CESifo Economic Studies, 2018, 1–37
doi: 10.1093/cesifo/ify002
Original article

Take Me to the Centre of Your Town! Using Micro-geographical Data to Identify Town Centres

Paul Cheshire*, Christian Hilber†, Piero Montebruno‡ and Rosa Sanchis-Guarner‖

*Department of Geography and Spatial Economics Research Centre (SERC/CEP), London School of Economics. e-mail: p.cheshire@lse.ac.uk, †Department of Geography and Spatial Economics Research Centre (SERC/CEP), London School of Economics. e-mail: c.hilber@lse.ac.uk, ‡Department of Geography, University of Cambridge. e-mail: pfm27@cam.ac.uk and ‖Imperial College Business School and SERC/CEP, London School of Economics. e-mail: r.sanchis-guarner@imperial.ac.uk

Abstract

We often talk about 'Town Centres' (TCs), but defining their location and extent is surprisingly difficult. Their boundaries are hard to pin down and intrinsically fuzzy. Nevertheless, policymakers often speak or act as if their definition was self-evident. The Dutch and later the British governments, for example, introduced very specific policies for them without ever clearly defining what or where they were. In this article, we propose a simple methodology to predict TC boundaries and extent. Using a range of micro-geographical data, we test our method for the whole of Great Britain in an attempt to capture all the dimensions of 'town centredness' in a 3D surface. We believe this is a contribution in its own right but is also an essential step if there is to be any rigorous analysis of TC or evaluation of policies directed at them. Our method should contribute to improve not just debates about cities, shopping hierarchies, and TCs but also to other more general debates where people and policy proceed ahead of any clear definition of what are the objects of interest. (JEL codes: L81, R12, R52)

Key words: town centre, planning, retail sector, land use

1. Introduction

Imagine you are anywhere in a city—London, Lyon, Berlin, and Wolverhampton—and you know that city well. Suddenly, someone comes up to you and asks, 'Could you tell me where the town centre is?' This could appear to be a simple, even a trivial, question, but it is not. In fact, in many instances, it proves to be surprisingly hard to answer. The aim of this article is to devise a method which could provide a response and not just a response but an answer which meets the criterion of being replicable. If you apply the method to a different town, your answer will be strictly comparable, and you would get the same answer asking different people so long as they applied the method.

This question has a particular salience, since, in many countries, there are influential urban policies that apply to 'Town Centres' (TCs).[1] But, if we cannot define the boundaries of these areas, not only can we not identify the actual areas the policies are supposed to apply to, we cannot evaluate any effects such 'Town Centre policies' may have on outcomes. Our aim in this article is to design, explain, apply, and test a method to answer this apparently trivial question. We are not concerned with *why* the TC is sought. Instead, we explore and provide an operational answer. We do this in the specific context of Britain but would suggest both the question and our approach have significant application elsewhere.

Our interest in identifying and predicting TC space arose as one part of an investigation into the effects of 'Town Centre First Policy' (TCFP) on shoppers' travel patterns (Cheshire et al. 2017), as adopted in England in 1996 (Department of the Environment 1996). This policy, remarkably similar to that applied in The Netherlands some 15 years earlier (Evers 2002), was intended to 'redirect development, not just in retailing but in all "key Town Centre uses," including leisure, office development and other uses, such as restaurants, to Town Centres', although the policy most notably affected the location of new retail development. So TC protection strengthened in England just as in the Netherlands it was becoming more flexible to support the competitiveness of the retail sector (Evers 2002). As was shown in Cheshire et al. (2015), TCFP policy did, indeed, have a substantial negative impact on total factor productivity in the English supermarket sector.

The avowed purpose of policies to support TCs was to maintain their 'viability' or, in the case of The Netherlands, to ensure that the distribution of retail outlets corresponded to the urban hierarchy. But in England TCFP was specifically introduced to facilitate 'linked shopping trips' and allow shopping trips to be undertaken using public transport—partly with the aim of reducing their carbon footprint but also for equity purposes: to protect access to shops of those without cars. Evaluation of such policies, therefore, necessarily requires information on patterns of shopping trips and changes in the extent to which shopping destinations are located in TCs.

To begin to assess the impacts of the TCFP—or any policy aimed at TCs—it is thus necessary to have definitions of where and what TCs are[2] and to be able to apply the same definitions to contexts where TCFPs were not introduced. TCFP was implemented, however, with no such definitions. While for England and Wales TCs were subsequently defined in research commissioned by the relevant government department (ODPM 2004),[3] these were not official nor are they enforced: 'It should be noted that these areas [Areas of Town Centre Activity] have no policy status and are not town centres for policy purposes – such centres will be designated in development plans' (ODPM and CASA 2002). To provide the tools for such an evaluation, the focus of the present article is to develop a method for predicting and estimating the location and extent of TC space in both England and Wales and in Scotland. In addition, we would expect our method to be widely applicable.

1 For example, Denmark, Germany, The Netherlands, or Sweden: see Guimarães (2016) for a recent survey of some of these.

2 The Scottish government produced a glossy 138 page handbook called the Town Centre Toolkit in 2015 without ever defining what a 'Town Centre' was.

3 ODPM, created in 2002 and predecessor of the DCLG.

To do this, we first obtained data on TCs as defined for 2000 from the Department for Communities and Local Government (DCLG).[4] Even with the caveat that they have 'no policy status', these 'official'[5] TC definitions are the most reliable and accurate definitions of TC space in England and Wales. They consist of GIS shapefiles for 1075 TCs, of which the majority are defined as 'Areas of Town Centre Activity' (ATCA) and 46 as 'Retail Cores' (RCs)—which overlap and are sub-centres of the ATCAs. From these shapefiles we obtain the centroids of the England and Wales TCs (called DCLG TCs in what follows). This identifies the central point in each town or city. Separately, we obtain a list of alternative TCs for all Britain, from the towns and cities list in the Ordnance Survey (OS) Gazetteer and locate their central points. Below we refer to these as OSC TCs (Ordnance Survey Cities Town Centres).

To predict the extent of the TCs around these two sets of locations, we use abundant small-scale geographical information, in a range of 1–3 km from the centroids. We calculated a long list of geographical and socio-economic factors that relate to TC activities, following closely the variables used by DCLG in the construction of their Index of Town Centre Activity (ODPM 2004). To assess the extent to which these factors accurately predict TC space, we regress them on the radius of the DCLG TCs (derived from the area of the shapefiles), to replicate as closely as possible the areas of these TCs for England and Wales. We then subject the results to robustness checks and, having satisfied ourselves as to the results, apply the estimated coefficients in a separate exercise to the set of locations (OSC TCs) available for all three countries of Great Britain to predict the size of their TCs. By doing this, we obtain a full set of estimated TC boundaries for all countries in Britain, and, in particular, Scotland, on a measure consistent with that used to identify the DCLG TCs for just England and Wales.

We believe this article makes three contributions. First, we show how important it is to have clear and replicable measures of TCs to be able to consistently evaluate policies aimed at these particular locations. This is an issue which both the interested academic and policy communities seem to have surprisingly overlooked. Second, we propose a simpler methodology than others available in the literature to predict the extent of TC space around a set of locations (as discussed in Section 2). Our method requires less data than others and uses straightforward regression techniques. Finally, we provide the necessary tools to implement a robust evaluation of policies applying to TC locations, in particular for the British context where these policies are very popular with planners and policymakers.

The rest of the article is organized as follows. In Section 2 we review the existing theories relating to TCs and how those, in turn, relate to work on the urban system. Then in Section 3, we discuss the definition of TCs and some existing methods to identify their location and boundaries. In Section 4, we describe the existing data on TCs for England and Wales. In Section 5, we explain our methodology to predict the location and extent of TCs for all of Great Britain. Section 6 presents the results and provides some statistics to check how well the method works. Finally, Section 7 concludes.

4 Data for 2004 can be accessed at https://data.gov.uk/dataset/english-town-centres-2004, but we have also had access to data for years 2000 and 2002 provided to us by DCLG. These data were originally created by the ODPM. Their methodology is described in ODPM (2004).

5 As we have said, there are only 'unofficial' estimations of TCs by the ODPM. Nevertheless we call these 'official'.

2. TCs and the Wider Urban System

One can draw on two main bodies of analysis, both trying to explain where TCs are and why they are important: central place theory (CPT) and gravitational theory. In the case of CPT, economists go first to Lösch (1940), although geographers might prefer the slightly earlier contribution of Christaller (1933). But both analyse essentially the same problem: Why does an urban system emerge and would emerge on even a flat and homogeneous plain? The essential mechanism is the tension between economies of scale and the costs of distance combined with the fact that some producers—farmers—are tied to the land and consume land in their production. Imagine a flat, fertile, and homogeneous plain with farmsteads dispersed over it. Over time some production gets concentrated in space because of economies of scale: so instead of all farmers brewing their own beer, for example, a brewery emerges serving the surrounding farms. The more important are economies of scale in any activity, the fewer will be the centres which end up producing that good other things equal. Similarly, the more significant are transport costs for any activity, the more centres will produce that good other things equal. So, we end up with a settlement pattern which has lots of brickworks and pubs but very few centres producing pharmaceuticals. The result is a hierarchy of places.

Translating this to the context of retail, we can think of the hierarchy of shopping centres. Many small places will offer convenience stores, but specialized fashion or department stores will be concentrated in a smaller number of larger shopping centres. In retail, as with other economic activities, there are economies of scale and a threshold market size necessary to support the activity. Rolls Royce dealers or bespoke tailors require large catchment areas (market sizes) to support them, so they are concentrated in fewer larger centres. If transport costs fall or the necessary minimum market size increases (the growth of Internet shopping may have increased the necessary minimum market size to support record or bookshops, e.g.), then there will tend to be an increase in concentration of retail in the larger centres: so the distribution of the 'hierarchy of shopping centres' will become more skewed.

CPT is a theory of a system of cities, of an urban hierarchy, and translates directly into a theory of a system of shopping centres. Some authors (Fujita et al. 1999) argue that CPT does not have testable assumptions and so should be only be considered as a descriptive theory. This argument is contested by researchers such as Denike & Parr (1970) who show there can be strict microfoundations for Christaller's model. In a similar vein, Dicken and Lloyd (1990) discuss testable hypotheses of the theory: in particular on the 'desire lines' (consumers' travel patterns or 'flows') within the hierarchy. Low-order goods (bread) generate short-distance and abundant 'desire lines' within a fine grid of central places, and high-order goods (furniture or cars) generate long-distance and fewer 'desire lines' within a coarse grid of central places.

We can think of CPT, therefore, as providing a theory of the system and hierarchy of shopping centres, but there is also a body of work which focuses on consumers' choices of where to shop and so on 'shopping trips'. As early as 1930, Reilly explored the location of retail (Reilly 1929, 1931). He presented a 'law of gravitation': areas of greater population ('mass') will generate more purchases in their centre, but their attraction will decay with the square of distance to any consumer or shopper. This theory was extended and refined by Huff (1963; 1964) taking as his inspiration, Newton's Law of Universal Gravitation. He described in a simple and powerful way the interactions between cities on a plain with dispersed population. This not only accounts for the length of shopping trips, increasing with the 'pull' of the shopping centre, the infrequency of that type of purchase, or a reduction in travel costs but also an emerging hierarchy of shopping centres of different sizes (Klaesson & Öner 2014).

Both these theories of cities and shopping trips can also be theories for TCs. Both can play a role in assessing the location, size, and distribution of TCs. In this article, we use an eclectic theory that draws heavily on both CPT and the gravity model approach. Specifically, we follow an econometric forecasting model initiated recently by Thurstain-Goodwin and Unwin (2000). We try to predict given TCs' locations, sizes, and distribution in one region using many variables, including proxies for 'mass' (population and area of retail as generally used in gravitational models) and 'desire lines' and hierarchies (drawing on CPT). Then, after verifying that there is a good fit, we predict the size of TCs in another region using the coefficients found in the first step.

3. What Is a TC and How Should It Be Identified?

As noted in the introduction, identifying the exact boundaries of TCs is a more challenging question to answer than it appears at first sight. TCs are not definite entities. They might not be located at the geometric or geographical centre of a city, and they might have fuzzy or indeterminate borders. The 'ideal' TC is not a point but is represented by a space of significant dimension. As the Oxford English Dictionary (OED) defines it: 'the central part or main business and commercial area of a town'. In general conversation, people might understand a TC to be the focal point of a city where main roads converge and people congregate. Historically the town or city centre was a place where citizens met or gathered: the place of the Italians' *passegiata*. Another function of a TC, captured in the OED definition, is as a space where jobs are concentrated, a shared workplace for people who live more spatially dispersed, and a centralized destination (workplace) for decentralized origins (households). Firms locate in TCs to be able to draw on a wider pool of labour. So, people commute to work in TCs. And the third main function of TCs is as a commercial hub, the space where people shop. 'High Streets' and market places are located in TCs.

