Representing the dwelling stock as 3D generic tiles estimated from average residential density

Hargreaves, Anthony

DOI:
10.1016/j.compenvurbsys.2015.08.001

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Hargreaves, A 2015, 'Representing the dwelling stock as 3D generic tiles estimated from average residential density', *Computers, Environment and Urban Systems*, vol. 54, pp. 280–300. https://doi.org/10.1016/j.compenvurbsys.2015.08.001

Computers, Environment and Urban Systems 54 (2015) 280–300

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Representing the dwelling stock as 3D generic tiles estimated from average residential density

A.J. Hargreaves

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

ARTICLE INFO

Article history:
Received 21 March 2015
Received in revised form 20 July 2015
Accepted 6 August 2015
Available online 24 October 2015

Keywords:
Urban modelling
Housing survey
Gamma distribution
Dwelling typology
Building-scale technologies

ABSTRACT

Forecasting the variability of dwellings and residential land is important for estimating the future potential of environmental technologies. This paper presents an innovative method of converting average residential density into a set of one-hectare 3D tiles to represent the dwelling stock. These generic tiles include residential land as well as the dwelling characteristics. The method was based on a detailed analysis of the English House Condition Survey data and density was calculated as the inverse of the plot area per dwelling. This found that when disaggregated by age band, urban morphology and area type, the frequency distribution of plot density per dwelling type can be represented by the gamma distribution. The shape parameter revealed interesting characteristics about the dwelling stock and how this has changed over time. It showed a consistent trend that older dwellings have greater variability in plot density than newer dwellings, and also that apartments and detached dwellings have greater variability in plot density than terraced and semi-detached dwellings. Once calibrated, the shape parameter of the gamma distribution was used to convert the average density per housing type into a frequency distribution of plot density. These were then approximated by systematically selecting a set of generic tiles. These tiles are particularly useful as a medium for multidisciplinary research on decentralized environmental technologies or climate adaptation, which requires this understanding of the variability of dwellings, occupancies and urban space. It thereby links the socioeconomic modeling of city regions with the physical modeling of dwellings and associated infrastructure across the spatial scales. The tiles method has been validated by comparing results against English regional housing survey data and dwelling footprint area data. The next step would be to explore the possibility of generating generic residential area types and adapt the method to other countries that have similar housing survey data.

1. Introduction

There has been an increasing emphasis on understanding the building stock and how to reduce the consumption of energy and production of waste (Kohler & Hassler, 2002). Much of this research has focused on the buildings themselves and predominantly their energy consumption (Kavgic et al., 2010). This has normally used typologies that correspond with the classification of national housing stock data such as dwelling types, age bands and fabric. Examples include building energy models (Firth, Lomas, & Wright, 2010; Cheng & Steemers, 2011) and studies of the building stock and energy efficiency such as McKenna, Merkel, Fehrenbach, Mehne, and Fichtner (2013); Ballarini, Corgnati, and Corrado (2014); Filogamo, Peri, Rizzo, and Giaccone (2014) and Mata, Sasic Kalagasidis, and Johnsson (2014).

However, there is increasing recognition that decentralized supply technologies are also important for helping to meet government environmental targets but uncertainty about whether properties have the space required for installation has been identified as a barrier to implementation (DECC, 2012, Sept. 20th). For example, the dimensions of gardens, roof space or cluster size will affect the feasibility of some technologies such as ground source heat pumps, rainwater harvesting and recycling. Modeling tools have been developed to assess the potential of environmental technologies (Hofierka & Kanuk, 2009; Girardin, Marechal, Dubuis, Calame-Darbellay, & Favrat, 2010; Lukac, Zlaus, Seme, Zalik, & Stumberger, 2013; Makropoulos, Natsis, Liu, Mittas, & Butler, 2008; Robinson et al., 2007) but these detailed simulations for relatively small areas require inputs on future urban form.

Planning policies, building regulations and incentive schemes are applied at national or regional level and require a long time scale and considerable investment to take effect. The outcome will depend on urban density, occupancies and whether dwellings are as existing or new build. Urban densities and occupancies will vary spatially within a city region as a result of the socio-economic pressures that drive the property market and shape urban form. Human factors are thought to account for a substantial amount of the variability of energy use in buildings. Yu, Fung, Haghighat, Yoshino, and Morofsky (2011) and Pereira and Assis (2013) showed how the increases over time in household energy consumption are spatially correlated with socio economic

E-mail address: a.j.hargreaves@bham.ac.uk.

http://dx.doi.org/10.1016/j.compenvurbsys.2015.08.001

changes in income. A forecasting capability is therefore needed at regional scale to test the longer term impacts and cost effectiveness.

Regional-scale land use-transport forecasting models can provide a detailed top-down simulation of the supply and demand for land and floor space at the building parcel scale (Abraham, Weidner, Gliebe, Willison, & Hunt, 2005). This included GIS-based micro-scale modeling of floor space types and rental values for land parcels, with floor space categorized according to general building types. This reproduces the spatial layout and overall floor space but does not represent the size and variability of buildings. This reliance on mapping limits the capability to forecast the future urban form. In another example, a regional scale macro-model was linked to an UrbanSim model (Waddell et al., 2003), which simulated neighborhoods as 2.25 ha grid cells chosen from a set of 25 development types further defined by a range of residential units and non-residential floor space to create typical contiguous urban areas. These models aim to represent the actual land parcels and this leads to difficulties matching the data sources which makes the models very resource intensive to create and operate over large areas within a macro-modeling framework.

Computer graphic simulation methods are available to study built-form and its potential for sustainable technologies (Meinel, Hecht, & Herold, 2009; Vanegas et al., 2010; Wiginton, Nguyen, & Pearce, 2010; Jacubiec & Reinhart, 2013). They rely on location specific inputs of road networks and user specified attributes of land parcels and building shapes using mapping and aerial imaging. The outputs can be similar in complexity to the actual physical built environment, thus limiting the practical size of study area. An alternative is to use a theoretical simulation of built form. Some are based on metaphors for the urban development process such as (Crompton, 2012) who compared the variability of buildings to the variability of Lego™ pieces. Others such as Tuhus-Dubrow and Krarti (2010) have used optimization methods to estimate the most energy efficient building form. These theoretical approaches are not subject to the livability and commercial constraints that shape actual dimensions and are not empirically validated. Neither the graphic or theoretical simulation methods have the capability to forecast urban densities and occupancies.

The most promising approach for linking across the spatial scales is to use a statistical method of representing the variability of land per dwelling. Zhou and Kockelman (2008) recognized the advantages of this approach but were unable to fit a statistical function to their single family residential parcel size data. This is possibly because they were studying only one city and using GIS data instead of housing survey data.

The following sections of this paper present a unique method of estimating the variability of dwellings and residential land from the mean residential density. The parametric variability of dwellings and associated land is then represented by systematically selecting a set of discrete one hectare 3D tiles. The final part of the paper validates this 'tiles' method and discusses how this enables the modeling of dwellings and environmental technologies to be integrated with a regional forecasting model. This attains the important and previously difficult to achieve objective of linking models across spatial scales.

It is expected that this paper will be of interest for spatial modeling and urban simulation, particularly for forecasting the impacts of building scale interventions such as sustainable technologies and climate change mitigation.

2. Method

2.1. Research context

This research was part of a project that tested spatial planning policies in combination with scenarios for decentralized sustainable technologies for London and its surrounding regions over a 30 year time horizon. The aim was to explore how the density and clustering of development would affect the potential of 'green' technologies for buildings, energy, transport, water and waste. This required a regional modeling framework with a forecasting capability.

2.2. Spatial interaction model

A regional Land Use Interaction and Social Accounting model (LUISA) was developed to forecast the spatial allocation of industry, employment, households and population. This was an aggregate static model based on input–output socioeconomic accounting tables linked to random utility discrete choice modeling of spatial allocation and travel behavior (Echenique, Grinevich, Hargreaves, & Zachariadis, 2013). Rents arise when there are constraints on the amount of production that can be assigned to a location. In order to balance demand and supply, the production prices and disutilities need to adjust, by generating rents. This process is dealt with endogenously within the model, using an iterative procedure. The model has bimodal accounting of monetary and non-monetary disutility. The total monetary disutility includes the rents and building construction costs, and the total non-monetary disutility includes qualitative aspects such as what consumers of land are willing to tolerate in order to have a lower monetary rent, for example living in a high rise building.

For practical purposes the case study area was divided into zones, within which locations were assumed to be homogenous. Once constructed, the model was calibrated so that its outputs matched the base year data on spatial production and consumption and prices per zone. The regional economic and demographic projections, future land availability and transport improvements are then input to the model to test scenarios for the forecast year. The forecasts included households and population per zone by socio-economic classification. The challenge was to convert these aggregate outputs per zone into a realistic estimate of the future dwelling stock and occupancies so that scenarios could be tested for decentralized sustainable technologies. It was important that this included the variability of land per dwelling because this affects the potential for sustainable technologies.

2.3. Analysis of the English House Condition Survey data

The English House Condition Survey (EHCS) is a detailed source of data on the English housing stock that includes both the building and plot dimensions. The 2007 EHCS contained 16,194 sample dwellings and the sampling takes into account the location and tenure. It comprises a household interview, a physical inspection and a market valuation. The physical inspection provides detailed information about the building dimensions and plot size, building fabric and service systems of each sample dwelling. The plot is the private land that belongs to the dwelling (generally referred to in the USA as the lot) or if the private land belongs to a small number of dwellings the plot includes the proportion attributed to the surveyed dwelling.

The variables used to categorize the dwellings can be found in the survey guidance document (DCLG 2007) and further information can be found on the Department for Communities and Local Government website (DCLG 2013). The EHCS variables chosen to categorize the dwellings for this analysis were the dwelling type, urban morphology, area type, region, and age band, which are described in Appendix A.

The EHCS data was firstly prepared for the purposes of analysis by calculating the plot area per dwelling based on the EHCS plot dimensions (Fig. 1).

The plot area was estimated as:

$$a_i = W * (F + B + R) \quad (1)$$

where:

a_i = *plot area of a house or an apartment block i* (m^2)
W = *plot width* (m)
F = *depth of land at the front* (m)
B = *depth of the building* (m)
R = *depth land at the rear* (m).

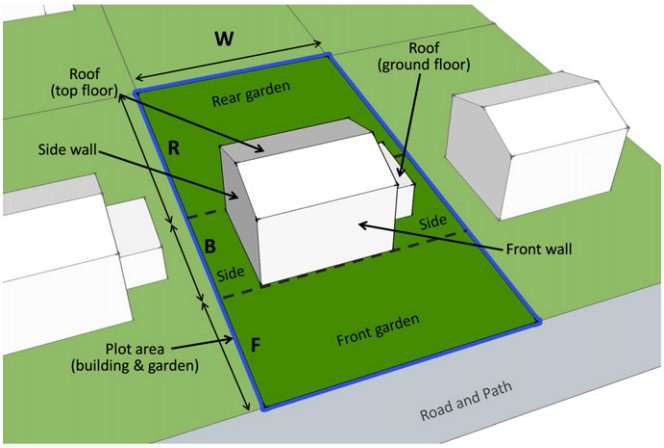


Fig. 1. Plot dimensions from EHCS data used for calculating the plot density.

The 'plot density' in dwellings per hectare was calculated as:

$$\mu_i = \left(u * 10^4\right)/a_i \tag{2}$$

μ_i = *plot density of dwelling i*

For a house	u = 1
For an apartment	u = number of apartments within the block

Plot density μ per dwelling was used as the metric instead of average dwelling density because the paper analyzes the variability of individual dwellings rather than neighborhoods. This plot density μ metric therefore gives a higher numerical value than the average residential area density normally used by planners because it excludes publically accessible areas and rights of way.

Each surveyed dwelling is allocated a grossing factor by the EHCS to convert the survey sample to an estimate of the English housing stock. The EHCS surveyed sample of dwellings was converted into an estimate of the English dwelling stock, as follows:

$$H_i = h_i \cdot w_i \tag{3}$$

where:

h_i = *Dwelling i surveyed by EHCS*
w_i = *EHCS grossing factor to convert the sample dwelling into the dwelling stock*
H_i = *Number of dwelling equivalent to surveyed dwelling* h_i.

Hence:

$$A_i = a_i \cdot w_i \tag{4}$$

where:

A_i *Total plot area of dwelling equivalent to surveyed dwelling* h_i.

