A parametric model of residential built form for forecasting the viability of sustainable technologies

Hargreaves, Anthony J

DOI:
10.1016/j.scs.2021.102829

A parametric model of residential built form for forecasting the viability of sustainable technologies

Anthony J. Hargreaves

School of Engineering, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Abstract

Renewable supply systems, such as rainwater harvesting and ground source heat pumps, have the potential to improve the sustainability and resilience of residential areas. However, their feasibility partly depends on the dimensions and plot areas of the dwellings, and there has been a lack of a suitable method of modelling these for future urban development. This paper therefore calibrates and validates a method of modelling how the plot areas and footprints of dwellings would vary in size. Its inputs are the total dwellings and their average density for each area type within a district. It could thereby complete a chain of urban modelling from regional to local scale for testing spatial planning scenarios. The results show that the calibration of the model is transferable between spatial scales and UK regions, and the results are validated against detailed GIS data. It is then developed into a novel parametric model to estimate how built form will affect the future potential of renewable supply systems, and this is demonstrated using rainwater harvesting as an example. It provides estimates of water-savings at district scale that are more reliable than the usual method of using discrete average dimensions per dwelling type.

Keywords:
Gamma distribution; housing typologies; spatial interaction; rainwater harvesting; renewable supply; urban modelling.

E-mail address: a.j.hargreaves@bham.ac.uk.
Published in Sustainable Cities and Society https://doi.org/10.1016/j.scs.2021.102829

1. Introduction

Renewable supply systems, such as rainwater harvesting systems and ground source heat pumps, have the potential to improve the sustainability and resilience of residential areas (Groppi et al., 2018; Wilcox et al., 2016). This could have substantial benefits because residential areas account for a large percentage of the urban land area. However, uncertainty about whether residential properties will have the space required for their installation is a barrier to implementation (DECC, 2012). The sizing of these technologies depends on the temporal balance between demand and supply. These demands depend on the building size, efficiency, and household characteristics (Cheng & Steemers, 2011); and the renewable supply depends on dimensions, such as the roof area and plot area (Jenkins et al., 2009; Campisano & Modica, 2012). Their performance also partly depends on location specific environmental factors, such as soil conditions, solar energy, and rainfall (Blum et al., 2010).

The cost effectiveness of these sustainable technologies is often delicately balanced between their overall cost and their savings on utility bills. Individually they are rarely cost effective on a commercial basis compared to conventional supply. However, collectively they can reduce environmental impacts, use of natural resources, and how much investment is needed in centralised systems (Campisano et al. 2017). They may therefore justify policy support through regulations and subsidies (Friedler & Hadari, 2006). Their installation is likely to be mainly as part of new build or refurbishment, and so their uptake will only be gradual. A decision on such support would therefore require a long-term perspective and so would need to consider spatial planning. This affects the future location and the built form of urban development, and hence the likely uptake and performance of sustainable technologies. Planning policies often change over time to reflect the changing priorities of policy makers. For example, the UK Planning Policy Statement (DCLG, 2005) constrained development to brownfield land at higher densities with the

aim of reducing urban expansion and car use. However, this was superseded by the National Planning Policy Framework (DCLG, 2012), which allowed development on greenfield sites to address a growing housing shortage and support economic recovery from the 2007-2008 financial crisis. An assessment of the future potential of building-scale sustainable technologies will therefore require sensitivity testing for a range of possible spatial planning policies over the forecast period (Hargreaves et al., 2019). Forecasting the growth and density of urban areas requires a '*spatial interaction model*' (McFadden, 1974; Williams, 1977) because the location choices of firms and households change over time depending on how planning policies and transport investment affect land values and access to employment and services. There has been increasing interest in using these models to forecast the resilience of spatial planning options and infrastructure (Ford et al., 2018). They are useful for understanding the future interactions between variables, and differences between options. The planning policy inputs to a spatial interaction model include the future location and amount of residential land, and forecasts the population, households, and floorspace per zone (Echenique et. al., 2013). The zones are usually local authority districts, or smaller areas, depending on the spatial detail of the model.

Linking regional-scale models to the building-scale has been a major unsolved challenge. Some studies have aimed to do this by representing the neighbourhood layout and buildings using pre-designed land parcels or cells. For example, UrbanSim used 2.25 ha grid cells (Waddell et al., 2003), but this results in difficulties matching and reconciling data and outputs of different types and scales (Abraham et al., 2005). Modelling at this parcel scale rather than at building scale also means that the uptake and yield of building-scale supply, such as rainwater harvesting, needs to be input based on assumptions about average values per parcel (Willuweit & O'Sullivan, 2013). The analysis of GIS data can establish relationships between urban morphology indicators and demands (Chen et al. 2020), but this does not provide enough detail of the dwellings to model building-scale supply systems and it lacks a forecasting capability. Some studies have focussed on simulating the local built form in more detail using rule-based models (Bach et al. 2018) but these rely on external inputs such as urban growth projections and planning parameters and so are unsuitable for testing future options for spatial planning and transport.

The housing stock is generally classified as a typology based on housing survey data. In the UK, the main types are detached, semidetached, end-terraced, mid-terraced, and flats (Orford & Radcliffe, 2007). Bottom-up models usually use dwelling types with dimensions that are an average of housing survey data. This is a way of simplifying research on heating demands. These are linearly correlated with floorspace per dwelling type (Cheng, & Steemers, 2008) and so factoring the demands by the number of dwellings provides an estimate for the housing stock. This method of using average dimensions per dwelling type has also been used for modelling the potential yield of renewable supply systems (Tarnawski et al. 2009; Domènech, & Saurí, 2011; Kim, & Furumai, 2012). However, these estimates would be inaccurate for the housing stock because renewable supply is not linearly correlated with dwelling dimensions. The reason is that the temporal balance of supply and demand per system partly depends on the area that a dwelling has available for the harvesting of renewable supply. This in turn affects the efficiency and cost effectiveness of these systems and hence their likely uptake (Friedler, & Hadari, 2006; Tarnawski et al. 2009; Campisano, & Modica, 2012).

Bottom-up simulation methods are being developed for exploring a range of sustainable built forms per dwelling type (D'Amico & Pomponi, 2019). However, there is a need for a top-down method of forecasting what built form would arise from spatial planning options. Utility companies, particularly water companies, have expressed a strong interest in using this type of model as a collaborative planning tool for constructive engagement with planning authorities on high-level scenario analysis. It could enable focused optioneering on the spatial demand for water and potential solutions for future sustainable supply and drainage.

A novel way of modelling the variability of built form within a top-down regional modelling framework was published in Hargreaves (2015). Its inputs were the number of dwellings and their average residential density, which would be available as forecasts from a spatial interaction model. It modelled the plot area (aka lot area) per dwelling as a probability distribution. This was then approximated using generic 'tiles' to represent dwelling types of different densities. The 'tiles' were able to simultaneously match the residential land area, number of dwellings and their

plot areas. This would have been very difficult to achieve if using a land parcel or grid cell method. The tiles method could achieve this because they are an abstract representation of land use and built form without being constrained to match a predesigned spatial layout (simulating the future layout of residential areas is unnecessary for testing sustainable technology systems for individual buildings). The method was developed by analysing English House Condition Survey (EHCS) data (DCLG, 2009). This survey had a large sample size per region, but this was still less than one in a thousand of the housing-stock. Hence, there are obvious uncertainties about how well the model would work at a more detailed scale when compared against GIS data.

Therefore, this study for the first time, calibrates and validates this method at district scale using detailed GIS data. It then develops this method into a novel parametric model of how built form affects the likely uptake and yield of sustainable technologies. This is demonstrated, using rainwater harvesting as an example, and its outputs are compared with those of the usual method of using discrete average dimensions per dwelling type. The paper will be of interest to researchers, policymakers, and practitioners of the planning, modelling, and design of buildings and infrastructure for sustainable urban development.

2. **Materials and methods**

2.1. *Modelling the shape of the plot density distribution*

The *'tiles-method'* presented in Hargreaves (2015) is based on modelling how *'Plot-density'* varies between dwellings. Plot-density is defined as,

$$\mu_i = h_i / a_i, \tag{1}$$

Where a_i is the *'Plot-area'* in hectares (the plot is the land immediately surrounding and belonging to a residential building) and h_i is the number of dwelling units on plot i (the Nomenclature is on page 23). For a house, $h_i = 1$, and for apartments, h_i is the number of apartments within the plot. Plot-density μ_i characterizes an individual dwelling by how much land belongs to the dwelling. This is conceptually different to the net density metric used by urban planners, which characterizes a residential area by the number of dwellings per hectare. The Plot-density metric is useful because it is inversely correlated with Plot-area and hence the space available for sustainable technologies. It is also consistent with the spatial interaction modelling of residential land and density per zone, which is based on the land available for housing plots, excluding public roads, paths, and open spaces. This innovative Plot-density metric can therefore link the regional-scale aggregate modelling of land use to the modelling of plot areas at the local-scale. As a first step, this uses the outputs of a spatial interaction model (or equivalent planning forecasts) to calculate the '*Mean plot-density'* $\bar{x}_j$,

$$\bar{x}_j = H_j / A_j, \tag{2}$$

Where A_j is the land-take by dwelling plots in hectares and H_j is the number of dwellings, per zone j. The probability density function of Plot-density per dwelling μ is modelled using the gamma distribution,

$$f(\mu; k, \theta) = \frac{1}{\theta^k . \Gamma(k)} \mu^{k-1} \cdot e^{-\frac{\mu}{\theta}} \; for \; \mu > 0 \text{ and } k, \theta > 0 \tag{3}$$

Where, k is the shape parameter, θ is the scale parameter, and Γ is the gamma function. The mean of this probability density function is, $\bar{\mu} = k \cdot \theta$. If $k \geq 1$ the gamma distribution is positively skewed, and Hargreaves, (2015) showed that $\bar{x}$ is the mode of the probability density function, (i.e., $\bar{x} = \arg_{max} \mu$) and so,

$$\theta = \bar{x} / (k - 1) \tag{4}$$

It is postulated that this theoretical gamma distribution can be fitted to the empirical '*Plot-density distribution'* extracted from GIS data. Also, that the shape parameter can be estimated per dwelling type as k_d and will be like the findings of the Hargreaves (2015) study. This will confirm that k_d can be calibrated for modelling purposes and is transferable between spatial scales and UK regions. The null hypothesis is that the fitted theoretical distribution and this empirical data are drawn from the same distribution. This is tested using the Kolmogorov Smirnov (K-S) one-sample test (Siegal and Castellan 1988). This looks for the largest difference, in absolute terms, between the theoretical cumulative distribution function and the empirical cumulative distribution function. If this test statistic '*ksstat*' exceeds the critical value, it is cause for rejection of the null hypothesis and implies that the fitted distribution is not doing an adequate job of modelling the empirical data. The Supplementary Material provides further details of the K-S test (Section B1), and a Matlab® programme written to fit the gamma distribution to the GIS data and calculate ksstat (Section B2).

2.2. *The tiles method*

The tiles method was used by (Hargreaves, 2015) to approximate the probability distribution per dwelling type. Each generic tile represents dwellings of the same type and Plot-area occupying one-hectare of residential land. There are ideally at least five tile types per dwelling type, each with a different Plot-density (Fig. 1). Tiles are then systematically selected to approximate the theoretical Plot-density distribution, and this can include fractions of tiles. They are best conceptualised as abstract objects within a spatial database. The Plot-density of each tile type is calibrated for the study area so that they each represent approximately the same percentage of dwellings per dwelling type. The dwelling dimensions per tile are the same as the EHCS data for dwellings of that type and Plot-density. Its areas of associated roads, pathways, and public land are based on GIS analysis. The number of dwellings per one hectare tile is its net residential density (for further details, see Hargreaves 2015). For consistency, all the densities reported in this paper are in Plot-density units.

The tiles are useful as a shared medium for multidisciplinary research on urban modelling of building-scale sustainable technologies and residential land use (Hargreaves et al., 2017; Hargreaves et al., 2019). They encapsulate the built-form inputs needed for modelling these technologies. The costs and savings per dwelling can then be summed for the study area. The tile types shown in Fig. 1 encompass the residential built form of English regions, but this can vary substantially within a study area. For example, apartments in central urban areas are often high-rise like tile types A4 and A5 but these tiles will not be representative of apartments in suburban areas, which are usually like tile types A1 and A2. Ideally, a set of tile types would be calibrated and designed for each area type so that there are enough tile types to closely approximate the Plot-density distribution. However, this would make the tiles-method too onerous in time and complexity. Hence, this study develops a novel parametric model that incorporates the advantages of the tiles-method but can more flexibly and closely fit the Plot-density distribution of different area types and is more practicable to use.

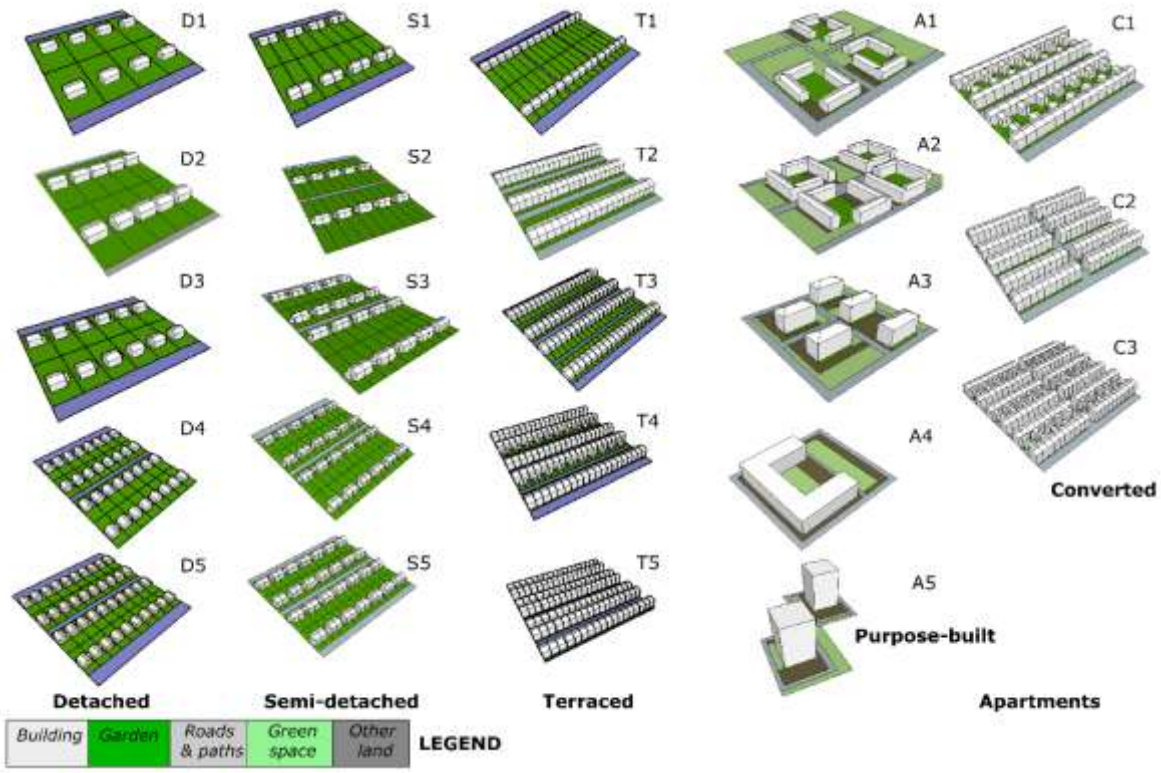

Fig. 1. Schematic examples of generic tiles.
Source: Hargreaves (2015)

2.3. Study areas

The areas selected for this study are four local authority districts in the West Midlands of England (Fig. 2). The West Midlands has been through a period of post-industrial decline and is currently undergoing economic regeneration (Green & Berkeley, 2006). It differs greatly from the Wider South East of England, which has a services and knowledge-based economy centred on London. Comparing the findings for the West Midlands with those in Hargreaves (2015) can therefore assess whether the method is transferable between English regions. Three of these districts, Sandwell, Solihull, and Walsall, are metropolitan boroughs on the periphery of the city of Birmingham, whereas Coventry is a smaller nearby freestanding city. Their populations are around 328k, 216k, 285k and 371k respectively (ONS, 2020). Most areas of Solihull are affluent suburbs whereas the other districts have areas of multiple deprivation (MHCLG, 2019). Hence, these four local authority districts differ historically, spatially, and socio-economically.

