Predicting residential building age from map data

Rosser, J.F.*[1], Boyd, D.S.[1], Long, G.[1], Zakhary, S.[1], Mao, Y.[1], Robinson, D.[2]

[1] University of Nottingham, Nottingham, UK

[2] University of Sheffield, Sheffield, UK.

* Corresponding author

Julian.Rosser@nottingham.ac.uk

University of Nottingham

Nottingham

NG7 2RD

Funding acknowledgements:

This work was supported by the Leverhulme Trust (grant number RP 2013-SL-015). Additional funding was provided from Ordnance Survey and 1Spatial Ltd.

Conflicts of interest:

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Predicting residential building age from map data

The age of a building influences its form and fabric composition and this in turn is critical to inferring its energy performance. However, often this data is unknown. In this paper, we present a methodology to automatically identify the construction period of houses, for the purpose of urban energy modelling and simulation. We describe two major stages to achieving this – a per-building classification model and post-classification analysis to improve the accuracy of the class inferences. In the first stage, we extract measures of the morphology and neighbourhood characteristics from readily available topographic mapping, a high-resolution Digital Surface Model and statistical boundary data. These measures are then used as features within a random forest classifier to infer an age category for each building. We evaluate various predictive model combinations based on scenarios of available data, evaluating these using 5-fold cross-validation to train and tune the classifier hyper-parameters based on a sample of city properties. A separate sample estimated the best performing cross-validated model as achieving 77% accuracy. In the second stage, we improve the inferred per-building age classification (for a spatially contiguous neighbourhood test sample) through aggregating prediction probabilities using different methods of spatial reasoning. We report on three methods for achieving this based on adjacency relations, near neighbour graph analysis and graph-cuts label optimisation. We show that post-processing can improve the accuracy by up to 8 percentage points.

Keywords: building age; energy modelling; prediction; classification optimisation; MasterMap; Digital Surface Model;

1 Introduction

Buildings are responsible for 32% of global energy usage and 19% of energy-related greenhouse gas emissions (Lucon *et al.* 2014). Residential building stocks can be a major factor in determining this demand. In the UK, for example, domestic buildings are estimated to be responsible for 29% of national energy use (BEIS 2017). Reducing the energy demand imposed by domestic buildings requires accurate modelling of the current energy use of the housing stock and precise targeting of effective retrofit measures (Shorrock and Dunster 1997, Shorrock *et al.* 2005, Davies and Oreszczyn 2012). One important aspect in this modelling of a residential building's energy demands is the capture of the time period in which it was constructed (i.e., its age or construction period). This importance is, largely, due to the fact that this period determines the morphology, façade design and the materials used, which impact on a building's energy performance when used for habitation.

Several studies have shown how the age or construction period of buildings is a key factor in modelling their energy usage (Firth *et al.* 2010, Tooke *et al.* 2014, Aksoezen *et al.* 2015, Nouvel *et al.* 2017). In particular, it can be used to determine the building regulations (if applicable) in place at the time of construction and provide a powerful proxy for estimating a range of attributes that greatly affect its energy performance, such as:

- the fabric of the built envelope (wall and roof types and associated the construction methods),
- glazing types and ratios,
- potential energy retrofit measures that could be implemented and the degree of disruption involved in their addition,
- storey heights.

The construction period is commonly used in conjunction with the built form of a property (e.g. semi-detached, terraced, detached) as the basis for creating housing archetypes that are used in housing stock energy modelling (such as Mavrogianni et al. (2012), Davila et al. (2016), Firth et al. (2010) and Sousa et al. (2017)).

Identifying when a building was constructed also has a variety of important applications outside of energy modelling. In the context of natural hazards, for example, building age is used to help in assessing the vulnerability of properties to earthquakes

(Steimen *et al.* 2004, Wieland *et al.* 2012). Similarly, the vulnerability of buildings to damage by flood events is linked to the property age, which acts as an indicator of the type of foundation, wall and roof construction (Fedeski and Gwilliam 2007). Models estimating landslide risk also incorporate construction period (Uzielli *et al.* 2015).

In the UK and other countries, digital databases of construction period (and thus building age) are not readily available. Furthermore, even when a dataset of building age is available for an area, perhaps created through manual digitising of historic mapping, it may not be suitable for assessing building energy use. For example, the data may not be up-to-date, or the choice of age band is not appropriate or sufficiently detailed for energy modelling purposes. Therefore, there is a need for a method that can automatically identify, at the per-building level and according to a classification that is meaningful for energy modelling, the construction period of properties that is applicable at the urban, regional or national scale.

As urban areas develop over time, the construction and style of buildings change. These changes can be apparent in the geometric shape and location of buildings, and lead to particular spatial characteristics in the neighbourhoods of built-up areas (Yu *et al.* 2010, Hermosilla *et al.* 2012). The geometry of the urban blocks can indicate the type of buildings found there (Wurm *et al.* 2016), and also help in enabling identification of areas corresponding to particular times in a city's development (Barnsley and Barr 1997, Hermosilla *et al.* 2014). In this work, we exploit this information to estimate a construction period of residential buildings. In particular, we demonstrate a predictive modelling approach to infer an age automatically based on a range of geometric and neighbourhood features, derived from commonly available datasets. We treat the age of the building as an era, classified according to five periods of urban construction. These include pre-1915, 1915 – 1944, 1945 – 1964, 1965 – 1979, and post-1980. These construction periods reflect those of the English Housing Survey (EHS) (DCLG 2015) - a common source of statistical data describing the UK residential building stock, which is used for energy modelling. For example, the EHS can provide details on the physical construction characteristics linked to geographic regions according to categories of age and form and has thus provided a basis for many housing stock energy models (Sousa *et al.* 2017).

Our contributions include: 1) a detailed description and analysis of a statistical method for inferring residential building construction period that is important for modelling energy usage; 2) improved statistical inferences via post-classification approaches based on spatial reasoning of the scene; 3) quantitative assessment of different accuracy metrics obtainable for different sample sizes and model combinations.

2 Related work

In this section, we review previous work on the identification of the age or construction period of stocks of buildings. In this we focus our review on the identification of per-building attributes, rather than the characterisation of urban structure (i.e., Hermosilla et al. (2014)).

Meinel et al. (2009) describe a system which integrates spatial data from mapping agencies with geo-referenced digital airborne imagery. The authors undertook a process of cross-referencing of the geometric features of building footprints and the neighbourhood structure against a large sample set building types and neighbourhood forms, thus enabling the classification of building types. However, their construction period focused on buildings newer than 1984 due to constraints on the available data.

Van Hoesen and Letendre (2013) analysed historical maps to determine building age for identifying suitable properties for implementing energy efficiency measures.

The authors note that land parcel information is frequently unavailable and thus undertake a digitisation of historical data to create a temporal database of construction. The difficulty with this kind of approach is the manual effort required to complete the digitisation, in this case stated to be 30-40 hours for 1000 footprints.

Automatic analysis of the footprint shape and context have also been demonstrated as an approach for inferring construction period in the UK context. Alexander et al (2009) report a high accuracy for a subset of the age bands evaluated by applying different techniques to topographic map data. However, their work does not consider the use of neighbourhood metrics, nor DSM data, which has become more widespread in recent years, as we propose in our work.