The frequency distribution of the plot densities μ_i of dwellings H_i is positively skewed and similar to the gamma distribution (as shown in Fig. 2). However, there is some 'lumpiness' in the distribution. If the

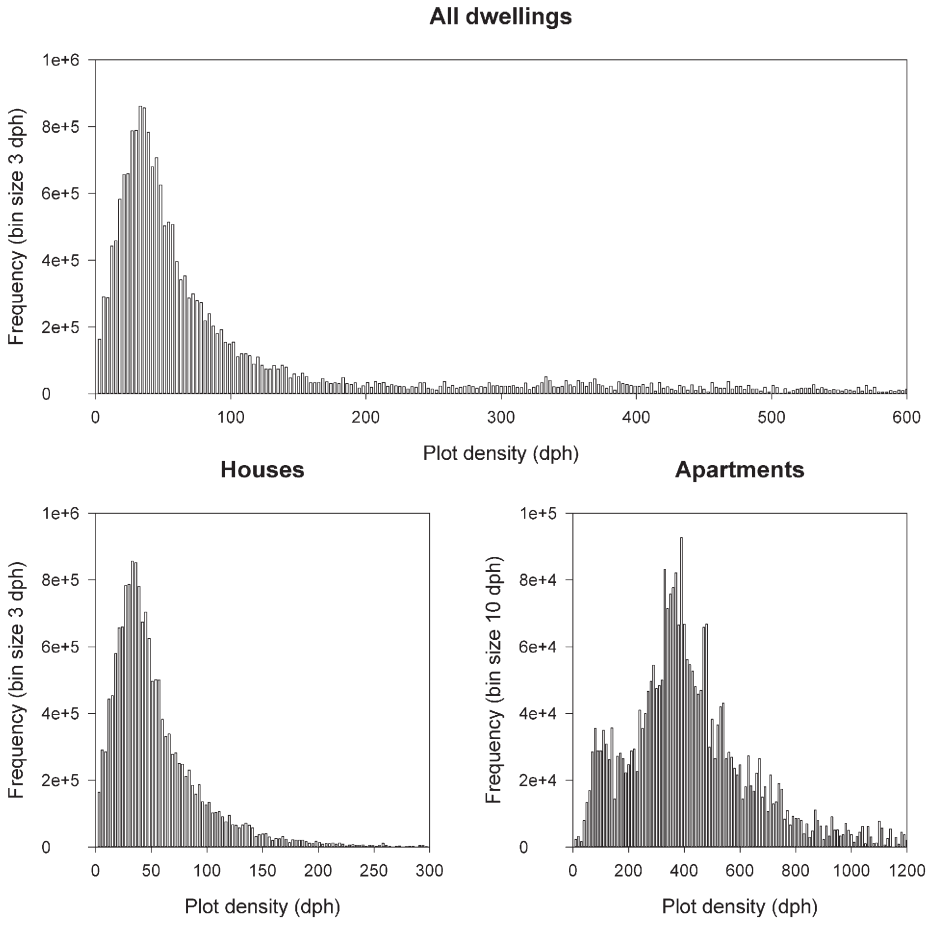


Fig. 2. Examples of frequency distributions of plot density μ.

Table 1
Dwellings H_i represented by the selected sample compared to total English dwellings.

Dwelling type	Dwellings represented in the analysis	Total English dwellings	Percentage
Detached	2,402,435	3,971,028	60%
Semi-detached	5,002,086	6,102,855	82%
Mid-terraced	1,628,060	2,184,612	75%
End-terraced	3,153,252	4,055,534	78%
Low density apartments	2,268,311	2,695,686	84%
High density apartments	189,821	317,697	60%
Bungalow	1,181,580	2,026,165	58%
Converted apartments	529,769	722,679	73%
Total dwellings	16,355,314	22,076,256	74%

dwellings are disaggregated into houses and apartments then the distributions have a broadly similar shape but with different mean and scale. It was found that further disaggregation into different dwelling types, area types and age band achieves density distributions that become progressively smoother and more distinct. This thereby increases the likelihood of fitting a density distribution function to the empirical data.

The variables that affect density were identified by analyzing the EHCS data using a generalized linear model (GLM). This estimated the significance of the correlation between the dependent variable and each independent variable whilst taking into account the variability of the other independent variables. The chosen specification of the GLM for this application used gamma regression with plot density μ as the dependent variable. The EHCS variables used as predictors were dwelling type, age band, area type, morphology and region. Further information about this type of model can be found in Chapter 8 of McCullagh and Nelder (1989).

This GLM analysis of the independent variables found that dwelling type, age band, area type, and urban morphology were all significantly related to the dependent variable of dwelling plot density. The region variable showed a 'north south divide' and the three northern regions (regions 1 to 3) had slightly higher densities than the rest of the country (around 2 dwellings per hectare higher). A possible reason is that a greater proportion of housing in the north was built for industrial workers, but this hasn't been investigated further. The three northern regions account for around one third of English housing and the difference in density is quite small so it was decided not to disaggregate the data by regions in order to avoid further reducing the sample sizes for analysis.

Importantly, the GLM analysis found no consistent trend in density per dwelling type over time. In fact, the analysis shows that densities have fluctuated with periods of lower than average density pre-1850 and from 1919 to 1980, and periods of higher than average density from 1850 to 1918, possibly due to industrialization, and from 1981 to 2007 possibly due to planning constraints. Hence, although plot densities per dwelling type have fluctuated over time, there is no evidence that a method derived from the analysis of existing housing stock cannot be used to forecast the density distributions of future housing stock. Overall residential densities have increased in England over this long timescale due to urbanization but this has mainly been achieved by having a greater proportion of apartments and terraced dwellings per area type, rather than by an increase in density per dwelling type.

Based on the above findings, the next step of the analysis aggregated the EHCS data into the possible combinations of age-band (9 bands), morphology (4 types), and area type (6 types). Each combination was regarded for this analysis as equivalent to an 'aspatial development type' c and was analyzed per dwelling type d. Dwellings were regarded as outliers and excluded if their plot density μ_i exceeded the upper quartile plus 1.5 times the inter-quartile range, or was less than the lower quartile minus the 1.5 times the inter-quartile range. The outliers were mainly dwellings with a plot size the same or smaller than its footprint, which may be due to it being part of a mixed-use building or of unusual construction. Many of these combinations of dwelling type and aspatial development type were either unusual or inconsistent and so only had small/zero samples per dwelling type (such as apartments in a rural area, or a rural area with urban morphology, etc.). Those combinations with a remaining sample h_{di} of less than 24 dwellings were excluded from the analysis. There were 108 remaining combinations of dwelling type and the aspatial development type for the next step of fitting a density distribution function. Table 1 shows that after applying the w_i grossing factors, 74% of the total English dwellings remained for analysis.

2.4. Fitting a probability distribution to the EHCS dwelling stock data

For each combination of dwelling type d and aspatial development type c, the frequency distribution of the plot density μ was found to have a similar shape to the gamma distribution for $k > 1$. The general probability density function (PDF) of the gamma distribution is:

$$f(\mu; k, \theta) = \frac{1}{\theta^k \Gamma(k)} \mu^{k-1} e^{-\mu/\theta} \quad \text{for } \mu > 0 \text{ and } k, \theta > 0 \tag{5}$$

where:

$\mu \sim \Gamma(k, \theta) \equiv Gamma\ (k, \theta)$
$k = shape\ parameter$
$\theta = scale\ parameter.$

The cumulative distribution function (CDF) is:

$$F(\mu|k, \theta) = \frac{\gamma(k, \mu/\theta)}{\Gamma(k)} \tag{6}$$

where γ is the lower incomplete gamma function:

$$\gamma(k, \mu/\theta) = \int_0^{\mu/\theta} t^{k-1} e^{-t} dt. \tag{7}$$

The gamma distribution has the convenient mathematical property that the mean $\overline{\mu}$ equals the product of the two parameters that define the distribution. Hence, the expected value of the theoretical mean plot density $\overline{\mu}$ is:

$$\overline{\mu} = k \cdot \theta. \tag{8}$$

Subsequent investigation of the literature found that the gamma distribution has been widely studied for its usefulness in curve fitting to data of extreme events such as accidents, climatology, and hydrology. For example, Ison, Feyerherm, and Dean (1971); and Husak, Michaelsen, and Funk (2006) studied the differences in the shapes of the gamma distributions fitted to empirical climate data to assess whether locations have irregular or extreme events by comparing the parameters k and θ as discussed in Wilks (1995).

An innovative feature of the 'tiles method' is that it uses the gamma distribution from the opposite perspective to the above studies. It firstly calibrates the shape parameter k using the EHCS data and then specifies the gamma distribution from the calibrated shape parameter $\hat{k}$ and mean density $\overline{x}$.

The shape parameter k is estimated using a maximum likelihood estimator for the gamma distribution (Thom, 1958). This includes the sample statistic E which is the difference between the natural log of the sample mean, and the mean of the logs of the data:

$$E = \ln(\overline{\mu}) - \frac{1}{n} \sum_{i=1}^{n} \ln(\mu_i) \tag{9}$$

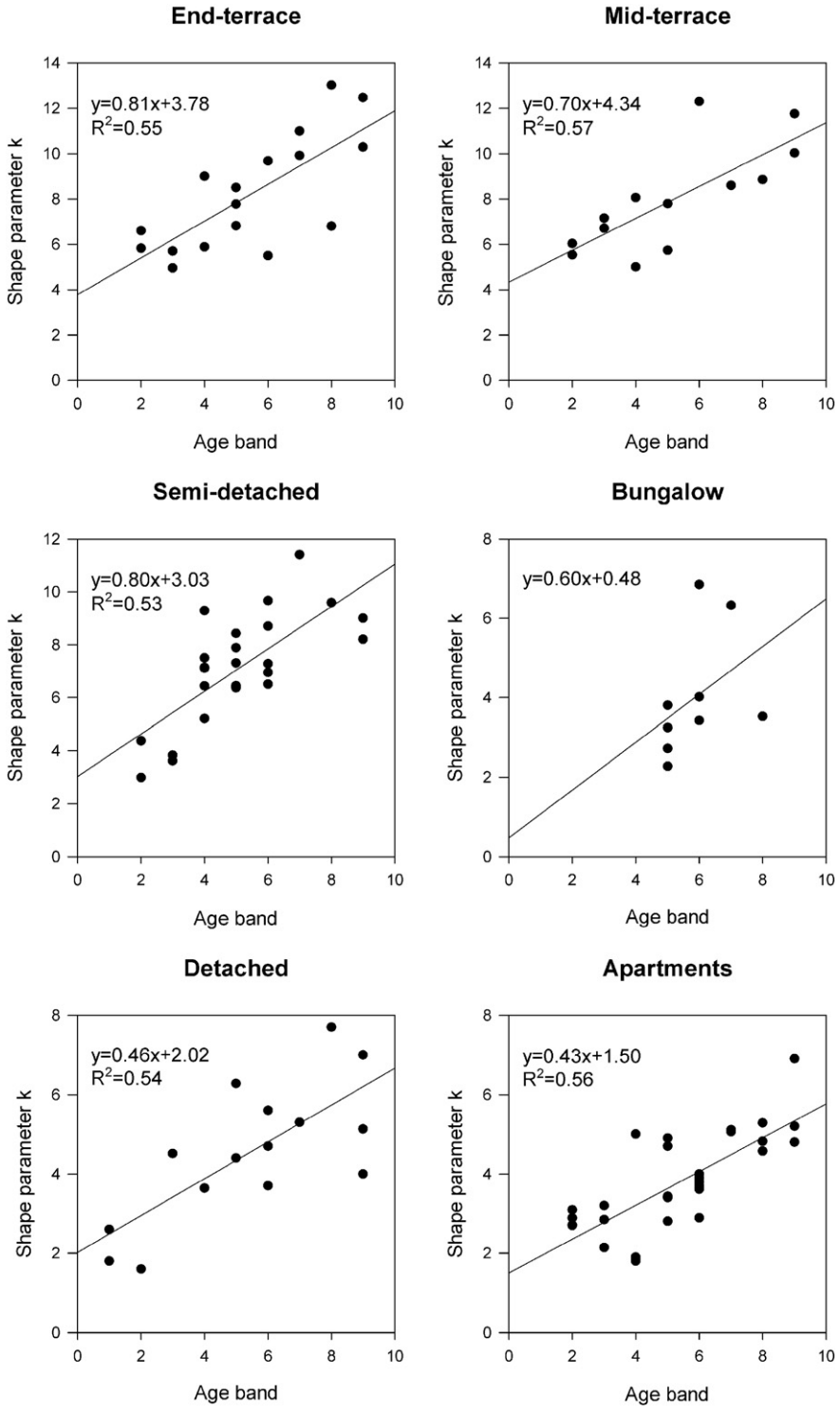


Fig. 3. Shape parameter k by age band for the main dwelling types.

$$k = \frac{1 + \sqrt{1 + 4E/3}}{4E}. \quad (10)$$

For the plot density frequency distribution, the shape parameter k can be conceptualized as the result of successive subdivisions of land into plots. The shape parameter k represents the degree of similarity between the plot sizes each time a plot sub-divides, whereas the scale parameter θ represents the amount of subdivision that has taken place. The smaller the shape parameter, the more positively skewed the distribution and the larger the shape parameter, the more similar it is to the normal distribution.