2.4. Data inputs for calibrating and validating the model

This section describes how GIS data has been extracted for calibrating and validating the model. This process identifies the dwelling plots and their attributes of Plot-area, dwelling type, age band, and area type because Hargreaves (2015) found that these variables affected the shape parameter *k*.

2.4.1. Address point and dwelling type data

The *'Address-point'* data provides the spatial location of dwellings. Ordnance Survey provided year 2016 'AddressBase Plus®' GIS data and 2016 Mastermap® for academic research. Almost all the Address-point data for Coventry, Solihull, and Walsall, included the '*Dwelling-type*' ('*Detached*', '*Semidetached*', '*Terraced*', or self-contained flat). The self-contained flats include maisonettes and apartments and are henceforth referred to as '*Apartments*.' These three districts are therefore used for calibrating *k* and are henceforth referred to as the '*Calibration-*

districts.' The Sandwell Address-points however did not include Dwelling-type data and so this District has been used for validation purposes only and is referred to as the '*Validation-district*.'

2.4.2. Cadastral land data

The Plot-areas are obtained from the cadastral land data, which is an open-source dataset (HM Land Registry, 2014) developed to comply with the EU INSPIRE Directive. It contains the position and indicative extent of freehold registered property, mapped as polygons. These polygon areas are therefore used as the Plot-areas. Housing still owned by the public sector is not included, which is about 20% of the housing stock.

2.4.3. Historic maps

The age of dwellings may affect their Plot-density distribution and so they have been classified by '*Age-band'*. For example, many of the older houses would have been built originally in towns and villages that have since been engulfed by the expanding conurbation and, although some residential areas may have originally been built to a masterplan, they change over time due to subdivision, extensions, and conversions (Whitehand et al., 1999). The classification by 'Age-band' has used maps from Historic Digimap® (see Supplementary Material, Section A2.2 and Fig. A1). The Age-band thresholds (Table 1) are chosen depending on when substantial revisions occurred to maps. These took place at different times in different Districts depending on how much change had taken place over intervening years. Therefore, the range of years per Age-band differs slightly between Districts. Each Age-band includes approximately a quarter of the dwellings of the District.

Table 1

The Age-band classification for each Calibration-district.

Age-band	Coventry	Solihull	Walsall
#4	1992-2016	1981-2016	1992-2016
#3	1960-1991	1955-1980	1963-1991
#2	1938-1959	1938-1954	1938-1962
#1	Built before 1938	Built before 1938	Built before 1938

2.4.4. Output Area type classifications

The type of area where dwellings are located may affect their Plot-density distribution. The dwellings are therefore classified by Output Area type. Output Areas are the smallest areas used by UK Office for National Statistics (ONS) for the aggregation and classification of data (they vary in size but on average have around 125 households). There are seven *'Output-Area types'*, and these correspond to the seven Supergroups of the year 2001 ONS classification. This clustered Output Areas into these Supergroups based on their built-form and socio-economic characteristics (see the Supplementary Material Section A1 for further details). Fig. 2 shows the Output Areas as choropleth mapping classified by their Output-Area type, and Fig. 3 shows aerial photos of examples of the Output-Area types, listed in approximate order of decreasing urbanisation (the Countryside type has too few dwellings to analyse and is not shown).

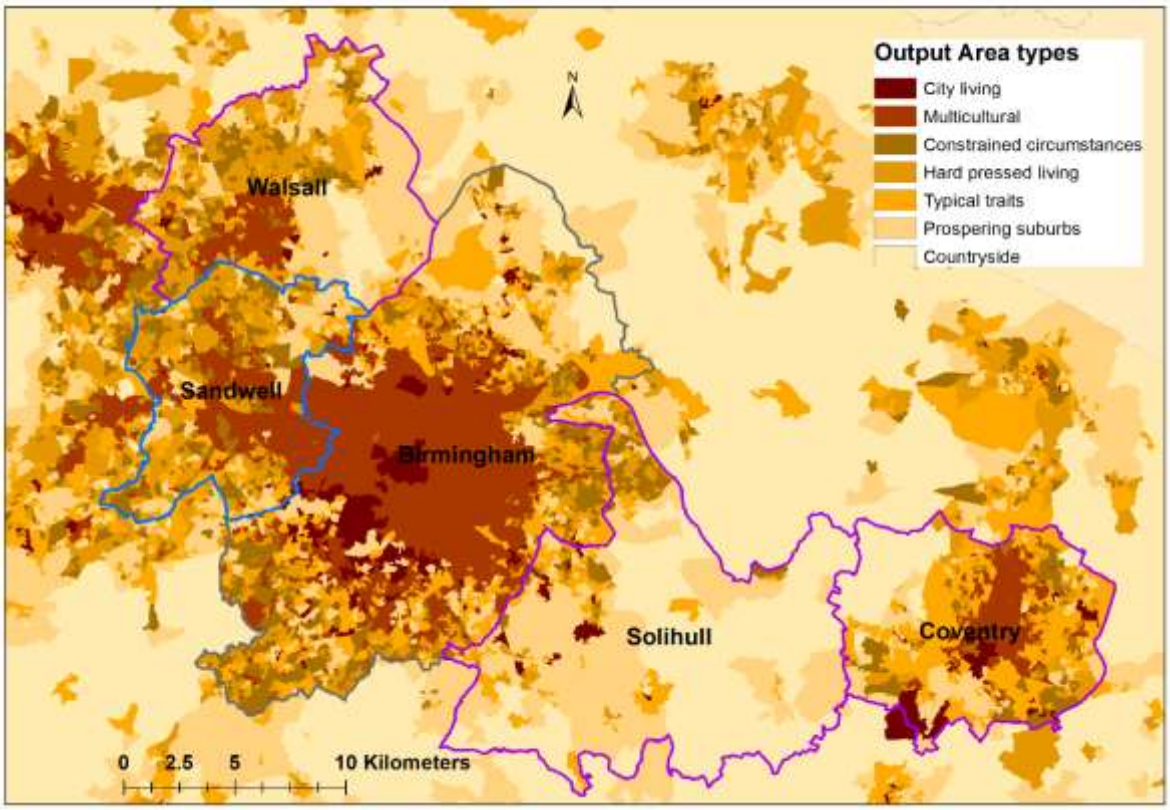

Fig. 2. The boundaries of the Calibration-districts (Coventry, Solihull, and Walsall) and Validation-district (Sandwell) and choropleth mapping of the Output Area classification Source: ONS, (2001).

2.4.5. Extracting the plot data from GIS maps

ArcGIS® tools have been used to spatially join the Address-points to the cadastral polygons and their Output-Area type, and to process and analyse the historic maps. If the polygons contain Address-points of either only a house, or only purpose-built Apartments the data is extracted for analysis. Details of the data extraction method can be found in the Supplementary Material; Section A2 for the calibration, and Section A3 for the validation. The attributes per plot for the calibration and validation of the model are indicated in Table 2 by (√). This shows that these attributes per plot for the validation do not include the Age-band or Dwelling-type as indicated by (x). The data extraction process results in a total sample size of 229,769 dwellings for the Calibration-districts, and of 89,284 dwellings for the Validation-district, which represent 67% and 69% of the total dwellings, respectively.

Table 2
Summary of the attributes per plot estimated from GIS data.

Districts	Attributes					
	Plot-area	Dwellings per plot	Output-Area type	Dwelling-type	Age-band	Footprint-area
Calibration-districts	√	√	√	√	√	x
Validation-district	√	√	√	x	x	√

Fig. 3. Examples of the Output-Area types.
Source: Aerial Digimap® year 2019.

Analysing the Address-points of the Apartments finds that there is no relationship between the number of Apartments per plot and the density of the plot and so the calibration has been carried out per plot i rather than per dwelling. This overcomes the obvious problem that using the Plot-density per Apartment would have produced a very 'lumpy' distribution at such a local scale because there is a large local variability in the number of Apartments per plot.

2.5. Calibration of the model

The upper part of the Fig. 4 flow-chart shows how the shape parameter and the functional relationships between variables are estimated to calibrate the model, and this is described in the following sub-sections.

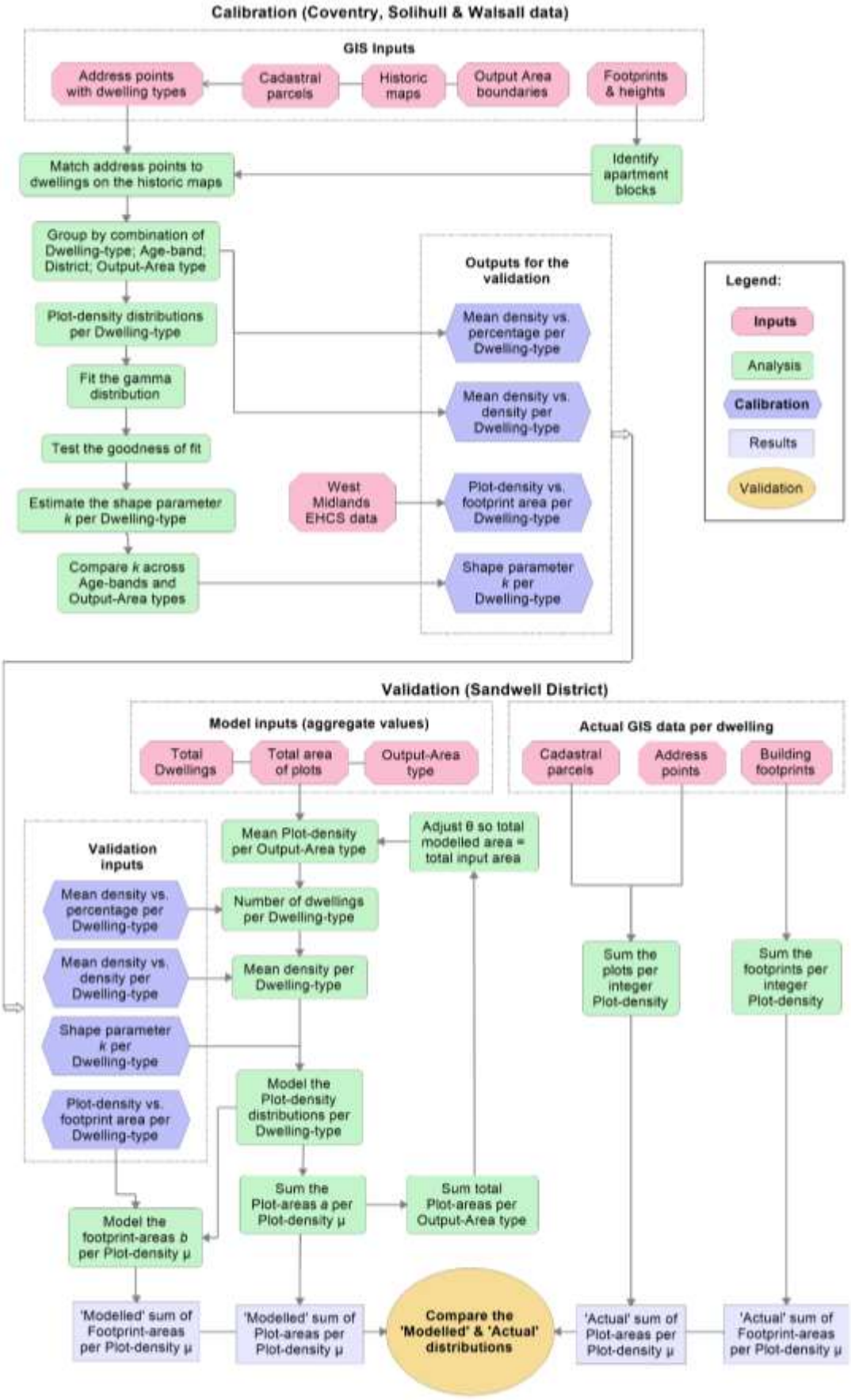

Fig. 4. Flow-chart of the calibration and validation processes.

2.5.1. Relationships between Plot-density, dwelling-type, and footprint area

The first step is to use the Mean plot-density $\bar{x}_j$ per zone (Eq. 2) to estimate the percentage of each Dwelling-type ρ_{dj} (see Fig. 5a), and the Mean plot-density per dwelling type $\bar{x}_{dj}$ (see Fig. 5b). This uses the Calibration-districts data because (Hargreaves, 2015) showed that these relationships differ between regions. This replicates the postprocessing of the forecasts of $\bar{x}_j$ by a spatial interaction model into the inputs needed for the Plot-density modelling.