Tooke et al. (2014) describe the application of a predictive modelling to LiDAR data in order to infer building ages in Vancouver, Canada, also with the aim of supporting energy modelling applications. Their approach extracted spatial metrics, which are used within a random forests regression model, to predict building age. However, their work does consider improving the accuracy of the prediction through spatial reasoning and analysis of the inference certainty, as we propose.

Biljecki and Sindram (2017) investigate the extent to which 3D city models can be used to determine building age. They experiment using different combinations of predictors, of building age extracted from a 3D city model such as building height, number of storeys, use and enclosed volume. Our work focuses on using 2D national mapping agency topographic data and DSM as sources, rather than a 3D city model.

In summary, we identify that relatively few approaches to determining either building age or construction period have been proposed. These existing methods involve either: (i) exhaustive manual digitisation – an approach which is not scalable to large urban areas; or (ii) rely on datasets that are not always available and do not consider the multiple sources of building and neighbourhood data available in the UK context; or (iii) do not consider the location or spatial context of a building as part of their inference process to assigning an age or construction period attribute.

We propose here the use of a supervised machine learning method to enable automatic inference of building age for energy modelling purposes, based on a variety of different types of geospatial data. We then develop and evaluate methods to analyse the model predictions and improve inference through a neighbourhood assessment of a building's adjacencies and relations.

3 Method

3.1 Study area

Our study area is the city of Nottingham, UK. The residential building stock of Nottingham exhibits a range of housing styles across age bands commonly found in the UK. Figure 1 shows the Nottingham City boundary, the sample locations for undertaking cross-validation model selection and testing and the detailed area of the neighbourhood sample (both discussed in further detail below).

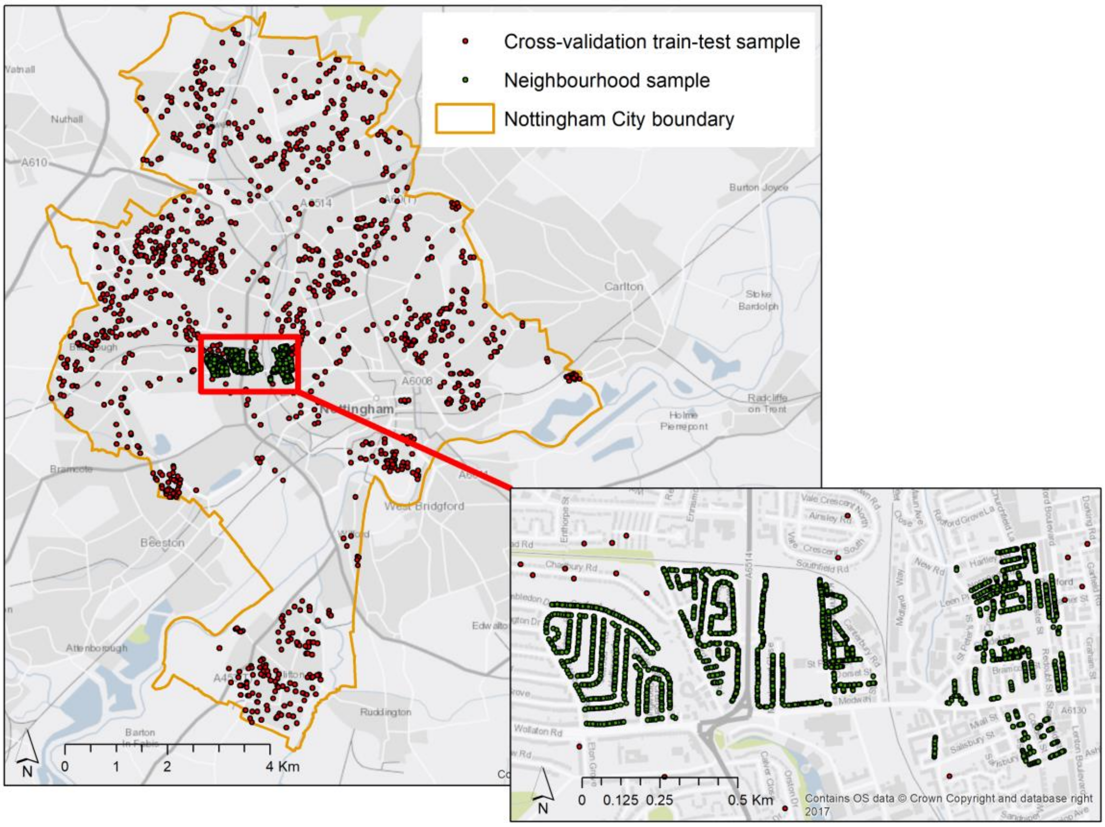

Figure 1. Nottingham city boundary area. Sample locations for cross-validation training and testing, and detailed section showing 1457 houses used for neighbourhood area testing (see section 4 for further details).

3.2 Data

Topographic mapping, in the form of building footprints, Digital Surface Model (DSM) and the statistical boundaries used by the UK national census were utilised for the predictive model development. Topographic mapping forms the core data for analysis from which building features were extracted.

We used Ordnance Survey (OSGB) MasterMap, which provides an accurate and geometrically detailed representation of building footprints (Ordnance Survey 2017). To create a dataset of only habitable building footprints (i.e no garages or sheds) we joined the MasterMap data to AddressBase Plus (Ordnance Survey 2016), a UK database of property address points, and removed polygons that were found to not have an address. OSGB also provides a building heights database (Building Height Attribute) of additional fields, which augment these footprints. These fields include a value for the building height as measured from ground-level, and ridgeline height of each footprint. As MasterMap is a primarily 2.5D product when enriched with the building heights, a 0.5m spatial resolution DSM, created in-house by OSGB using photogrammetry of aerial imagery, was used to provide additional 3D details. For census boundaries, the Office of National Statistics portal provided 2011 Output Area polygons (ONS Geography 2011).

3.2.1 Training and cross-validation data

In order to train and test the supervised classification model, a city-wide sample of buildings, their morphological and neighbourhood attributes and their age is required, for ground-truth purposes. Five periods of construction including pre-1915, 1915 – 1944, 1945 – 1964, 1965 – 1979, and post-1980 were adopted for the classification. These age bands reflect those found in the EHS, apart from a small difference in splits between the first two periods (the EHS uses pre-1919 and 1919-1944), which was necessary as these were the bands provided in the available source data. A spatial dataset of the age and location of all local authority owned and built properties (provided by Nottingham City Council) formed a starting point for creating this sample. In Nottingham, the local authority has built over 50,000 homes since 1901, which represents over 40% of the city's housing stock. However, the dataset contained relatively few old properties (those built before 1915) and few modern properties (built after 1980). To compensate for this sampling bias, we augmented the data with pre-1915 and post-1980 properties drawn from a random household visit survey conducted in 2014 and made available for this work (Long *et al.* 2015). To create a final training-validation dataset, we randomly selected properties across each age band using proportions obtained from a study that characterises the entire Nottingham housing stock (Long *et al.* 2015). As this produced a sample of 1096 properties across the city -

80% was used for cross-validation model selection and training and 20% was used as a validation dataset, as summarised in Table 1. The locations of this sample are shown in Figure 1.