2.5. Kolmogorov–Smirnov (K–S) goodness of fit test

The estimated gamma distribution is compared against the empirical density distribution using the Kolmogorov–Smirnov (K–S) one-sample goodness of fit test (Siegal & Castellan, 1988). The statistical test needed to be more stringent than the standard K–S test because the distribution parameters were estimated using the EHCS data and

Table 2
Summary of the calibrated shape parameter $\hat{k}$ for different ages and types of dwelling.

Dwelling type[a]	Estimate of shape parameter $\hat{k}$			
	Pre-1945	1945–1974	1975–1990	Post–1990
Mid-terraced houses End-terraced houses Semi-detached houses	6	8	10	11
Detached houses Bungalows Low rise apartments High rise apartments	3	4	5	5.5

[a] Converted apartments are not included because only 6 of the variable combinations had large enough sample sizes, which is insufficient to reliably estimate the shape parameter.

the same data was used to test the goodness of fit. A more stringent K–S statistic was therefore used based on a version of the Lilliefors test using critical values for this K–S statistic to assess goodness of fit of gamma distributions that were originally published in Crutcher (1975) and reproduced in Wilks (1995).

Both the K–S and Lilliefors tests utilize the following test statistic:

$$g_n = max_\mu |F_n(\mu) - F(\mu)| \tag{11}$$

where $F_n(\mu)$ is the empirical cumulative probability, which is estimated as $F_n(\mu_i) = {}^{i}/_{n}$ for the i'th smallest data value in the sample n of h_i and F(μ) is the theoretical cumulative gamma distribution function (CDF) evaluated at μ_i. Thus the K–S test statistic g_n looks for the largest difference, in absolute terms, between the empirical and the fitted CDFs for the sample of size n.

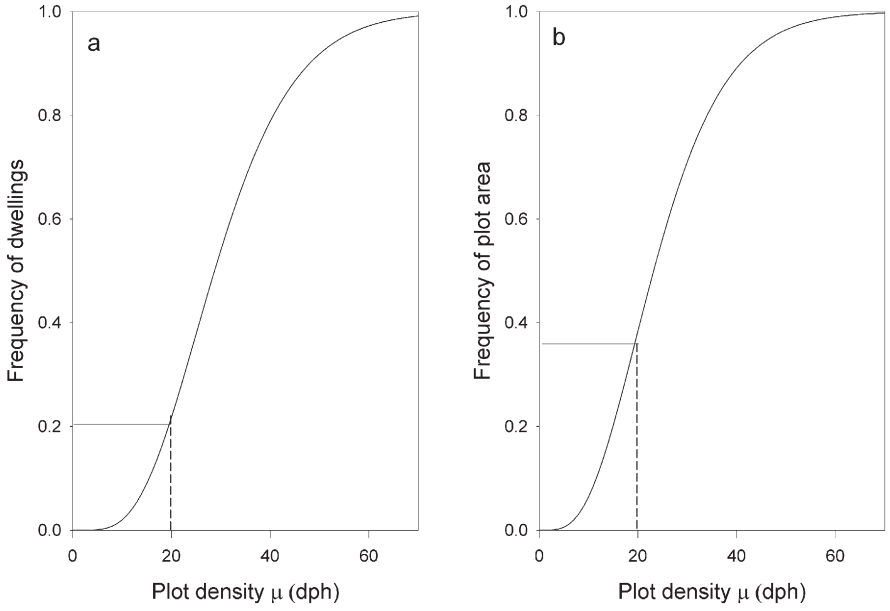


Fig. 4. a: Detached dwellings $\bar{x} = 25$ dph, $\hat{k} = 5.5$ (built since 1990s).

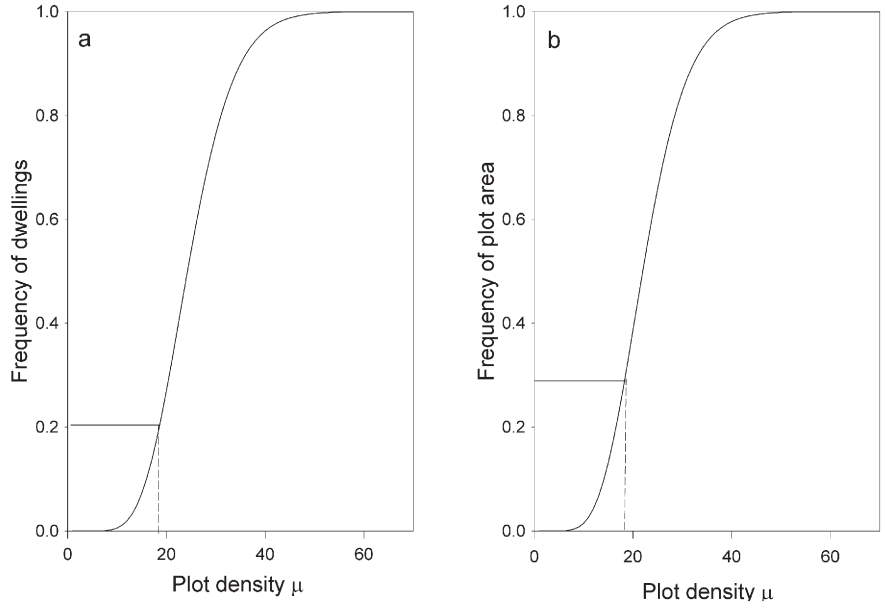


Fig. 4. b: Semi-detached dwellings $\bar{x} = 25$ dph, $\hat{k} = 11$ (built since 1990s).

The null hypothesis is that the observed data are drawn from the chosen theoretical distribution. If the discrepancy g_n exceeds the critical value then this is cause for rejection of the null hypothesis and implies that the theoretical distribution is not doing an adequate job of modeling the empirical density distribution.

3. Results of fitting the gamma distribution to the data

3.1. Results of the K–S test

Appendix B summarizes for each combination of dwelling type and independent variables the sample size, number of outliers, estimated gamma distribution parameters, and the significance of the K–S statistic. This shows that the gamma distribution was a good fit for almost all of the combinations and in most cases the null hypothesis could not be rejected. Of the total 108 fitted gamma distributions, the null hypothesis could be rejected at the 20% level in 47 cases; (i.e., with only 80% confidence that the sample was not drawn from the fitted theoretical gamma distribution) at the 10% level in 18 cases, and at the 5% level in 9 cases and at the 1% level in 17 cases. Hence there were only 17 remaining out of the 108 cases where there was 99% confidence that the dwellings were not drawn from a gamma distribution and these were spread relatively evenly across the dwelling types. It is surprising that the gamma distribution fits the empirical data so well given that plot sizes vary

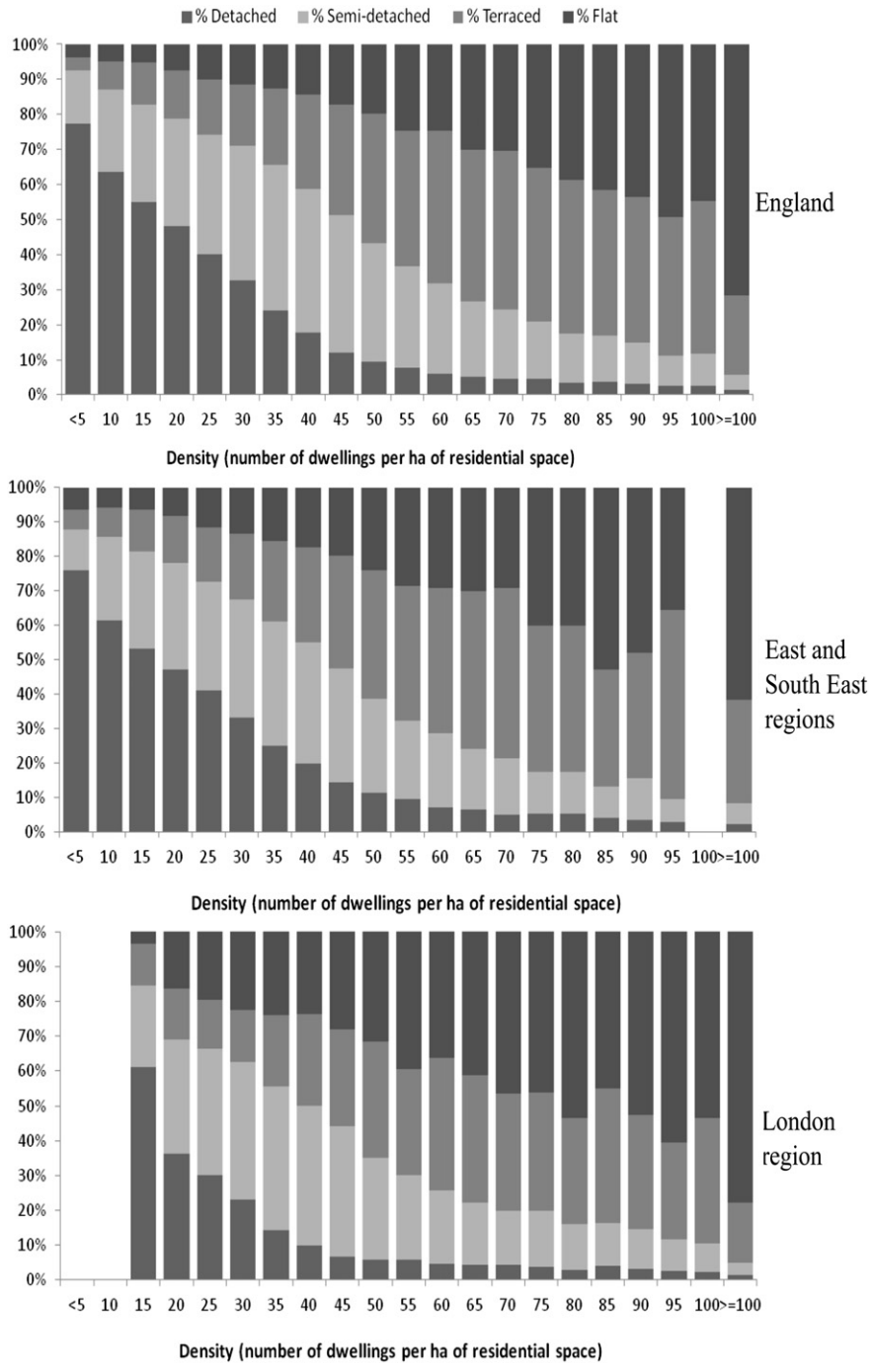


Fig. 5. Percentage of dwelling types per density band.

greatly between dwellings. This high success rate in fitting the gamma distribution and its convenient mathematical properties made it a suitable basis for representing the distribution of residential plot densities.

3.2. The calibrated shape parameters of the gamma distribution

The shape parameters of the fitted gamma distributions from Appendix B are plotted in Fig. 3 and show a clear relationship between the shape parameter k and the age band of the dwelling: The older the dwelling, the smaller the shape parameter. This indicates that newer dwellings have a more uniform division of plot size whereas older dwellings have a less equitable distribution of plot size.

The shape parameters for detached houses, bungalows and apartments are approximately half the size of those for semi-detached and terraced houses indicating that apartments and detached houses have a greater variability in plot size for a given mean density. Apartments are often built in areas where space is constrained and can vary greatly in density per plot, and similarly detached houses and bungalows can have large variability in plot size, whereas plot sizes for semi-detached and terraced houses tend to be more uniform. Table 2 summarizes these calibrated shape parameters $\hat{k}$ for the different dwelling types.

The analysis found no relationship between the shape parameter k per dwelling type and either the area types or the morphologies. However, the aspatial neighborhood types with samples large enough for analysis had a relatively narrow range of area types per dwelling type, for example apartments were in urban areas, whereas detached dwellings were in suburban and rural areas.

3.3. Estimating the gamma distribution from the mean density

For a given dwelling type and location, the mean density $\overline{x}$ is:

$$\overline{x} = \frac{H}{A}. \tag{12}$$

It can be intuitively deduced that $\overline{x}$ is the mode of the plot densities μ (i.e., the maximum value of the probability density function of μ).