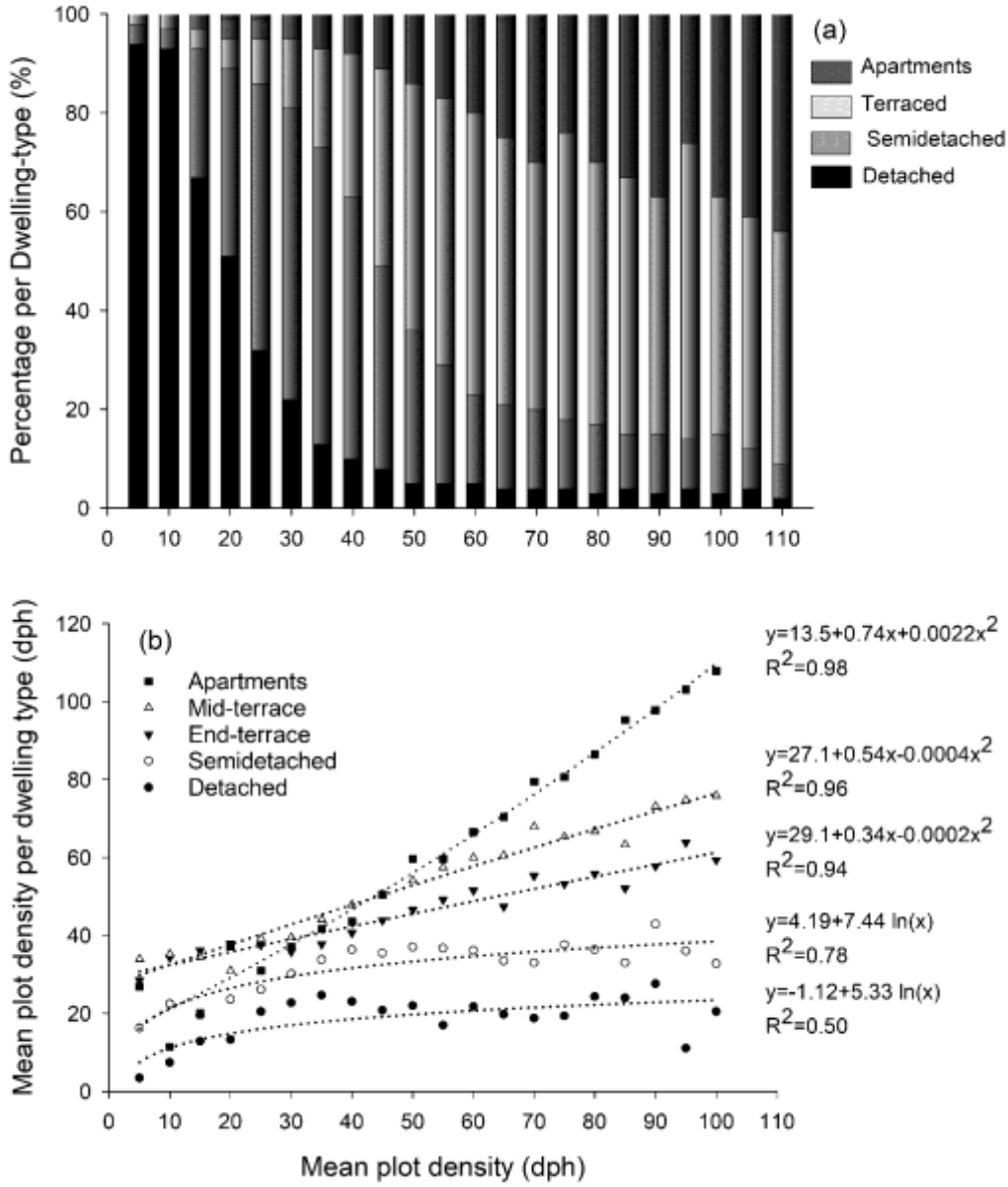

Fig. 5. The relationships between the Mean plot-density and (a) the percentage of dwellings per Dwelling-type, (b) the Mean plot-density per Dwelling-type.
Sources: (a) ONS 2001 Census aggregate data (Edition: 2005) and General Land Use Database (DCLG, 2007); (b) GIS data, of the Calibration-districts.

Similarly, EHCS data of West Midlands urban areas are used to estimate the relationship between Plot-density per dwelling μ_d and the building footprint area per dwelling b_d (Fig. 6). The Footprint-area reduces as Plot-density increases indicating a strong correlation between the Plot-area and house dimensions, which greatly enhances the usefulness of the model.

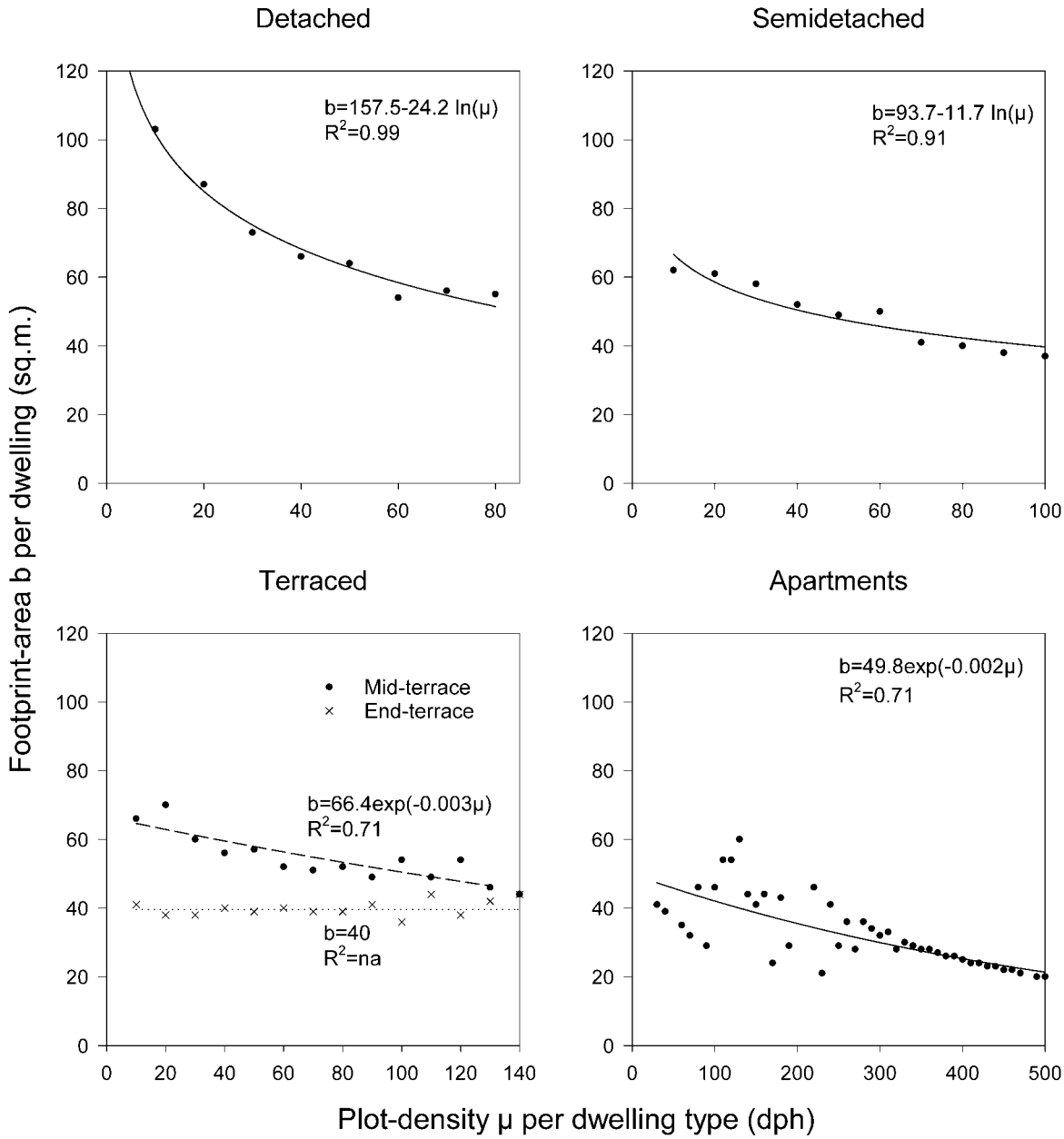

Fig. 6. The relationship between average Footprint-area and Plot-density per dwelling. Source: EHCS data of urban areas in the West Midlands region (DCLG, 2009).

2.5.2. Calibrating the shape parameter

The next step is to estimate the shape parameter k_d per Dwelling-type (Eq. 3). The sample of plots i for houses is partitioned into mutually exclusive datasets corresponding to each combination of Dwelling-type, District, Age-band, and Output-Area type. There are 5 x 3 x 4 x 7 variables for houses (Table 3) which results in 420 datasets. There are relatively few Apartment blocks, so this data has only been partitioned by Output-Area type and by only two Age-bands (#1 to #3; and #4) resulting in 7 x 2 = 14 datasets for Apartments. The gamma distribution is fitted to each dataset that has a sample size of plots i greater than 26. A theoretical Plot-density probability distribution is fitted to each dataset and the results for k_d are then compared for consistency to assess if it can be calibrated for modelling purposes. Note that for countries with insufficient GIS data for calibrating k_d, it may be possible to extract the necessary data from satellite imagery, mapping, and LIDAR data using image processing and machine learning (Hecht et al., 2019).

Table 3
Variables for partitioning the dwellings into datasets for calibrating the model.

Dwelling-types	Districts	Age-bands	Output-Area types
Detached	Coventry	#4. 1990's-2016	City Living
Semidetached	Solihull	#3. 1960's~1990's	Multicultural
End-terrace	Walsall	#2. 1938~1960's	Constrained Circumstances
Mid-terrace		#1. Before-1938	Hard Pressed Living
Apartments			Typical Traits
			Prospering Suburbs
			Countryside

2.6. The validation method

It is postulated that the calibrated model can be applied to similar Districts and its outputs will match the empirical Plot-area and Footprint-area data. This is tested using the Validation-district data so that the validation is independent of the calibration. The method is described below and illustrated in the lower part of Fig. 4.

The model is designed to be nested within an integrated modelling framework so that it could complete a chain of urban modelling from regional scale to building scale (Fig. 7). The validation therefore only uses the inputs that would be available as outputs per zone of a spatial interaction model. These would be the total amount of land available for the plots and the forecast number of dwellings. This provides a stringent test of the validity and robustness of the Plot-density model for forecasting built-form. The boundaries of the zones of a spatial interaction model would be chosen to aggregate areas of a similar type, and so the Output-Area types j are used as a proxy for these zones. The validity of the model is tested for the following three ways of estimating the shape parameter k as an average value (weighted by the number of dwellings per dataset) of either; all the datasets ($\ddot{k}$); the datasets per Dwelling-type ($\ddot{k}_d$); or the datasets that have the same combination of Dwelling-type and Output-Area type ($\ddot{k}_{dj}$). The following method of testing how well the model fits the empirical data is repeated for each of these ways of estimating *k*. The results are then compared to assess the robustness of the model. The inputs to the Plot-density model for the validation therefore consist of:

a) Total land-take of dwelling plots A_j
b) Total number of dwellings H_j
c) Shape parameter calibrated as either $\ddot{k}$, $\ddot{k}_d$, or $\ddot{k}_{dj}$
d) Calibrated relationships between $\bar{x}$ and ρ_d
e) Calibrated relationships between $\bar{x}$ and $\bar{x}_d$

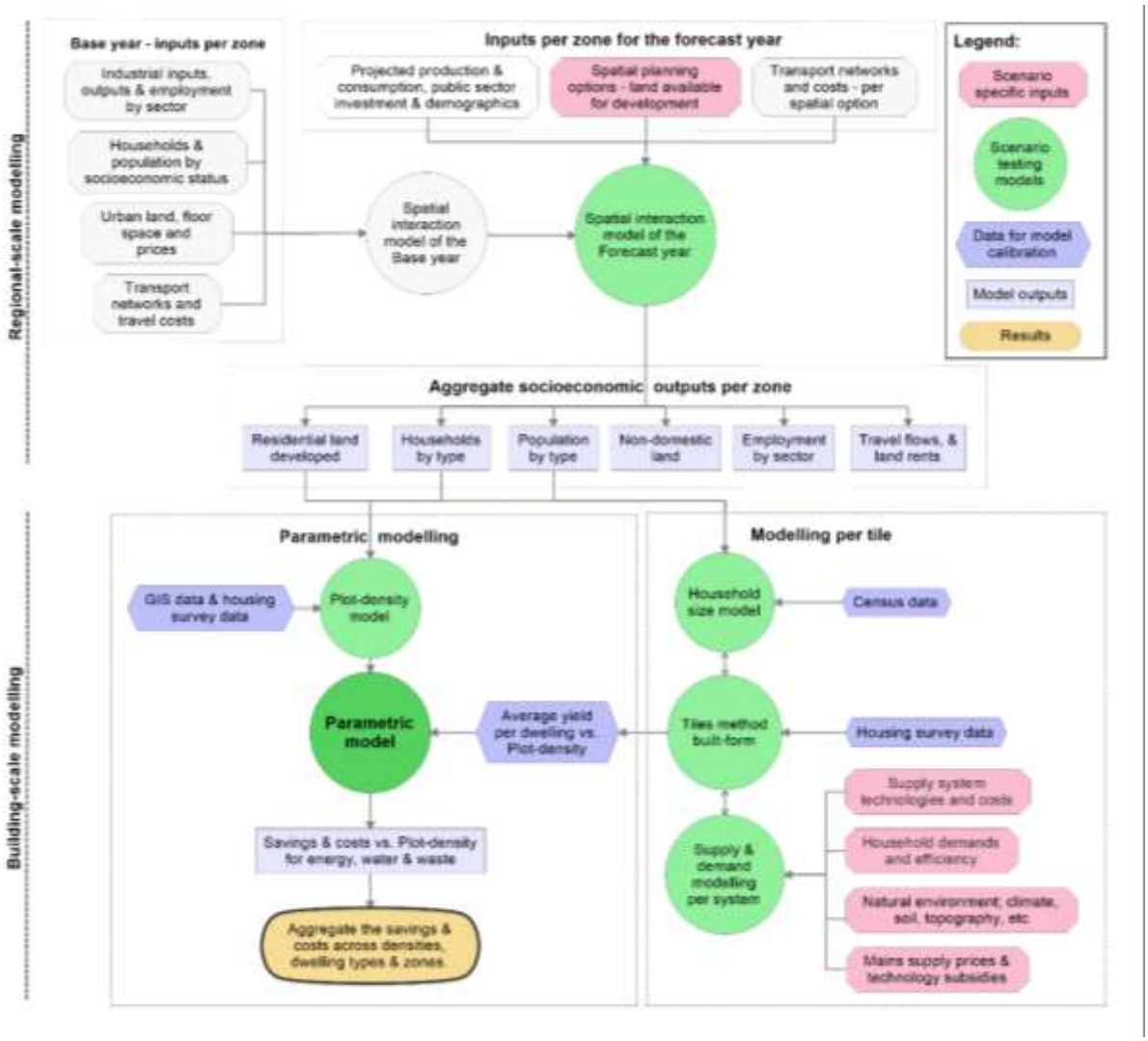

Fig. 7. Flow-chart of how the model enables the scenario testing of building-scale sustainable technologies within a regional-scale modelling framework.