But the space that represents a TC not only need not be at the geometric centre of a city, it does not have a unique shape. It would only be like that in a location that is constructed according to a rigidly imposed, utopian planning scheme, where all the uses and functions identified would be neatly and exclusive concentrated in only the TC, and TCs would have some uniform shape. Real TCs, in real cities, are much more messy and diverse, sometimes two or three blocks in the centre of a small town and sometimes very extensive. For example, Central London's DCLG 'designated' TC extends over 44 sq km, centred around Trafalgar Square, and includes many retail sub-centres, areas focused on business, and other specialized areas such as 'theatre land' or entertainment zones with a concentration of restaurants and nightlife. The diversity of real TCs certainly adds to choice and likely generates greater productivity and welfare. Left to choose for themselves, businesses and individuals will usually find superior locations to those decided on by urban planners, although there are significant qualifications resulting from externalities in land use that individualistic decision makers will tend to ignore.

If we are to reliably identify TC areas, then we ought to give due weight to the location of all the main functions discussed above to identify the location, size, and shape of the TC. All three aspects of TCs tend to be problematic theoretically and empirically. Centres do not need to be at the centroid of the city or some set of central jurisdictions. The observed shapes of TCs are motley and uneven. Size is also contentious. Empirically, in this article, we try to predict radiuses using a model with over 65 explanatory variables that capture all the multiple dimensions of 'town centredness'.

Attempts to provide operational definitions of TCs in Britain have been lead historically by what is now the DCLG (Thurstain-Goodwin and Unwin (2000); ODPM and CASA (2002); ODPM (2004); and more recently Dolega et al. (2016)). ODPM and CASA (2002) start by discussing a TC definition that depends on the perspective of a particular stakeholder. For instance, a taxi driver would have a different definition of a TC to a planner. For the taxi driver, the areas with the highest footfall can be determinants, while for a planner, the future evolution of the area might be a priority. Moreover, ODPM and CASA (2002) make the definition of TCs relative to other features of a city, creating an open approach from which they can build their model to define TCs.

The result is that their TCs are necessarily diverse. For some TCs the priority would be 'a retail core, and office centre and an area of high building density', while for others, 'a concentration of visitor attractions and associated retail outlets' would be the focus (Thurstain-Goodwin and Unwin 2000). What is meant by this is that it is essential to include multiple dimensions and functions, not just focus on one dimension of 'town centredness'. This implies that TCs are 'indeterminate objects' with fuzzy borders, extremely difficult to define and agree upon. We can add that an operational definition should be implemented with consistency over an entire set of cities because the identification of a TC remains problematic. For example, Wolverhampton's TC has a distinct ring road—some emergency services use it as a boundary, but administrative boundaries have been set in a much more extensive area reflecting a longer-term strategic vision of how the TC should evolve (ODPM and CASA 2002).

Typically, humans can easily detect an outlier, but not as easily notice when observations are clustered (Everitt and Hothorn 2011). Estimating *kernel density functions* can help identify clusters of 'objects'. These generate surfaces similar to mountainous terrain. This is called 'smoothing' and permits discrete and clustered data to be transformed into these mountain ranges. The kernel counts the number of observations in a given two-coordinate space as a histogram would, but it uses the number of observations to amplify a pulse function (rectangular, triangular, or normal most commonly) (Everitt and Hothorn 2011). Thus, waves effectively transform the discrete information of the numbers and intensities of the points into peaks and valleys. The key parameter is the bandwidth, which can be adjusted (Everitt and Hothorn 2011).

A very small bandwidth creates a single point to be counted independently, resulting in a spiky, disaggregated graph. An even smaller bandwidth provokes equal-sized extra-large pulse functions independent of each other if the observations are not located in exactly the same place. A very high bandwidth includes all points in a uniform one-shaped tiny image equal to the generating pulse kernel. Figure A1 (modified from Everitt and Hothorn 2011) shows an example of a one-dimensional normal kernel function for extremely low, low, optimal, and extremely high bandwidths. So, to be useful a researcher estimating kernel density functions needs to find a Goldilocks bandwidth neither too high nor too low. Many techniques have as a result been elaborated for finding such appropriate bandwidths. Then comes the next vital step: slicing the surfaces to get the curves or contour maps which are much easier to interpret. Thus, clustering can be detected by higher mountains, and areas, where data points are scarce, can be detected by lower ones.

Thurstain-Goodwin and Unwin (2000) define an index of intensity of 'town centredness' using the dimensions of property, economy, diversity, and visitor attractiveness. Because the categories are different in units, they employ a *z*-score normalization. The model is populated by points at the Unit Post Code (UPC) level (full postcodes), shaping town centredness as a mass function that is sliced for visualization. The intensity of the functions helps to delimit the border of the TCs, the visualization of which is the point of the study. The ODPM reports (ODPM and CASA 2002; ODPM 2004) are based on this methodology.

A catchment area is an area that draws in some group—customers or workers, for example. A gravity model adds some forces of attraction and repulsion. Gravity models are simple but can be empirically well-behaved and make good predictions. In the case of a retail centre, gravity models typically use square footage of retail space as a measure of size and travel time between retail centres for distance. The so-called 'Huff model' (Huff 1963) uses square footage as a directly proportional proxy of the number of products a consumer would find in each shopping centre and time as an inversely proportional proxy of the cost (including opportunity costs) of travelling to the given retail centres. Then, the more products there are and the greater quantity of a given product that is sold—represented by the square footage dedicated to a given kind of product—the greater the probability of visiting the given retail centre. And the lower the cost—measured as time—the greater the probability of visiting a given retail centre. The model has in the numerator the linear probability of the consumer choosing the retail good of a given type and in the denominator the sum of the linear probabilities of choosing all types of retail goods.

The Liverpool group, Dolega et al. (2016), discusses a method of defining TCs based on catchment areas. In summary, their method consists of replicating a catchment area for multiple stores. They use the Huff-model (Huff 1963, 1964, 2003) mentioned above. In this the probability, P_{ij}, that a consumer located at i chooses to shop at retail centre j is:

$$P_{ij} = \frac{A_j^{\alpha} D_{ij}^{-\beta}}{\sum_{j=1}^{n} A_j^{\alpha} D_{ij}^{-\beta}},$$

where:

A_j is a measure of attractiveness of retail centre j, as square footage.
D_{ij} is the distance from location i to shop j.
α is the attractiveness parameter to be estimated.
β is the distance decay parameter to be estimated.

Until recently the estimation of these parameters did not have known properties of large samples. Huff (2003) suggests it is necessary to explore alternative models similar to those presented in this article. In addition, Dolega et al. (2016) suggest that calibration at a national level would be superior to a local or subnational one. We also include a national-level estimation in our model.

The approach we take is more pragmatic and, in spirit, closer to Thurstain-Goodwin and Unwin (2000). We take the extent of the DCLG-defined TCs (their area-imputed radius) as 'true' on average and collate a long list of explanatory factors that we believe correlate with TC activities and characteristics to predict the TC radius. Then, having satisfied ourselves that the method provides sufficiently high goodness of fit, we use the estimated coefficients from this prediction to extrapolate out-of-sample and apply the coefficients to a different set of locations. Details of the data used for the estimates and the details of the method are explained more fully in the next two sections.

4. The Existing TC Data for England and Wales

As explained above, the first step of our methodology relies on the use of a given set of TC locations that we believe are reasonable approximations, as accurate a set of measures as is available: those identified by DCLG for England and Wales and as defined for 2000. Thurstain-Goodwin and Unwin (2000) and ODPM and CASA (2002) set out a methodology to identify what they call ATCAs, generalized to all locations in England and Wales in ODPM (2004). In the 2000 data, there are 1029 ATCAs, and additionally, within these ATCAs there are 46 RCs, giving a total of 1075 TCs for England and Wales.

The ATCAs are defined areas containing concentrations of 'town centre activity' aiming to be consistent with the theoretical basis summarized in Section 2. Both the hierarchy and the mass of TC activity are taken into consideration by the list of variables chosen to represent the point information with a kernel function. These 3D surfaces with heights reflecting TC activity are then sliced to form contour maps or level curves that represent locations with the same degree of TC activity. For instance, the concentration of employment is a direct measure of the mass of TC activity in gravity theory. At the same time, the postcode centrality structure is a direct measure of the CPT hierarchy.

The ATCAs were first constructed in a so-called Feasibility Study (DETR 1998)[6] using information on seven variables or elements: turnover, activities and facilities, pedestrian gateways, diversity, lack of resident population, intensity of use, and visitor attractions. In the follow-up London Pilot Study (ODPM and CASA 2002), these components were reduced to just three: economy, diversity, and property/intensity of use. Economy includes activities frequently found in TCs, such as retailing (convenience, comparison, and service retail); commercial offices; public administration; restaurant and licenced premises; arts, culture and entertainment; hotels; and public transport. This calculation implies the use of a set of very detailed values on variables reflecting employment (economy and diversity) and floor space (property), with a slightly less important use of turnover. The Office for National Statistics (ONS) contributed to the Inter-Departmental Business Register on employment and turnover for individual businesses, while the Valuation Office Agency—VOA—supplied an extensive commercial and industrial property floor space database.

The model identifies concentrations of the type of activities and patterns of property likely to be found in TCs where there are high levels of employment in economic activities common to TCs (including retail, offices, and leisure activities), a diversity of these activities, and a high density of office and retail floor space. Estimates are mapped at the detailed unit-level postcode to produce a surface of economic activity. Cutting through the peaks in the activity at a prescribed level for the whole of England and Wales gives the ATCA boundaries. Intuitively, combining employment and retail floor space data, a 3D data surface was constructed for different locations in England and Wales where the tallest peaks identified the largest concentrations of retail activity. Then, contours were drawn around these peaks, and the resulting areas were identified as ATCAs. In a second step, the data were cross-validated using external sources to make sure they corresponded to the main centres of activity in England and Wales.

Even if the ODPM/DCLG ATCAs are not intended to be operational for robust policy evaluation, since they correspond to revealed TC space and not planners' TCs as used for purposes of policy, they are the best definitions available to us, and their identification is based on high-quality data for very small geographical units. However, for the purposes of

6 Department of the Environment, Transport, and the Regions.

the evaluation (Cheshire et al. 2017), there exists an important limitation. This critical limitation is that these TCs are not defined for Scotland, and to evaluate the impact of TCFP, one needs to be able to compare developments in TCs in England and Wales, where the policy was strictly applied, to those in Scotland, where it was not. At the same time, we cannot replicate the exact methodology of ODPM/DCLG using data for Scotland because either these data are not readily available to us (e.g. the postcode-level information on different activities) or they do not exist for Scotland (e.g. the VOA data). Given these reasons, we opted to exploit the information on the size of the TCs that we can derive from the England and Wales set in the DCLG data, and combine it with a very rich data set on small geography explanatory factors (including socio-economic and topological features) that can successfully explain the variation in TC space we observe in the data.

5. Identifying TC Space for All Locations in Great Britain: Methodology

We combine data at small geographical scales from multiple sources to predict the extent of TCs for the whole of Great Britain. The main aim behind our methodology is to find a way to replicate 'as close as possible' the TC definitions available for England and Wales (ODPM 2004) and to be able to apply it to obtain TC boundaries in all cities in Britain. There are seven steps in our process:

1. **Select DCLG 2000 TC sample (DCLG TCs):** We start the process by exploring the DCLG list of TCs for England and Wales for the year 2000. From their observed surfaces we find the radius representing all the TCs as circular.[7] Then we select the samples for the regressions in Step 4. Of the 1075 TCs (1029 ATCAs and 46 RCs), we select two main samples: (i) all ATCAs; (ii) ATCAs and, for Central and West London, the RCs. From these samples, we drop the TCs which we consider cannot be used in the estimations.[8] To identify these, we use the information from the National Survey of Local Shopping Patterns (NSLSP) on the location of (grocery) shops in 1998 (more details are provided below).[9] The final samples

7 Of course, not all TCs need to be circular; although in the DCLG sample, they mostly are. Circles are one of the most efficient shapes to serve an area. In the case of TCs, because they occupy only a small fraction of the overall UK landscape, there is no need to impose a more efficient shape, such as a hexagon, to 'fill up space'.

8 As we use variables defined over 1 km of the centroid of the TC, those which do not have values defined within that radius were dropped from the sample.