The mode of the gamma distribution has the following property if $k > 1$:

$$\overline{x} = \theta(k-1). \tag{13}$$

Alternatively, this can be shown by integration of the probability density distribution $f(\mu; k, \theta)$. The number of dwellings cancels out and the denominator is the integral of the dwelling frequency distribution divided by plot density μ, as shown below;

$$\overline{x} = \frac{H}{A} = \frac{\int_0^{\infty} PDF(\mu;k,\theta)}{\int_{\mu=0}^{\infty} \frac{PDF(\mu;k,\theta)}{\mu}} = \frac{1}{\int_{\mu=0}^{\infty} \frac{PDF(\mu;k,\theta)}{\mu}}$$
$$= \frac{1}{\int_{\mu=0}^{\infty} \frac{1}{\theta^k \Gamma(k)} \mu^{k-2} e^{-\mu/\theta}}. \tag{14}$$

And this results in Eq. (13) above
Hence:

$$\theta = \frac{\overline{x}}{(k-1)}. \tag{15}$$

Substituting Eq. (8) into Eq. (15) gives the following new convenient relationship that will allow the mean plot density to be calculated from the mean area density:

$$\overline{\mu} = \overline{x} + \theta. \tag{16}$$

Hence, the scale parameter θ can firstly be estimated from Eq. (15) because $\overline{x}$ is forecast by an urban model such as LUISA and the shape parameter k has been empirically calibrated as $\hat{k}$ using the EHCS data. Therefore $\overline{\mu}$ can then be estimated from Eq. (16) and the theoretical gamma distribution can be fully specified as a calibrated function of plot density. Note that $\overline{\mu}$ and $\overline{x}$ are different density metrics; $\overline{x}$ is the conventional method of dwellings divided by the sum of the residential plot areas, whereas $\overline{\mu}$ is the mean value of the plot densities of the individual dwellings.

3.4. Deriving the distribution of the plot area per dwelling type

The plot area per dwelling is:

$$a = {}^{1}/_{\mu}. \tag{17}$$

Hence from the PDF of the dwellings, Eq. (5), the area per dwelling is:

$$f(\mu|\, k,\, \theta) = \frac{1}{\theta^k \Gamma(k)} \mu^{k-2} e^{-\mu/\theta} \quad \text{for } \mu > 0, \theta > 0, \quad \text{and } k > 2. \tag{18}$$

The mean plot area $\bar{a}_d$ per dwelling from Eq. (14) is:

$$\overline{a}_d = \int_{\mu=0}^{\infty} \frac{1}{\theta^k \Gamma(k)} \mu^{k-2} e^{-\mu/\theta} = \frac{1}{(k-1)\theta}. \tag{19}$$

Hence, the PDF of the plot area per dwelling over the plot density range μ is:

$$f(\mu|\, k,\, \theta) = \frac{a_d}{\overline{a}_d}$$
$$= \frac{(k-1)}{\theta^{k-1}\Gamma(k)} \mu^{k-2} e^{-\mu/\theta} \quad \text{for } \mu \text{ and } \theta > 0, \quad \text{and } k > 2. \tag{20}$$

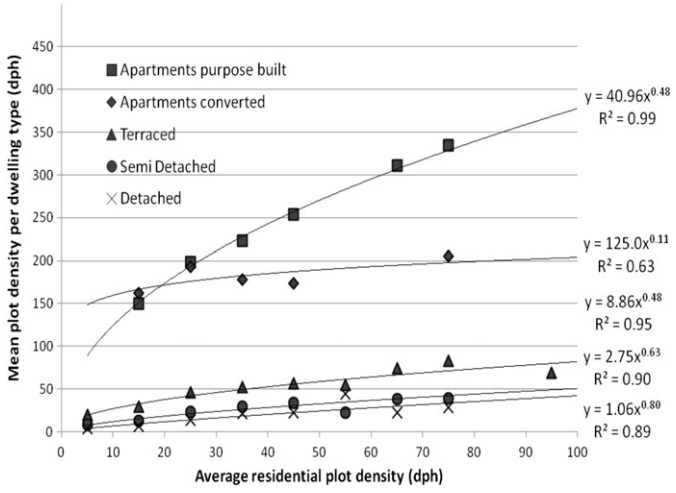


Fig. 6. Estimate of mean density per dwelling type from the average residential density.

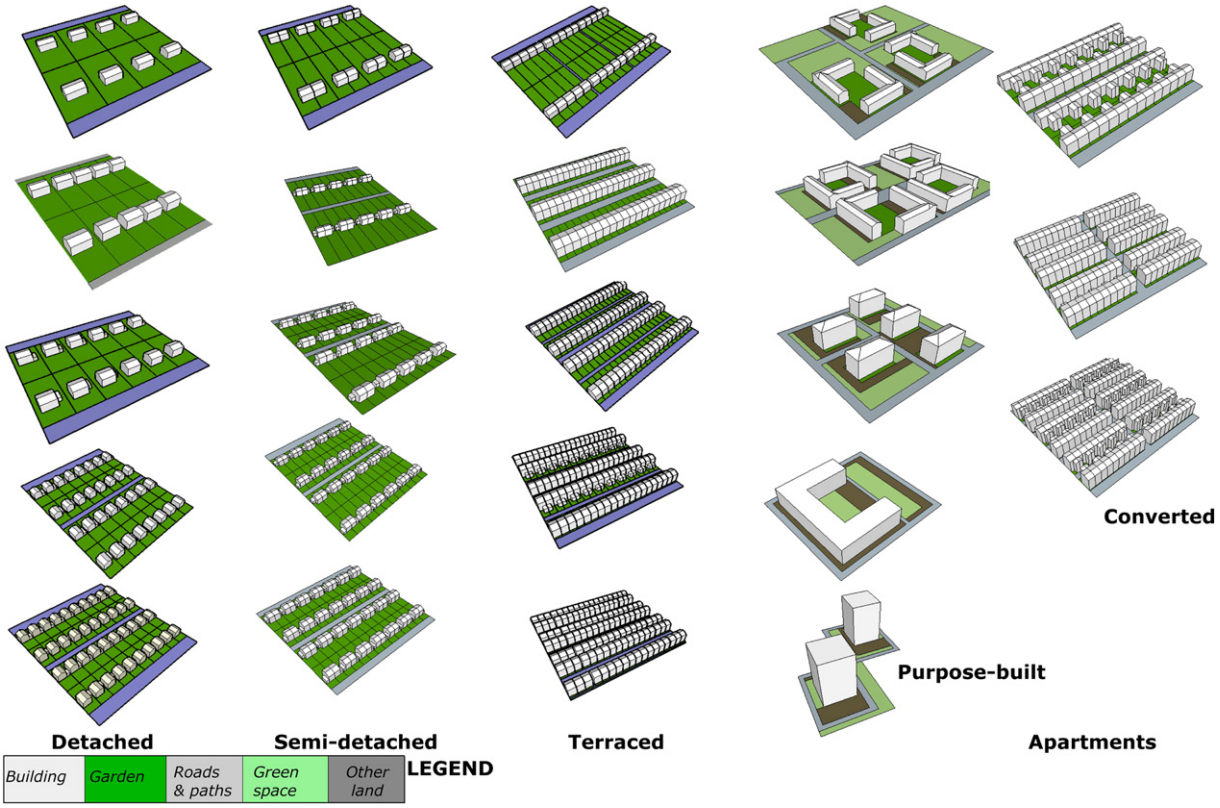


Fig. 7. Schematic illustration of the tiles.

The following integration produces the CDF of plot area per dwelling:

$$F(\mu|k,\theta) = \int_{x=0}^{x=\mu} \frac{(k-1)}{\theta^{k-1}\Gamma(k)} x^{k-2} e^{-x/\theta} \cdot d(x) = \frac{\mu^{k-1} e^{-\mu/\theta}}{\theta^{k-1}} \cdot \frac{\gamma(k,\mu/\theta)}{\Gamma(k)}. \quad (21)$$

It is interesting to compare this CDF of plot area (Eq. 21) with the CDF of dwelling frequency (Eq. 5). Fig. 4a shows this comparison for detached dwellings and Fig. 4b for semi-detached dwellings. Both are shown for the same mean plot density $\bar{x}$ but these dwelling types differ on their shape parameter $\hat{k}$ (Table 2). The 20% lowest density detached dwellings would occupy 36% of the plot area for detached dwellings whereas the 20% lowest density semidetached dwellings would occupy only 30% of the plot area. This illustrates how $\hat{k}$ can indicate the equity in plot sizes per dwelling type. Those with a smaller $\hat{k}$, such as detached dwellings and apartments, have a less equitable division of plot sizes than terraced and semi-detached dwellings.

4. Using discrete tiles to represent the calibrated density distribution

The next stage was to develop a method of using the preceding calibrated functions to convert the forecasts of average density per zone of an urban model, such as LUISA, into discrete representations of the dwellings and plot sizes.

4.1. Disaggregating dwellings into dwelling types and densities

The EHCS dwelling types were aggregated into the minimum number of distinct types. These were detached, semi-detached, terraced houses, purpose-built and converted flats/apartments (with bungalows divided between detached and semi-detached houses). This was done to reduce the number of tile types and amount of work involved in designing and modeling the tiles.

4.1.1. Estimate the proportions of each dwelling type per zone

The first step estimated the percentage of each dwelling type for a given mean area density $\bar{x}$ using a similar method to Mitchell, Hargreaves, Namdeo, and Echenique (2011). This combined the 2001 Census dwellings data with the residential land areas of the Generalized Land Use Database (GLUD) that is based on Ordnance Survey Mastermap™ (DCLG 2005). Fig. 5 shows how these percentages varied with plot density and the East and South East of England regions are similar to the average for England, whereas London has a greater proportion of apartments for each density band.

Note however that the GLUD data only classifies the dwelling footprints and domestic gardens as residential land whereas the EHCS data is based on a manual survey that measures the residential plot. There are therefore some disparities between GLUD and EHCS metrics especially for high density urban centers. For example, if a mixed-use building has a non-domestic unit on the ground floor then GLUD classifies the whole building as non-domestic. Nevertheless, the GLUD data is broadly consistent with the plot density metric for most of the case study area.

The relationships illustrated in Fig. 5 were then represented as empirical equations per region so that the forecast number of dwellings could be split into the number of dwellings by type H_{dj} based on the forecast mean residential density $\bar{x}_j$ per zone j.

4.1.2. Estimate the mean density per dwelling type from mean dwelling density

The next step estimated the mean density of each dwelling type from the overall mean density $\bar{x}$. This analysis aggregated the EHCS data by region, morphology and area type giving 216 possible 'aspatial' location types l ($9 \times 4 \times 6$). However, there were some invalid

combinations, such as rural morphologies in London region, and so only 164 'aspatial' location types had data. The estimated mean density per dwelling type d in aspatial location l is:

$$\bar{x}_{dl}{}' = \frac{\sum_{i=1}^{n} H_{dli}}{\sum_{i=1}^{n} A_{dli}} \tag{22}$$

where:

H_{dli} = *number of dwellings i of type d in aspatial location type l*
A_{dli} = *plot area of dwellings i of type d in aspatial location type l.*

The overall mean density of the aspatial location type l is:

$$\bar{x}_{l} = \frac{\sum_{d=1}^{n} H_{dl}}{\sum_{d=1}^{n} A_{dl}}. \tag{23}$$

Fig. 6 shows that there are clear relationships between the overall mean residential density $\bar{x}_l$ and the estimated mean residential density per dwelling type $\bar{x}_{dl}{}'$. These were represented as empirical equations to convert average density into density per dwelling type. (The correlation for converted apartments is not as good as the other dwellings because they are of very variable construction but these are a small proportion of total dwellings and so this makes little difference to the results.)

These density estimates per zone of the urban model were then proportionally adjusted so that the resulting residential area matches the land input constraints per zone j;

$$A_j = \sigma_j \sum_{d=1}^{n} {}^{H_{dj}}\!/_{\bar{x}_{dj}}{}' \tag{24}$$

$$\bar{x}_{dj} = \sigma_j{}^{-1} \bar{x}_{dj}{}' \tag{25}$$

where:

A_j = *input residential area to urban model for zone j*
σ_j = *estimated adjustment factor*
$\bar{x}_{dj}$ = *adjusted mean density for dwellings type d.*

4.2. The generic tiles

The tiles are a new innovative method of transforming the parametric distribution of plot densities into a discrete 3D representation of built form. They have been created primarily as a medium for multidisciplinary research on urban planning, buildings and decentralized environmental technologies. Fig. 7 shows examples of the tiles which are generic forms that range from low to high density for each dwelling type. Appendix C shows of two of the tiles in more detail.

The tiles were designed using the EHCS data on the dwelling dimensions, building fabric, floor space, occupancies and plot sizes (DCLG 2009).

4.3. Designing the plot density of each tile type

The method of selecting a set of tiles is conceptually equivalent to fitting a histogram to a gamma distribution as illustrated in Fig. 10. The frequency distribution of plot densities is represented by the numbers of tiles selected from a pre-designed set, which can include fractions of tiles. This is similar in principle to approximate integration where in this case the subinterval is defined by the plot density boundaries of the tile types. The following procedure is carried out for each location j and dwelling type d.