The Mean plot-density $\bar{x}_j$ (Eq. 2) is used to estimate ρ_{dj} and $\bar{x}_{dj}$ (Section 2.5.1 and Fig. 5). Then the number of dwellings per Dwelling-type is estimated as, $h_{dj} = H_j \cdot p_{dj}/100$; and the scale parameter $\grave{\theta}_{dj}$ is estimated from $\bar{x}_{dj}$ using Eq. (4). The Plot-density distribution is modelled as, $\grave{h}_{dj\mu} = h_{dj} \cdot f(\mu; \ddot{k}, \grave{\theta}_{dj})$, where $f(\mu; \ddot{k}, \grave{\theta}_{dj})$ is the gamma distribution (Eq. 3). The modelled Plot-areas are then calculated from Eq. (1) as, $\grave{a}_{dj\mu} = \grave{h}_{dj\mu}/\mu$ and their total land-take is, $\grave{A}_j = \sum_{d=1}^{d=5} \sum_{\mu=0}^{\mu=\infty} \grave{a}_{dj\mu}$. The scale parameter is proportionally adjusted by using a factor σ_j so that the modelled $\grave{A}_j$ equals the input A_j and so, $\hat{\theta}_{dj} = \bar{x}_{dj} \cdot \sigma_j/(\ddot{k} - 1)$. This increases the robustness of the method by fitting the shape of this distribution to the total land-take of the dwelling plots A_j. The resulting novel *'Plot-density model'* is,

$$\hat{h}_{dj\mu} = h_{dj} \cdot f(\mu; \ddot{k}, \hat{\theta}_{dj}) \tag{5}$$

This is used to calculate $\hat{a}_{dj\mu}$ these are summed for the Dwelling-types and Output-Area types to give the modelled *'Plot-area distribution'* of the District $\hat{a}_\mu$. This then is compared with the empirical Plot-area distribution, $A_\mu = \sum_{i=1}^{i=n} a_{i\mu}$ per integer Plot-density μ. The Matlab® programme in the Supplementary Material Section C has been written to calculate 'ksstat' and the p-value to test the goodness of fit.

Similarly, the modelled *'Footprint-area distribution'* is, $\hat{b}_\mu = \sum_{j=1}^{j=7} \sum_{d=1}^{d=5} b_{d\mu} \cdot \hat{h}_{dj\mu}$ where $b_{d\mu}$ is calculated from the regression functions in Fig. 6. This is compared with the empirical Footprint-area distribution, $B_\mu = \sum_{i=1}^{i=n} B_{i\mu}$, The goodness of fit of the modelled distribution is then tested in the same way as for the Plot-areas above.

2.7. Modelling the future potential of building-scale sustainable technologies

As explained earlier, there would be advantages in creating a parametric model that combines the Plot density model with the tiles-method for modelling sustainable technologies. The following novel parametric model therefore uses the tiles-method to calibrate functional relationships between Plot-density and sustainable supply and then combines these functions with the Plot-density model (see Fig. 7). This is demonstrated below using rainwater harvesting (RWH) systems as an example. A previous regional-scale study (Hargreaves et al., 2019) modelled how spatial planning options would affect the future potential of alternative water supply, such as either RWH, or greywater recycling. The inputs to the RWH model are the built-form dimensions per tile and location-specific data of; rainfall, system cost, household-size, water demands, mains water price, and policy support, such as subsidies and regulations. It models the temporal demand and supply of rainwater for toilet flushing and the storage tank capacity of the RWH system. If a system would be cost effective, the RWH model outputs the yield, which is the amount of rainwater that is used by the household and results in a water saving. The water savings per dwelling (in litres/dwelling/day) are calculated as an average for all dwellings in the area and so this partly depends on the estimated uptake of RWH systems. These '*average water-savings*' per dwelling are plotted against Plot-density in Fig. 8. The symbols represent each tile type and a regression function $w_{d\mu} = f(\mu_d)$ is fitted to these values.

The tile types represent built-form of Dwelling-types at different Plot-densities and so are an important intermediate step of calibrating how the average water-savings would vary with Plot-density per Dwelling-type. The water-savings function w_d declines as Plot-density μ_d increases, mainly because higher-density dwellings generally have less roof area per dwelling for harvesting rainwater, and partly because they have less floorspace and so smaller households and less water demand. If there is less water-saving it is less cost effective to install a RWH system (Campisano & Modica, 2012). Hence, fewer dwellings are likely to have RWH and so the shape of the function declines more steeply as density increases. Conversely, it rises less steeply for low density dwellings because these tend to have a larger roof area and so rainwater supply is more likely to exceed demand. These functions differ between Dwelling-types due to differences in their built form. For example, the relationship between their Plot-area and building footprint-area (Fig. 6), and how much of the footprint-area is covered by an upper-roof suitable for rainwater collection. The Apartments have greater water-savings than houses at higher densities because they are more suitable for communal systems, which can be more cost effective than individual household systems. These non-linear functions are combined with Eq. (5) to create the following novel *'Parametric Model'* of district-scale average water-savings,

$$\widehat{W}_{dj\mu} = w_{d\mu} \cdot \hat{h}_{dj\mu} \tag{6}$$

The results of this novel Parametric Model are compared with those of the '*Discrete Method*' that uses an average dwelling per Dwelling-type. These average dwellings are designed to have the average floorspace per Dwelling-type, $\bar{f}_d = \sum_{i=1}^{i=n} f_{di} / \sum_{i=1}^{i=n} h_{di}$ estimated from the year 2009 EHCS data of urban areas in the West Midlands region (DCLG, 2009). These data are then used to estimate the Plot-area of each Dwelling-type of average floorspace. Hence, their corresponding discrete Plot-density $\ddot{x}_d$, is very similar to $\bar{x}_d$. The values of $\ddot{x}_d$ correspond to the vertical dotted lines on Fig. 8. For example, $\ddot{x}_d = 23\ dph$ for Detached houses and the corresponding $w_d = 65$ l/d/d.

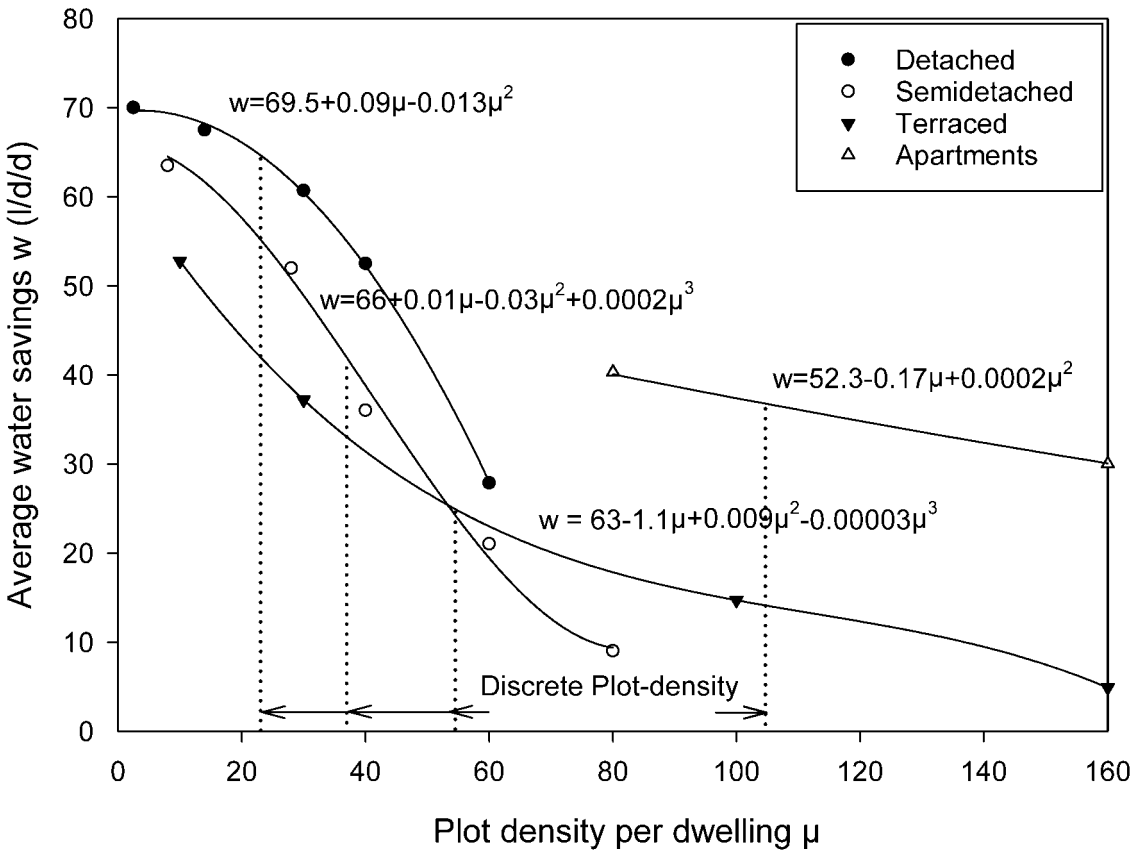

Fig. 8. Relationship per Dwelling-type between the average water-savings of rainwater harvesting and Plot-density (this depends on the study area, and the modelling, policy, and retrofitting assumptions, and is for illustrative purposes only).
Source: Outputs of the model used by Hargreaves et al., (2019).

3. Results

3.1. *The goodness of fit of the theoretical distribution to the GIS data*

The K-S test results are summarised in Fig. 9a, which shows the percentage of positive results for each combination of Dwelling-type and Age-band (a positive result means that the null hypothesis cannot be rejected at the 1% probability level). Details of the results can be found in the Supplementary Material Section B3.

In most cases, the null hypothesis cannot be rejected as shown by the high percentage of positive results. However, the main exceptions are Detached housing built in Age-band #4, and Mid-terraced housing built in Age-bands #1 to #3. Their empirical Plot-density distributions are concentrated within a narrower range than a gamma distribution. Detached houses are the most desirable dwelling type for home buyers, but local authorities have introduced targets in recent decades for minimum average densities of around 30 dph to reduce sprawl, and this may explain why their Plot-densities in Age-band #4 are mainly concentrated within a narrow range (20 to 40 dph). Many Terraced houses were built en-masse to regular plans as low-cost housing for industrial workers (e.g., Fig. 3b) during Age-bands #1 to #3, and this may explain why their Plot-densities are mainly concentrated within a narrow range (40 to 60 dph).

The shape parameter *k* tends to differ between the Output-Area types as shown in Fig. 9b. For example, the 'Prospering Suburbs' and 'Countryside' areas generally have smaller values of *k* than other areas, signifying a less equitable distribution of residential land. Conversely, the 'Hard Pressed Living' and 'Constrained Circumstances' areas generally have a larger *k* than other areas. They were mostly built by the public-sector to supply the housing needs of lower income families and the plot areas are more equitably distributed.

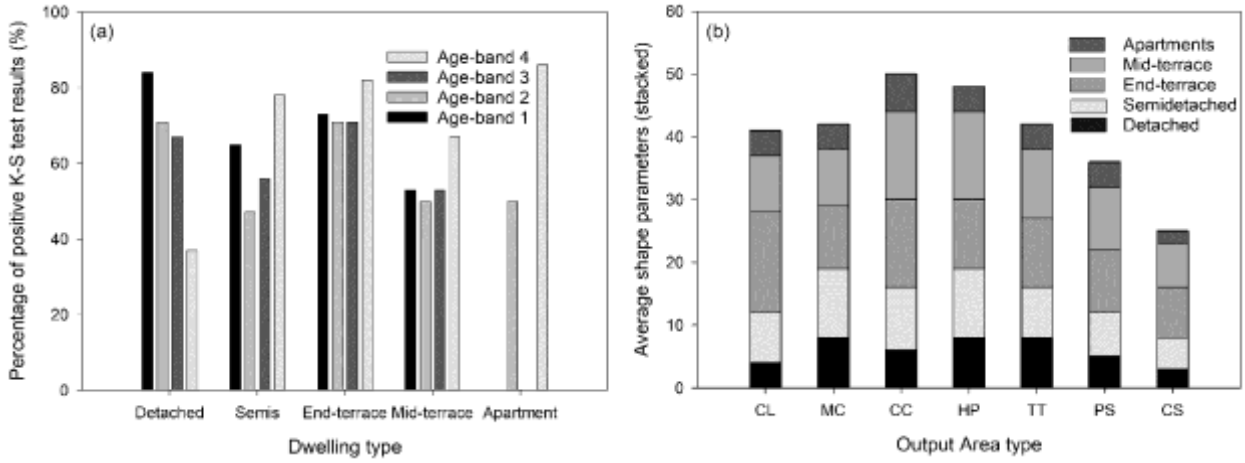

Fig. 9. (a) Percentage of positive results of the K-S goodness of fit test per combination of Dwelling-type and Age-band; (b) Average shape parameter per combination of Dwelling-type and Output-Area type (see Fig. 3).

The analysis found that the shape of the empirical Plot-density distribution usually becomes more like the gamma distribution if the sample of dwellings is partitioned by Dwelling-type, Output-Area type, and Age-band. A possible reason is that dwelling type and area type are two of the most salient factors in the search process of home buyers (Schirmer et al., 2014). In the UK, the cost of the plot is a substantial proportion of the house price. Developers may therefore intuitively sub-divide their land so that the plot areas available for a particular type of dwelling and local area type, broadly corresponds to the distribution of household income. This is positively skewed (Salem & Mount, 1974), as is the inverse-gamma distribution. However, it would need further research to confirm this hypothesis.

The other contender for modelling the plot densities was the lognormal distribution. Like the gamma distribution, it is positively skewed with support of zero to infinity. However, the analysis found that it does not fit the empirical data as well as the gamma distribution (with over 20% fewer positive K-S test results), and so it was not used any further.

3.2. The shape parameter

The shape parameter per Dwelling-type k_d is plotted on Fig. 10 per Age-band for each dataset that has a sample size $i > 200$ for houses and $i > 26$ for Apartment blocks (results for smaller datasets are excluded because their estimated shape will be less reliable). The values of k_d and their upward trend over time are like those in Hargreaves (2015) for all Dwelling-types except Mid-terrace houses. These have the same upward trend over time but range from 10 to 13 compared to a range of 7 to 11 in the previous study.

This close correspondence gives confidence that the method is valid, and that k_d is consistent and transferable between UK regions and spatial scales. It differs between Dwelling-types and is smaller (i.e., more positively skewed) for Apartments and Detached, probably because their Plot-density is less constrained by the layout of neighbouring plots.