Table 1. Sample descriptions, class sizes, and splits used for training, validation and testing.

Sample description	pre-1915	1915 – 1944	1945 – 1964	1965 – 1979	post-1980	Split	Usage description
1096 points - city-wide sample	229	280	259	220	108	80%	Cross-validation model selection, hyper-parameter tuning, learning curve training data, final model training.
						20%	Test set for user and producer accuracy calculation, selected CV model learning curve testing and trained CV model testing.
1457 points - neighbourhood sample	301	463	363	203	127	100%	Neighbourhood test set for post-classification spatial reasoning testing

3.2.2 Neighbourhood testing sample

In addition to the city-wide training and cross-validation samples, evaluation of the predictive model and subsequent post-classification spatial analysis utilised a specific residential area sample. The area was chosen due to its inclusion of residential properties across all age bands. Non-residential buildings and blocks of flats were removed from this sample and no buildings within the area were included in the city-wide training and cross-validation samples (to ensure effective validation). This created a sample of 1457 properties for the neighbourhood evaluation, the locations of which are shown in Figure 1.

3.3 Predictors

To infer age, features based on building (see section 3.3.1) and neighbourhood characteristics (see section 3.3.4) were derived from the vector-based topographic mapping, DSM and census boundary datasets. Our experiments utilised fifteen predictors in total, summarised in Table 2.

Table 2. Features used for characterizing the construction period of buildings

Predictor	***Description***
Area (m)	*Area of building footprint*
Perimeter (m)	*Length of a building outline*
NPI	*Normalized Perimeter Index*
footprintVertices	*A count of the number of vertices that make up the building footprint*
relHEaves (m)	*Height of building eaves relative to the ground*
relHMax (m)	*Maximum height of building relative to the ground*
roofPitchMean (m)	*Average slope of the roof pitch DSM pixels falling within the building footprint*
roofHeightStdDev (m)	*Standard deviation of DSM pixels falling within the building footprint*
builtForm	*The building is detached, semi-detached, terraced, end-terraced or complex.*
numberInBlock	*A count of the number of buildings that are in a block (i.e. are topologically related)*
sharedPerimeter (m)	*Length of the building outline shared between touching building footprints*
non-sharedPerimeter (m)	*Length of the building outline not shared between touching building footprints*
exteriorRatio	*Ratio of the length of the exposed building outline divided by the total footprint perimeter*
builtUpAreaRatio	*A ratio of built-up area size (m) to the size (m) of the census Output Area*
voronArea (m)	*Area of the Voronoi parcel estimated around the building's centroid*

3.3.1 Building footprint geometry

The area of the building footprint (m^2) is computed as this can provide some indication of its age. For example, over the last 30 years in the UK and the US contexts, old (e.g. pre-1919) houses are noted to be small relative to modern standards, and a tendency of increasing house size over time has been documented (Brown and Steadman 1991, Wilson and Boehland 2005). However, in more recent years, this is also paralleled by construction of homes with smaller floor areas (RIBA 2015).

Perimeter measures pertaining to the amount of exposed wall of each building were also calculated. These included the overall footprint perimeter (m), length of shared wall (m), length of exposed wall (m) and the ratio of exposed wall to the total perimeter.

We utilise a measure of polygon complexity to characterise the shape of the building footprints. The Normalised Perimeter Index (NPI) has been shown to help with inferring building heights (Biljecki *et al.* 2017) and identification of building types (Wurm *et al.* 2016), which in turn can provide evidence for a particular construction period. NPI provides a characterisation of the compactness of the building footprint shape and is calculated using $\frac{2\sqrt{a\pi}}{p}$ (where a and p are the footprint area and perimeter respectively). Example measures of NPI are shown in Figure 2. As an additional measure of shape complexity, the number of vertices that make up the footprint shape was also recorded.

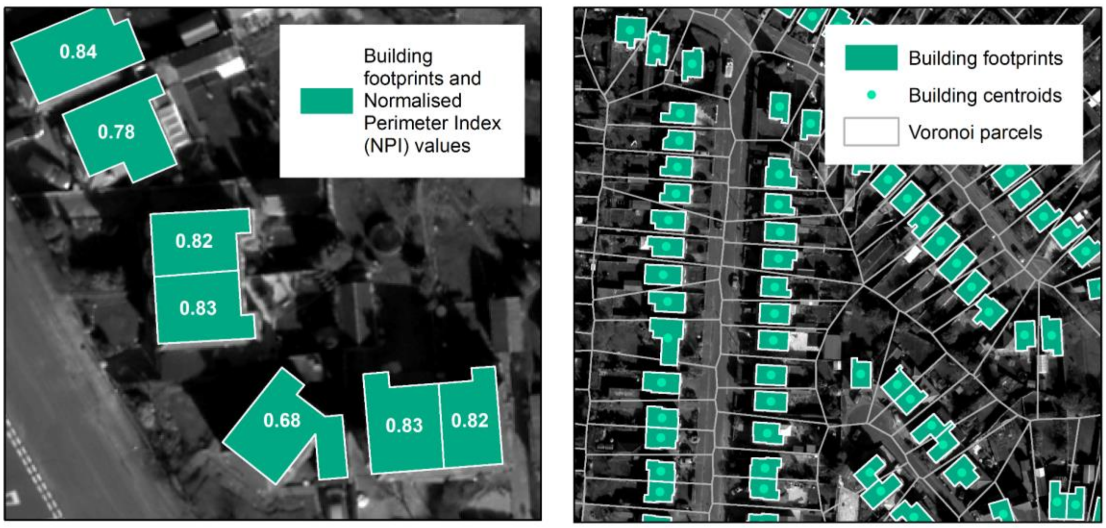

Figure 2. Example building footprints with their Normalised Perimeter Index (NPI) values (left) and example footprint centroids (right). The Voronoi parcels depicted are discussed in further detail in section 3.3.4.

3.3.2 *Building roof and height details*

The three-dimensional form of buildings is an expression of their architectural style and can thus provide a meaningful indication of their age. We use two measures relating to the building height (m), namely relative eaves height (approximately ground level to eaves height) and maximum height (approximately ground level to the highest feature including dormers or chimneys). This is sourced from the OSGB MasterMap building height attributes. In addition to these height measures, we use two measures relating directly to the roof geometry. The first is to estimate the pitch of the roof by calculating the mean slope (degrees) within each building footprint. The second aims to quantify complexity in the roof by calculating the standard deviation of the height values (m) within each building footprint. These two roof predictors are sourced from the DSM. Examples of these predictors are shown in Figure 3.