The gamma distribution for dwellings H_d in zone j was specified by inserting the mean density $\bar{x}$ and the appropriate value of $\hat{k}$ (Table 2) into Eqs. (15) & (16) to calculate μ and θ. The gamma distribution CDF $F(\mu|\hat{k},\theta)$ of the plot area frequency (Eq. 21) was then used to calculate the probability that the plot area is of tile type t as follows:

$$p_{at} = \text{CDF } f\left(\mu_t|\hat{k},\theta\right) - \text{CDF } f\left(\mu_{t-1}|\hat{k},\theta\right) \tag{26}$$

where:

p_{at} = *probability that plot area is of tile type t*
μ_t = *upper boundary of the plot density subinterval of tile type t.*

(Note that for t_{max} the probability of $\mu < \mu_t = 1$.)

Hence the total plot area of tiles of type t is:

$$a_t = A \cdot p_{at}. \tag{27}$$

The CDF $f(\mu|\hat{k},\theta)$ of dwelling frequency (Eq. 6) was used to calculate the probability that a dwelling is of tile type t as follows:

$$p_{ht} = \text{CDF } f\left(\mu_t|\hat{k},\theta\right) - \text{CDF } f\left(\mu_{t-1}|\hat{k},\theta\right) \tag{28}$$

where:

p_{ht} = *probability that a dwelling is of file type t.*

Hence, the number of dwellings of tile type t is:

$$h_t = H_d \cdot p_{ht}. \tag{29}$$

Table 3
Density of the tile boundaries.

Dwelling type	Tiles boundaries μ_t (dph)					
	Tile 1	Tile 2	Tile 3	Tile 4	Tile 5[a]	
D–Detached	0	13	18	24	33	100
S–Semi–detached	0	23	29	36	46	150
T–Terraced	0	52	65	81	106	200
A–Purpose built apartment	0	200	265	338	440	2000
C–Converted apartment	0	154	228	1000	na	na

[a] Each dwelling type was allocated an upper boundary to represent the highest feasible density of the dwelling type, to improve the reliability of the mean tile density calculation.

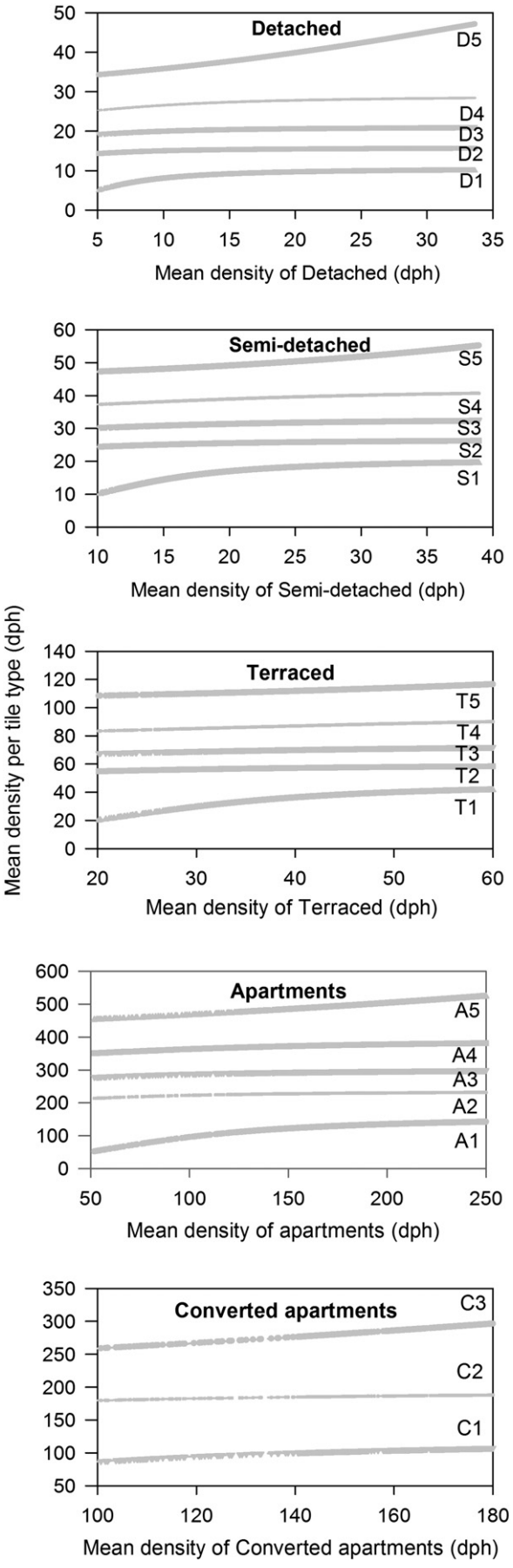


Fig. 8. Mean density per tile type plotted against the mean density of the dwelling type.

Hence, the mean density of the subinterval for tile type t is an output from the above empirically calibrated parametric functions:

$$x_t = \frac{h_t}{a_t}. \tag{30}$$

The accuracy of the tiles method depends on specifying a discrete tile density that is as close as possible to the mean density x_t of the tile subinterval. The sum of the tiles will then closely match the target for available plot area.

Values of x_t were calculated using GLUD Output Area data in the Wider South East of England case study area. The Output Areas (OAs) are the smallest areas at which the UK Office for National Statistics provides geographical data (around 125 households per OA). Using such small areas provided a rigorous test of the tiles method. The OAs were selected if they were within an urban boundary and had less than 2% non-domestic land (19,770 of the OAs in the WSE met these criteria). This ensured that most of land per OA was for residential use and thereby reduced the inaccuracies that mixed-use buildings cause to the GLUD estimates of residential land.

The boundaries of the tile subintervals were adjusted so that each tile type represented approximately the same proportion of dwellings for the case study area. The finalized boundaries are shown in Table 3.

$$p_{ht} \cong \frac{1}{n_t} \tag{31}$$

where:

n_t = *number of tile types.*

Fig. 8 shows the mean densities per tile type x_t versus the mean density of the dwelling type $\bar{x}_d$ per OA. It can be seen that x_t was relatively constant for the intermediate density tiles and was approximately the mid-point of the density subinterval and so each could be approximated by a discrete tile density.

However, the mean density was not constant for the lowest and highest density subintervals because there was a decreasing probability that the target density would be matched at these extremes by a mixture of tile types. The upper and lower subintervals represent the most extreme generic built form per dwelling type and the only way to achieve a more extreme density is to vary either its plot size or number of storeys. Table 4 shows the discrete tile densities selected to represent the mean density of each tile type.

Table 5 shows an example of using the tiles for semi-detached houses of mean density $\bar{x} = 30$ dph (based on an input plot area of 10 ha and forecast of 300 dwellings). If their average age is 1945 to 1974 then their calibrated shape parameter $\hat{k} = 8$ from Table 2. The scale parameter $\theta = 4.3$ from Eq. (13). Hence, the mean plot density is $\bar{\mu} = 34.3$ dph from Eq. (16), which thereby fully specifies the calibrated gamma distribution. Fig. 9 shows the CDF of the plot area per dwelling (Eq. 21) with the tile subintervals from Table 3. The tiles can be used in two alternative ways to convert the parametric distribution to the discrete tiles per dwelling type, depending on whether the aim is to exactly match actual dwelling forecast or the actual inputs on plot area. The method to match the number of dwellings uses the CDF of plot area, and the

Table 4
Plot densities of the 23 domestic tiles used to validate the tiles method.

Dwelling type	Plot density x_t per tile type (dph)				
	Tile 1	Tile 2	Tile 3	Tile 4	Tile 5
D—detached	8	15	21	28	45
S—Semi-detached	18	25	32	40	55
T—terraced	40	57	72	90	118
A—purpose built apartment	140	230	290	370	540
C—converted apartment	100	180	300	na	na

Table 5
Example of the tiles method for $\bar{x} = 30$ dph, $\hat{k} = 8$.

Dwelling type	Tiles (dph)					
	Tile 1	Tile 2	Tile 3	Tile 4	Tile 5	
Tile Boundaries μ_t (dph)	0	22	29	36	46	150
CDF $(\mu_t \mid \hat{k}, \theta)$ plot area	0.00	0.29	0.51	0.73	0.91	1.00
CDF $(\mu_t \mid \hat{k}, \theta)$ dwellings	0.00	0.17	0.36	0.60	0.84	1.00
x_t (from Table 4)	15	25	32	40	55	
p_{at} (eq. 26)	0.29	0.22	0.22	0.18	0.09	
p_{ht} (from eq. 28)	0.17	0.19	0.24	0.24	0.16	Total
Estimate of plot area A'_d	3.4	2.3	2.2	1.8	0.9	10.5 ha (+5%)
Estimate of dwellings H'_d	43.5	55	70.4	72	49.5	290.4 (−3%)

method to match the plot area uses CDF of dwellings (Eq. 6). In this example, the results are within 5% for plot area and −3% for dwellings which are typical for the method.

The plot area is estimated as:

$$A'_d = \sum_{t=1}^{t=n} \frac{H_d \cdot p_{ht}}{x_t}. \quad (32)$$

The dwellings are estimated as:

$$H'_d = \sum_{t=1}^{t=n} A_d \cdot p_{at} \cdot x_t. \quad (33)$$

Fig. 10 illustrates how the above example compares with the gamma distribution PDF of the dwellings (Eq. 5). The tile densities are shown as broken lines.

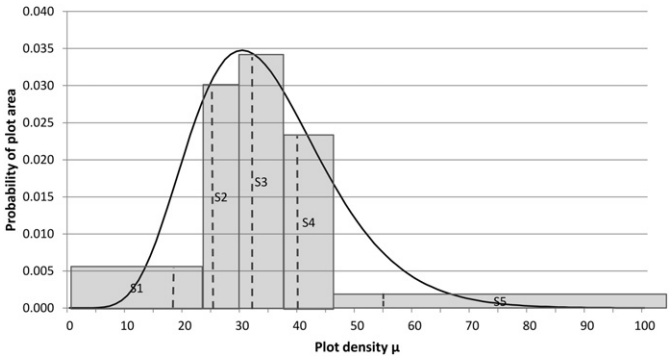


Fig. 10. Comparison between the tiles and the PDF for semi-detached ($\bar{x} = 30$ dph, $\hat{k} = 8$).

5. Validation results for the tiles method

5.1. Comparing the estimates of plot area using the tiles with the GLUD data

The total residential plot area was estimated using the tiles as:

$$A_{dt} = \sum_{d=0}^{d=n} \sum_{t=0}^{t=n} \frac{h_t}{x_t} \quad (34)$$

where:

A_{dt} = total residential plot area of dwelling type d.

Fig. 11 shows that the tiles give a very close estimation to the actual GLUD data on residential land per Output Area. The only exceptions are the Output Areas larger than around 10 ha, which are lower density areas that often have unusually large properties with outbuildings.

5.2. Comparing the regional distribution of generated tiles against empirical EHCS data

The next step of the validation process generated a set of tiles for the East of England and compared the results with the East of England EHCS housing stock data. The EHCS data was firstly disaggregated into 21 aspatial location types based on combinations of the 6 area types and 4 morphologies (3 of these combinations had no dwellings). The only inputs for the tile generation process were the average plot density and total number of dwellings of each aspatial location type. Everything else was then calculated using the previously described tiles method to generate a set of tiles for each of the 21 aspatial location types. Fig. 12 compares the empirical plot density distributions of the EHCS data for

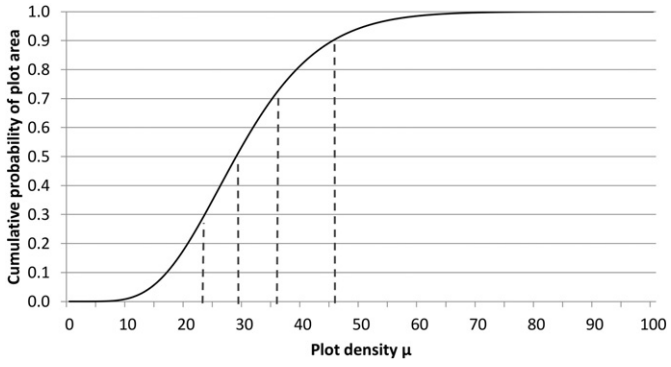


Fig. 9. Example of a CDF of plot area per dwelling for semi-detached ($\bar{x} = 30$ dph, $\hat{k} = 8$).

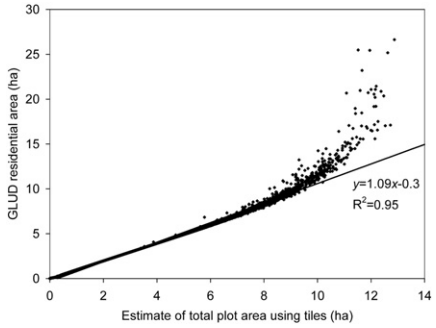


Fig. 11. Comparison of the total plot area of the tiles vs. GLUD residential area.

the East of England with the plot density distribution from the tiles and shows that the tiles method generates a realistic distribution of plot density per dwelling type.