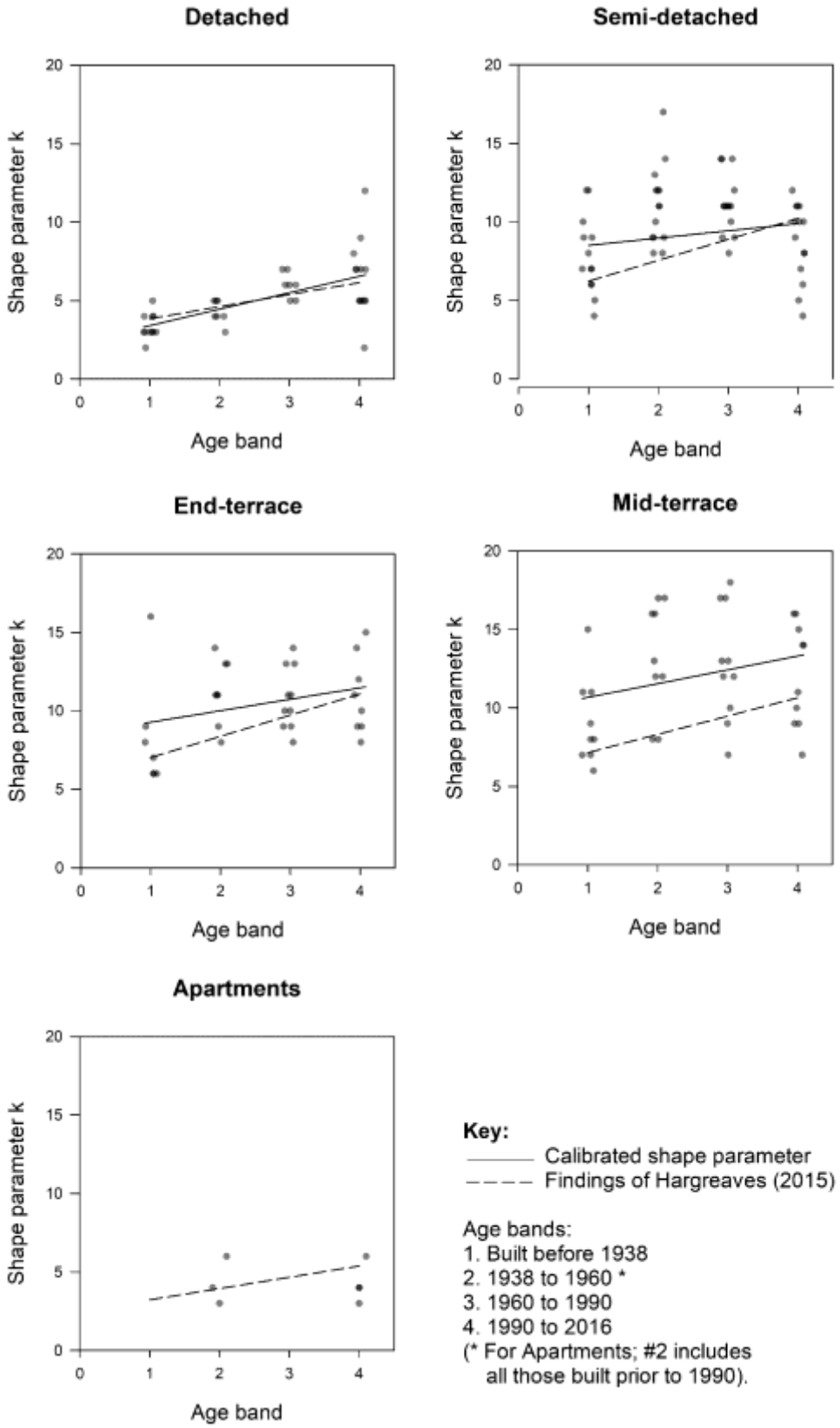

Fig. 10. The shape parameter of each Dwelling-type per Age-band, and comparison with the results of Hargreaves (2015).

3.3. Validation results

The validation process tests how well the modelled Plot-area distribution $\hat{a}_\mu$ fits the empirical Plot-area distribution A_μ, and how sensitive the goodness of fit is to different ways of calibrating the shape parameter $\ddot{k}$. Similarly, it also compares the modelled Footprint-area distribution $\hat{b}_\mu$ with the empirical Footprint-area distribution B_μ.

Fig. 11i shows the results for the Validation-district when using $\ddot{k}$, the average value for all Dwelling-types ($\ddot{k} = 9$). This shows that the apex for the modelled values is lower than, and misaligned with, the empirical values.

Fig. 11ii shows the results when using $\ddot{k}_d$, the average per Dwelling-type (which equals 5, 10, 10, 11 and 4 for Detached, Semidetached, End-terrace, Mid-terrace, and Apartments respectively). This improves the goodness of fit, compared to using $\ddot{k}$, as can be seen by comparing the test statistics between Fig. 11iia (ksstat = 0.09 and p-value = 0.51) and Fig. 11ia (ksstat = 0.11 and p-value = 0.27). The findings for Footprint-areas (Fig. 11b) are similar although, as expected, the fit is not quite as good.

A further sensitivity test was carried out using $\ddot{k}_{dj}$, the average per combination of Dwelling-type and Output-Area type, but this made no further improvement (ksstat = 0.09) and so is not shown.

It can be seen in Fig. 11iia that the empirical plot-areas slightly exceed the modelled areas in the μ range of around 30 to 40 dph. This is due to the 'Hard-Pressed Living' Output-Area type, which accounts for around 25% of the dwellings in this District and includes large areas that were built as social housing with very uniform plot layouts. If this Output-Area type is excluded, the empirical and modelled distributions match much more closely in this density range but the overall goodness of fit remains unchanged (ksstat = 0.09). The other difference is that the empirical areas below around 5 dph exceed the modelled values. This is due to atypical low-density dwellings, such as houses with adjoining buildings or non-residential land in the same ownership as the house, large historic houses, or dwellings with large plots because they are part of the grounds of a school or public house.

Hence, the Plot-density model provides good results, so long the shape parameter is estimated per Dwelling-type. These results are obtained using only total dwellings and total plot-area as inputs and so they validate use of the Plot-density model at District scale. The model has the following inherent features that contribute to its robustness; the gamma distribution is specified by $\bar{x}$, $\ddot{k}$, and θ, and these are functionally interrelated; the values of the scale parameter θ can therefore be proportionally adjusted so that the total modelled area matches the empirical total available area; and although the shape of the distribution varies markedly between Age-bands and Output-Area types, these perturbations tend to balance out when the distributions are combined together.

3.4. The potential water-savings of rainwater harvesting

The novel Parametric Model (Eq. 6) is demonstrated by using it to estimate the potential water-savings of RWH for the Validation-district. The results are shown on Fig. 12 as parametric distributions and the modelled dwellings have been normalised to unity so that the results can be more easily compared between Output-Area types. They are compared in Fig. 12i between the 'Prospering Suburbs' and 'City Living' area types, and for ease of presentation, the results for End-terraced and Mid-terraced have been combined. The thin lines are the probability per Dwelling-type calculated as, $\check{h}_{dj\mu} = \hat{h}_{dj\mu} / \sum_{d=1}^{d=4} \hat{h}_{dj\mu}$, and the thickest line shows their sum, which is the probability distribution per average dwelling $\check{h}_{j\mu}$ (plotted against the left-hand y-axis). The shape of each distribution is clearly correlated with the average density of the Output-Area type (30 dph for 'Prospering Suburbs', and 52 dph for 'City Living').

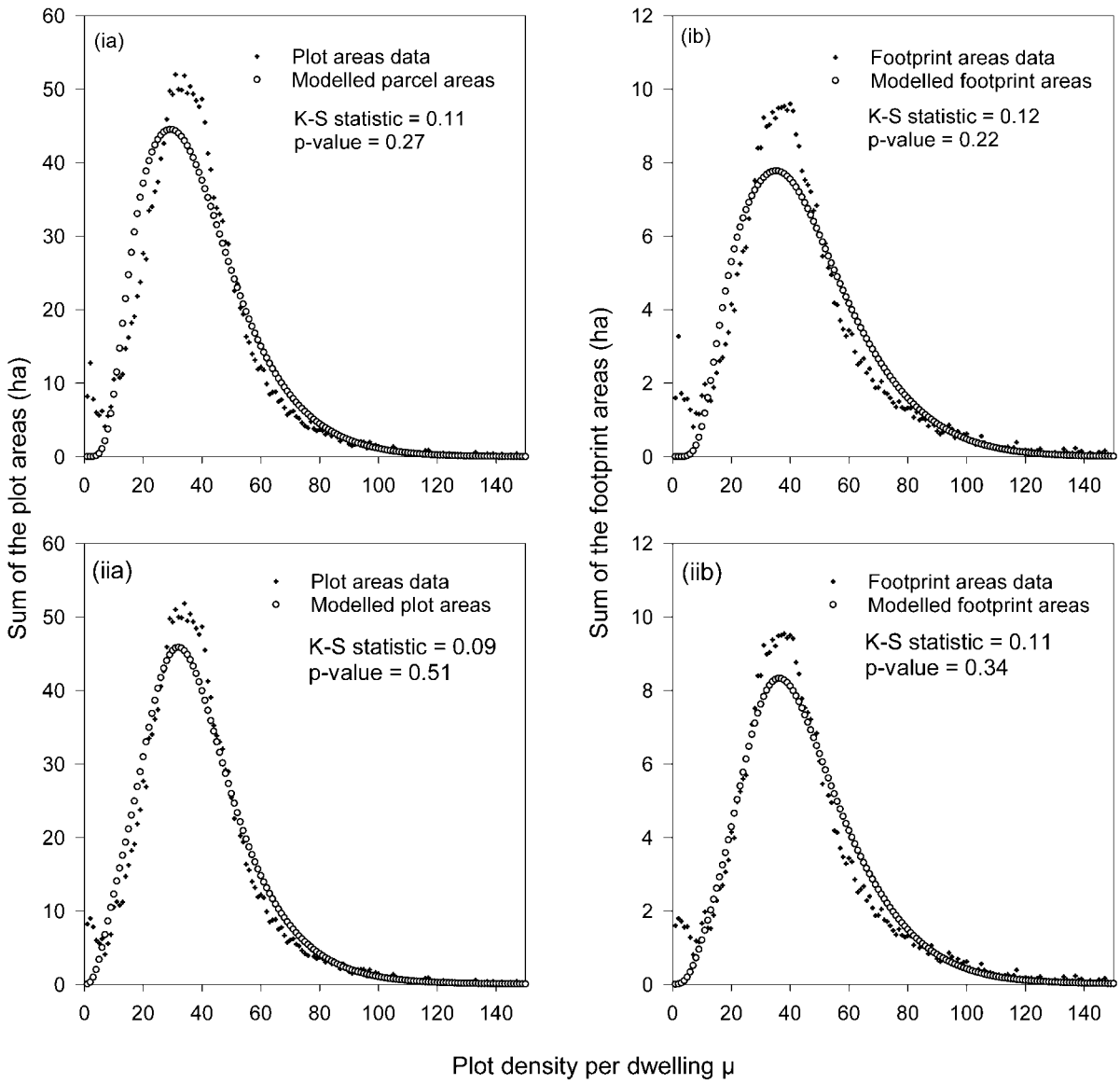

Fig. 11. Validation results using (i) an average shape parameter for all dwellings, (ii) an average shape parameter per Dwelling-type; for (a) sum of the Plot-areas, (b) sum of the Footprint-areas.

Fig. 12ii shows the parametrically modelled water-savings that correspond to these dwelling probability distributions. The thin lines show the water savings contributed per Dwelling-type, $\breve{W}_{dj\mu} = w_{d\mu} \cdot \breve{h}_{dj\mu}$, where $w_{d\mu}$ is the function of average water savings (Fig. 8). The thickest line shows their sum which is the Plot-density distribution of water savings per average dwelling $\breve{W}_{j\mu}$ (plotted against the left-hand y-axis). The average water savings per average dwelling, $\breve{W}_j = \sum_{\mu=0}^{\mu=\infty} \breve{W}_{j\mu}$.

These results are compared with those of the Discrete Method, which are shown as bars on these charts. It uses the same calibration of $\bar{x}_j$ versus p_{dj} as the Parametric Model so that both methods are directly comparable. The dwellings are again normalised to sum to unity and so, $\breve{h}_{dj} = p_{dj}/100$, (Fig. 5a). The Discrete Method uses the same average density per Dwelling-type $\ddot{x}_d$ for the whole study area (as can be seen by comparing Fig. 12ia and Fig. 12ib). The water-savings contributed per Dwelling-type, $\breve{W}_{dj} = w_d \cdot \breve{h}_{dj}$, are shown as bars on Fig. 12ii, where w_d is the function of average water savings for Plot-density $\ddot{x}_d$. The average water savings per dwelling, $\breve{W}_j = \sum_{d=1}^{d=4} \breve{W}_{dj}$.