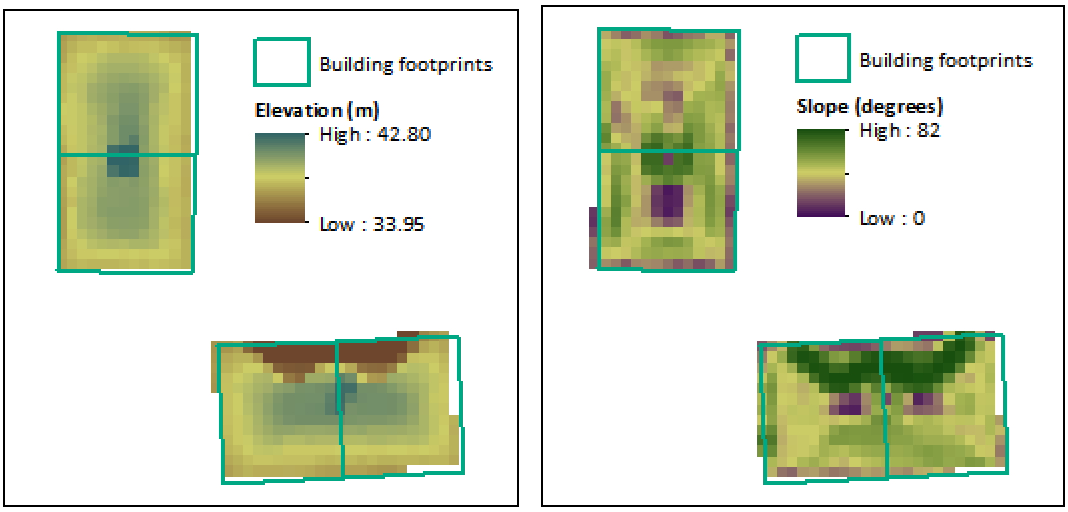

Figure 3. Roof elevation used to derive standard deviation of roof heights in order to infer roof complexity (left) and roof slope, used to derive mean roof pitch (right).

3.3.3 *Building topology*

We include several features relating to the topology of the property. The first is the built form (i.e. whether the construction is terraced, end-terraced, semi-detached or detached). This feature was determined using a topological query of each footprint where the number of neighbours determines the form, (see Figure 4, left). We follow the implementation of Beck et al (2018) to first classify buildings based on touching neighbours (i.e. 0 is detached, 1 is semi-detached or end-terraced, 2 is mid-terraced), and then separate the end-terraced buildings into their own class where the number properties in a whole block of touching polygons is greater than two. In addition to the built form, a block level metric defining the number of buildings in a topological block (i.e. the number of buildings that form a group of buildings that are touching) is used (see Figure 4, right).

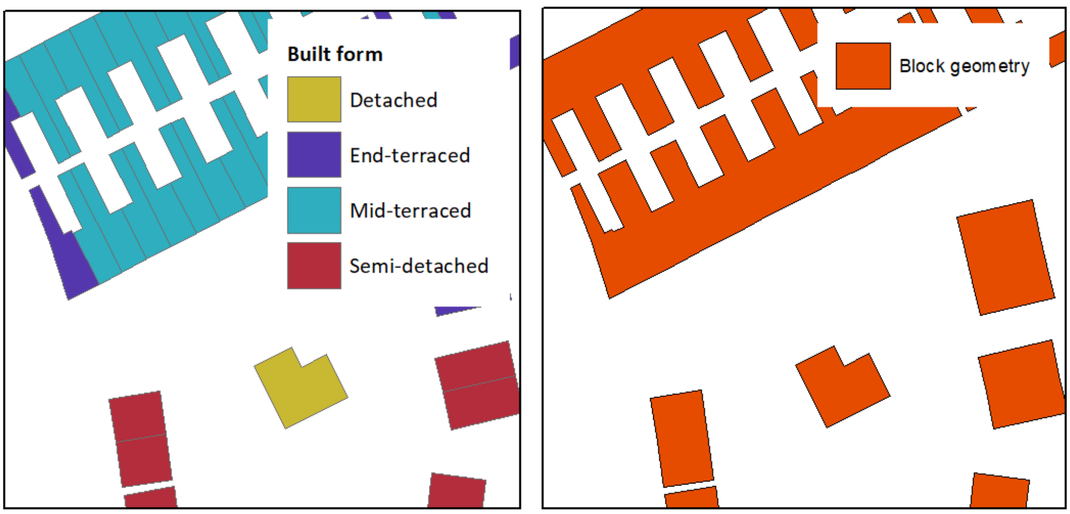

Figure 4. Built form classification showing detached, end-terraced, mid-terraced and semi-detached properties (left) and block level geometry used to determine number of buildings that are touching (right).

Two measures pertaining to the lengths (m) of the building's outline that are touching were calculated. These include the length of the shared wall and the length of the exposed wall. Lastly, the exposed wall length divided by the perimeter was computed in order to define an exterior wall ratio value.

3.3.4 Neighbourhood attributes

The density of buildings can give a particular character to a neighbourhood or local area. To provide an indication of the degree to which the neighbourhood is urbanised, a measure of built-up area was calculated. Here we adopt the census Output Area (OA) as a neighbourhood zone to which a built-up area statistic can be attached. The built-up area is defined as the total area of buildings (both residential and non-residential) within each OA, normalised by area of the OA. The amount of open space surrounding a building can also provide an indication of the type of development within an area. For example, neighbourhood details based on the Voronoi relations have been shown to help identify usage and occupancy of buildings (Huang *et al.* 2013). The lot area (i.e. property parcel) was found to be a strong predictor of age in previous work (Tooke *et al.* 2014). We use the Voronoi diagram constructed over the centroids of each footprint that has an address (thus including both residential and non-residential properties) to estimate a land parcel surrounding a building, and calculated its area (m^2). An illustration of the Voronoi parcels is shown in Figure 2.

3.4 Model implementation

3.4.1 Predictive model

We implement the statistical model for learning the predictors as a random forest multi-class classification problem. Random forests are a widely used ensemble-based supervised learning method suitable for both classification and regression problems (Breiman 2001). Breiman (2001) demonstrates that random forests are robust to noise, are computationally efficient to train, handle both categorical and continuous data, provide a measure of variable importance and unlike Support Vector Machines and multi-layer perceptron types of learning models, have the advantage of estimating class probability when used for inference. In summary, as an ensemble method random forests extend a decision tree-type estimator. Decision trees model the classification task as a tree defined as recursive binary splits of the predictor variables at important values. To determine the splitting, a measure of the node (leaf) impurity is required, typically using either entropy or Gini importance. We use the Gini importance to determine splitting which calculates how much a feature reduces the impurity resulting from the split. Decision trees offer an appealing method for learning due to their invariance to transformations on the predictors (thus obviating the need for feature normalisation). They perform an internal feature selection as part of their training procedure, yet they also suffer from high variance and perform poorly in comparison to other estimators (Hastie *et al.* 2009). Tree-based bootstrap aggregation or bagging extends the decision tree approach by training a set of trees on subsets of the data derived from bootstrap sampling (sampling with replacement). During inference, the average prediction is taken across all trees, yielding a more powerful estimator than a single model. In random forests, the splitting of each tree is based on a random sample (m) of the available predictors, thus de-correlating individual trees and making the bagged average prediction more reliable and less variable (James *et al.* 2007).

We adopt the random forest package of the Apache Spark (v2.1) distributed computing framework (spark.apache.org), which adopts the Breiman (2001) implementation and enables rapid, parallelised model testing and evaluation (Zaharia *et al.* 2010).

