morning peak and ‘night-time economy’ on Fridays and Saturdays.

(a)



(b)

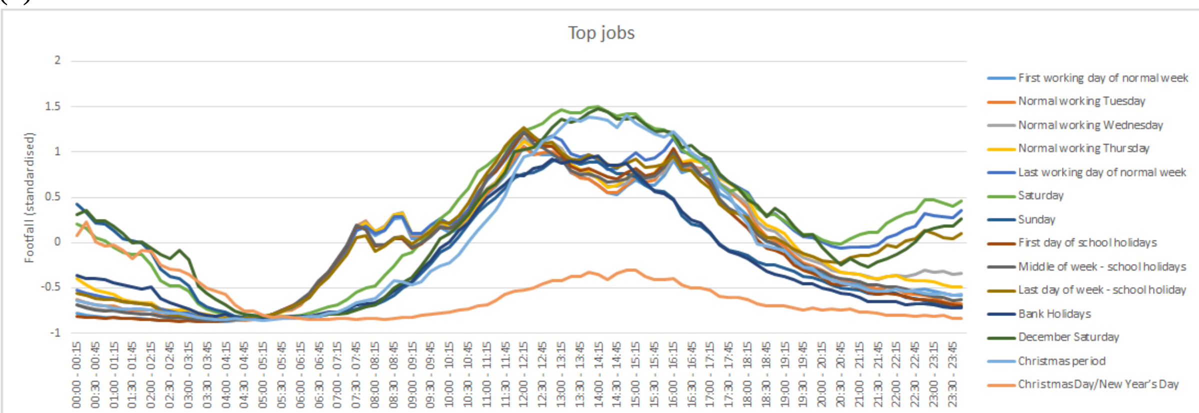


Figure 3 Time profiles - standardised footfall per 15 minute interval by Traffic England day type for sensors located in workplace zones classified as (a) Traditional high streets (b) Top jobs

Figure 4 shows the refined spatial extents of the retail groups in Bristol city centre. The Broadmead shopping district is clearly identifiable in the north west of the image, classified as Shop until you drop. The Asda Bedminster Superstore appears as Geolytix band D to the south of the River Avon.