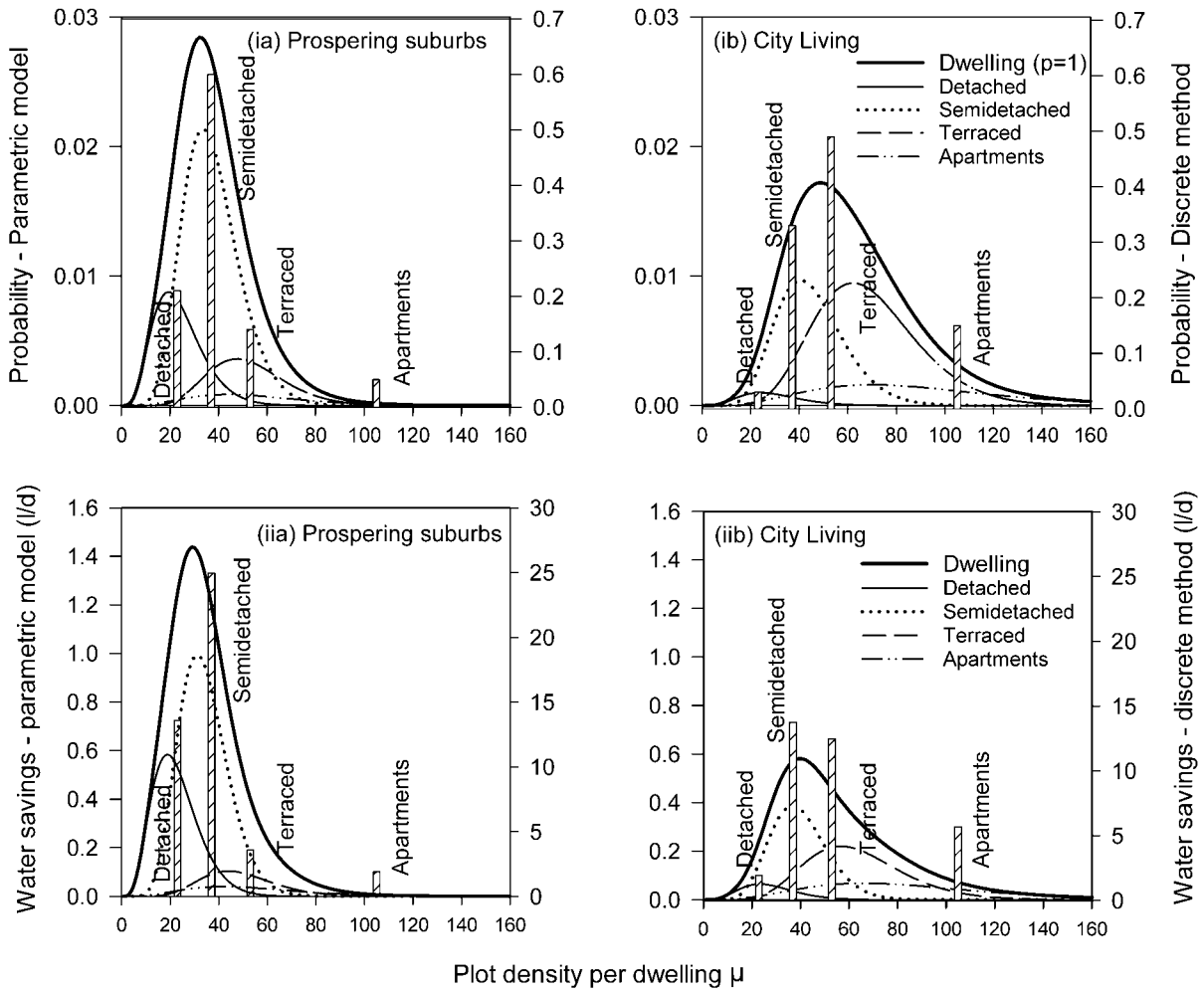

Fig. 12. Comparison of the Parametric Model and Discrete Method between; (i) Plot-density probability distributions per Dwelling-type normalised to sum to unity, (ii) average water-savings corresponding to the above Plot-density distributions; for (a) a lower-density Output-Area type, (b) a higher-density Output-Area type.

Comparison of the results shows that the Parametric Model provides a better understanding than the Discrete Method of which dwellings would contribute the most water savings, and how this would differ between the Output-Area types. For example, most of the water savings for 'Prospering suburbs' would be attributable to Semidetached and Detached houses with a Plot-density of less than around 55 dph (Fig. 12iia). These insights could help planners, urbanists, and engineers to coordinate planning policies and design guidance to enhance the potential water savings. For example, by planning future suburban areas so that these dwelling types within this Plot-density range are considered for RWH. Conversely these insights could help to identify areas, such as 'City-living', that would be less suitable for RWH. Alternative water saving systems, such as greywater recycling, could then be considered, along with supportive measures such as the clustering of dwellings so that these areas are more suitable for communal systems.

Fig. 13 compares the Parametric Model and the Discrete Method on the average water-savings $\breve{W}_j$ per average dwelling for each Output-Area type. This also shows the proportion attributable to each Dwelling-type. Both methods produce similar estimates of average water-savings per dwelling for the lowest density Output-Area type (Prospering Suburbs). However, the Discrete Method overestimates the water-savings for the other Output-Area types. This overestimation increases with density and is 15% for the highest density Output-Area type (City Living), and averages 10% for the Validation-district.

The reasons why the Discrete Method overestimates the water savings are that; (a) it has the major shortcoming that, unlike the Parametric Model, it does not represent how Plot-density per

Dwelling-type is correlated with the average density of the area; (b) the functions of water-savings decline more steeply as density increases (Fig. 8) and, unlike the Parametric Model, it is unable to model these non-linear relationships; (c) the modelled Plot-density distribution per Dwelling-type is positively skewed, and so $\ddot{x}_d$ does not represent a typical dwelling. The mean, $\bar{\mu}_d$ of the parametric distribution would be a more appropriate metric (Section 2.1). For example, $\bar{\mu}_d = 52\ dph$ for 'Terraced' houses in Prospering Suburbs, but in City 'Living' areas, $\bar{\mu}_d = 69\ dph$. This is substantially larger than the constant $\ddot{x}_d = 53\ dph$ used by the Discrete Method for Terraced houses. Hence, the Discrete Method overestimates the water-savings for all Output-Area types except the one of lowest density.

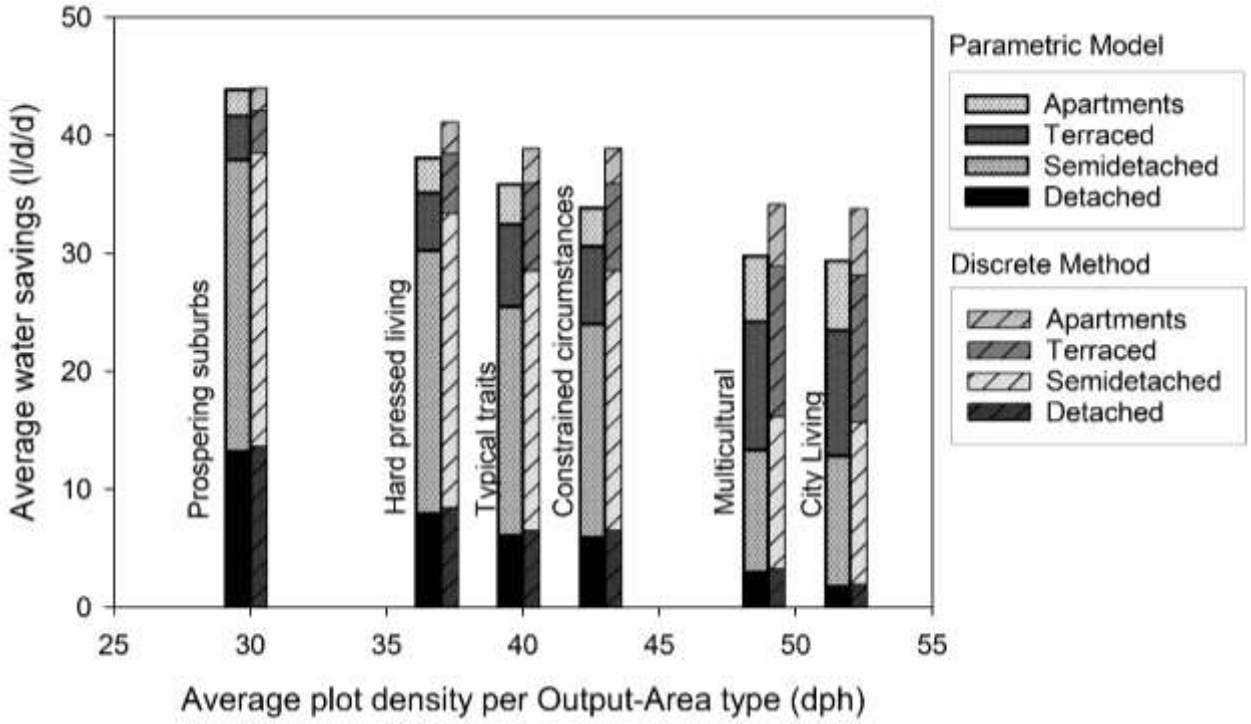

Fig. 13. Comparison between the Parametric Model and the Discrete Method on their estimates of water-savings per Output-Area type, normalised to an average dwelling and showing the proportion attributable to each Dwelling-type.

4. Discussion

The Parametric Model provides a better understanding than the Discrete Method of future residential built form and can produce more accurate estimates of the likely water savings of rainwater harvesting. Its advantages also apply to the other sustainable technologies because the space available for their installation, the area for harvesting renewable supply, and the demands related to floorspace, are crucial factors affecting their feasibility. Once calibrated and documented for a study area, this type of model could enable policy makers, utility companies, and practitioners to test scenarios without needing sector-specific expertise in modelling sustainable technology systems. Those scenarios that are the most promising could then be analysed by experts in more detail. The model could be useful for formulating local planning and urban design policies. Its outputs could also be useful as inputs to the simulation modelling of the spatial layout of urban areas.

Further research would be needed to test if this modelling method is valid for other countries. It is more likely to be applicable to countries where housing development is demand-led, with planning constraints on the availability of land, so that the distribution of plot area per dwelling is related to the willingness to pay of households. Otherwise, if the land has low commercial value, or is supplied and regulated by the public sector, the plot area per dwelling type is likely to be more narrowly and symmetrically distributed. The differences in the shape of the Plot-density

distribution between Output-Area types and Age-bands can plausibly be explained by changes in the governance of residential development and transport accessibility over time. Further research to confirm these insights would be useful to inform the calibration of the model for the scenario testing of future urban development.

5. Conclusions

There is a high level of consistency between the findings of this study and those of Hargreaves (2015), even though these two studies used different type of data, spatial scale, and study area. This gives confidence that the shape parameter per Dwelling-type is transferable between different UK regions, and spatial scales of modelling.

There is a good fit between the modelled Plot-area distribution and the empirical Plot-area distribution (p-value=0.51). This confirms that the method is valid for modelling at District-scale. Similarly, the modelled Footprint-area distribution closely matches the empirical Footprint-area distribution (p-value = 0.34), which gives confidence that the method can model the variability in dwelling dimensions. This has been achieved using only an average shape parameter per Dwelling-type, and the total number and average density of dwellings as the inputs. This demonstrates this robustness of this Plot-density model.

The Plot-density model can be combined with functional relationships between built form and building-scale supply to create a novel Parametric Model, as demonstrated using rainwater harvesting as an example. The findings show that the Parametric Model can more accurately estimate the water-savings than the usual method of using discrete average dimensions per Dwelling-type. The Discrete Method would overestimate the water-savings by up to 15% for higher density areas, and by an average of 10% for the Validation-district as a whole, and so would be unsuitable for option appraisal.

The findings give confidence that the Parametric Model could be embedded within a top-down strategic modelling framework to forecast how spatial planning scenarios would affect the future potential of building-scale sustainable technologies. This could help urban planners, policy makers, and practitioners to better understand how spatial planning policies would affect the future potential of sustainable technologies. This could enable constructive engagement between utility companies and planning authorities on coordinating spatial planning, building regulations and policy support to improve the future sustainability and resilience of urban areas.

Acknowledgments

The author gratefully acknowledges the financial support of the Engineering and Physical Sciences Research Council, UK under the ‘Liveable Cities’ programme grant [EP/J017698]; and the ‘Self Repairing Cities’ research grant [EP/N010523/1]. The author is also grateful to Ordnance Survey for kindly providing the AddressBase Plus® and Mastermap® data for academic use. The historic maps were available from Digimap® for academic use (© Crown copyright and Landmark Information Group Limited, 2019) and the cadastral polygon data from the UK Land Registry (© Crown copyright and database rights 2019 Ordnance Survey 100026316).

Nomenclature

A = Total input area of the dwelling plots
a = Plot-area
a_μ = Plot-area distribution
B = Empirical footprint areas
b = Modelled footprint areas
d = Dwelling-type
h = Number of dwellings
h_μ = Plot-density distribution
i = Plot
j = Output-Area type
k = Shape parameter of the gamma distribution
$\ddot{k}$ = Calibrated shape parameter
ksstat = K-S test statistic
μ = Plot-density
$\bar{\mu}$ = Mean of the gamma distribution
ρ_d = Percentage per Dwelling-type
θ = Scale parameter of the gamma distribution
$\bar{x}$ = Mean Plot-density
$\ddot{x}_d$ = Discrete Plot-density per Dwelling-type
w = Water-savings

Accents:

$\grave{}$ = Parametrically modelled
$\hat{}$ = Parametrically modelled and adjusted
$\check{}$ = Normalised to unity

Acronyms:

EHCS = English House Condition Survey
ONS = UK Office for National Statistics
RWH = Rainwater harvesting

References

Abraham, J.E., Weidner T., Gliebe J., Willison C., Hunt J.D. (2005). Three Methods for Synthesizing Base-Year Built form for Integrated Land Use-Transport Models. Transportation Research Record: Journal of the Transportation Research Board 1902, (2005), 114-123.

Bach, P.M., Deletic, A., Urich, C., McCarthy, D.T. (2018). Modelling characteristics of the urban form to support water systems planning. Environmental Modelling & Software, 104, 249-269.

Blum, P., Campillo, G., Münch, W., Kölbel, T. (2010). CO2 savings of ground source heat pump systems a regional analysis. Renewable Energy, 35(1), 122-127.

Campisano, A., Butler, D., Ward, S., Burns, M.J., Friedler, E., De Busk, K., Lloyd, N., Fisher-Jeffes, L.N., Ghisi, E., Rahman, A., Furumai, H., Han, M., (2017). Urban rainwater harvesting systems: research, implementation, and future perspectives. Water Research, 115, 195-209.

Campisano, A., Modica, C. (2012). Optimal sizing of storage tanks for domestic rainwater harvesting in Sicily. Resources, Conservation and Recycling, 63, 9-16.

Chen, H.C., Han, Q., de Vries, B., (2020). Urban morphology indicator analysis for urban energy modeling. Sustainable Cities and Society, 52, p.101863.

Cheng V., Steemers, K. (2011). Modelling domestic energy consumption at district scale: A tool to support national and local energy policies. Environmental Modelling & Software, 26, 10, 1186-1198.