3.4.2 *Prediction optimisation*

Most statistical learning methods, including random forest classification, operate on a model where labelled vectors of feature values are used for training and testing. When applying a trained model for practical purposes, the model inference does not take account of the spatial nature of the phenomena in question - in this case an estimated class of building age. For the case of building age inference, we identify that an automatic analysis of the classifier prediction, which adopts some form of spatial reasoning about the scene, can improve upon per-building estimates made by the classifier alone. For example, in the case of two semi-detached houses linked by an adjoining wall, it is highly probable that such houses were built at the same time. The same is also probable in the case of a block of terraced houses. Furthermore, this spatial autocorrelation of the building age is not limited to structures that are touching. Housing construction is likely to have been undertaken by developers who build a set of houses across a block or plot of land, and this means clusters of nearby similarly aged residential buildings can often be identified in suburban and city centre residential areas.

We test three approaches to exploiting spatial autocorrelation to improve our building age prediction. These techniques utilise the probability value of each age class produced by the random forest classifier. Typically, when conducting inference with a random forest classifier, the class with the highest probability is assigned for a single query data point. Instead, we aggregate the per-building class prediction probabilities across groups of buildings. Determining these groups depends upon defining spatial relations between the objects. Here we use both *touching* and *nearness*.

Our first approach is to take the mean of each class prediction across all buildings in a block. Then, the class with the maximum mean probability is assigned to each building in the block.

The second approach extends the first to buildings that are near to each other. Using the same Voronoi diagram, described above in section 3.3.4, we define near neighbour relations between buildings. These nearby buildings correspond to the nodes of the Delaunay triangulation dual graph of the Voronoi diagram. Using the relations defined by this graph, for each building we can take the mean of the class probabilities over both itself and its neighbours, and then assign the class with the maximum mean probability to that building.

The third approach extends the second approach to an energy minimization problem, which is solved using graph cuts (Boykov *et al.* 2001, Boykov and Kolmogorov 2004, Kolmogorov and Zabih 2004, Fulkerson *et al.* 2009). The graph cuts method has been widely used in image segmentation and smoothing applications, where spatial coherence over pixels helps to identify the features depicted. Schindler (2012) reviews its application in land cover remote sensing classification, for example. The method generalises to other graph structures (non-imagery) problems such as grouping of buildings for 3D city model simplification (Wang *et al.* 2015). The graph cuts method is an approach to finding the minimum cut or maximum flow path across a graph. It can be used to find the set of labels that minimizes an energy function comprised of a term corresponding to the observed data (known as unary) and a term corresponding to an adjacency graph (known as pairwise or boundary).

$$E(L) = U(L) + \lambda B(L) \quad (1)$$

Where L is the labelling for the nodes, $U(L)$ is the unary term, $B(L)$ is the pairwise term, and λ is a weighting factor (which here is set to 15). We adopt similar energy functions to the image segmentation problems described by Fulkerson et al. (2009) and Hernandez-Vela (2012), using the confidence in the classifier predictions. In particular, we use the random forest prediction probabilities of each class to define the unary term:

$$\mathrm{U}(L) = \sum_{i \in \mathcal{P}} -\log(c_i) \tag{2}$$

Where $\mathcal{P}$ is the set of nodes and c_i is the vector of age class probabilities for node i which has been inferred by the classifier.

In image segmentation problems, the pairwise term is usually defined as a 4-pixel or 8-pixel neighbourhood around each pixel. In our case, however, we use the adjacencies defined by the Voronoi neighbours to define the graph. Iterating over the adjacencies allows us to define a cost for each edge based on the Euclidean n-space distance between the class probabilities of the corresponding points.

$$B(L) = \sum_{\{i,j\} \in \mathcal{N}} \left\| x_i - x_j \right\| \cdot \delta\left(L_i L_j\right) \tag{3}$$

$$\delta\left(L_i L_j\right) = \begin{cases} 1 \; if \; L_i \; \neq L_j \\ \quad 0 \end{cases} \tag{4}$$

where x is the vector of class probabilities for a node (i,j). Here the smoothing term $\delta\left(L_i L_j\right)$ fulfils the role of an indicator function which encourages similarity in the building labelling.

4 Results

4.1 Model selection and tuning

We test various combinations of predictor to investigate the model accuracy obtainable in different scenarios of data availability. Furthermore, random forests require specification of several hyper-parameters for effective training and inference. The number of predictors selected for determining tree splits (m), the number of trees that make up the forest, the maximum depth to which each tree may branch and the binning used when discretizing continuous features are important in determining the performance of the classifier. Although large numbers of trees can improve performance, studies have shown that such improvements can level off (Hastie *et al.* 2009, Oshiro and Perez 2012) while the time to fit the model increases. Simultaneously, the depth of the tree can improve predictive performance but can increase overfitting. To evaluate each of our models and the corresponding hyper-parameters, we applied a grid search and cross-validation evaluation. To determine *m*, from all available predictors (p), we use $m = \sqrt{p}$, as suggested by Breiman (1984) in all model cases. For the remaining hyper-parameters, we used a 5-fold cross-validation grid search on 80% of features randomly selected from the city-wide sample, in order to identify values close to optimal, according to a particular evaluation metric. Given our multi-class prediction problem, we use accuracy for the metric, which reports the number of true positives normalised by the number of data points. For the grid search parameters, maximum tree depths of 3, 4 and 5, maximum binning values of 35, 45, 60 and 80 and forests of size 30, 40, 60, 80, 100 and 120 were used.

Table 3 illustrates combinations of features used and the accuracy for the best performing model, as identified by the grid search. Table 3 indicates that relatively strong classifier performance is achievable even with a limited set of predictors. For example, Model 6, which excludes the neighbourhood features, achieves 74% cross-validation accuracy and Model 8, which excludes the DSM derived metrics, achieves 72%.

The model using all features achieved the highest cross-validation accuracy (75%), with hyper-parameters of 40 trees, a maximum depth of 5 and a maximum binning of 60. This model was selected for further analysis. For practical evaluation of the classifier we use confusion matrices, see section 4.2 and 4.3.

Table 3. Models with different combinations of predictors (features) and cross-validation accuracy.

Predictor	*Model #*								
	1	2	3	4	5	6	7	8	9
Area (m)	o	o	o	o	o	o	o	o	o
Perimeter (m)		o	o	o	o	o	o	o	o
NPI			o	o	o	o	o	o	o
footprintVertices			o	o	o	o	o	o	o
relHEaves (m)				o	o	o		o	o
relHMax (m)				o	o	o		o	o
roofPitchMean (m)					o	o			o
roofHeightStdDev (m)					o	o			o
builtForm						o	o	o	o
numberInBlock						o	o	o	o
sharedPerimeter (m)						o	o	o	o
non-sharedPerimeter (m)						o	o	o	o
exteriorRatio						o	o	o	o
builtUpAreaRatio							o	o	o
voronArea (m)							o	o	o
Overall CV Accuracy	0.46	0.6	0.66	0.69	0.72	0.74	0.7	0.72	0.75

Plotting the learning curve of the classifier allows for identification of bias and variance, and thus the trade-off between underfitting and overfitting, in the model. Furthermore, the learning curve shows the effect of sample size on the efficacy of prediction. Using the best performing model (Model 9) and hyper-parameters identified from the 5-fold cross validation above, iterative increases of 100 data points for the training set size was tested. For this evaluation, 80% of the city-wide sample was used for training and the remaining 20% was used for testing, as shown in Table 1. The training and testing accuracy (i.e. $accuracy(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1} 1(\hat{y}_i = y_i)$ where $\hat{y}_i$ is the predicted value and y_i is the corresponding true value) were computed for each sample to create the learning curve for the model shown in Figure 5. We can see that the testing accuracy appears to plateau at approximately 77%, close to the model's training accuracy of approximately 80%.