5.3. Comparing the land areas of the generated tiles against Census Output Area data

The tiles were then compared with GLUD dwelling footprint data per Census Output Area. This provided an independent validation because GLUD dwelling footprint data were not used for either the tile design or tile generation process. The dwelling footprint areas of the tiles were summed per Output Area and compared with the GLUD residential footprint area. GLUD was found to consistently overestimate the footprint area by around 4 m^2 which was found to be due to the inclusion of outbuildings, such as garages and garden sheds. The validation process therefore deducted 4 m^2 per dwelling from the GLUD residential footprint data before making the comparison with the estimates from the tiles.

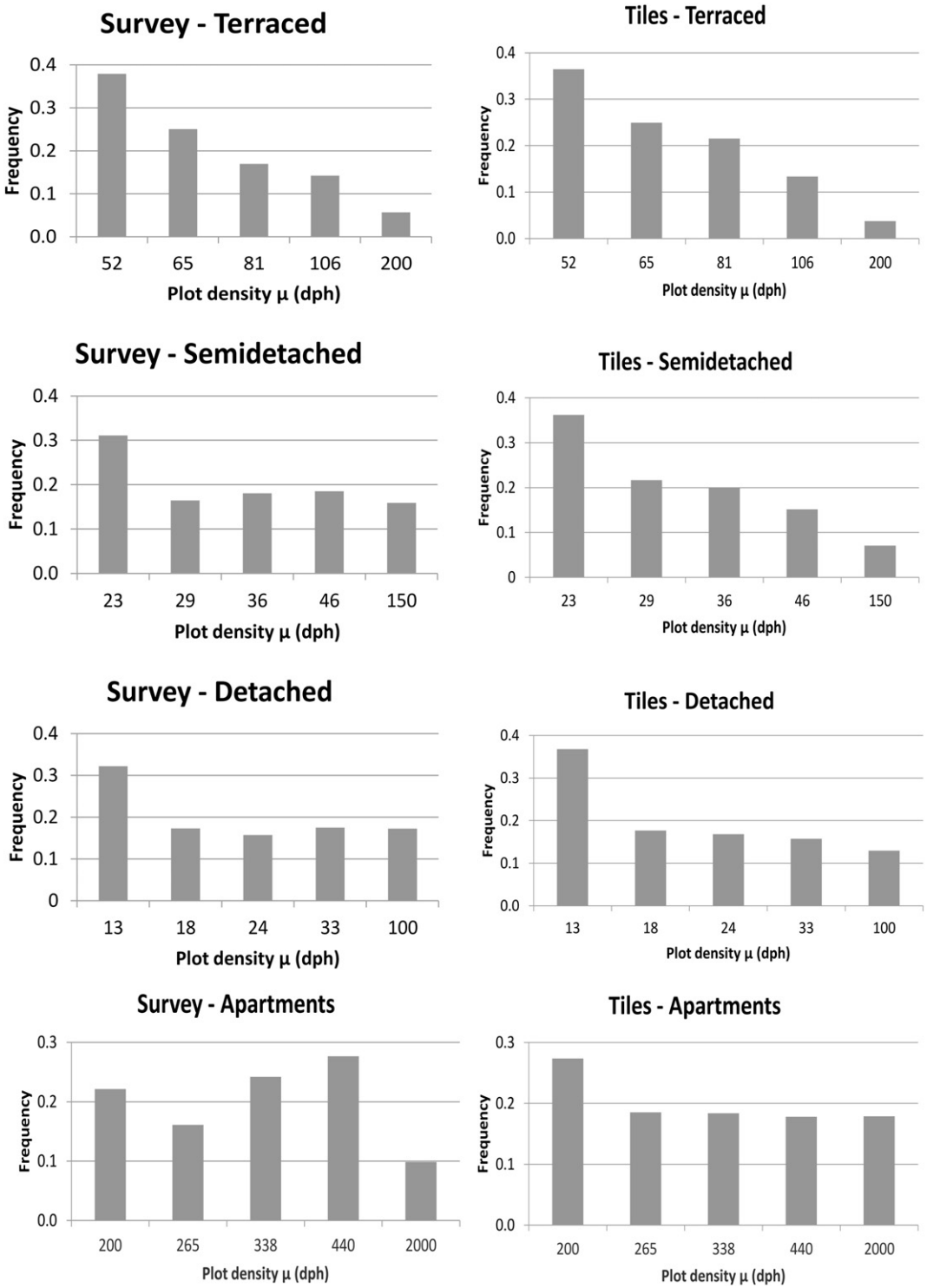


Fig. 12. Results of the validation for the East of England region.

The validation was carried out in 4 stages shown in Fig. 13:

a) Using only the total residential plot area and total number of dwellings per Output Area as the input to generate the tiles. The footprint areas were estimated based on the average footprint per dwelling type from the EHCS data.
b) Same as (a) above but using data on the dwelling type percentages instead of an estimate — this shows that using the actual percentages, if available, makes little difference to the correlation.
c) Same as (a) above but using the actual footprint areas per tile type — this shows how distinguishing between sizes of dwellings of the same type greatly improves the correlation even when only using an estimate of dwelling percentages.
d) Same as (c) above but using the actual percentage per dwelling type — this gives a further improvement in the correlation and $R^2 = 0.82$ is surprisingly good for these small Output Areas.

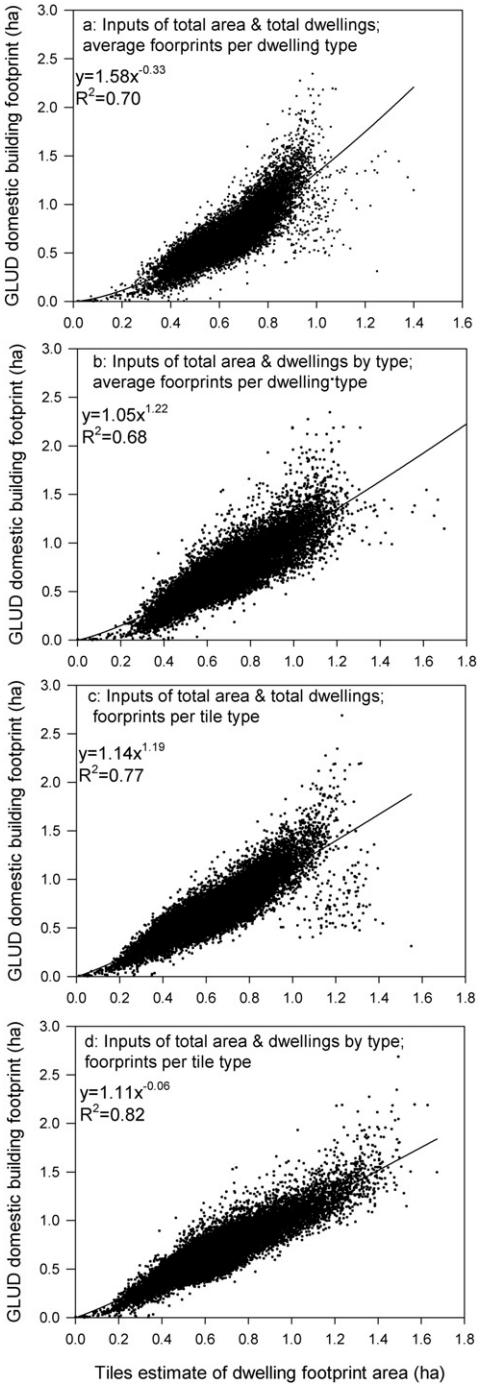


Fig. 13. Results of the validation using the GLUD area of residential building footprints per Output Area.

This shows that the tiles method produces a much more accurate estimate of dwelling footprint areas than using average footprint per dwelling type. This is reassuring because footprint areas are a proxy for floor space and roof space, which are important for investigating energy demands and the potential of sustainable technologies such as solar energy and water harvesting.

6. Using the tiles for integrating regional scale and building scale modeling

The residential land areas in the LUISA regional forecasting model were based on GLUD residential land which as explained earlier, has broadly consistent metrics with EHCS and so the average densities per zone of the LUISA model could be converted directly into plot densities.

However urban planners normally measure the densities based on total residential area, which may include pathways, parking, communal space and roads. These additional areas were therefore subsequently added to the design of the tiles so that the number of tiles selected equals the total residential area in hectares. This resulted in two alternative measures of density; the plot density per tile x_t that was used in the preceding method to calculate the number of dwellings per tile type; and the residential density z_t shown in Table 6 that was used to convert these dwellings into the number of one-hectare tiles;

$$z_t = \frac{a_t . x_t}{(a_t + r_t)} \tag{35}$$

where:

z_t = tile residential density (dph)
a_t = *plot area per dwelling of file type t* (m^2)
r_t = *the additional residential land per dwelling tile type t* (m^2)

$$n_t = \frac{h_t}{z_t} \tag{36}$$

where:

n_t = *number of tiles of type t* (*each is one hectare*).

One hectare was chosen as the tile size partly for convenience of accounting so that the number of tiles equals the residential area, but also because if they were any smaller it would be difficult to visually illustrate its typical housing layout. The additional areas such as roads, paths, green space and residential parking were estimated from Google Satellite maps and Ordnance Survey mapping. Fig. 14 shows how the percentage of these different residential land types varies with the mean plot density x_t. At low densities x_t is similar to z_t because the

Table 6
Tile densities of the 23 domestic tiles.

Dwelling type	Plot density z_t per tile type (dph)				
	Tile 1	Tile 2	Tile 3	Tile 4	Tile 5
D—detached	7	12	17	22	34
S—semi-detached	15	20	25	30	40
T—terraced	29	39	47	57	70
A—purpose built apartment	67	94	110	130	167
C—converted apartment	69	122	203	na	na

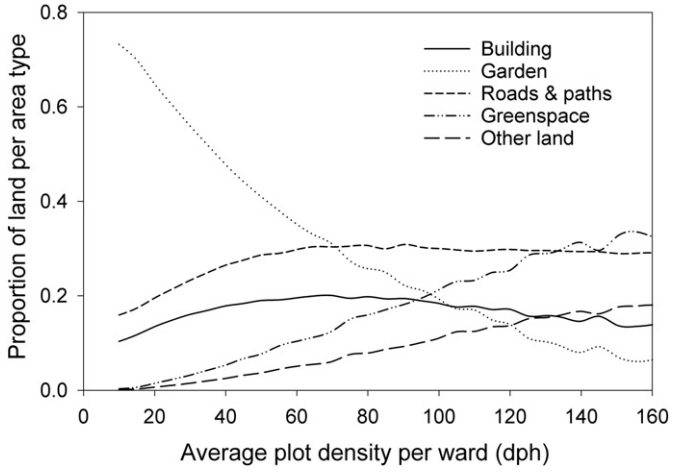


Fig. 14. The average proportion of residential land versus the average plot density $\bar{x}$ of the tiles.

buildings and gardens account for most of the tile area but as the density increases the rest of the residential land such as roads, paths, green space and other land become a larger percentage of the tile area, and so x_t becomes much larger than z_t.

A dataset was produced per tile of consumption, emissions and costs using building-scale models for energy, water and waste for combinations of scenario-specific variables, such as technology scenario and uptake; climate; development type (existing area, redevelopment or new land); area type (central, urban, suburban or rural), occupancy characteristics and whether the dwellings were as existing, retrofitted or new build. The tiles were generated per zone to match each scenario forecast of the regional scale model and then the tile data was aggregated.

It would be very useful if this tiles method could be further developed to directly estimate the distribution of residential density rather than plot density. To investigate this possibility Fig. 15 compares the mean plot density $\bar{x}$ with the mean tile density $\bar{z}$ for the main dwelling types. This relationship is almost linear especially for the detached and semi-detached dwellings. It is not quite so linear for terraced dwellings and apartments but this is probably due their greater range of dwelling types and if disaggregated into end-terrace and mid-terrace and low-rise and high-rise apartments then this may increase their linearity.

This broadly linear relationship suggests that the plot densities could be transformed into tile densities and still fit a gamma distribution. If the shape parameter k was then recalibrated it may be possible to estimate the distribution of residential density z directly from $\bar{z}$ and generate larger tiles as generic neighborhood types. This would be helpful for exploring the interactions between bottom-up urban design and top-down socio-economic modeling of city regions, but is beyond the scope of this current paper.