9 The NSLSP is a yearly survey run by CBRE covering over a million households in the UK. Each sampled household is asked about their socio-economic characteristics, where they live and where they undertake their main shopping for a series of goods (groceries and household white goods). The data we obtained correspond to the grocery shopping locations and were aggregated spatially. It consists of an origin (postal sector)-destination (store) matrix of shopping trips. Postal sector areas are aggregations of postcodes and correspond to small areas (there are 12,000 in Britain). For the purposes of this article we used the shopping destination data to obtain a list of main (grocery) stores identified in 1998 as a grocery shopping destination in the NSLSP data. We can use this to infer how relatively important a TC is. This is illustrated in Figure 2. In addition, we used another CBRE-supplied data set on the location of retail units called RETLOC (RETail LOCations) in the main text, which includes information about all grocery stores and not only on those identified by the sample of households in the NSLSP as their grocery shopping destinations.

have between 810 and 950 TCs located in England and Wales. The mean radius of these is slightly less than 250 m. We then create centroids from the shapefiles of these TCs.

2. **Identify alternative TC locations for all Great Britain (OSC TCs):** We define an alternative list of TC centroid candidates using the towns and cities information in the OS Gazetteer towns and cities. Initially, there are 1315 towns and cities in Great Britain as a whole. As in the case of the DCLG TCs, the list is further trimmed when we combined it with the spatial data around the centroids. The exact location of some of these town and city centroids was 'relocated' by looking at where popular map navigation tools (such as Open Street Map or Google Maps) located the city centroid.
3. **Collection of data around the centroids of the DCLG and OSC TCs:** We collect abundant information at very small geographical scales (the largest is the Output Area and the smallest is postcode units) for the areas around the centroids of the DCLG and OSC TCs. The main results (presented in Section 6) use information around 1 km of the centroid, but we also calculated all the models using information around 2 and 3 km.[10] We believe that these long lists of socio-economic and topological features around 1 km of the centroid are sufficient satisfactorily to predict the extent of TC space around these centroids (remember the average DCLG TC radius is around 250 m). We obtained information on multiple variables (over 100) and 66 were used for the regressions of Step 4. The list of variables and their data sources appears in Table 1 (and in detail in Table A2).
4. **Estimation of the factors determining the extent of TC space:** For the DCLG TC samples selected in Step 1, we estimated several models where we explained the (log) radius of the TC as a function of the large set of explanatory variables around 1 km of the centroid of the TC. Inspired by the original ODPM models (2004), we use explanatory variables related to different town centredness dimensions (shop density and location, employment density and diversity, local amenities, socio-economic characteristics of the resident and working populations, infrastructure endowments, geographical location, physical barriers, etc.). The results of these regressions are shown in Table 2 and discussed in the next section. The majority of estimates are significantly different from 0, and the models have high goodness-of-fit statistics (R^2 between 0.78 and 0.88).
5. **Validation of the results (within DCLG sample):** The first step to validate our results is to check if the predictions correlate with the actual values for the in-sample. We use the coefficients estimated in Step 4 to predict the (log) radius of the DCLG TCs, both for the whole sample (1001) and for the samples used in each of the models estimated (referred as sub-samples in the tables). We both summarize and correlate the actual and predicted radius (and derived area) and use this to check the internal validity of the methodology. The results are shown in Tables 3 (and Table A3) and 4 and are discussed in the next section. They show that the statistical moments and the correlation between the actual and predicted (log) radius and area of the TCs are reasonably similar/high.
6. **Application of the model to predict TC space around the OSC TCs:** The results from Step 5 give us sufficient confidence that the models are satisfactorily accurate in their prediction of the extent of TCs for different values of the explanatory factors. We, therefore, proceed to apply the estimated betas from Step 4 to the 'out-of-sample' list of

10 These models had less predictive power, so we favoured the ones using 1 km. The average radius of a TC in England and Wales is 250 m, so we expect that values of the variables beyond 1 km of the TC centroid will have little power in predicting the extent of TC space. In fact, the heterogeneity of values beyond 1 km reduced the predictive power of the models.

Table 1. List of explanatory variables included in our model

Variable	Data sources
Number of shoppers, shops, and location of these	*CBRE: RETLOC and NSLSP data in 1998*
Number and employment in sector 52 (retail)	*Annual Business Inquiry (ABI) accessed via NOMIS*
Population, residential employment, workplace-based employment, share of occupations (in workplace employment) of different levels (employment diversity), residential socio-economic characteristics (age structure, unemployment rates, labour market, commuting patterns)	*Census, 2001*
Transport infrastructure (roads, rail, buses)	*Ordnance Survey (OS) Strategi 2009, Open Street Map*
Cultural amenities (libraries, museums, art galleries, theatres, cinemas), consumption amenities (bars, restaurants), and historical amenities (landmarks, tourist info, local government)	*OS Strategi, OS Points of Interest (POI)*
Postcode centrality structure (sectors, districts, and towns)	*ONS National Statistics Postcode Directory (NSPD) and Wikipedia*
Geographical location (distance to the coast, river, rail, town hall, natural park/woodland)	*OS Strategi, POI, Wikipedia*
Topological features: terrain elevation (m) and slope (degrees)	*OS Panorama 50x50*
Nightlight brightness intensity (96–97 average)	*NOAA-NGDC*[11]
Postcode centrality structure (sectors, districts and towns)	*NSPD*

Notes: List of the abbreviations used and more details on the variables provided in Tables A1 and A2.

OSC TCs and calculate the predicted (log) radius and area for these locations. This generates a set of estimated surrogate TC shapefiles to cover all the TCs of Great Britain. We can compare the predicted radius for the two sets of TCs (DCLG and OSC) for the sample which is available in both data sets (e.g. England and Wales together and England and Wales separately). This is done in the first two rows of Table 5 and shows that the values of the DCLG sample and our OSC TC predicted values are similar.

7. **Comparison of socio-economic variables within the DCLG and OSC TCs**: The DCLG TCs and our predicted OSC TCs differ in two dimensions: their particular size for a given set of explanatory factors (which we fit in Step 4) and their specific location. The precise places where the OSC and DCLG centroids are located can differ, and, in particular, there is no comparison group for Scotland. To overcome this, in Step 7 we calculate several socio-economic descriptive statistics (population, number of addresses, number of shoppers, etc.) within the boundaries (or a small distance of them) of the two sets of TCs. The summary statistics for these are shown in the remaining rows of Table 5. These allow us to check whether, even when located at slightly in different places, the underlying economic factors within TC boundaries are comparable in the two samples and, additionally, to explore how different the Scottish TCs are compared to those in England and Wales.

11 National Oceanic and Atmospheric Administration; National Geophysical Data Center, non-censored version.

Table 2. TC extent prediction model results

DCLG 2000 TC sample	(1)	(2)	(3)	(4)	(5)	(6)
Type of TC	ATCA	ATCA	ATCA	ATCA	ATCA	ATCA
Fuzzy TC border	10 m	100 m	500 m	10 m	100 m	500 m
Central and West London	ATCA	ATCA	ATCA	RC	RC	RC
Number of observations	812	870	931	824	882	949
Adjusted R^2	0.878	0.868	0.810	0.841	0.834	0.779
R^2	0.888	0.878	0.824	0.854	0.846	0.795
	(1)	(2)	(3)	(4)	(5)	(6)
Log of grocery shoppers	0.020***	0.024***	0.031***	0.021**	0.025***	0.032***
(NSLSP98), 1 km of centroid	(0.007)	(0.007)	(0.009)	[0.009]	[0.009]	[0.009]
Log number of shops	0.024***	0.026***	0.021***	0.022***	0.024***	0.018***
(RETLOC/NSLSP98), 1 km of centroid	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Log average distance to shops	−0.441***	−0.401***	−0.684***	−0.435***	−0.396***	−0.631***
(RETLOC98), 1 km of centroid	(0.123)	(0.123)	(0.144)	(0.131)	(0.133)	(0.146)
Log average distance to shops	−0.177***	−0.161***	−0.283***	−0.175***	−0.158***	−0.261***
(RETLOC98), 1 km of centroid^2	(0.054)	(0.055)	(0.064)	(0.059)	(0.060)	(0.065)
Log average distance to shops	−0.021***	−0.019***	−0.034***	−0.021***	−0.019**	−0.031***
(RETLOC98), 1 km of centroid^3	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Log average distance to shops	−0.051	−0.080	−0.149	−0.046	−0.090	−0.155
(NSLSP98), 1 km of centroid	(0.095)	(0.095)	(0.115)	(0.093)	(0.096)	(0.114)
Log average distance to shops	−0.015	−0.021	−0.045	−0.011	−0.024	−0.046
(NSLSP98), 1 km of centroid^2	(0.043)	(0.043)	(0.052)	(0.043)	(0.044)	(0.053)
Log average distance to shops	−0.002	−0.002	−0.004	−0.002	−0.003	−0.004
(NSLSP98), 1 km of centroid^3	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)
Log distance to closest shop	0.054***	0.046***	0.013	0.059***	0.050***	0.018*
(NSLSP98), 1 km of centroid	(0.010)	(0.010)	(0.011)	(0.011)	(0.010)	(0.011)

(continued)

Table 2. Continued

DCLG 2000 TC sample	(1)	(2)	(3)	(4)	(5)	(6)
Number of units retail sector	0.001***	0.001***	0.001***	0.001***	0.001***	0.002***
(ABI98), 1 km of centroid	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of units retail sector	−0.000*	−0.000**	−0.000	−0.000***	−0.000***	−0.000***
(ABI98), 1 km of centroid^2	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Employment retail sector	−0.000	−0.000	−0.000	−0.000	−0.000*	−0.000**
(ABI98), 1 km of centroid	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Employment retail sector	0.000	0.000*	0.000	0.000**	0.000***	0.000***
(ABI98), 1 km of centroid^2	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log density of total population	0.362	0.211	0.457*	0.323	0.180	0.325
(Census01), 1 km of centroid	(0.254)	(0.230)	(0.251)	(0.251)	(0.228)	(0.238)
Log density of total population	−0.024	−0.015	−0.032*	−0.020	−0.012	−0.022
(Census01), 1 km of centroid^2	(0.016)	(0.015)	(0.016)	(0.016)	(0.015)	(0.016)
Log workplace employment	0.286***	0.284***	0.282***	0.281***	0.281***	0.279***
(Census01), 1 km of centroid	(0.024)	(0.024)	(0.028)	(0.026)	(0.026)	(0.031)
Share wemple high occupations	−1.507***	−1.647***	−1.238**	−2.391***	−2.444***	−2.130***
(Census01), 1 km of centroid	(0.481)	(0.465)	(0.573)	(0.505)	(0.494)	(0.553)
Share wemple medium occupation	−1.838***	−1.784***	−1.308**	−2.577***	−2.434***	−2.032***
(Census01), 1 km of centroid	(0.474)	(0.457)	(0.576)	(0.522)	(0.507)	(0.578)
Share wemple low occupations	−2.292***	−2.433***	−2.104***	−3.155***	−3.201***	−3.009***
(Census01), 1 km of centroid	(0.490)	(0.473)	(0.581)	(0.515)	(0.503)	(0.559)
Average commuting distance	0.013	0.012	0.012	0.018	0.017	0.018
(Census01), 1 km of centroid	(0.009)	(0.008)	(0.009)	(0.011)	(0.011)	(0.011)
Average commuting distance	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
(Census01), 1 km of centroid^2	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

(continued)