D'Amico, B., Pomponi, F. (2019). A compactness measure of sustainable building forms. Royal Society Open Science, 6 (181265).
DCLG. (2012). The National Planning Policy Framework. Department for Communities and Local Government, HM Government, UK.
DCLG (2009). English house condition survey 2007: Annual Report. Department of Communities and Local Government Publications, HM Government, UK. http://doc.ukdataservice.ac.uk/doc/6449/mrdoc/pdf/6449ehcs_annual_report_2007.pdf
DCLG. (2005). Planning policy statement 1; delivering sustainable development. Department for Communities and Local Government, HM Government, UK.
DCLG (2007). Generalised land use database statistics for England 2005; Feb. 2007, Department for Communities and Local Government, HM Government, UK. http://webarchive.nationalarchives.gov.uk/20120919132719/http:/communities.gov.uk/documents/planningandbuilding/pdf/154941.pdf
DECC (2012). Renewable Heat Incentive: Consultation on proposals for a domestic scheme. London, UK: Department of Energy & Climate Change, London, UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/236015/8446.pdf
Domènech, L. Saurí, D. (2011). A comparative appraisal of the use of rainwater harvesting in single and multi-family buildings of the Metropolitan Area of Barcelona (Spain): social experience, drinking water-savings and economic costs. Journal of Cleaner production, 19(6-7), 598-608.
Echenique, M.H., Grinevich, V., Hargreaves, A.J., Zachariadis, V. (2013) LUISA: A Land Use Interaction with Social Accounting model; presentation and enhanced calibration method. Environment and Planning B, 40 (6) 1003-1026.
Ford, A., Dawson, R., Blythe, P., Barr, S. (2018). Land-use transport models for climate change mitigation and adaptation planning. Journal of Transport and Land Use, 11(1), 83-101.
Friedler, E., Hadari, M., 2006. Economic feasibility of on-site greywater reuse in multi-storey buildings. Desalination 190 (1-3), 221-234.
Green, A. Berkeley, N. (2006). The West midlands: the 'Hinge' in the middle. The Rise of the English Regions? (184-195). Routledge.
Groppi, D., de Santoli, L., Cumo, F., Garcia, D.A. (2018). A GIS-based model to assess buildings energy consumption and usable solar energy potential in urban areas. Sustainable Cities and Society, 40, 546-558.
Hargreaves, A.J., Farmani, R., Ward, S., Butler, D. (2019). Modelling the future impacts of urban spatial planning on the viability of alternative water supply. Water Research, 162, 200-213.
Hargreaves A.J., Cheng V., Deshmukh S., Leach M., Steemers K. (2017). Forecasting how residential urban form affects the regional carbon savings and costs of retrofitting and decentralized energy supply. Applied Energy, 186(3), 549-561.
Hargreaves A.J. (2015). Representing the dwelling stock as 3D generic tiles estimated from average residential density, Computers, Environment and Urban Systems, 54, 280-300.
Hecht, R., Herold, H., Behnisch, M., Jehling, M. (2019). Mapping Long-Term Dynamics of Population and Dwellings Based on a Multi-Temporal Analysis of Urban Morphologies. ISPRS International Journal of Geo-Information, 8(1), 1-21.
HM Land Registry. (2014). INSPIRE Index Polygons spatial data https://www.gov.uk/guidance/inspire-index-polygons-spatial-data
Jenkins, D. P., Tucker, R., Rawlings, R. (2009). Modelling the carbon-saving performance of domestic ground-source heat pumps. Energy and Buildings, 41(6), 587-595.
Kim, J., Furumai, H. (2012). Assessment of rainwater availability by building type and water use through GIS-based scenario analysis. Water Resources Management, 26(6), 1499-1511.
MHCLG. (2019). The English indices of deprivation 2019. Ministry of Housing, Communities and Local Government, HM Government, UK. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019
McFadden D. (1974). Conditional logit analysis of qualitative choice behavior, in Frontiers in Econometrics, Ed. P Zarembka. Academic Press, New York, 105–142.

ONS. (2020). Population estimates for England, Wales, Scotland, and Northern Ireland, mid-2019. Office for National Statistics, UK. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration

ONS. (2001) National Statistics 2001 area classification for output areas. Office for National Statistics, UK. https://webarchive.nationalarchives.gov.uk/20160120194634/http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/national-statistics-area-classifications/national-statistics-2001-area-classifications/cluster-summaries/Output-Areas/index.html

Orford, S., Radcliffe, J. (2007). Modelling UK residential dwelling types using OS Mastermap data: A comparison to the 2001 census. Computers, Environment and Urban Systems, 31(2), 206-227.

Salem, A.B., Mount, T.D. (1974). A convenient descriptive model of income distribution: the gamma density. Econometrica: Journal of the Econometric Society, 42(6), 1115-1127.

Schirmer, P.M., van Eggermond, M.A., Axhausen, K.W. (2014). The role of location in residential location choice models: a review of literature. Journal of Transport and Land Use, 7(2), 3-21.

Siegal, S., Castellan N. (1988). Nonparametric statistics for the behavioral sciences, 2nd Edition. McGraw Hill, New York.

Tarnawski, V.R., Leong, W.H., Momose, T., Hamada, Y. (2009). Analysis of ground source heat pumps with horizontal ground heat exchangers for northern Japan. Renewable Energy, 34(1), 127-134.

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M., Ulfarsson, G. (2003). Micro-simulation of urban development and location choices: Design and implementation of UrbanSim. Networks and Spatial Economics, 3, (1), 43-67.

Whitehand, J.W.R., Morton, N.J., Carr, C.M.H. (1999). Urban morphogenesis at the microscale: how houses change. Environment and Planning B: Planning and Design, 26(4), 503-515.

Wilcox, J., Nasiri, F., Bell, S., Rahaman, M.S. (2016). Urban water reuse: A triple bottom line assessment framework and review. Sustainable Cities and Society, 27, 448-456.

Williams H. (1977). On the formation of travel demand models and economic evaluation measures of user benefits. Environment and Planning A, 9, 285–344.

Willuweit, L., O'Sullivan, J. J., (2013). A decision support tool for sustainable planning of urban water systems: Presenting the Dynamic Urban Water Simulation Model. Water Research, 47(20), 7206-7220.

Supplementary Material

A parametric model of residential built form for forecasting the viability of sustainable technologies

A. Data inputs

A1. Output-Area type classification

The Output-Area types are based on the year 2001 Output Area classification (ONS, 2001) because this is the closest available classification year to the mid-point of the most recent Age-band (#4). The seven Output-Area types correspond to the seven 'Supergroups' of this Output Area classification. The classification method, described in Vickers et al., (2005), clustered Output Areas into the seven Supergroups using 41 variables that included different built form and socio-economic factors. They characterised the Supergroups by their distinguishing variables on which they were either far below or far above the national average, as described in Table A1.

A2. Extracting the plot data of the Calibration-districts using GIS tools

A2.1 Selecting the dwellings

The Address-points are spatially intersected with the Output-Area boundaries to join the Address-point data to the Output-Area type. The data extraction process then spatially intersects this Address-point data with the cadastral polygon boundaries (HM Land Registry, 2014) and selects those polygons that contain one or more residential Address-point. This data is then processed to exclude polygons that contain either; more than one house; apartments and house/s; or commercial Address-point/s. These polygons would have needed to be analysed individually to identify the plot area per dwelling, which would have been too time consuming.

Of the remaining polygons, those that are intersected by only a single house Address-point are identified as the Plot-area of a house, and the Address-point data is joined to the Plot-area and extracted for the study. Those intersected by Apartment Address-points are subject to further GIS analysis, as follows. They are spatially intersected with building footprints of 2016 Mastermap®. If a polygon contains a building Footprint-area that is less than 100m^2 or intersected by less than three residential Address-points, it is assumed to contain houses that have been converted into flats, rather than purpose-built apartments and is therefore excluded. The polygons that remain are identified as the Plot-areas of Apartment blocks, and their data is extracted for the study. They are then spatially joined to the building height data of 2016 Mastermap® to identify whether they are low-rise or high-rise Apartments. The English House Condition Survey (DCLG (2009) classifies Apartments as low-rise if the building is up to 15m high (4 storeys or less). The Address-point data of the Apartments is joined to the Plot-area and building height data and extracted for the study.

A2.2 Identifying the Age-band

The dwellings have been classified by 'Age-band' using Ordnance Survey maps from Historic Digimap®. The 1938 maps are 2500 scale (County Series 3rd Revision), and the later historic maps are 1250 scale (National Grid 1st Edition and Revisions). The most recent maps are 1250 scale, year 2016, Mastermap®. The historic maps are converted from raster data to vector polygons and then cleaned before carrying out the following GIS analysis to attribute each Address-point an Age-band. The results are shown in Fig. A1.

To identify the Age-band of houses, the polygons are regularised as circles (including a spatial tolerance of 4m because the hand-drawn dwelling footprints of old maps can be slightly inaccurate in shape and location). Those circles within a size range of 15m^2 to 130m^2 are selected on each historic map as possibly being the building footprint of a house. If an Address-point of a house spatially intersects one of these circles on the oldest map (i.e., year 1938) it is attributed the Age-band #1. The process is repeated for the oldest of the remaining maps and if an Address-point intersect a circle that has not already been attributed Age-band #1, it is attributed Age-band #2. This process is repeated until all the Address-points for houses have been attributed an Age-band.

The allocation of the Apartments to Age-bands is more complex because footprints of similar size and shape to Apartment blocks could be industrial buildings on sites that have since been redeveloped as Apartments. Therefore, the actual building footprints of Apartments on the 2016 Mastermap® are compared with those polygons on the historic maps that spatially are intersected by Apartment Address-points. If a polygon has a similar area to the current building footprint, it is identified as an Apartment block and allocated the appropriate Age-band, (using the sequential method described above).

Table A1
The Output-Area types and their variables that are above or below the national average.[1]
(Source: ONS, 2001).

Name	Variables with proportions far below the national average	Variables with proportions far above the national average
City Living	Detached housing [2]: Households with non-dependent children [3]: Children age 5-14 [4].	Apartments: Rent (private) [5]: Single person household (not a pensioner) [6]: Higher education qualification [7]: Born outside the UK [8].
Multicultural	Detached housing.	Apartments: Rent (private): Rent (public) [9]: Public transport to work: Born outside the UK [10].
Constrained Circumstances	Detached housing: 2+ car household [11]: Higher education qualification.	Apartments: Rent (public).
Hard Pressed Living [12]	Apartments: Higher education qualification.	Terraced housing: Rent (public).
Typical Traits	Rent (public).	Terraced housing.
Prospering Suburbs	Apartments: Terraced housing: Rent (private): Rent (public): No central heating [13].	Detached housing: 2+ car household.
Countryside	Apartments: Population density [14]: Public transport to work.	Detached housing: 2+ car household: Work from home [15]: Agriculture/fishing employment [16].

Notes:
[1] Based on range-standardised difference from the UK mean; a variable is 'far below average' if it has a difference of more than 0.15 below the mean, and 'far above average' it has a difference of more than 0.15 above the mean.
[2] Percentage of all household spaces which are detached.
[3] Percentage of households comprising one family and no others with non-dependent children living with their parents.
[4] Percentage of resident population aged 5-14.
[5] Percentage of households that are resident in private sector rented accommodation.
[6] Percentage of households with one person who is not a pensioner.
[7] Percentage of people with a higher education qualification.
[8] Percentage of people not born in the UK.
[9] Percentage of households that are resident in private sector rented accommodation.
[10] Percentage of people aged 16-74 in employment who usually travel to work by public transport.
[11] Percentage of households with 2 or more cars.
[12] This Output Area Supergroup was originally called 'Blue collar communities' but was renamed 'Hard Pressed Living' in the ONS 2011 classification.
[13] Percentage of occupied household spaces without central heating.
[14] Population density (number of people per hectare).
[15] Percentage of people aged 16-74 in employment who work mainly from home.
[16] Percentage of people aged 16-74 in employment working in agriculture and fishing.

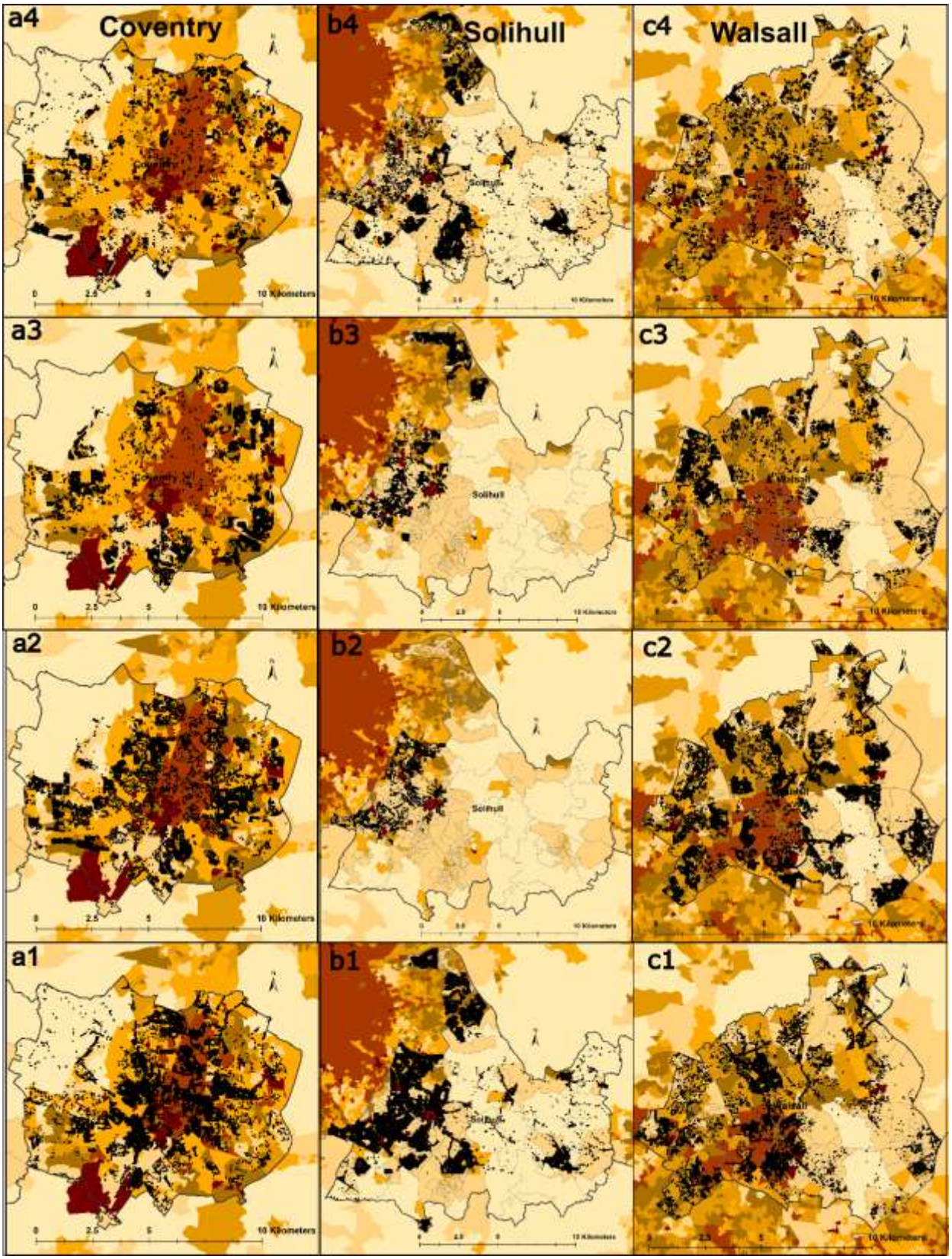

Fig. A1. Address-points (shown as black dots) for (a) Coventry, (b) Solihull & (c) Walsall; of dwellings built during Age-bands (#1), (#2), (#3) & (#4); (see Fig. 2 in the main text for the legend to the background choropleth map).