7. Discussion and conclusions

This paper has presented a new innovative method of analyzing housing survey data to explore the variability of dwellings plot sizes. Housing development in England is largely commercially driven but subject to planning constraints and so land is relatively expensive. The provision and adaptation of the housing stock is therefore responsive

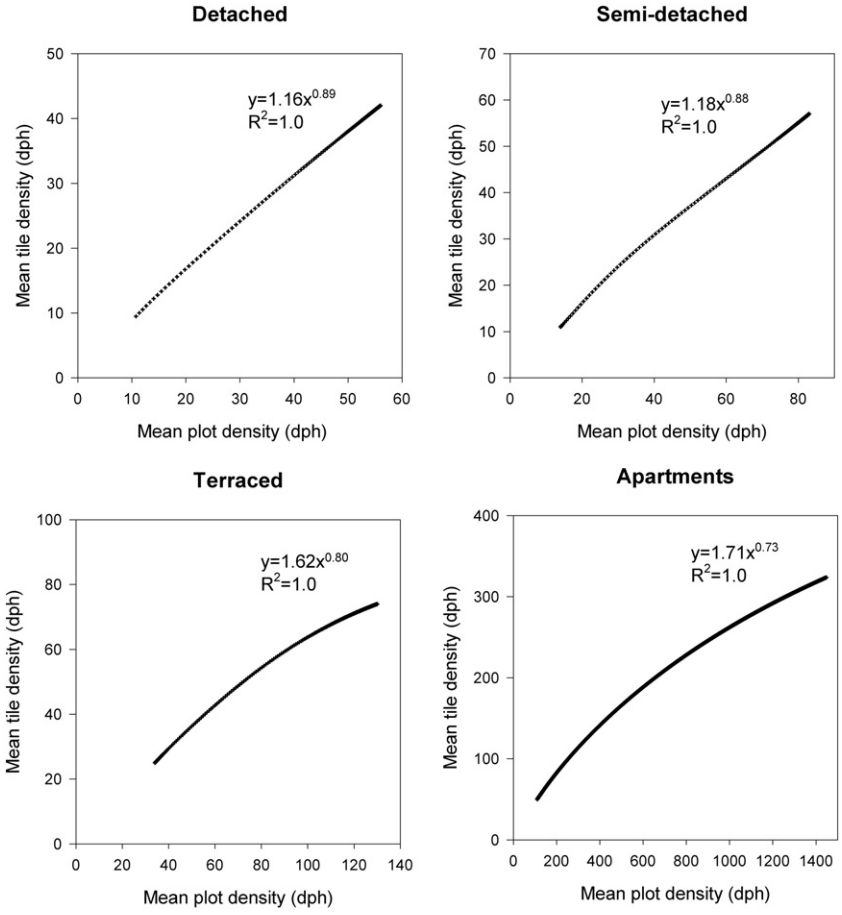


Fig. 15. Comparison of mean plot density and mean tile density for each main dwelling type.

to buyer demands and may partly explain why the distributions of plot densities of each dwelling type have a relatively consistent shape that can be approximated by the gamma distribution.

Older dwellings were found to have a greater variability in plot size than newer dwellings and the following possible reasons would need further investigation. Dwellings may have become more uniform over time due to a tightening of planning regulations and housing standards, whereas older dwellings are more likely to be the result of property conversion and thereby become more diverse. It is also likely that the variability of plot size per dwelling type is correlated with variability in household income. Greater mobility from widespread car ownership has allowed more social segregation between area types whereas previously neighborhoods had a wider mix of income levels and consequently a wider variation in plot sizes per dwelling type.

In reality the density distribution of dwellings does not necessarily follow the gamma distribution on a location specific basis because the built environment is so variable. However, the tile estimates become very similar to the data when aggregated over larger areas. Some alternative distributions were tried but none outperformed the gamma distribution on providing a consistently good fit to the data. The method could in principle be used for the parametric simulation of dwellings and plot sizes but it is more practical for multidisciplinary research to convert the density distributions into discrete predesigned 3D tiles.

This tiles method is a useful extension to urban forecasting models by allowing the average density forecast per zone to be converted into a representation of the dwelling stock and residential land. This captures the variability in garden size, roof areas and floorspace that is needed to estimate the likely uptake and performance of decentralized sustainable technologies for energy, water and waste management and provides more accurate estimates than could be achieved using more conventional methods such as dwelling typologies, mean densities, and floor area ratios. The tiles also reflect qualitative aspects, such as garden size, number of party walls and storeys, which may be useful for modeling housing choice. Also, their data on building heights and land cover could be a useful input to the forecasting of urban climate and flood risk.

Adding more tile types would increase the accuracy of estimating the dwelling stock. However, this needs to be balanced against the extra work of designing each tile and modeling its building efficiency, demand and supply characteristics. Some extra tiles would need to be added to more accurately represent mixed-use buildings in urban centers such as inner London. The tiles method is being adapted to non-domestic buildings and so it may eventually be possible to combine the domestic, non-domestic and mixed-use tiles within a single estimation process.

Further more detailed validation would be useful by land parcel size and if possible by dwelling age bands. The next steps are to adapt the method to other countries that have similar housing survey data and also to explore the possibility of generating larger generic tiles of residential areas.

Acknowledgments

The author is grateful for the assistance of Dr Vicky Cheng, Technical University of Munich who designed the dwelling dimensions of each tile type to match its plot density. The research was carried out as part of the ReVISIONS (*Regional Visions of Integrated Sustainable Infrastructure Optimized for NeighborhoodS*) research project funded by the UK Engineering and Physical Sciences Research Council (EPSRC) as part of the Sustainable Urban Environments program, grant reference number EP/F007566/1. The English House Condition Survey data was provided by the UK Department of Communities and Local Government. Mastermap™ was provided for academic use by Ordnance Survey.

Appendix A. English House Condition Survey variables used for the analysis

Details of how these were measured can be found in the EHCS Surveyors' Briefing Manual (DCLG 2007).

Table A1
Dwelling types.

No.	Dwelling type
1	End terrace
2	Mid terrace
3	Semi-detached
4	Detached
5	Bungalow
6	Converted flat/apartment
7	Purpose built apartment, low rise
8	Purpose built apartment, high rise

Table A2
Morphology.

No.	Rural urban morphology (COA)
1	Urban >10 k
2	Town and fringe
3	Village
4	Hamlet & isolated dwellings

Table A3
Area types.

No.	Nature of area
1	Urban—commercial city/town center
2	Urban—urban
3	Urban—suburban residential
4	Rural—rural residential
5	Rural—village center
6	Rural—rural

Table A4
Definition of the area type in the EHCS Surveyor Briefing Manual.

No.	Nature of an area
1	Commercial city/town center — this is the area that would constitute part/all of the center of a city or town. Areas do not have to be run down to be coded as city or town center. It is likely that these areas will have a high percentage of commercial properties such as shops and businesses.
2	Urban — this is the area around the core of towns and cities, and also older urban areas which have been swallowed up by a metropolis. Areas would be largely but not exclusively residential.
3	Suburban residential — this is the outer area of towns or cities, and would include large, planned housing estates on the outskirts of towns or larger areas of older residential stock.
4	Rural residential — these can be free standing residential areas or suburban areas of villages, often meeting the housing needs of people who work in nearby towns and cities.
5	Village center — these are traditional English villages or the old heart of villages which have been suburbanized.
6	Rural — these areas are predominantly rural e.g. agricultural with isolated dwellings or small hamlets.

Table A5
Regions.

No.	Government office region
1	North East
2	Yorkshire and The Humber
3	North West
4	East Midlands
5	West Midlands
6	South West
7	East of England
8	South East
9	London

Table A6
Age bands.

No.	
1	Pre-1850
2	1850–1899
3	1900–1918
4	1919–1944
5	1945–1964
6	1965–1974
7	1975–1980
8	1981–1990
9	Post-1990

Appendix B. Shape parameters *k* and goodness of fit tests for the gamma distribution

Table B1
Results for semidetached and detached houses.

Dwelling type	Age band	Area type	Morphology	EHCS sample size *n*	EHCS dwellings N	No. of outliers	Mean plot density $\bar{x}$	Shape parameter *k*	K–S max (g_n)	K–S min (g_n)	Significance
Detached	1	6	3	27	45,606	4334	9	2.6	0.142	−0.036	****
	1	6	4	56	101,142	12,491	8	1.8	0.133	−0.031	*
	2	3	1	27	50,974	2985	34	1.6	0.158	−0.067	****
	3	3	1	24	47,069	6775	18	4.5	0.090	−0.035	****
	4	3	1	139	262,408	1879	20	3.6	0.040	−0.077	**
	5	3	1	133	251,149	11,838	20	6.3	0.044	−0.031	****
	5	4	1	26	46,112	0	13	3.4	0.075	−0.084	****
	6	3	1	166	321,590	12,056	27	5.6	0.036	−0.078	*
	6	4	2	38	67,548	0	26	3.7	0.036	−0.101	****
	6	4	3	29	56,307	4330	22	4.7	0.162	−0.087	***
	7	3	1	96	178,582	6093	27	5.3	0.051	−0.085	***
	8	3	1	187	356,793	10,625	31	7.7	0.026	−0.047	****
	9	3	1	237	578,965	17,441	33	7.0	0.041	−0.060	*
	9	3	2	25	57,555	0	30	5.1	0.107	−0.150	****
	9	4	2	30	75,292	3810	32	4.0	0.100	−0.082	****
Semi-detached	2	2	1	29	56,742	6074	44	4.4	0.152	−0.031	***
	2	3	1	58	115,281	5623	43	3.0	0.07	−0.051	****
	3	2	1	33	58,944	2710	57	3.6	0.077	−0.093	****
	3	3	1	86	161,551	10,223	47	3.8	0.058	−0.054	****
	4	2	1	186	263,098	3679	41	7.1	0.057	−0.028	***
	4	3	1	852	1,342,522	35,901	38	7.1	0.031	−0.002	**
	4	3	2	36	46,647	2275	30	7.5	0.071	−0.067	****
	4	4	1	26	42,479	0	36	9.3	0.11	−0.083	****
	4	4	2	28	39,509	2279	27	5.2	0.141	−0.133	****
	4	4	3	36	40,460	315	24	6.4	0.043	−0.078	****
	5	2	1	128	150,172	5318	41	6.4	0.104	−0.094	
	5	3	1	916	1,205,354	25,811	38	8.4	0.041	−0.009	
	5	3	2	74	88,101	4692	35	7.3	0.109	−0.023	*
	5	4	1	45	58,866	2528	31	7.9	0.11	−0.057	****
	5	4	2	67	87,408	358	35	6.4	0.084	−0.146	
	5	4	3	55	65,707	0	31	6.4	0.077	−0.107	****
	6	2	1	34	44,638	1300	50	6.9	0.138	−0.016	***
	6	3	1	295	534,875	23,377	42	9.7	0.045	−0.037	****
	6	3	2	37	56,288	0	46	6.5	0.114	−0.033	****
	6	4	1	17	32,349	4620	35	8.7	0.138	−0.053	****
	6	4	2	31	55,169	0	39	7.3	0.098	−0.085	****
	7	3	1	104	179,430	8943	45	11.4	0.075	−0.051	****
	8	3	1	102	166,062	6240	53	9.6	0.051	−0.045	****
	9	2	1	34	47,492	758	62	8.2	0.120	−0.029	****
	9	3	1	123	234,160	18,194	55	9.0	0.066	−0.059	****

Key to Significance: [g_n] is within the K–S test critical value that the dwelling sample is drawn from the fitted gamma distribution.
**** 20% level.
*** 10% level.
** 5% level.
* 1% level.

Table B2
Results for terraced houses.

Dwelling type	Age band	Area type	Morphology	EHCS sample n	EHCS dwellings N	No. of outliers	Mean plot density $\bar{x}$	Shape parameter k	K–S max (g_n)	K–S min (g_n)	Significance
Mid-terraced	2	2	1	105	300,098	17,624	144	6.0	0.069	−0.011	****
	2	3	1	70	197,099	8203	142	5.5	0.108	−0.035	**
	3	2	1	97	76,506	6484	148	6.7	0.144	−0.078	
	3	3	1	80	138,748	8336	132	7.2	0.08	−0.096	**
	4	2	1	80	151,658	2511	65	5.0	0.099	−0.061	**
	4	3	1	248	164,517	5781	62	8.1	0.018	−0.074	
	5	2	1	41	130,931	2238	67	5.7	0.108	−0.117	****
	5	3	1	208	145,107	3056	57	7.8	0.039	−0.056	***
	6	3	1	102	121,261	2519	81	12.3	0.06	−0.034	****
	7	3	1	49	113,266	1452	73	8.6	0.094	−0.137	*
	8	3	1	109	70,137	6476	96	8.9	0.051	−0.076	***
	9	2	1	29	43,095	0	92	11.8	0.123	−0.015	****
	9	3	1	79	44,479	4162	92	10.0	0.018	−0.085	****
End-terraced	2	2	1	33	441,549	15,124	419	6.6	0.105	−0.027	****
	2	3	1	46	386,688	5098	395	5.8	0.091	−0.010	****
	3	2	1	28	344,375	8464	595	5.0	0.105	−0.062	***
	3	3	1	80	144,663	11,340	560	5.7	0.12	−0.029	*
	4	2	1	97	139,674	13,714	406	9.0	0.094	−0.051	**
	4	3	1	43	170,816	10,952	591	5.9	0.143	−0.035	*
	5	2	1	148	318,771	8562	446	8.5	0.06	−0.050	
	5	3	1	355	380,476	16,440	323	7.8	0.075	−0.028	
	5	3	2	32	359,792	9269	717	6.8	0.086	−0.080	***
	6	2	1	166	223,977	4161	428	5.5	0.079	−0.055	*
	6	3	1	361	64,047	2151	412	9.7	0.062	−0.020	
	7	2	1	25	90,211	1511	234	9.9	0.215	−0.072	*
	7	3	1	118	65,685	1739	452	11.0	0.048	−0.040	
	8	2	1	211	25,744	552	391	6.8	0.049	−0.100	
	8	3	1	24	30,897	0	558	13.0	0.057	−0.090	****
	9	2	1	96	45,959	4086	489	10.3	0.061	−0.067	****
	9	3	1	187	33,091	0	440	12.5	0.035	−0.022	***

Key: see Table B1.