Table 2. Continued

DCLG 2000 TC sample	(1)	(2)	(3)	(4)	(5)	(6)
Percentage commuters by foot/bike	0.110	0.125	0.216	0.307**	0.305**	0.383**
(Census01), 1 km of centroid	(0.142)	(0.145)	(0.157)	(0.144)	(0.146)	(0.155)
Percentage commuters public trans	0.018	0.083	0.107	0.294**	0.344***	0.279**
(Census01), 1 km of centroid	(0.114)	(0.114)	(0.128)	(0.120)	(0.119)	(0.130)
Average household size	0.100*	0.030	0.042	0.141**	0.079	0.067
(Census01), 1 km of centroid	(0.057)	(0.064)	(0.071)	(0.065)	(0.068)	(0.076)
Average age of resident population	0.019***	0.013*	0.009	0.028***	0.022***	0.018**
(Census01), 1 km of centroid	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Percentage of population aged 18–44 years	0.297	0.184	−0.064	0.733**	0.626*	0.571
(Census01), 1 km of centroid	(0.273)	(0.280)	(0.333)	(0.325)	(0.330)	(0.368)
Total unemployment rate	0.819***	0.815***	0.853***	1.178***	1.160***	1.298***
(Census01), 1 km of centroid	(0.285)	(0.290)	(0.326)	(0.293)	(0.295)	(0.341)
Share of students in population	−0.933***	−0.890***	−0.663***	−1.043***	−0.983***	−0.842***
(Census01), 1 km of centroid	(0.183)	(0.180)	(0.204)	(0.233)	(0.222)	(0.247)
Share of retirees in population	−0.699	−0.476	−0.274	−0.842*	−0.630	−0.333
(Census01), 1 km of centroid	(0.450)	(0.453)	(0.475)	(0.478)	(0.477)	(0.501)
Log of km of all roads	−0.005	0.004	0.007	−0.001	0.004	0.013
(Strategi 2009), 1 km of centroid	(0.022)	(0.022)	(0.025)	(0.024)	(0.024)	(0.025)
Number of bus stations	−0.001	−0.000	−0.001	0.005	0.004	0.004
(POI 2015), 1 km of centroid	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.010)
Number of tube/tram stations	0.059**	0.057**	0.001	−0.048**	−0.047**	−0.056***
(POI 2015), 1 km of centroid	(0.023)	(0.025)	(0.026)	(0.022)	(0.021)	(0.019)
Number of rail stations	0.003	−0.000	−0.012	0.000	−0.002	−0.010
(POI 2015), 1 km of centroid	(0.010)	(0.010)	(0.014)	(0.015)	(0.015)	(0.015)

(continued)

Table 2. Continued

DCLG 2000 TC sample	(1)	(2)	(3)	(4)	(5)	(6)
Number of libraries	−0.014	−0.011	−0.032**	−0.028**	−0.026**	−0.045***
(POI 2015), 1 km of centroid	(0.010)	(0.011)	(0.015)	(0.012)	(0.012)	(0.014)
Number of museums	0.004	−0.004	−0.000	0.021*	0.011	0.010
(Strategi 2009), 1 km of centroid	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.014)
Number of art galleries	−0.010*	−0.009*	−0.010*	−0.005	−0.004	−0.009
(POI 2015), 1 km of centroid	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Number of cinemas and theatres	0.017***	0.016***	0.017**	0.018**	0.017**	0.019**
(POI 2015), 1 km of centroid	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)
Number of discos and nightclubs	0.004	0.003	0.002	0.022***	0.021***	0.021***
(POI 2015), 1 km of centroid	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Number of landmarks	0.027*	0.020	0.005	0.035**	0.030**	0.019
(Strategi 2009), 1 km of centroid	(0.015)	(0.015)	(0.021)	(0.014)	(0.014)	(0.021)
Number of cafes, restaurants and pubs	0.001*	0.001**	0.001**	−0.001**	−0.001*	−0.001**
(POI 2015), 1 km of centroid	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of B&B, hotels and motels	0.000	0.000	0.000	0.000	0.000	0.000
(POI 2015), 1 km of centroid	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of youth hostels	−0.007	−0.008	−0.006	−0.012	−0.013*	−0.008
(POI 2015), 1 km of centroid	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Number of local government sites	0.015***	0.016***	0.013	0.012	0.013*	0.009
(POI 2015), 1 km of centroid	(0.005)	(0.005)	(0.010)	(0.007)	(0.007)	(0.011)
Number of tourist info offices	0.012	0.015	0.023*	0.012	0.014	0.025*
(Strategi 2009), 1 km of centroid	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)	(0.014)
Number of visitor centres	−0.002	−0.001	0.004	−0.005	−0.004	0.006
(Strategi 2009), 1 km of centroid	(0.007)	(0.006)	(0.008)	(0.008)	(0.008)	(0.009)

(continued)

Table 2. Continued

DCLG 2000 TC sample	(1)	(2)	(3)	(4)	(5)	(6)
Share of central addresses	−0.000	0.002	−0.011	0.001	0.004	−0.014
(pcsect in pctown), 1 km of centroid	(0.031)	(0.030)	(0.031)	(0.031)	(0.029)	(0.030)
Share of central addresses	0.034**	0.038**	0.030	0.034*	0.038**	0.028
(pcdistr in pctown), 1 km of centroid	(0.017)	(0.017)	(0.019)	(0.019)	(0.019)	(0.020)
Log distance in km to first	−0.002	−0.004	−0.009	−0.002	−0.003	−0.008
postcode in the closest postal town	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)
Log of distance to closest town hall	−0.003	−0.003	−0.001	−0.005	−0.005	−0.004
(Wikipedia), in km	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Log of distance to closest rail or tube station	−0.002	−0.000	−0.010	−0.006	−0.004	−0.011
(POI 2015), in km	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)
Log of distance to closest point in the coastline	−0.005	−0.000	0.008	−0.003	0.001	0.007
(Strategi 2009), in km	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)
Log of distance to closest river or lake	−0.008	0.000	0.008	−0.013*	−0.005	0.002
(Strategi 2009), in km	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
Log of distance to closest park or woodland	0.001	−0.003	−0.003	0.001	−0.004	−0.003
(Strategi 2009), in km	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Standard dev of elevation	0.001	0.000	0.001	0.002	0.001	0.002
(Panorama), 1 km of centroid	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mean of elevation	−0.010*	−0.009*	−0.010*	−0.005	−0.004	−0.009
(Panorama), 1 km of centroid	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Maximum of elevation	0.002*	0.001	0.002**	0.001	0.001	0.002*
(Panorama), 1 km of centroid	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Range of elevation	−0.002	−0.001	−0.002**	−0.002	−0.001	−0.002**
(Panorama), 1 km of centroid	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

(continued)

Table 2. Continued

DCLG 2000 TC sample	(1)	(2)	(3)	(4)	(5)	(6)
Standard dev of terrain slope	0.016	0.018	0.022	0.015	0.017	0.017
(Panorama), 1 km of centroid	(0.017)	(0.016)	(0.017)	(0.018)	(0.017)	(0.017)
Mean of terrain slope	0.011	0.009	0.012	0.008	0.007	0.010
(Panorama), 1 km of centroid	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)
Maximum of terrain slope	0.001	−0.010	0.011	−0.020	−0.019	0.001
(Panorama), 1 km of centroid	(0.127)	(0.124)	(0.122)	(0.131)	(0.129)	(0.127)
Range of terrain slope	−0.004	0.008	−0.013	0.017	0.017	−0.003
(Panorama), 1 km of centroid	(0.127)	(0.124)	(0.122)	(0.131)	(0.129)	(0.127)
Standard dev of lights brightness	0.005	0.003	0.006	−0.004	−0.004	0.000
(NASA 96-97), 1 km of centroid	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)	(0.006)
Mean of lights brightness	0.000	0.001	0.001	0.000	0.001	0.001
(NASA 96-97), 1 km of centroid	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Max of lights brightness	0.001	0.000	0.000	0.000	−0.000	0.000
(NASA 96-97), 1 km of centroid	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Range of lights brightness	−0.002	−0.001	−0.003	0.001	0.001	−0.001
(NASA 96-97), 1 km of centroid	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Sum of lights brightness	−0.000	−0.000*	−0.000*	−0.000	−0.000	−0.000
(NASA 96-97), 1 km of centroid	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant term (excluded: occupation	−4.977***	−3.948***	−5.621***	−4.656***	−3.798***	−4.972***
other and commuter by motor vehicle)	(1.138)	(1.063)	(1.252)	(1.143)	(1.078)	(1.193)

Notes: Log stands for natural logarithm and km for kilometres. Details on the definition of the variables are given in Table A2.

Table 3. England and Wales TC radius/area (actual and predicted), and number of TCs, DCLG 2000 sample

	TC derived radius in km					TC Area in square km				
Variable	Observed	Mean	Standard deviation	Minimum	Maximum	Observed	Mean	Standard deviation	Minimum	Maximum
Observed value	*1001*	*0.247*	*0.153*	*0.113*	*3.187*	*1001*	*0.265*	*1.047*	*0.040*	*31.905*
Average all samples	1001	0.261	0.165	0.107	2.973	1001	0.308	0.994	0.036	28.018
Sample 1 betas	1001	0.273	0.196	0.104	3.095	1001	0.355	1.223	0.034	30.091
Sample 2 betas	1001	0.270	0.195	0.106	3.076	1001	0.350	1.203	0.035	29.718
Sample 3 betas	1001	0.260	0.178	0.105	3.051	1001	0.312	1.059	0.034	29.245
Sample 4 betas	1001	0.257	0.148	0.105	2.526	1001	0.276	0.718	0.034	20.040
Sample 5 betas	1001	0.255	0.153	0.107	2.711	1001	0.277	0.809	0.036	23.088
Sample 6 betas	1001	0.249	0.163	0.106	3.382	1001	0.278	1.184	0.035	35.929
Sub-sample 1	812	0.256	0.157	0.104	3.095	812	0.283	1.090	0.034	30.091
Sub-sample 2	870	0.250	0.153	0.106	3.076	870	0.270	1.044	0.035	29.718
Sub-sample 3	931	0.243	0.147	0.111	3.051	931	0.254	0.996	0.039	29.245
Sub-sample 4	824	0.251	0.120	0.105	1.040	824	0.243	0.306	0.034	3.397
Sub-sample 5	882	0.246	0.118	0.107	1.070	882	0.233	0.302	0.036	3.600
Sub-sample 6	949	0.239	0.112	0.108	1.030	949	0.218	0.281	0.036	3.336

Notes: The DCLG sample for which we have values of the variables around 1 km of the centroids corresponds to 1001 TCs in England and Wales. Rows Sample 1 betas–Sample 6 betas apply the coefficients estimated in Columns 1–6 of Table 2 to these 1001 locations. Sub-samples refer to applying the estimated coefficients of each column of Table 2 to the particular sample used in that estimation, for example, the number of TCs corresponds to the number of observations in each of these regressions.

Table 4. Correlation coefficients of real versus predicted values for radius and area, DCLG 2000 sample

	All (1001)			Sub-sample		
	All	*England*	*Wales*	*All*	*England*	*Wales*
TC radius in km						
Average all samples	0.755	0.757	0.866			
Sample 1	0.722	0.718	0.912	0.970	0.970	0.975
Sample 2	0.725	0.721	0.912	0.966	0.966	0.973
Sample 3	0.803	0.800	0.923	0.948	0.948	0.967
Sample 4	0.867	0.866	0.912	0.943	0.942	0.977
Sample 5	0.870	0.869	0.913	0.940	0.938	0.973
Sample 6	0.883	0.882	0.921	0.915	0.913	0.966
TC area in km^2						
Average all samples	0.928	0.928	0.865			
Sample 1	0.810	0.810	0.858	0.994	0.994	0.974
Sample 2	0.811	0.811	0.857	0.993	0.993	0.974
Sample 3	0.914	0.914	0.873	0.988	0.988	0.958
Sample 4	0.940	0.941	0.863	0.937	0.937	0.977
Sample 5	0.950	0.950	0.862	0.928	0.927	0.975
Sample 6	0.973	0.973	0.868	0.893	0.892	0.955
Sample sizes						
Average all samples	1001	944	57			
Sample 1	1001	944	57	812	768	44
Sample 2	1001	944	57	870	820	50
Sample 3	1001	944	57	931	876	55
Sample 4	1001	944	57	824	780	44
Sample 5	1001	944	57	882	832	50
Sample 6	1001	944	57	949	894	55

To illustrate the logic behind our methodology, Figure 1 shows a flowchart depicting the seven steps explained above and the relationship between them. In the next section, we apply these steps to our data and discuss the results and the validation checks carried out.