A2.3 Attributes per plot

The above methods result in the residential plots having the following attributes needed for analysis; Dwelling-type; Age-band; Output-Area type; building height; dwelling units per plot; and Plot-area. This is the '*empirical*' GIS data for the calibration.

Terraced houses have been classified as End-terraced or Mid-terraced. This uses the distance matrix vector analysis tool in QGIS® to measure the distance to the nearest two Address-points. Those with only one Address-point within 8m are identified as End-terraced and those with two within 8m as Mid-terraced. This gives similar percentages to the English House Condition Survey data of approximately one-third End-terraced and two-thirds Mid-terraced. There are too few high-rise Apartments in the study areas for analysis and so only the data of low-rise Apartments is used for the study. There were also some unusually high-density outliers, such as plots that have been divided into more than one cadastral polygon due to a historical right of way through the garden to nearby properties. These outliers have been excluded before carrying out the K-S goodness of fit test.

A3. Extracting the plot data of the Validation-district using GIS tools

The method of extracting the data for the Validation-district aims to be consistent with that used for the Calibration-districts, but it is not attributed an Age-band. However, the residential Address-point data for this District does not include the Dwelling-type. Therefore, the residential Address-points are spatially intersected with the building footprint polygons of the 2016 Mastermap®. Those footprints that contain only a single residential Address-point are identified as houses. The remainder are analysed in a similar way to Section A2.1 to identify purpose-built Apartments. The Address-points are intersected with the Cadastral polygons, and the Output-Area boundaries to join them to the Plot-area and Output-Area type data. The above method results in the following data needed for the analysis; the number of houses and low-rise apartments and their Plot-area per Output-Area type. This is the empirical GIS data for the validation.

B. Results of the K-S goodness of fit test of the gamma distribution

B1. The method of testing the goodness-of-fit

The theoretical gamma distribution is fitted to empirical Plot-density distribution by maximum likelihood estimation of the values of k and θ (Thom, 1958). The empirical Plot-density distribution is the number of dwellings per integer value of Plot-density μ. The null hypothesis is that the fitted theoretical distribution and this empirical data are drawn from the same distribution. The Kolmogorov Smirnov (K-S) one-sample test (Siegal and Castellan 1988) looks for the largest difference ('ksstat'), in absolute terms, between the theoretical cumulative distribution function and the empirical cumulative distribution function.

This statistical test needs to be more stringent than the standard K-S test because the parameters of the fitted distribution are estimated from the same data that are used to test the goodness of fit. These critical values are published in Crutcher (1975). A Matlab® programme (Section B2) has been written to fit the gamma distribution to this GIS data and calculate ksstat, which is then compared with the critical value. Tables B1 to B5 summarise the estimates of shape parameter k and the significance α of the K-S test results.

B2. Matlab® code for fitting the gamma distribution to the GIS data to estimate the shape parameter k and test the goodness-of-fit

```
% Delete the outliers
% The Plot data input format is: Column 1 = ID code of the dataset: Column 2 = the Plot-density
% of the dwelling from GIS data: Column 3 = grossing factor (this study grossing=1 because it
% uses all available data rather than data sampling).
clear
load Plot_data.csv
save plot_data
ID=Plot_data(1,1);
m=size(Plot_data,1);
for i=1:m
  if i==1
    for j=1:Plot_data(i,3)
      g(j)=Plot_data(i,2);
    end
  else
    for j=1:Plot_data(i,3)
      g(j+sum(Plot_data(1:(i-1),3)))=Plot_data(i,2);
    end
  end
end
grossing=g';
s=sum(grossing(:,1));
n=size(grossing,1);
% Estimating k and θ using Maximum Likelihood Estimation
sg=sum(grossing);
X=sg/n;
A=log(X);
lg=log(grossing);
b=sum(lg);
B=b/n;
D=A-B;
k=(1+(1+4*D/3)^0.5)/(4*D);
theta=X/k;
sorted=sort(grossing,'ascend');
% K-S one-sample test
test_cdf = [sorted,gamcdf(sorted,k,theta)];
[h,p,ksstat,cv] = kstest(sorted,'CDF',test_cdf);
% Output the results
T=table(ID,k,theta,ksstat,n);
writetable(T,'C:\Users\Name\Documents\Matlab_model\mySpreadsheet.xlsx','FileType','spreadsheet');
% end
```

B3. Results of the goodness of fit test and estimates of the shape parameter k

Table B1
Results for Age-band #4 for houses.

		Coventry				Solihull				Walsall			
	Area type	*k*	$\bar{x}$	n	*α*	*k*	$\bar{x}$	n	*α*	*k*	$\bar{x}$	n	*α*
Detached	1	8	30	219		7	26	313	***	7	29	613	
	2	2	10	71		5	15	29	*****				
	3	2	5	167	*****	2	6	956	*	5	20	228	
	4	5	20	2279		5	18	12139		5	22	3307	
	5	6	23	141	*	6	26	149		7	30	283	****
	6	7	28	1156		5	21	1263		12	30	902	
	7	9	34	219						6	29	135	*
Semi-detached	1	12	49	215	****	10	40	757		10	43	712	*
	2	8	37	37	**								
	3	5	22	67	*****	4	23	443	*	5	26	70	*
	4	8	31	708	****	6	28	3549	**	7	38	1065	
	5	10	47	139	*****	11	40	499		8	47	394	****
	6	9	40	1562	*****	5	31	1117	*	10	43	515	
	7	11	53	287	*					11	49	336	**
End-terraced	1	11	53	150	*****	14	58	699		9	52	365	
	2	16	52	43	*****								
	3					3	30	61	*****	14	44	26	*****
	4	16	48	99	*****	5	39	182	***	10	49	145	*****
	5	15	58	135	*****	12	52	479		15	57	154	*
	6	9	45	993	*****	8	44	201	*****	15	49	203	*****
	7	10	56	351	*****					9	53	97	***
Mid-terraced	1	10	55	158	*	16	64	1476		15	63	458	
	2	11	52	40	*****	7	58	55	*				
	3					5	44	133	*****	9	55	46	**
	4	6	40	139	*****	7	51	498		16	62	120	*****
	5	12	59	156		16	59	635		14	69	227	
	6	9	52	1113	*	11	58	802	*****	14	63	271	***
	7	9	66	429	****					10	69	128	*****

Key:

k= shape parameter; $\bar{x}$ =mean Plot-density; n = sample size; *α* = significance of the K-S test result that the empirical dwelling data and the fitted distribution are drawn from the same distribution:

***** 20% level; **** 15% level; *** 10% level: ** 5% level; * 1% level.

Table B2
Results for Age-band #3 for houses.

		Coventry				Solihull				Walsall			
	Area type	*k*	$\bar{x}$	n	*α*	*k*	$\bar{x}$	n	*α*	*k*	$\bar{x}$	n	*α*
Detached	1	10	33	79	*					7	24	354	*
	2	3	9	74	**					6	21	27	*****
	3	5	22	52	*								
	4	5	20	1921		6	18	3267		6	23	3495	
	5	9	29	70	****	6	23	34	*****	5	23	121	*
	6	6	24	522	*	5	22	378		7	27	603	
	7	6	28	26	*****					7	29	49	*****
Semi-detached	1	11	37	558	*	9	36	972		11	37	1440	
	2	14	34	153	*****	16	24	40	***	13	36	43	
	3												
	4	11	31	2726		11	29	4471		12	36	3436	
	5	8	38	630	*****	11	34	842	*****	14	40	496	*
	6	11	34	2932	*	9	28	1417	*****	14	38	1759	
	7	10	35	319	*****					14	45	293	*****
End-terraced	1	13	50	427	**	10	46	468		11	50	669	
	2												
	3												
	4	9	39	249	*****	7	34	70	*****	11	39	91	**
	5	11	49	385		14	52	923		13	50	302	*
	6	10	41	1606	***	10	40	143	*****	9	43	214	*****
	7	8	50	329	*****					13	55	46	*****
Mid-terraced	1	12	52	662		13	55	774		17	59	883	
	2	4	41	27	*****								
	3												
	4	7	37	346	**	8	47	131	****	7	45	130	
	5	13	53	516		18	55	1070		20	55	320	*****
	6	9	48	2174	****	12	51	309	*	17	60	329	
	7	10	56	446	*****					13	57	71	***

Key:
See Table B1.

Table B3
Results for Age-band #2 for houses.

		Coventry				Solihull				Walsall			
	Area type	*k*	$\bar{x}$	n	*α*	*k*	$\bar{x}$	n	*α*	*k*	$\bar{x}$	n	*α*
Detached	1	8	30	58						5	21	436	**
	2	2	9	69	*****	18	17	47	*	6	15	43	*****
	3	1	4	28	*****					4	14	117	
	4	3	17	616	****	4	13	1130		4	16	3593	
	5	9	31	70						5	20	186	*****
	6	4	20	500	**	4	18	95	*****	5	21	547	****
	7	5	26	76	*					6	25	107	*****
Semi-detached	1	14	36	1209		12	30	782	****	13	33	6477	
	2	7	26	127	*****	19	26	107	*	9	21	38	***
	3	19	31	57	*****					11	27	188	
	4	9	27	1718		8	26	2978	****	8	27	5188	
	5	17	39	632		12	27	548		12	34	2312	
	6	10	33	3806	*	11	32	1669		9	31	1984	
	7	11	38	755	*					9	32	1324	***
End-terraced	1	13	41	918	*	11	37	60	**	11	40	2581	**
	2	13	42	70	***								
	3												
	4	11	36	93	*****					15	37	101	
	5	13	45	496	*****	8	44	26	*****	9	40	605	*****
	6	11	40	3674		13	33	88	*****	14	41	509	*
	7	8	46	1127						11	38	385	
Mid-terraced	1	17	46	973		11	39	98	*****	13	46	2858	
	2	12	49	96									
	3												
	4	4	32	124	**					9	36	143	*****
	5	12	45	593	*****	17	57	36	**	12	48	689	*
	6	16	47	6327		17	39	306	***	16	49	693	
	7	8	57	2115						8	45	482	

Key:
See Table B1.

Table B4
Results for Age-band #1 for houses.

		Coventry				Solihull				Walsall			
	Area type	*k*	$\bar{x}$	n	*α*	*k*	$\bar{x}$	n	*α*	*k*	$\bar{x}$	n	*α*
Detached	1	5	25	50		6	22	45	*****	4	18	458	**
	2	1	3	66	*	5	14	38	*****	3	11	31	
	3	1	2	134	*	2	5	205	*****	3	11	83	****
	4	3	14	734	**	3	11	2527	*	3	15	1136	
	5	3	17	46	*					4	23	213	**
	6	3	18	525	***	4	15	361	*****	5	23	519	*****
	7	3	23	160	***					3	23	210	*****
Semi-detached	1	12	37	576	**	12	33	350		10	32	3382	
	2	4	13	190		11	25	147	*****	6	22	37	*****
	3	5	22	116		4	20	149	*****	6	23	70	*****
	4	7	24	1082	*	8	24	3908		7	27	1538	*
	5	10	32	144	*****	13	29	105	**	7	33	831	*
	6	9	33	3334		9	29	1498	**	6	31	1086	*****
	7	5	41	998						6	38	739	*
End-terraced	1	16	47	720		6	44	45	*****	8	39	1009	
	2	8	61	101	*								
	3												
	4	5	33	92	*	6	33	62	*****	7	36	85	*****
	5	13	45	170	*****	9	46	66	**	7	42	325	*
	6	9	44	4015		7	37	76	*****	6	45	331	*****
	7	6	61	3127						6	53	1070	*****
Mid-terraced	1	15	50	752		11	54	114	*	7	49	1410	
	2	6	70	228		11	68	25	*****				
	3	2	24	27	*****								
	4	3	23	116	*	4	41	106	*****	6	43	82	****
	5	11	49	253		9	43	100		9	53	482	**
	6	11	52	8245		5	48	188	*	8	56	774	**
	7	8	76	7800						7	68	4359	

Key:
See Table B1.

Table B5
Results for Age-band #4 and Age-bands #1 to #3 for Apartments.

Apartments	All Districts Age-band #4				All Districts Age-bands #1 to #3			
Area type	*k*	$\bar{x}$	n	*α*	*k*	$\bar{x}$	n	*α*
1	4	56	166	*****	5	64	121	*****
2	4	61	155	*****	3	78	132	
3	2	17	37	*****				
4	3	44	267	*****	5	47	124	****
5	6	65	428	*****	6	63	272	
6	4	57	451		4	62	311	**
7	4	79	331	*****	3	83	898	

Key:
See Table B1.

C. Matlab® code to validate the model by testing the goodness of fit of the modelled variables to the GIS data

```
% Data input format is: Column 1 integer value of Plot-density: Column 2 'empirical' variable:
% and Column 3 'modelled' variable.
clear
load Plot_data.csv
save plot_data
density=Plot_data(:,1);
xcol=Plot_data(:,2);
ycol=Plot_data(:,3);
xaxis=density';
x=xcol';
y=ycol';
[m,n]=size(x);
% This uses the 2-sample K-S test because the model has been calibrated independently from
% the Validation-district data.
[h,p,ks2stat] = kstest2(x,y,"alpha",0.05);
T=table(h,p,ks2stat,n);
writetable(T,'C:\Users\Name\Documents\Matlab_model\mySpreadsheet.xlsx','FileType','spreadsheet');
% end.
```

References

Crutcher, H.L. (1975). A note on the possible misuse of the Kolmogorov-Smirnov test. Journal of Applied Meteorology, 14, (8) 1600-1603.

DCLG (2009). English House Condition Survey 2007: Annual Report. Department of Communities and Local Government Publications, HM Government, UK. http://doc.ukdataservice.ac.uk/doc/6449/mrdoc/pdf/6449ehcs_annual_report_2007.pdf

HM Land Registry. (2014). INSPIRE Index Polygons spatial data https://www.gov.uk/guidance/inspire-index-polygons-spatial-data

ONS. (2001) National Statistics 2001 Cluster summaries of Output Areas. Office for National Statistics, UK. https://webarchive.nationalarchives.gov.uk/20160120194634/http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/national-statistics-area-classifications/national-statistics-2001-area-classifications/cluster-summaries/Output-Areas/index.html

Siegal, S., Castellan N. (1988). Nonparametric Statistics for the Behavioral Sciences, 2nd Edition. McGraw Hill, New York.

Thom, H. C. S. (1958). A note on the gamma distribution. Monthly Weather Review, 86, 117–122.

Vickers, D., Rees, P., Birkin, M. (2005) Creating the national classification of census output areas: data, methods, and results. Working Paper 05/2, University of Leeds, White Rose Publishing http://eprints.whiterose.ac.uk/5003/