Table B3
Results for bungalows and purpose built and converted apartments.

Dwelling type	Age band	Area type	Morphology	EHCS sample n	EHCS dwellings N	No. of outliers	Mean plot density $\bar{x}$	Shape parameter k	K–S max (g_n)	K–S min (g_n)	Significance
Bungalows	5	2	1	33	30,841	–	53	2.3	0.138	−0.054	***
	5	3	1	267	332,343	13,688	34	6.1	0.071	−0.065	*
	5	3	2	24	23,289	3870	37	2.7	0.07	−0.06	*
	5	4	3	34	34,124	1075	25	3.8	0.188	−0.095	*
	6	3	1	182	233,786	–	37	4.0	0.127	−0.048	
	6	4	2	35	44,560	551	32	3.4	0.103	−0.117	****
	6	4	3	35	49,117	–	27	6.9	0.131	−0.151	***
	7	3	1	71	84,370	9524	34	6.3	0.086	−0.115	***
	8	3	1	76	93,262	6430	44	3.5	0.1	−0.078	**
Low rise apartments	2	2	1	33	276,504	7862	419	2.7	0.105	−0.076	****
	3	2	1	46	329,254	5306	395	2.1	0.091	−0.033	****
	4	1	1	28	181,938	10,990	595	5.0	0.105	−0.077	***
	4	2	1	80	87,963	1912	560	1.9	0.12	−0.085	*
	4	3	1	97	195,456	10,080	406	1.8	0.094	−0.016	**
	5	1	1	43	219,146	5301	591	4.7	0.143	−0.096	*
	5	2	1	148	165,342	15,573	446	3.4	0.06	−0.114	
	5	3	1	355	139,548	9026	323	3.4	0.075	−0.139	
	6	1	1	32	109,026	6401	717	3.7	0.086	−0.152	***

(continued on next page)

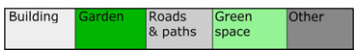
Table B3 (*continued*)

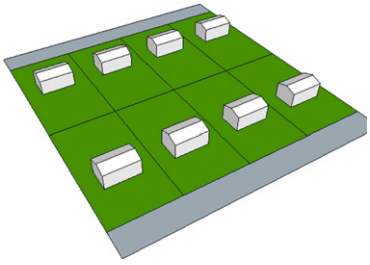
Dwelling type	Age band	Area type	Morphology	EHCS sample n	EHCS dwellings N	No. of outliers	Mean plot density $\bar{x}$	Shape parameter k	K–S max (g_n)	K–S min (g_n)	Significance
	6	2	1	166	124,540	6945	428	4.0	0.079	−0.063	*
	6	3	1	361	115,324	1054	412	3.6	0.062	−0.078	
	6	4	2	25	122,048	8113	234	2.9	0.215	−0.000	*
	7	2	1	118	67,856	5937	452	5.1	0.048	−0.126	
	7	3	1	211	44,433	0	391	5.1	0.049	−0.103	
	8	1	1	24	45,327	1411	558	5.3	0.057	−0.082	****
	8	2	1	96	34,126	2508	489	4.8	0.061	−0.036	****
	8	3	1	187	25,936	0	440	4.6	0.035	−0.071	***
	9	1	1	26	35,540	1679	487	5.2	0.146	−0.045	****
	9	2	1	82	23,211	1154	513	4.8	0.065	−0.069	****
	9	3	1	117	27,647	602	427	6.9	0.044	−0.028	****
High rise	5	2	1	74	73,664	442	1666	2.8	0.099	−0.056	***
apartm'ts	5	3	1	35	58,872	4357	1416	4.9	0.158	−0.005	**
	6	2	1	92	34,463	0	1676	3.9	0.065	−0.097	*
	6	3	1	51	28,291	670	1733	3.8	0.18	−0.027	
Converted	2	1	1	37	202,881	0	144	2.7	0.056	−0.136	***
apartments	2	2	1	143	87,718	0	135	3.1	0.045	−0.061	****
	2	3	1	50	80,629	0	107	2.9	0.048	−0.068	****
	3	2	1	59	77,015	0	104	3.2	0.086	−0.059	****
	3	3	1	54	48,241	0	111	2.8	0.037	−0.044	****

Key: see Table B1.

Appendix C. Examples of tile dimensions

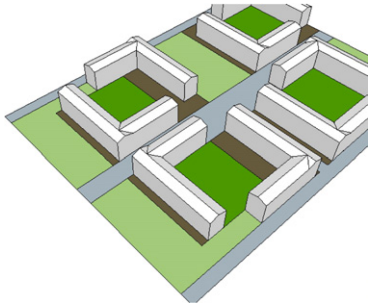
Table C1
Tile D1 — detached houses.





Density in dwellings per hectare (dph)			
Plot density per dwelling μ		8	
Tile density z		7	
Land use (%)			
Domestic building footprints		8	
Domestic gardens		78	
Roads and paths		14	
Green space		0	
Other land (such as parking areas)		0	
Basic dimensions per dwelling			
Building footprint (m^2)		117	
Floor area (m^2)		234	
Building height (m)		6 (2 storeys)	
Wall area (m^2)	Front: 81	Side: 59	
Roofs (m^2)	Top floor: 117	Bottom floor: 0	Pitched roof: 100%
Garden (m^2) 1131	Front: 382 (60% soft)	Rear: 633 (80% soft)	Side: 116 (50% soft)

Table C2
Tile A2 — purpose built apartments.



Density in dwellings per hectare (dph)			
Plot density per dwelling μ		230	
Tile density *z*		94.2	
Land use (%)			
Domestic building footprints		20	
Domestic gardens		21	
Roads and paths		26	
Green space		23	
Other land (such as parking areas)		10	
Basic dimensions per apartment block			
Building footprint (m^2)		679	
Number of apartments		32	
Floor area per apartment (m^2)		68	
Floor area per block (m^2)		2168	
Building height (m)		12 (4 storeys)	
Wall areas (m^2)	Main facade: 566	Minor facade: 457	
Roofs (m^2)	Top floor: 679	Pitched roof: 90%	Flat roof: 10%
Garden (m^2)	713	Courtyard	65% soft (35% hard)

Each tile type shows a typical arrangement of the dwelling type and its associated residential area for the tile density but the layouts may vary.

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*, 114–123.

Ballarini, I., Corgnati, S. P., & Corrado, V. (2014). Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project. *Energy Policy, 68*, 273–284.

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.

Crompton, A. (2012). The entropy of LEGO. *Environment and Planning B: Planning and Design, 39*, 174–182.

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

DCLG (2013). *English housing survey*. Department of Communities and Local Government (https://www.gov.uk/government/organisations/department-for-communities-and-local-government/series/english-housing-survey last accessed 11th October 2013).

DCLG (2009). *English House Condition Survey 2007: Annual report*. London, UK: Department of Communities and Local Government Publications.

DCLG (2007). *English House Condition Survey surveyor briefing manual: Part 1 completing the survey form*. London, UK: Department of Communities and Local Government.

DCLG (2005). Generalised land use database statistics for England 2005; Feb. 2007, UK. http://webarchive.nationalarchives.gov.uk/20120919132719/http:/communities.gov.uk/documents/planningandbuilding/pdf/154941.pdf (last accessed 06/07/2015)

DECC (2012, Sept 20th). *Renewable Heat Incentive: Consultation on proposals for a domestic scheme*. London, UK: Department of Energy & Climate Change.

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: Planning and Design, 40*(6), 1003–1026.

Filogamo, L., Peri, G., Rizzo, G., & Giaccone, A. (2014). On the classification of large residential buildings stocks by sample typologies for energy planning purposes. *Applied Energy, 135*, 825–835.

Firth, S. K., Lomas, K. J., & Wright, A. J. (2010). Targeting household energy efficiency measures using sensitivity analysis. *Building Research & Information, 38*(1), 25–41.

Girardin, L., Marechal, F., Dubuis, M., Calame-Darbellay, N., & Favrat, D. (2010). EnerGis: A geographical information based system for the evaluation of integrated energy conversion systems in urban areas. *Energy, 35*(2), 830–840.

Hofierka, J., & Kanuk, J. (2009). Assessment of photovoltaic potential in urban areas using open-source solar radiation tools. *Renewable Energy, 34*(10), 2206–2214.

Husak, G. J., Michaelsen, J., & Funk, C. (2006). Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. *International Journal of Climatology, 27*(7), 935–944.

Ison, N. T., Feyerherm, A. M., & Dean, Bark L. (1971). Wet period precipitation and the gamma distribution. *Journal of Applied Meteorology, 10*, 658–665.

Jakubiec, J. A., & Reinhart, C. F. (2013). A method for predicting city-wide electricity gains from photovoltaic panels based on LiDAR and GIS data combined with hourly Daysim simulations. *Solar Energy, 93*, 127–143.

Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., & Djurovic-Petrovic, M. (2010). A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment, 45*, 1683–1697.

Kohler, N., & Hassler, U. (2002). The building stock as a research object. *Building Research & Information, 30*(4), 226–236.

Lukac, N., Zlaus, D., Seme, S., Zalik, B., & Stumberger, G. (2013). Rating of roofs' surfaces regarding their solar potential and suitability for PV systems, based on LiDAR data. *Applied Energy, 102*, 803–812.

Makropoulos, C. K., Natsis, K., Liu, S., Mittas, K., & Butler, D. (2008). Decision support for sustainable option selection in integrated urban water management. *Environmental Modelling and Software, 23*, 1448–1460.

Mata, É., Sasic Kalagasidis, A., & Johnsson, F. (2014). Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. *Building and Environment, 81*, 270–282.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall Ltd0-412-31760-5.

McKenna, R., Merkel, E., Fehrenbach, D., Mehne, S., & Fichtner, W. (2013). Energy efficiency in the German residential sector: A bottom-up building-stock-model-based analysis in the context of energy-political targets. *Building and Environment, 62*, 77–88.

Meinel, G., Hecht, R., & Herold, H. (2009). Analyzing building stock using topographic maps and GIS. *Building Research and Information, 37*(5–6), 468–482.

Mitchell, G., Hargreaves, A. J., Namdeo, A., & Echenique, M. H. (2011). Land use, transport and carbon futures: The impact of spatial form strategies in three UK urban regions. *Environment and Planning A, 43*(9), 2143–2163.

Pereira, I. M., & Assis, E. S. (2013). Urban energy consumption mapping for energy management. *Energy Policy, 59*, 257–269.

Robinson, D., Campbell, N., Gaiser, W., Kabel, K., Le-Mouel, A., Morel, N., et al. (2007). SUNtool — A new modelling paradigm for simulating and optimising urban sustainability. *Solar Energy, 81*(9), 1196–1211.

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

Thom, H. C. S. (1958). A note on the gamma distribution. *Monthly Weather Review*, *86*, 117–122.
Tuhus-Dubrow, D., & Krarti, M. (2010). Genetic-algorithm based approach to optimize building envelope design for residential buildings. *Building and Environment*, *45*(7), 1574–1581.
Vanegas, C. A., Aliaga, D. G., Wonka, P., Muller, P., Waddell, P., & Watson, B. (2010). Modelling the appearance and behaviour of urban spaces. *Computer Graphics Forum*, *29*(1), 25–42.
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.
Wiginton, L. K., Nguyen, H. T., & Pearce, J. M. (2010). Quantifying rooftop solar photovoltaic potential for regional renewable energy policy. *Computers, Environment and Urban Systems*, *34*(4), 345–357.
Wilks, D. S. (1995). *Statistical methods in the atmospheric sciences*. San Diego, USA: Academic Press Inc0-12-751965-3.
Yu, Z., Fung, B., Haghighat, F., Yoshino, H., & Morofsky, E. (2011). A systematic procedure to study the influence of occupant behaviour on building energy consumption. *Energy and Buildings*, *43*(6), 1409–1417.
Zhou, B., & Kockelman, K. M. (2008). Micro-simulation of residential land development and household location choices: Bidding for land in Austin Texas. *Transportation Research Record: Journal of the Transportation Research Board*, *2077*(1), 106–112.