6. Regression Results and Validity Tests

The first step of our methodology concerns the selection of the samples of the TC locations used in the estimations of the models that predict TC extent. The DCLG 2000 TC data set originally contained 1075 units. When we calculate the variables included in the estimation of Step 4, within 1 km of the centroid, a number of TCs are dropped from the sample because the values of some of these factors do not exist within that distance radius

Table 5. Predicted size and socio-economics for OSC and DCLG's TCs (500 fuzzy boundary tolerance)

	England and Wales		England		Wales		OSC TC areas	
Variable	OSC	DCLG	OSC	DCLG	OSC	DCLG	ALL GB	SCOT
Number of observations	*752*	*931*	*687*	*876*	*65*	*55*	*861*	*109*
Predicted TC radius in km	0.231	0.246	0.236	0.248	0.187	0.210	0.222	0.161
Predicted TC area in km^2	0.215	0.257	0.222	0.263	0.135	0.161	0.201	0.109
NSLSP 1998 shops	2.20	2.37	2.23	2.37	1.92	2.33	2.20	2.17
NSLSP 1998 shoppers	20,910.63	24,085.03	21,560.91	24,416.07	14,037.62	18,914.91	20,003.98	13,748.96
NSPD address counts	3183.63	4175.27	3291.85	4266.92	2039.77	2715.62	3044.47	2084.39
NSPD small businesses counts	419.88	479.08	433.57	487.45	275.25	345.84	398.52	251.14
NSLSP 1998 shops per km^2	1.26	1.31	1.26	1.30	1.26	1.46	1.29	1.50
NSLSP 1998 shoppers per km^2	11,387.03	12,928.03	11,651.57	13,032.77	8591.06	11,292.24	11,100.73	9125.53
NSPD address counts per km^2	1762.30	2130.30	1806.99	2162.26	1289.95	1621.23	1716.77	1402.70
NSPD small businesses counts per km^2	208.51	217.23	212.86	218.54	162.48	196.36	200.70	146.79
Population	6004.47	8192.70	6221.02	8389.54	3715.80	5158.30	5716.36	3728.64
Residential employment	2699.94	3708.18	2821.36	3817.85	1416.57	2017.62	2563.13	1619.28
Workplace employment	7438.53	8828.14	7792.53	9088.64	3697.03	4812.55	6986.72	3869.62
Workplace high occupations	3292.30	3972.25	3474.86	4113.57	1362.80	1793.74	3072.93	1559.44
Share of work employment high occupations	0.350	0.363	0.353	0.366	0.310	0.327	0.345	0.314
Ratio work employment over population	1.05	0.93	1.07	0.94	0.79	0.84	1.01	0.75
Population per km^2	3718.12	4487.26	3785.34	4563.91	3007.71	3320.25	3634.28	3055.87
Residential employment per km^2	1658.80	2019.03	1706.66	2066.52	1153.01	1295.91	1616.10	1321.48
Work employment per km^2	3501.33	3366.00	3611.98	3417.46	2331.83	2582.36	3340.43	2230.32

Notes: The first three columns compare mean values of the variables for the TCs in the common sample countries (England and Wales). The last column provides the mean values of the variables for the whole GB sample and the areas in Scotland. TC stands for Town Centre, OSC corresponds to Ordnance Survey cities sample and DCLG to the ODPM/DCLG Town Centres; GB stands for Great Britain, km stands for kilometres, and NSLSP stands for National Survey of Local Shopping Patterns.

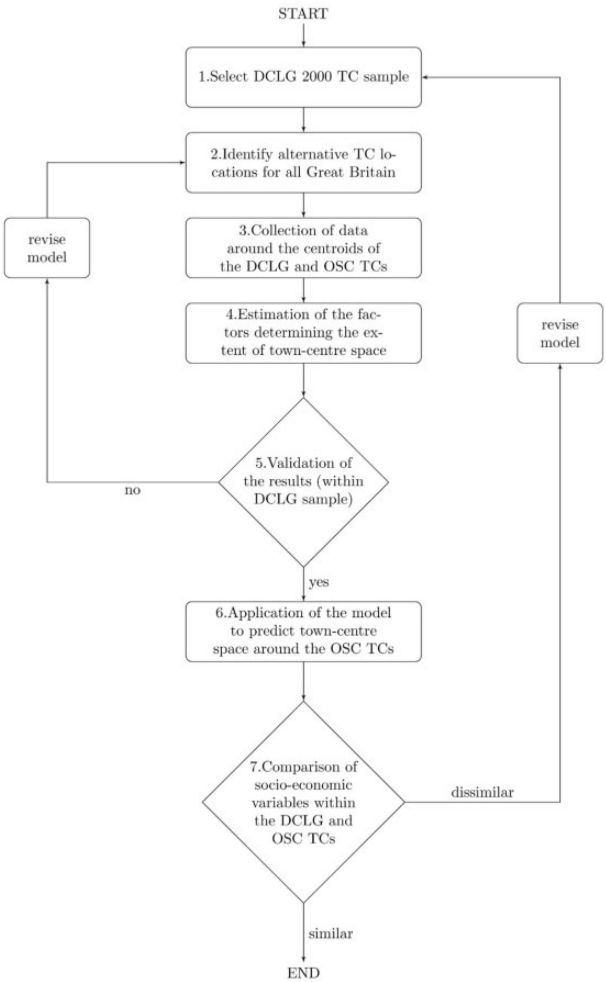

Figure 1. Flow chart illustrating the seven steps of our methodology.

(in particular, for the 1 km data set, we are left with 1001 TCs).[12] In addition, the two types of TC in the DCLG sample, ATCAs and RCs, overlap, so we do not want to use all of the polygons in our regressions. The samples used in the estimations presented in Table 2 differ in the set of DCLG TCs which are used. In Columns (1)–(3) of Table 2, we use all the ATCAs and none of the RCs. However, for Central and West London, the ATCAs are very large, and they mask the richness of small sub-centres (or towns) within them, as is depicted in Figure A2. In Columns (4)–(6) of Table 2, we use the ATCAs in all England and Wales, but for the Central and West London areas, we use the RCs (dark areas).[13]

12 The same occurs in the OSC sample—from 1315 we are left with 964 locations when we use the values of the explanatory variables

13 In all the maps the background geographic areas are the postal sectors.

Within each of these two samples (all ATCAs or ATCAs and London RCs), we introduce an additional criterion to select which TCs to include in the estimations of Step 4. Some of these TCs are certainly very small (25% of the ATCAs have an area of less than 0.08 sq km and an implied radius of less than 160 m). The NSLSP 1998 data allow us to map a set of approximately 4700 shops which consumers identify as their main grocery shopping destinations. Given the very large size of the NSLSP sample—more than 1 million households a year—it seems reasonable to identify TCs which do not contain any of these shops within a certain distance of their boundaries as 'less important'. This is illustrated in Figure 2: for areas around Manchester and Glasgow, we plot (tiny triangles) the NSLSP shops in 1998. We calculate, for both the DCLG and the OSC samples, the number of shops (and shoppers that choose those shops) within different distances of the TC boundary. We can choose an *ad hoc* threshold beyond which we consider the shop too far to be part of that TC.[14] A TC can have shops strictly inside its boundaries, within some allowed close distance of its boundary (fuzzy) or beyond an allowed distance of the boundary.[15] In the full results, we used six distance tolerance levels (fuzzy boundaries): 0 (at least one shop completely within the TCs), 10, 100, 250, 500 m, and 1 km. Without loss of generality, for the regression results provided in the article, we focus on 10, 100, and 500 m. The use of this restriction is what makes the sample size in Columns 1–6 differs from one another. It is worth noticing that the stricter we are with the criterion of at least one NSLSP shop in the TCs, the higher is the explanatory power of the models of Table 2. In Steps 2 and 6, we also use the fuzzy boundary criterion to select which OSC TCs are relevant in our final samples.

In Step 3, we select a large number of explanatory factors to predict the extent of TCs. We choose factors that we believe relate to TC activities. This step involves the collection of potentially relevant variables; GIS work to geographically match the data; and then choosing what variables to include in the final empirical model mainly on the basis of intuition and goodness of fit. This is akin to a forecasting and descriptive process, so we do not pay serious attention to multicollinearity but to the overall validity and explanatory power of the prediction models.

The specific list of variables used in the regressions of Step 4 is inspired by previous attempts in the construction of British TCs.[16] In particular, in the construction of the Index of Town Centredness discussed in the documents and papers that describe the construction

14 We consider the boundaries of the DCLG TCs to be subject to some level of measurement error. Therefore, being very strict about the location of NSLSP shops with respect to the TC boundaries would result in dropping many TCs from the samples. For this reason we adopt a flexible position and try using three different thresholds when selecting the TCs in the different estimating samples.

15 In the map for Manchester we observe all these cases: first inside the ATCA area of Eccles, there are four shops (tiny triangles), while the Trafford Centre has a nearby shop outside its ATCA area but probably inside both a 100 and 500m buffer of its boundary. Finally, Oldham Road, close to the Manchester metropolitan area, has no nearby shop, so it should be dropped from our sample if one of our many restrictions applies. In the map for Glasgow, there are no ATCA areas —because there are no 'official' areas defined for Scotland, only our predictions. These are shown as dark circles around Glasgow and Renfrew. Inside Glasgow's predicted TC, there are six shops (tiny triangles) and one nearby probably at a buffer distance of 10,100 and 500 m.

16 Most relevantly those of ODPM and CASA (2002), ODPM (2004) and Thurstain-Goodwin and Unwin (2000), but also Dolega et al. (2016) and Pavlis, Dolega and Singleton (2017).

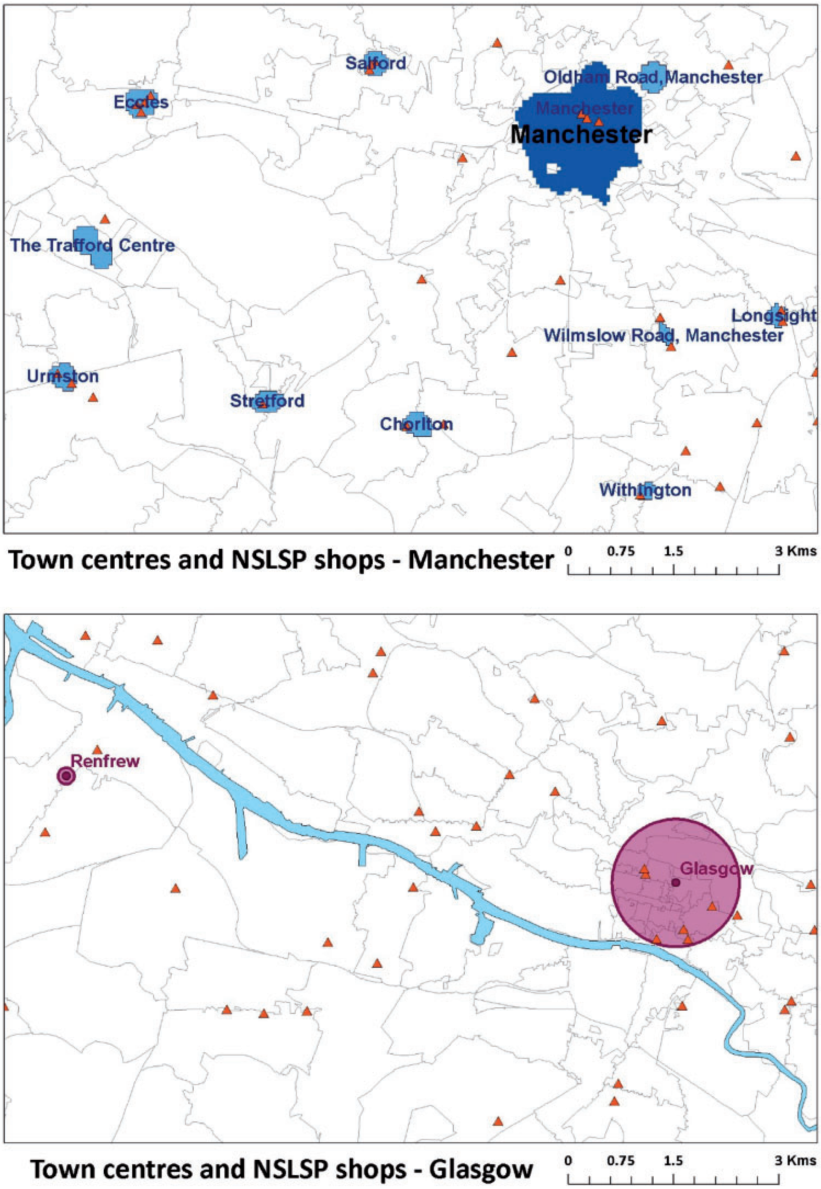

Figure 2. Shops inside and outside an ATCA, Manchester, and Glasgow areas.

of the DCLG TCs (which we try to replicate in our methodology), the authors identify four types of factors that characterize TCs: the economy (type and intensity of economic activities), the diversity (of activities carried out in TCs), visitor attractions (transport, retail, and local amenities), and the property of the buildings (intensity and use of land/floorspace of different activities). Unfortunately, we do not have access to the full set of variables used by ODPM/DCLG, so we collected as wide a range as possible—including variables not used in ODPM/DCLG—that we believe capture the four dimensions specified above.

In addition, we use some features related to the physical geography of the TCs and their relative geographical location such as their elevation and distance to the coastline. We also exploit the postcode hierarchy, which in the UK traditionally relates to historical TCs.[17] The different sets of explanatory variables and their main sources are summarized in Table 1.