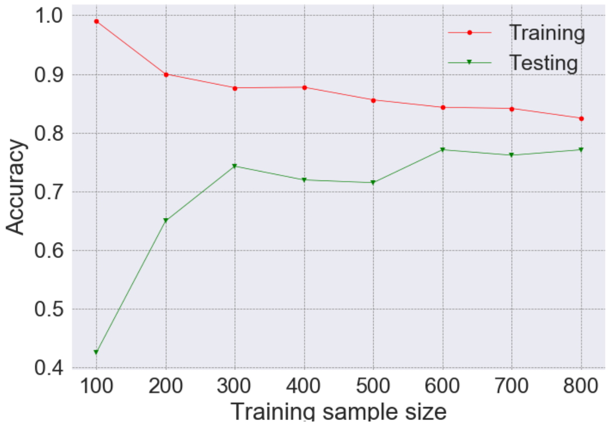

Figure 5. Learning curve for training sample size increasing by 100 per step.

4.2 Model performance, feature importance and cross-validation

The feature importance metric, shown in Figure 6, is exposed by the Apache Spark classification module and is computed according to the average of its importance (in our case, Gini importance) across all trees in the ensemble, with the resulting vector normalized to sum to 1 (Hastie *et al.* 2009). Figure 6 shows the measures of footprint complexity (NPI), footprint size (footprintArea) and average roof pitch (roofPitchMean) are the strongest predictors for construction period. The measure of built-up area (builtUpAreaRatio) and perimeter also appear to be useful discriminators. In contrast, the number of vertices (footprintVertices), the measure of the roof shape complexity (roofHeightStdDev), the roof eaves height (relHEaves) and the estimated size of the land around the property (voronArea) are relatively poor as indicators of age.

A confusion matrix of the model performance offers insight into the predictive power across different age classes. Figure 7 shows the confusion matrix (row-wise normalised) calculated on the 20% city-wide test set. The results show effective identification of the pre-1915 (82%), 1945 – 1964 (88%) and 1965 – 1979 (78%) age bands. The classifier scores worse on 1915 – 1944 (67%) and post-1980 (50%) and there is a notable discrepancy with the differentiation between post-1980 and pre-1915 properties; this latter reflecting the relative scarcity of data in these categories.

In addition to the overall accuracy of 77%, we compute user's accuracy (precision) as 70%, and producer's accuracy (recall) of 73%. The kappa statistic, which takes into account chance matches between the predicted and true classes, of 0.7 indicates substantial agreement (Landis and Koch 1977).

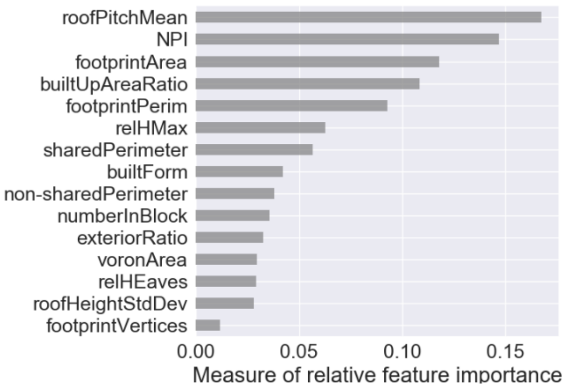

Figure 6. Ranked mean decrease in Gini importance of the fifteen predictor variables.

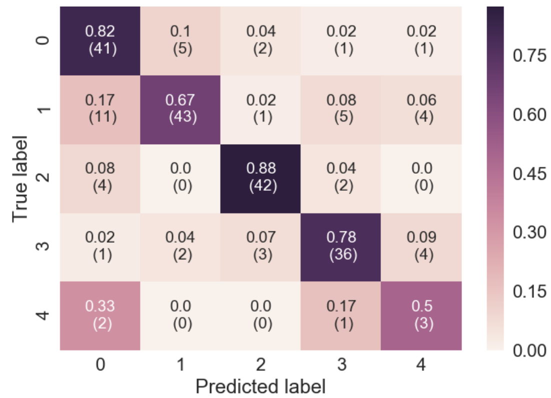

Figure 7. Confusion matrix of prediction accuracy (row-wise normalised) based on city-wide sample with building counts per class in brackets

4.3 Neighbourhood specific sample

Here we evaluate the classifier on the neighbourhood sample of 1457 houses, and apply the techniques described in section 3.4.2 to improve upon the age band estimate provided by the predictive model alone.

Table 4. User's and producer's accuracy (UA and PA), overall accuracy (OA) and kappa statistic.

Model	***UA***	***PA***	***OA***	***Kappa***
Prediction only	57	57	64	0.54
Block averaging	59	58	65	0.55
Near-neighbour averaging	61	60	68	0.59
Graph-cuts label optimisation	56	61	72	0.64

Table 4 presents the different accuracy metric of the neighbourhood sample, and Figure 8 shows the confusion matrices. These show the prediction model estimate (overall accuracy = 64%), block averaged predictions (overall accuracy = 65%), near-neighbour (Voronoi) averaging (overall accuracy = 68%) and graph-cuts label optimisation (overall accuracy = 72%). Considering the user's and producer's accuracy and kappa, we see that the different methods of adjusting the predictions generally improve the performance. However, there is a notable drop in user's accuracy (precision) when applying the graph-cuts approach; though this is offset by an increase in producer's and overall accuracy. Considering the confusion matrices in Figure 8, this is due to an increase in the misclassification of post-1980 properties as pre-1915.