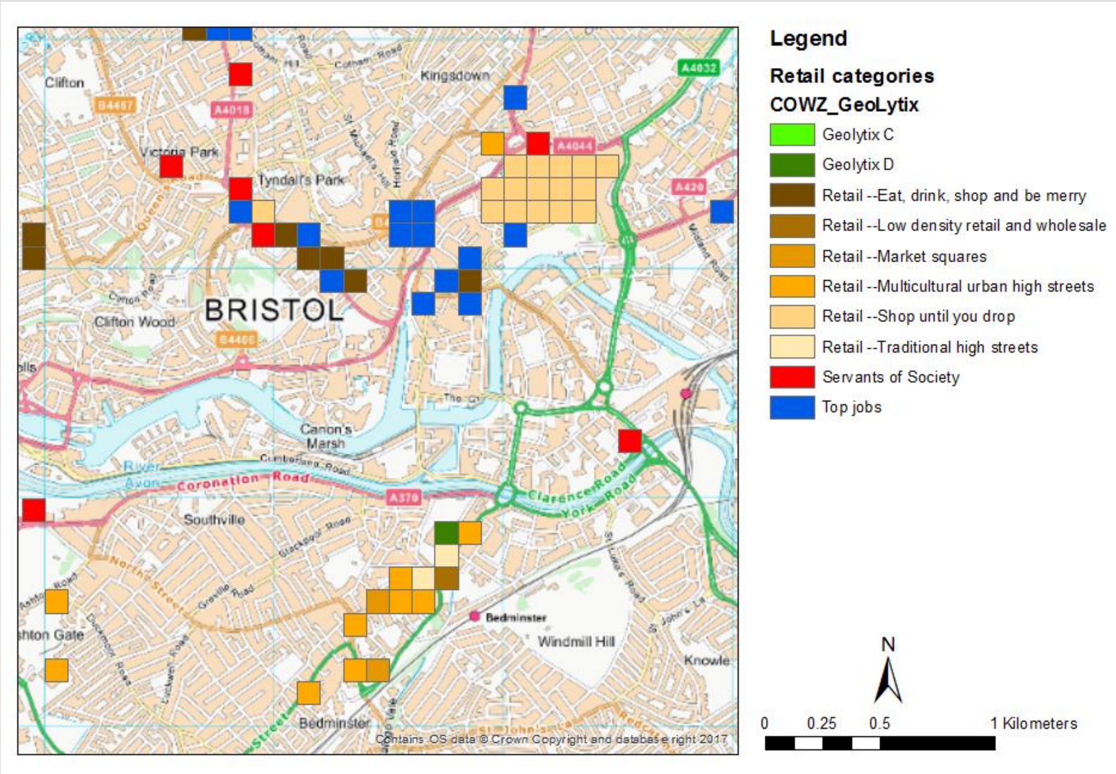


Figure 4 Refined spatial extent of retail groups in Bristol city centre (100m grid)

4 Next steps

These locations, time profiles and magnitudes for the retail sector will be used as inputs to SurfaceBuilder247, together with equivalent data for other sectors (residential, workplaces, health, education, leisure, transport, plus background layers including road traffic). Time- and date-specific population estimates will then be generated and employed within the partner case studies. Finally, the challenges involved in implementing the Pop247NRT methods within and beyond the partners' sectors will be discussed at a stakeholder workshop.

5 Acknowledgements

ESRC Award ES/P010768/1. Co-investigators: Tom Charnock (PHE), Nick Gibbins (University of Southampton), Glen Hart (Dstl), William Holmes (HSE). Project team: Jeremy Austin, Julia Branson, Andrew Campbell-Sutton, Gemma Gubbins, Andrew Harfoot, Duncan Hornby, Jason Sadler.

6 Biography

Samantha Cockings is Associate Professor in Socioeconomic Applications of GIS at the University of Southampton. Her research interests include automated zone design, spatiotemporal population modelling and applications of GIS to health.

David Martin is Professor of Geography at the University of Southampton. His research over the last 30 years has spanned many aspects of population representation in GIS, including automated zone design methods for census outputs, and grid-based population mapping with an increasing focus on spatiotemporal dynamics.

References

Ahola T, Virrantaus K, Krisp J, Hunter G (2007) A spatio-temporal population model to support risk assessment and damage analysis for decision-making *International Journal of Geographical Information Science* 21 (8), 935–53

Alexis-Martin B (2016). *RADPOP: A New Modelling Framework for Radiation Protection*. Thesis submitted for the degree of PhD, University of Southampton, pp.196

Batista e Silva F, Craglia M, Freire S, Rosina K, Lavalle C, Marin M, Schiavina M (2016). *Enhancing activity and population mapping: Exploratory research project interim report*. JRC Technical Report no. EUR 28335 EN. doi:10.2791/056071

Martin D, Cockings S, Leung S (2015). Developing a flexible framework for spatiotemporal population modelling. *Annals of the Association of American Geographers*, 105 (4), 754-772.

Renner K, Schneiderbauer S, Pruß F, Kofler C, Martin D, Cockings S. Spatio-temporal population modelling as improved exposure information for risk assessments tested in the Autonomous Province of Bolzano. *International Journal of Disaster Risk Reduction*, doi: 10.1016/j.ijdrr.2017.11.011 [In press, available online]

Smith A, Martin D, Cockings S (2016). Spatio-temporal population modelling for enhanced assessment of urban exposure to flood risk. *Applied Spatial Analysis and Policy*, 9, 145, doi:10.1007/s12061-014-9110-6

Smith G, Fairburn J (2008) *Updating and improving the National Population Database to National Population Database 2. Research Report 678*. Health and Safety Executive, London. http://www.hse.gov.uk/research/rrpdf/rr678.pdf (Accessed 11/01/18)