The variables we use include factors related to the concentration of retail and shopping activity (two data sets from CBRE: the NSLSP and Retail Locations (RETLOC)); size of the retail sector (units and employment); socio-economic and workplace-based factors (including diversity of employees by occupation);[18] infrastructure endowments; local amenities (cultural, consumption, institutional); postcode centrality (based on the order of the postcodes within the postal sector, district and town);[19] location (distance to social and natural amenities); topological features (elevation and slope); and nightlights brightness intensity.[20] We calculated these features around 1–3 km of both the DCLG and OSC TCs, but in this article, we focus on the results using 1 km. To account for non-linearities, some of the variables are included in levels and also with second- and third-order polynomials.

In Step 4 we use all the variables from Table 1 to predict the (log) radius of the DCLG TCs, and after checking how good the fit is (in Step 5), we apply the estimated coefficient to data around the OSC TCs. Formally, our prediction is in two steps. First, we estimate the extent of TCs regressing the explanatory variables (such as shoppers, socio-economic, etc.) on the TC radius using the DCLG England and Wales TCs sample (DCLG TCs):

$$\log(TC\ radius_{DCLG}) = \alpha + \beta_{1,DCLG} shoppers_{DCLG} + \beta_{2,DCLG}\ socioeconomic_{DCLG} + \ldots + \varepsilon. \tag{1}$$

The results of the regressions on the six DCLG 2000 samples explained above are provided in Table 2. Most of the estimates are significantly different from 0 (and by groups, all the sets of explanatory variables are jointly significant) and the goodness of fit of the models is very high (R^2 between 0.78 and 0.88). This suggests that our models predict the extent of TCs relatively well.

Having estimated these models, we can save the resulting coefficient values and apply them to different values of the explanatory variables. We do that in Steps 5 and 6. In Step 5, we apply

17 See for example https://www.bph-postcodes.co.uk/guidetopc.cgi

18 Diversity of activities in TCs is one of the key factors determining town centredness in the ODPM methodology. As much as we would like, the data we can use to take into account the diversity of employment around the TC locations (based on the Census 2001 workplace statistics) do not have detailed information on sectors, just occupations. We try to capture diversity in the regressions in two ways: by constructing a normalised Herfindahl (HH) index using the nine occupation categories (which estimates one coefficient) or by flexibly including the share of each occupation on total employment as separate variables, either all nine categories or grouped in a few categories, which allows us to estimate one coefficient per group. In the final results presented in Table 2, we use the second option, and all the coefficients are highly significant. Adding an HH index does not significantly increase the explanatory power of the models.

19 Postal sectors, districts, and towns are aggregations of postcode units in the UK. Their letters and numbers relate to their 'centrality'. For more information, see https://en.wikipedia.org/wiki/Postcodes_in_the_United_Kingdom.

20 We experimented adding additional topological features related to land use (EEA Corine data) and other natural boundaries (share of land in water bodies and green spaces) but none add any further explanatory power to the models.

the coefficients to the DCLG sample to compare the predicted and actual TC radius (and area) for the estimating samples. In Step 6, we apply the coefficients out-of-sample to the set of OSC TCs to predict the extent of TC space for the new set of TC locations. Formally, we calculate the prediction by multiplying the estimated $\hat{\beta}$ s to a different set of locations (OSC TCs):

$$\log(TC\ \widehat{radius}_{OSC}) = \hat{\alpha} + \hat{\beta}_{1,DCLG}\ shoppers_{OSC} + \hat{\beta}_{2,DCLG} socioeconomic_{OSC} + \ldots \quad (2)$$

Table 3 (and Table A3 for England and Wales separately) summarizes the actual and predicted values for the radius and area of the DCLG TCs for the whole sample (1001, 1075 TCs minus 74 TCs without shops within 1 km of the centroid) for each of the six specifications of Table 2 and for the average of the six predictions. In the bottom panel, for each model, we show the summary statistics of the predictions when we restrict the observations to the sample used in each of the estimated models. By comparing the numbers in each row with the actual values in the first row, we can see that on average, the actual values are very similar to the actual TC values. In Table 3 we provide correlations between the actual and predicted values for the same samples for both England and Wales and separately by country (Table A3). The correlations are again very high, and, in some cases (especially for the predicted area), they are almost equal to 1.

Once we obtain the coefficient in Step 4 and validate the model in Step 5, in Step 6 we apply them to the data around the OSC centroids and calculate their predicted radius of the OSC TCs. This allows us to create buffers around the OSC to draw the extent of the OSC TCs in a map. Figures 3–5 illustrate the method. The DCLG TCs are depicted as solid, irregularly shaped areas, and the OSC TCs centroids are depicted as dark points. The background geographical boundaries correspond to the postal sectors.

Figures 3 and 4 show the steps of the prediction method in three boxes, one for Manchester (in England) and one for Cardiff (in Wales). Box A shows the solid, irregularly shaped DCLG's 'main' 2000 TCs around Manchester (Figure 3) and Cardiff (Figure 4). Then in Box B, the dark points show the OSC TCs centroids of towns and cities around Manchester and Cardiff. Finally, in Box C, our predictions of the extent of TCs around these centroids are seen as shaded dark-bordered circles surrounding Manchester and Cardiff. As explained, these predictions have been obtained by applying the estimated coefficients from Table 2 on the data around the OSC centroids. We can see that for these two cases, the location and extent of both DCLG and OSC TCs are very similar.

Figure 5 shows the predictions around Edinburgh and Glasgow, where there is no DCLG counterpart, since these cities are located in Scotland. As expected, the size of the circles of the two major Scottish cities is larger than those in the neighbouring smaller towns.

There could be several not mutually exclusive reasons accounting for differences between the TCs produced by DCLG and our OSC-predicted samples: (i) the number or location of what are considered towns or cities might differ, (ii) we do not have a comparison group for Scotland, so we cannot check how well the model is doing there; and (iii) the shape of TCs differs (the OSC TCs are circular by construction, while the DCLG TCs can have different shapes, e.g. following a street). However as already discussed and shown in Tables 3 and 4, when we compare the actual and predicted values of the area and radius of the TCs to the DCLG sample, they are actually very similar.

Even if the estimated values of the R^2s and the in-sample validations make us confident that we can successfully predict the radius within a TC centroid, we could still be getting

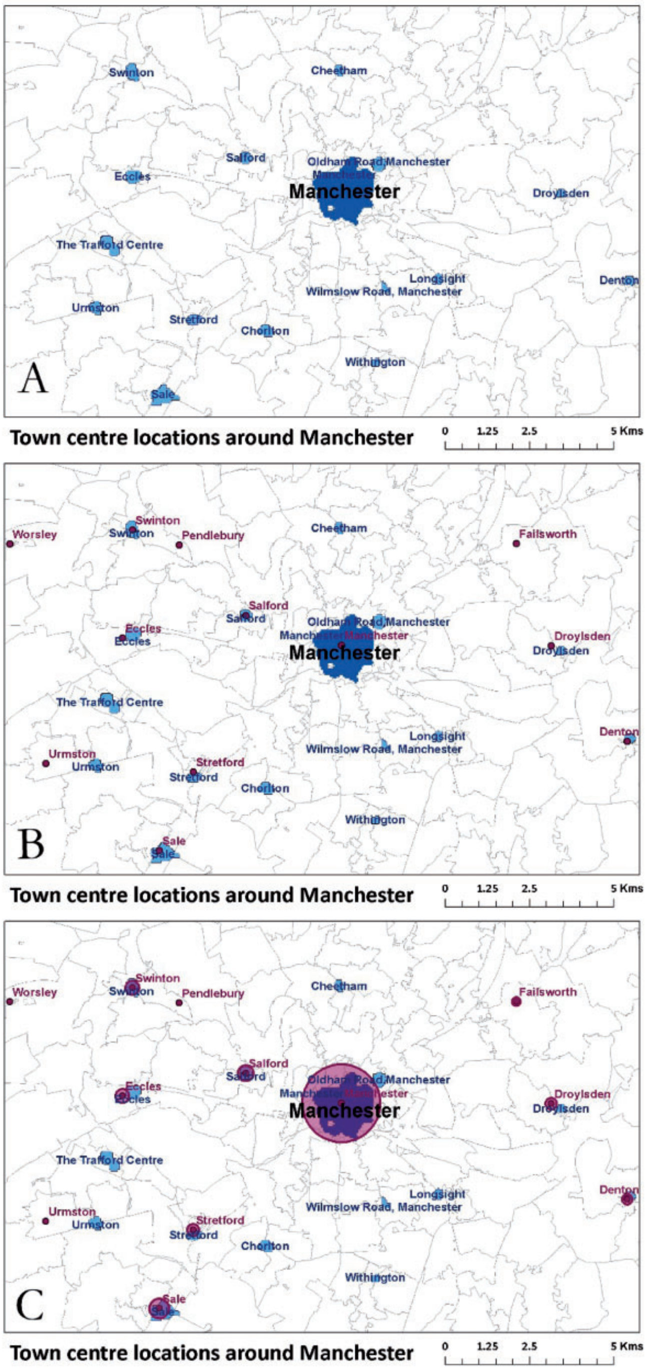

Figure 3. TCs prediction, Manchester, England (step-by-step).

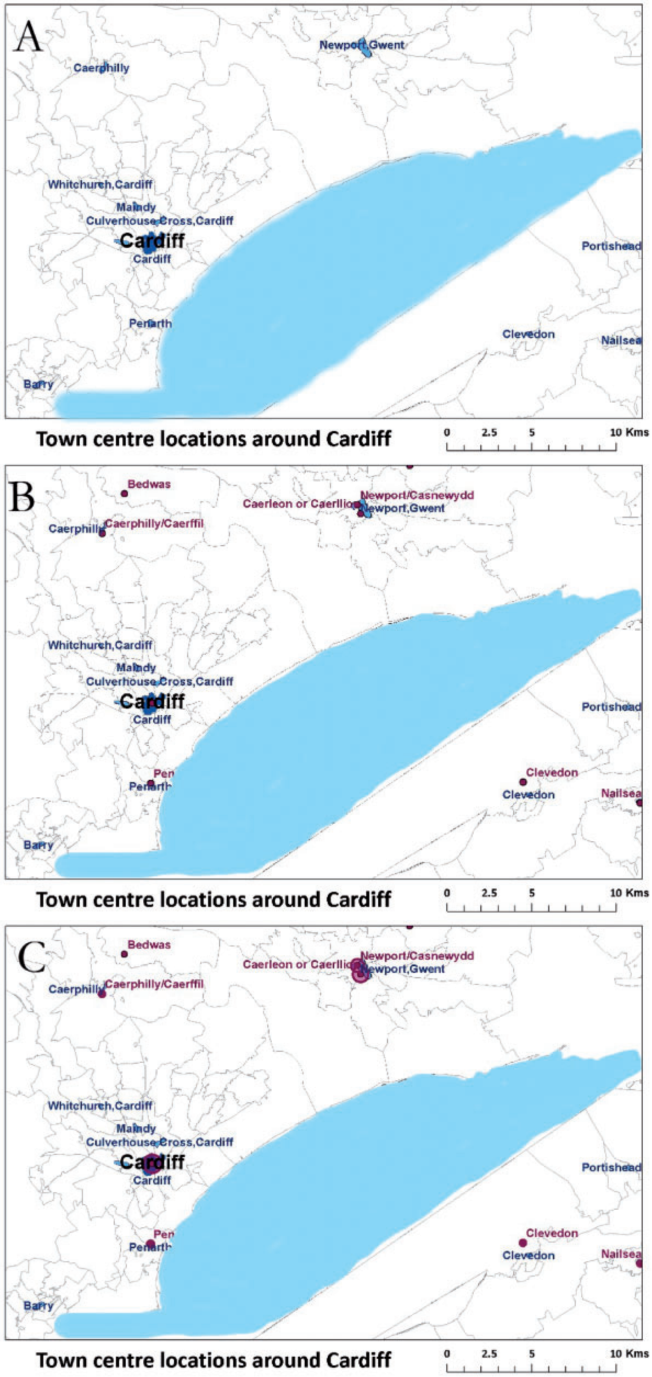

Figure 4. TCs prediction, Cardiff, Wales (step-by-step).