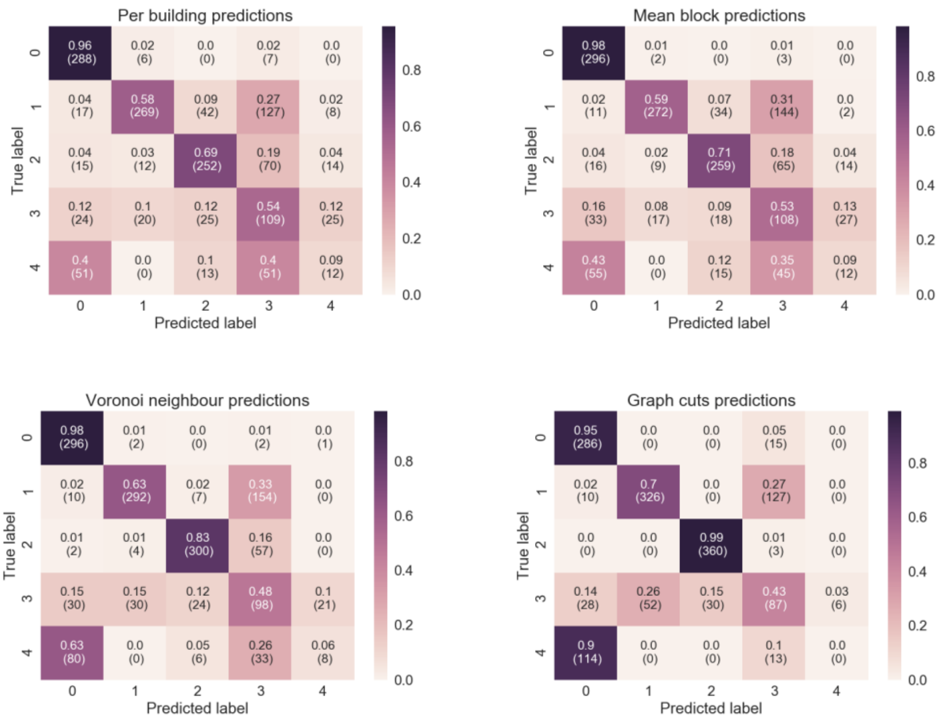

Figure 8. Normalised (row-wise) confusion matrices for the neighbourhood sample. Prediction model estimate (top-left), averaged prediction confidences by building block (top-right), averaged prediction confidences by Voronoi adjacencies (bottom-left), graph-cut label optimisation using Voronoi adjacencies (bottom-right).

In Figure 9 and Figure 10 we show maps of the ground-truth and predicted construction periods for the neighbourhood sample. Comparing the ground-truth and per-building inference, we can see that the method appears to capture a reasonable overall distribution of property ages across the scene. However, some areas such as the housing found at the centre of the map are estimated incorrectly. The map of block averaged predictions is seen to improve the accuracy in some areas, particularly where long pre-1915 housing terraces include a property with an anomalous age. The near-neighbour averaging and graph-cuts label optimisation smooth the predicted ages, particularly in the Western and Eastern sides of the map. The graph-cuts approach in particular, which scores highest overall, does however incorrectly group many buildings together into the pre-1915 band, while conversely correctly expanding the cluster of 1946-1964 detached properties.

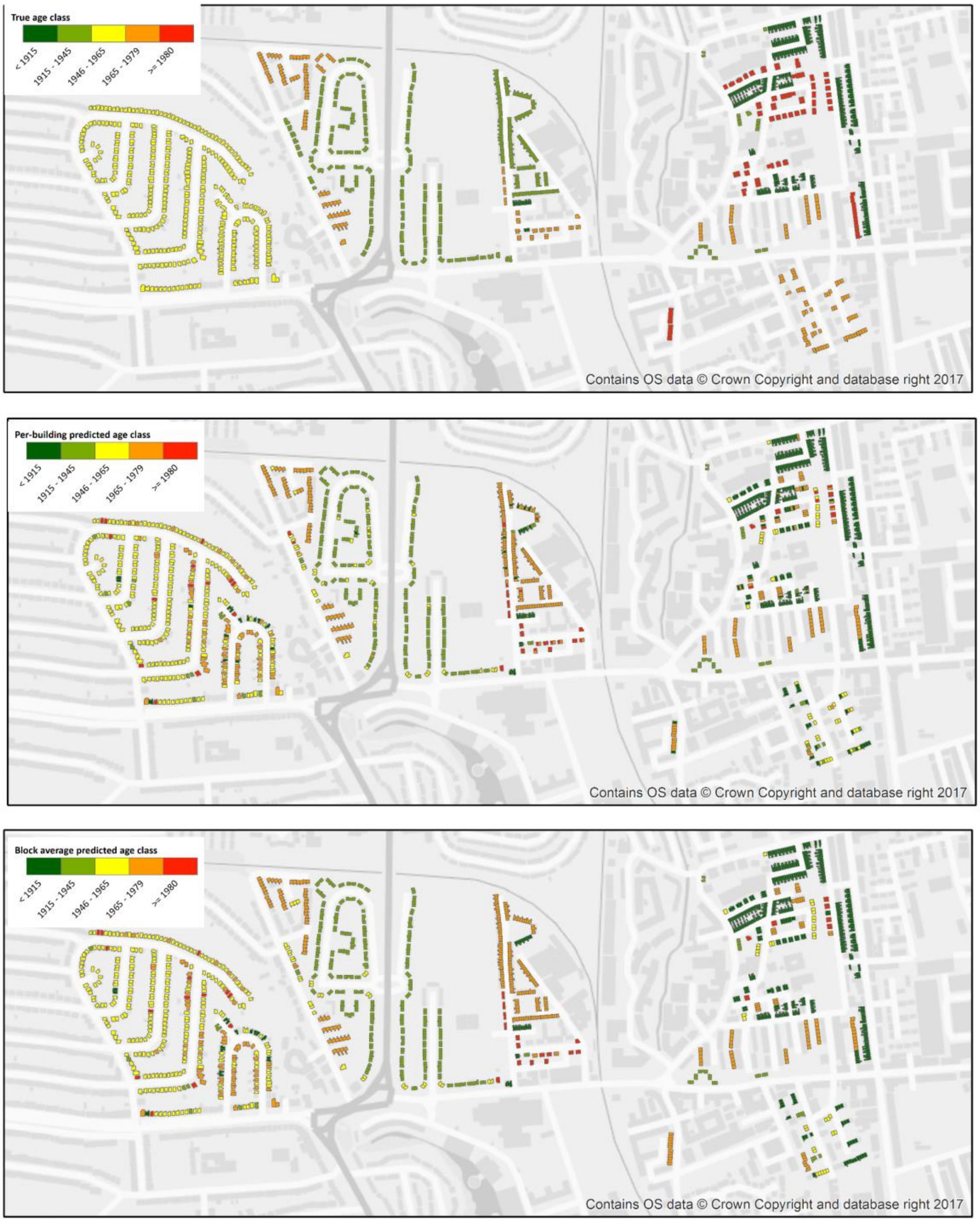

Figure 9. Neighbourhood sample showing true age class (top), per-building predicted age class (middle) and block averaged predicted age class (bottom).

Figure 10. Predicted age class based on Voronoi average (top) and graph-cuts label optimisation (bottom)

5 Discussion

In this work, we demonstrated the assignment of a construction period of a residential property given certain evidence that can be automatically extracted from maps. Using geometric measures of the building footprint, a digital surface model and neighbourhood statistics as a source of variables, a predictive model is shown to achieve reasonable accuracy (5-fold cross-validation = 75%, and 77% test accuracy) and this estimation can effectively be improved through an automatic post-classification analysis. A data-driven model for estimating building age will never achieve perfect accuracy. As-built survey with extensive manual work is the only absolutely reliable way to assign all buildings a correct construction period. However, as this is prohibitively expensive, a data-driven approach can serve as a useful tool for energy modelling practitioners undertaking city-wide analysis, and provide guidance on the locations of properties where the construction period has been inferred with a high degree of confidence.