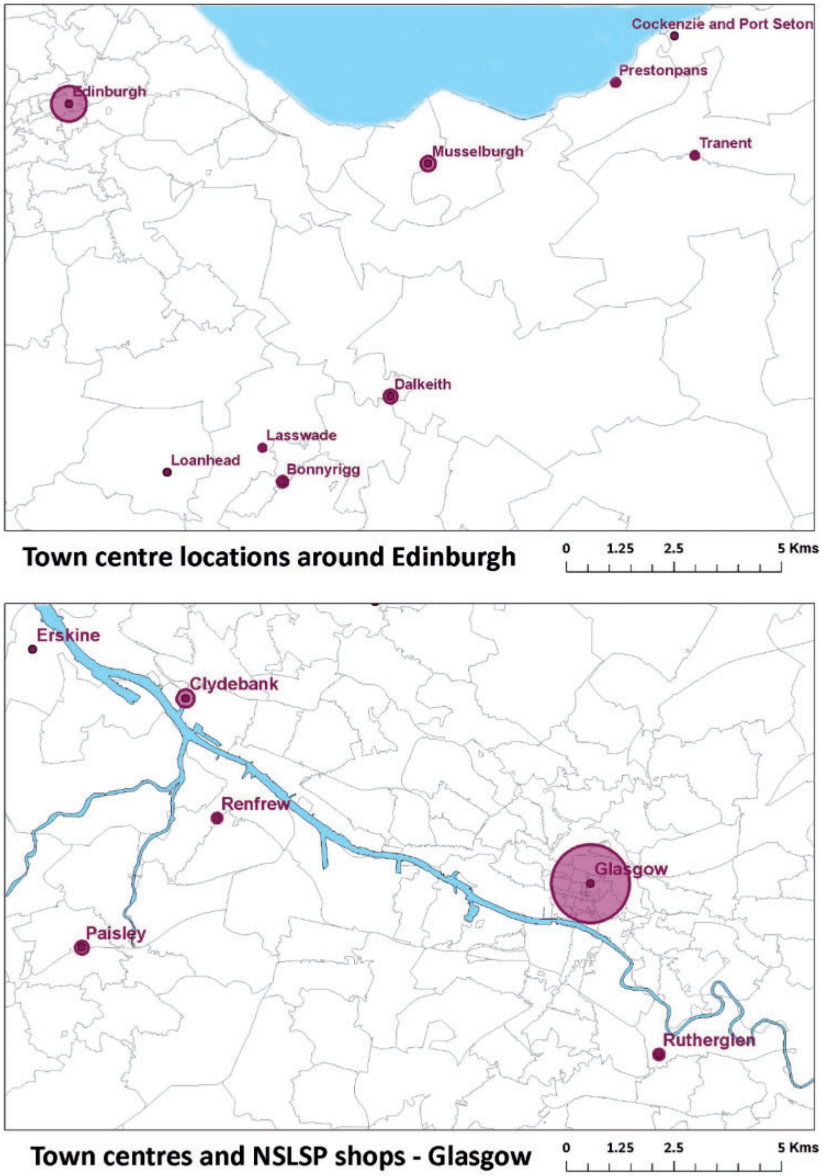

Figure 5. TCs prediction, Edinburgh, and Glasgow.

the 'location' of the TCs wrong if the OSC centroids are not sited in the same place as actual TCs. For this reason, in Step 7 we provide a final validation exercise: we compare the socio-economic characteristics of the TCs in the actual DCLG and OSC-predicted samples, first for the countries where we have information for both (England and Wales) and then, for completeness, for Scotland and the whole of Britain. The results of this exercise are shown in Table 5. The table shows the average value for a set of socio-economic and shopping variables using both the DCLG and the OSC samples (we use the criterion of one shop within 500 m of the boundary to select our TCs). These values were obtained combining data from Table 1 and information of the location and extent of the TCs (the original DCLG 2000 shapefiles and the buffered OSC TCs using the average prediction for the six models of Table 2).

The average value of the variable is provided both for its level and for the by-square-kilometre values (to normalize by the size of the TCs and make them more comparable). The DCLG TCs seem to be slightly larger than the OSC ones, especially in Wales, but in general both samples are quite similar. The number of TCs also differs, with more TCs in England in the DCLG sample and fewer in Wales. The last columns show the values for the sample for the whole of Great Britain and for Scotland alone. The Scottish values seem to be somewhere in between the English and the Welsh ones, but they do not look extremely different from the average British or English and Welsh values. In a nutshell, the statistics in Table 5 suggest that the socio-economic and shopping density values of the DCLG and our OSC samples are quite comparable and so we can be reasonably confident that our methodology yields estimates of TCs for all three countries of Great Britain very similar to those of DCLG for England and Wales alone. This opens the door to rigorous analysis of the evolution of TCs in Scotland compared to those in England and Wales and so to an evaluation of policies introduced on one country but not other(s).

7. Conclusions

A TC is in a sense the opposite of a pole of inaccessibility. But it is more than that. A TC is a spatial pattern, so it is a recognizable regularity of the urban landscape. Given this one would think it should be central to the research interests of economic geographers. But this interest has not been apparent. In this article, we argue strongly the case for opening the black box of town centredness. Micro-geographical data are now readily available and should be used. In this article we propose and apply a method to exploit this type of data to define the location and extent of TCs in Britain.

This article starts with an apparently naïve question: How can one identify a TC in a given city? The answer proves not to be so simple. To answer it we find we need a whole new method. TC policies have been around for several decades, in many European countries (apart from Britain we mainly discuss the case of The Netherlands). These policies seem to have been applied with less rigour than rhetoric. We cite the case of a handbook for Scottish TCs in which there is no definition at all of what a ‘Town Centre’ is, where it is to be found, or how it is to be defined. There are pictures but no maps or definitions. Our research tries to bridge this gap, proposing a new methodology to locate, identify, delimit, and determine the radius of TCs. Calibrating our model on TCs defined by ODPM (2004), we test our method in a full Great Britain setting, but it is easily transferred to other locations or countries because of its reproducibility and ease of calculation.

In this article, we apply a method for predicting the location and extent of TC space to all of Great Britain. Our method relies on four assumptions. The first assumption is that the DCLG TC definitions are good approximations of the true TCs for England and Wales. The second assumption is that the underlying socio-economic and geographical factors within a radius of around 1 km of the TC centroids are effective determinants of TC space. The validity of this assumption can be assessed by looking at the goodness-of-fit statistics of our models predicting the extent of the DCLG’s TCs and at the evidence provided in Table 3–5. The third assumption is that the OS list of towns and cities provides a reliable set of potential TC locations. The final assumption is that the determinants of TC space in Scotland do not systematically differ from those in England and Wales, both in observed and unobserved characteristics relevant to defining TCs. If all these assumptions hold, we can satisfactorily apply the coefficients on socio-economic and geographical variables estimated in Table 2 to Britain-wide data to yield estimates of the location and extent of TCs for England, Wales, and, in particular, Scotland. Equally, so long as the critical assumptions hold, the methodology could be adapted to identify TCs in other countries.

While this study gives an answer to the question of the extent of TCs and so allows one to estimate where their centroids are located, there is no such thing as *the* answer. As our robustness checks and data validations suggest, the method can be considered 'successful' with a correlation of actual to predicted radius of 0.75–0.99 depending on the sample. Our predictions for England and Wales match the actual DCLG ATCA 2000 quite accurately. In Scotland, its direct accuracy cannot be judged because there are no 'official' DCLG TCs—to offset for which is one of the purposes of this study. However, the exploration of socio-economic and shopping density values in and very close to the TCs defined with both methodologies suggests that they provide a very similar picture. Overall, we judge that our method is promising and certainly provides a useful tool to be applied for the evaluation of TCFP, and more generally, for the evaluation of any policy that applies to TCs.

Our final aim is policy discussion and evaluation. Having workable and agreed definitions of TCs and their boundaries is a necessary step if we are to have an open, consistent, and reliable discussion or evaluation of relevant policy. TCs as a distinct spatial pattern of modern cities deserve this effort. In this article, we hope we are demonstrating a replicable method for the analysis of this particular spatial organization which will help in policy development and analysis. At least, the discussion both in the UK and in Europe signals an urgent need to first consider town centredness seriously as a precondition to policy analysis and debate.

TCs, their extent and the hierarchy of TCs, are, as we argue in Section 2, closely related to, indeed an extension of, CPT and gravity models. Many of our assumptions are borrowed directly from these two intellectual traditions, but some come from a more empirical approach where several of the recent papers we discuss have shown the way. We hope future policy debates may incorporate our primary aim: that we should have agreed definitions of things before we launch discussion, let alone policy for them; and perhaps borrow or adapt our methodology. TCs should be recognized as real entities with real shapes, with real areas and real boundaries, capable of real descriptions and definitions.

Finally, with this article we aim to contribute to improving not just debates about cities and TCs but to other debates where people and policy proceed ahead of any clear definition of what it is they are analysing or generating policy for. This has very much been the case with TCs (as it has with other concepts relating to urban development such as 'sprawl'), but they are not abstruse ideas, and we hope this article has shown that they can be clearly and unambiguously defined and identified and that they are basically material, applied, and experimental in nature.

Acknowledgements

The authors thank SERC/CEP for funding (ESRC grant ES/M010341/1), and R.S.-G. acknowledges support from the British Academy. The authors thank Mr Mark Teale from CBRE and DCLG for providing them with data. ABI data accessed via NOMIS under BRES notice NTC/BRES15-P0572. The authors benefited from comments and suggestions from participants at the SERC Work-in-progress research seminar, at the CESifo Conference on the Use of Geocoded Data in Economic

Research and at the University of Birmingham City-REDI seminar. The authors finally thank Steve Gibbons and Daniel Arribas-Bel for comments and an anonymous referee for helpful criticisms. Any errors remain our own.

References

Cheshire, P. C., C. A. L. Hilber, and I. Kaplanis (2015), "Land Use Regulation and Productivity-Land Matters: Evidence from a Uk Supermarket Chain", *Journal of Economic Geography*, **15**, 43–73.

Cheshire, P. C., C. A. L. Hilber, and R. Sanchis-Guarner (2014), "Do English Planning Policies Make Shopping More Sustainable?", Working paper, IEB Urban Economics Workshop, Barcelona, May 2014. http://www.ieb.ub.edu/files/PapersWSUE2014/Sanchis.pdf (last accessed 26 February 2018).

Christaller, W. (1933), *Die Zentralen Orte in Süddeutschland*. G. Fischer, Jena.

Denike, K. G. and J. B. Parr (1970), "Production in Space, Spatial Competition, and Restricted Entry", *Journal of Regional Science*, **10**, 49–64.

Department of the Environment (1996), *Planning Policy Guidance: Town Centres and Retail Developments. Revised PPG6*. HMSO, London.

Dicken, P. and P. E. Lloyd (1990), *Location in Space: Theoretical Perspectives in Economic Geography*, 3rd edn. Harper and Row, New York.

DETR (1998), *Town Centres: Defining Boundaries for Statistical Monitoring - Feasibility Study*. HMSO, London.

Dolega, L., M. Pavlis, and A. Singleton (2016), "Estimating Attractiveness, Hierarchy and Catchment Area Extents for a National Set of Retail Centre Agglomerations", *Journal of Retailing and Consumer Services*, **28**, 78–90.

Everitt, B. and T. Hothorn (2011), *An Introduction to Applied Multivariate Analysis with R*. Springer, New York.

Evers, D. (2002), "Window on The Netherlands. The Rise (and Fall?) of National Retail Planning", *Tijdschrift Voor Economische En Sociale Geografie*, **93**, 107–13.

Fujita, M., P. R. Krugman, and A. Venables (1999), *The Spatial Economy: Cities, Regions and International Trade*. MIT Press, Cambridge, Massachusetts.

Guimarães, P. (2016), "Revisiting Retail Planning Policies in Countries of Restraint of Western Europe", *International Journal of Urban Sciences*, **20**, 361–80.

Huff, D. L. (1963), "A Probabilistic Analysis of Shopping Center Trade Areas", *Land Economics*, **39**, 81–90.

Huff, D. L. (1964), "Defining and Estimating a Trading Area", *Journal of Marketing*, **28**, 34–8.

Huff, D. L. (2003), "Parameter Estimation in the Huff Model", *ArcUser*, 34–6.

Klaesson, J. and Ö. Öner (2014), "Market Reach for Retail Services", *The Review of Regional Studies*, **44**, 153–76.

Lösch, A. (1940), *Die Räumliche Ordnung Der Wirtschaft*. G. Fischer, Jena.

ODPM (2004), *Producing Boundaries and Statistics for Town Centres, England and Wales 2000 Interim Report*. Office of the Deputy Prime Minister, London.

ODPM and CASA (2002), *Producing Boundaries and Statistics for Town Centres - London Pilot Study Summary Report*. The Stationery Office, London.

Pavlis, M., L. Dolega, and A. Singleton (2017), "A Modified DBSCAN Clustering Method to Estimate Retail Center Extent", *Geographical Analysis*, doi: 10.1111/gean.12138.

Reilly, W. J. (1929), *Methods for the Study of Retail Relationships*. University of Texas, Austin.

Reilly, W. J. (1931), *The Law of Retail Gravitation*. Knickerbocker Press, New York.

Thurstain-Goodwin, M. and D. Unwin (2000), "Defining and Delineating the Central Areas of Towns for Statistical Monitoring Using Continuous Surface Representations", *Centre for Advanced Spatial Analysis*, **18**, 1–11.

Appendix

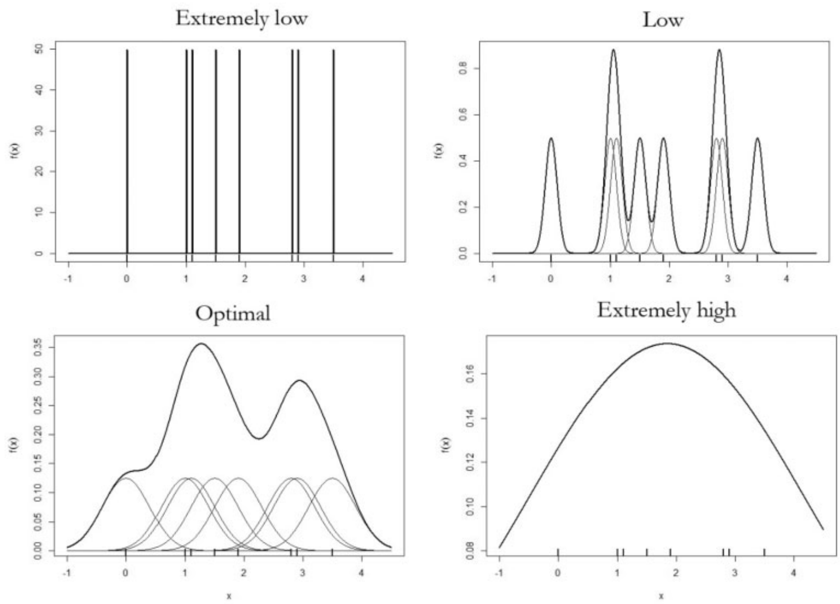

Figure A1. Bandwidth selection (modified from Everitt and Hothorn 2011).

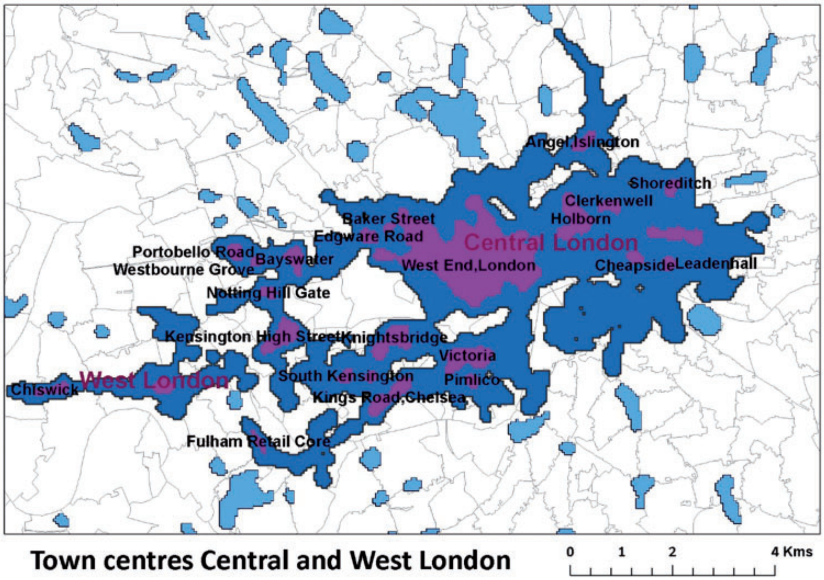

Figure A2. ATCA and RCs in London in the DCLG sample.

Table A1. List of abbreviations

ABI	Annual Business Inquiry
ATCAs	Areas of Town Centre Activity
BRES	Business Register and Employment Survey
CASA	Centre for Advanced Spatial Analysis
CBRE	CBRE Group., Inc.
CEP	Centre for Economic Performance
City-REDI	City-Region Economic and Development Institute
CPT	Central Place Theory
DCLG	Department for Communities and Local Government
DETR	Department of the Environment, Transport and the Regions
EEA	European Environment Agency
ESRC	Economic and Social Research Council
GB	Great Britain
GIS	Geographic Information System
NASA	National Aeronautics and Space Administration
NGDC	National Geophysical Data Center
NOAA	National Oceanic and Atmospheric Administration
NOMIS	Official Labour Statistics Portal (from ONS)
NSLSP	National Survey on Local Shopping Patterns
NSPD	National Statistics Postcode Directory
ODPM	Office of the Deputy Prime Minister
ONS	Office for National Statistics
OS	Ordnance Survey
OSC	Towns and cities list in the Ordnance Survey Gazetteer
POI	Points of Interest
RCs	Retail Cores
RETLOC	Retail Locations
SERC	Spatial Economics Research Centre
TC(s)	Town Centre(s)
TCFP	Town Centre First Policies
UPC	Unit Post Code
UK	United Kingdom
VOA	Valuation Office Agency

Table A2. Details on variables used in the estimates of Table 2 (all within 1 km of centroid)

Variable	Description	Data source
Number of (grocery) shoppers	Shoppers in shops located around the centroid from shopping patterns survey	NSLSP and RETLOC 1998
Number of (grocery) shops	Number of grocery shops located around the centroid from both CBRE surveys/directories	NSLSP and RETLOC 1998
Average distance to (grocery) shops	Average distance from centroid to all shops around it from both CBRE surveys/directories	NSLSP and RETLOC 1998
Distance to closest (grocery) shop	Distance from centroid to closest shop	NSLSP 1998
Number of units in retail sector	Establishments in sector 52 (SIC 1992)	NOMIS ABI 1998
Employment in retail sector	Employment in sector 52 (SIC 1992)	NOMIS ABI 1998
Density of total population	Total residential population in a given area	UK Census 2001
Workplace employment	Workplace-based total employment	UK Census 2001
Share of workplace employment in different occupation levels (diversity)	High (managerial/professional), medium (intermediate/supervisory), low (routine)	UK Census 2001
Average commuting distance	Average distance in kilometres	UK Census 2001
% of commuters by foot or bike	Percentage using bike or walking to work	UK Census 2001
% of commuters using public transport	Percentage using bus/tube/rail/taxi	UK Census 2001
Average household size	Average number of members in household	UK Census 2001
Average age of the resident population	Average age of all resident population	UK Census 2001
% of population aged 18–44 years	Percentage of residents aged 18–44 years	UK Census 2001
Total unemployment rate	Total unemployed as percentage of active population	UK Census 2001
Share of students in population	Share of students in population	UK Census 2001
Share of retirees in population	Share of retired workers in population	UK Census 2001
Kilometres of all-type roads	Motorways, primary, A & B, and minor roads	OS Strategi 2009
Number of bus stations	Number of major bus stations (not bus stops)	OS Points of Interest 2015
Number of tube/tram stations	Number of tube and tram stations	OS Points of Interest 2015
Number of rail stations	Number of rail station (not tube or tram)	OS Points of Interest 2015

(continued)

Table A2. Continued

Variable	Description	Data source
Number of libraries	Number of public libraries	OS Points of Interest 2015
Number of museums	Number of museums open to the public	OS Strategi 2009
Number of art galleries	Number of art galleries	OS Points of Interest 2015
Number of cinemas and theatres	Number of cinemas and theatres	OS Points of Interest 2015
Number of discos and nightclubs	Number of discos and nightclubs	OS Points of Interest 2015
Number of landmarks	Landmark point (historic building/castles, etc.)	OS Strategi 2009
Number of cafes, restaurants, and pubs	Number of cafes, restaurants, and pubs	OS Points of Interest 2015
Number of B&B, hotels, and motels	Private accommodation establishments	OS Points of Interest 2015
Number of youth hostels	Youth hostel establishments	OS Points of Interest 2015
Number of local government sites	Number of government-related buildings	OS Points of Interest 2015
Number of tourist information offices	All tourist offices (including seasonal)	OS Strategi 2009
Number of visitor centres	All visitor centres which are not tourist offices	OS Strategi 2009
Share of central addresses (postal sector in postal town)	Central addresses (those whose postal sector number is 1) in postal town: total number	NSPD and Wikipedia
Distance to first postcode in postal town	First postcode in the postcode hierarchy	NSPD and Wikipedia
Distance to closest town hall	Town hall belonging to district administration	Wikipedia
Distance to closest rail or tube station	Closest tube or rail (one or the other)	OS Points of Interest 2015
Distance to closest point in coastline	Closest distance to the coastline	OS Strategi 2009
Distance to closest river or lake	Closest distance to water bodies	OS Strategi 2009
Distance to closest natural park or woodland	Closest distance to natural national parks or woodland (green spaces)	OS Strategi 2009
Elevation/terrain slope statistics	Standard deviation, mean, max, range of terrain elevation (altitude), and slope (degrees)	OS Panorama 50x50 m
Nightlight brightness statistics	Standard deviation, mean, max, range, and sum of nightlight brightness (non-censored version)	NASA and NOAA-NGDC

Table A3. England and Wales TC radiuses, actual and predicted

	TC derived radius in km					TC area in km^2				
Variable	Observed	Mean	Standard deviation	Minimum	Maximum	Observed	Mean	Standard deviation	Minimum	Maximum
England										
Observed value	944	0.248	0.156	0.113	3.187	944	0.270	1.077	0.040	31.905
Average all samples	944	0.263	0.168	0.107	2.973	944	0.316	1.021	0.036	28.018
Sample 1 betas	944	0.276	0.200	0.104	3.095	944	0.366	1.257	0.034	30.091
Sample 2 betas	944	0.274	0.199	0.106	3.076	944	0.360	1.237	0.035	29.718
Sample 3 betas	944	0.263	0.181	0.105	3.051	944	0.320	1.089	0.034	29.245
Sample 4 betas	944	0.259	0.150	0.105	2.526	944	0.282	0.738	0.034	20.040
Sample 5 betas	944	0.257	0.155	0.107	2.711	944	0.283	0.831	0.036	23.088
Sample 6 betas	944	0.252	0.166	0.106	3.382	944	0.285	1.218	0.035	35.929
Sub-sample 1	768	0.258	0.160	0.104	3.095	768	0.289	1.120	0.034	30.091
Sub-sample 2	820	0.252	0.156	0.106	3.076	820	0.276	1.075	0.035	29.718
Sub-sample 3	876	0.246	0.150	0.111	3.051	876	0.260	1.025	0.039	29.245
Sub-sample 4	780	0.253	0.121	0.105	1.040	780	0.247	0.311	0.034	3.397
Sub-sample 5	832	0.248	0.119	0.107	1.070	832	0.237	0.308	0.036	3.600
Sub-sample 6	894	0.241	0.113	0.108	1.030	894	0.222	0.287	0.036	3.336

(continued)

Table A3. Continued

Variable	TC derived radius in km					TC area in km^2				
	Observed	Mean	Standard deviation	Minimum	Maximum	Observed	Mean	Standard deviation	Minimum	Maximum
Wales										
Observed value	57	0.218	0.096	0.113	0.613	57	0.178	0.198	0.040	1.180
Average all samples	57	0.216	0.095	0.128	0.570	57	0.175	0.199	0.052	1.022
Sample 1 betas	57	0.219	0.097	0.125	0.585	57	0.180	0.206	0.049	1.077
Sample 2 betas	57	0.217	0.098	0.128	0.590	57	0.178	0.209	0.051	1.095
Sample 3 betas	57	0.213	0.091	0.122	0.534	57	0.168	0.182	0.047	0.895
Sample 4 betas	57	0.219	0.098	0.128	0.588	57	0.181	0.209	0.052	1.087
Sample 5 betas	57	0.217	0.098	0.127	0.589	57	0.178	0.210	0.050	1.089
Sample 6 betas	57	0.213	0.090	0.123	0.532	57	0.167	0.179	0.048	0.888
Sub-sample 1	44	0.218	0.093	0.125	0.538	44	0.218	0.093	0.125	0.538
Sub-sample 2	50	0.212	0.089	0.128	0.538	50	0.212	0.089	0.128	0.538
Sub-sample 3	55	0.208	0.082	0.122	0.495	55	0.208	0.082	0.122	0.495
Sub-sample 4	44	0.218	0.093	0.128	0.538	44	0.218	0.093	0.128	0.538
Sub-sample 5	50	0.212	0.089	0.127	0.535	50	0.212	0.089	0.127	0.535
Sub-sample 6	55	0.207	0.081	0.123	0.494	55	0.207	0.081	0.123	0.494

Notes: Samples define as explained in notes of Table 3.