Analysing the incorrect predictions made by the classifier (see Figure 6) indicates that a large proportion are within one age band of the true construction period. This degree of error is not likely to have a significant impact, estimated to be less than 10% based on analysis of gas usage by dwelling age reported in the Homes Energy Efficiency Database statistics (Hamilton *et al.* 2013). However, in the relatively unabundant case of modern buildings (post 1980), over 30% are incorrectly assigned as being constructed before 1915. This would lead to substantial errors when modelling the energy use of these buildings, which are generally less efficient than modern properties (Hamilton *et al.* 2013). In general, the performance of the classifier on predicting ages of modern building is shown to be challenging. The wide variation in age prediction for these buildings is likely due to their sparsity in the training set, leading to insufficient discrimination in the model for effective inference. Further work could incorporate a larger sample of modern buildings. In the UK, at least, poor performance for this age band may be partly ameliorated by the recently released Valuation Office Agency price paid data (landregistry.data.gov.uk), which enables identification of properties constructed during or after 1995. This dataset could form a useful set of additional training data or be used to automatically correct inferred classes.

Further investigation of the contribution of individual features may also help. Research has suggested that estimates of feature importance in random forests may be biased when variables are of mixed type or scale (Strobl *et al.* 2007). Therefore, analysis of predictor relevance could be undertaken by adopting permutation importance, rather than the Gini criterion, or by adopting an alternative implementation of the random forest (e.g. conditional random forests).

One aspect that we note in this work is the difference in performance for the trained predictive model validated against city-wide data versus the neighbourhood validation. The likely cause for this variation lies in the relatively unusual style of 1920s properties in the centre of the neighbourhood area which are likely to be under-represented in the training data, resulting in their misclassification as 1965 – 1979 properties. Recall that there were no training sample points within the neighbourhood area boundary in order to avoid biasing the predictive model and to effectively evaluate post-classification improvement. In practical deployment of this approach, a random sample across all city areas would be recommended.

With respect to the label optimisation based on graph cuts, we highlight that alternative or more detailed cost functions might further improve the approach. For example, the pairwise cost might utilise a more complex graph structure of the spatial relations, such as 1st order and 2nd order adjacencies, and a spatial Euclidean distance could be incorporated to weight neighbours differently based on their proximity. Cross-validation, expert judgement or a prior measure of spatial autocorrelation (such as Moran's I) could inform such a distance weighting. Similarly, the label cost $\delta(L_i L_j)$ could encode further knowledge about the feasibility of grouping certain buildings, such as relaxing the constraints for particular ages to appear adjacent to each other (e.g. if modern re-developments are particularly common in old sectors of a city). Lastly, it should be noted that the overall weighting of the pairwise term λ may be adjusted for particular scenes based on expert prior knowledge. For example, a higher weight can be used to enforce larger groupings and vice versa.

Our approaches for improving class prediction are evaluated using the random forest classifier prediction probabilities. However, it is important to note that similar label optimisation approaches could be applied to another type of machine learning classifier, providing that an assessment of class certainty or probability associated with the inference is part of the predictive model. For example, many other ensemble methods (e.g. AdaBoost) calculate probabilities as part of their inference (Breiman 2001); artificial neural network models can be trained to generate class probabilities (Bishop 1995) and Bayesian networks provide a degree of belief associated with a class prediction (Jensen 1996).

The work presented here utilised a sample of local properties for training the classifier. For application in other urban areas, a local sample would likely yield the best performance. However, a further interesting avenue for research could compare the performance of a local sample versus a model trained using a national, regional or alternative sample area. This may lead to greater insights on generality of features to other areas.

We tested models using features extracted from geospatial data sources that are readily available in the UK. However, our approach could in principle be applied in other countries where equivalent data is available from mapping agencies or perhaps from Volunteered Geographic Information or crowdsourcing campaigns. In urban areas, OpenStreetMap (OSM) building footprints can have a high-level of completeness and geometric similarity to authoritative map data, although commonly lack attribute level data (Fan *et al.* 2014). Use of this data to infer building age offers a promising line of research and a potential method of OSM enrichment. Similarly, further work on using alternative features, which might be found in other countries could be undertaken. For example, city models can include volume or gross-floor area attributes based on true 3D building geometry. In this work, singular height values per property are used but these may overestimate sizes of properties where the building plan varies across different floor levels.

6 Conclusion

In this paper, we developed and tested an automatic two-stage methodology for inferring the age of residential buildings for the purposes of energy modelling, using geospatial data that is readily available in the UK context. We use measures of building shape and neighbourhood characteristics that form predictors in a random forest classifier for estimating a per-building age class band. Both raster (a high-resolution Digital Surface Model) and vector-based map data (topographic mapping) are used to derive these measures. The predictive model achieves a cross-validation accuracy score of 75% and validation set score of 77%. We also describe and test methods of processing per-building age inferences using the topology and spatial relations of the urban structure to enforce likely assumptions on the classifier's prediction. We show that correct assignment of construction periods can be improved by up to 8 percentage points through post-prediction optimisation of the class labels.

7 Acknowledgements

The OS MasterMap data were made available by Ordnance Survey (via the Digimap Licence). The photogrammetric Digital Surface Model was made available by Ordnance Survey.

8 References

Aksoezen, M., Daniel, M., Hassler, U., and Kohler, N., 2015. Building age as an indicator for energy consumption. *Energy and Buildings*, 87, 74–86.

Alexander, D.K., Lannon, S., and Linovski, O., 2009. the Identification and Analysis of Regional Building Stock Characteristics Using Map Based Data. *Building Simulation 2009, Glasgow, Scotland July 27-30*, 1421–1428.

Barnsley, M.J. and Barr, S.L., 1997. Distinguishing urban land-use categories in fine spatial resolution land-cover data using a graph-based, structural pattern recognition system. *Computers, Environment and Urban Systems*, 21 (3–4), 209–225.

Beck, A., Long, G., Boyd, D.S., Rosser, J.F., Morley, J., Duffield, R., Sanderson, M., and Robinson, D., 2018. Automated classification metrics for energy modelling of residential buildings in the UK with open algorithms. *Environment and Planning B: Urban Analytics and City Science*.

BEIS, 2017. Energy Consumption in the Uk. *Department for Business, Energy and Industrial Strategy*, (July).

Biljecki, F., Ledoux, H., and Stoter, J., 2017. Generating 3D city models without elevation data. *Computers, Environment and Urban Systems, Under review*, 64, 1–18.

Biljecki, F. and Sindram, M., 2017. Estimating building age with 3D GIS. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-4/W5 (October), 17–24.

Bishop, C.M., 1995. *Neural networks for pattern recognition*. Oxford University Press.

Boykov, Y. and Kolmogorov, V., 2004. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 26 (9), 1124–1137.

Boykov, Y., Veksler, O., and Zabih, R., 2001. Efficient Approximate Energy Minimization via Graph Cuts. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 20 (12), 1222–1239.

Breiman, L., 1984. *Classification and Regression Trees.* New York: Routeledge.

Breiman, L., 2001. Random forests. *Machine Learning*, 45 (1), 5–32.

Brown, F. and Steadman, J., 1991. The morphology of British housing: an empirical basis for policy and research. Part I: Functional and dimensional characteristics. *Environment and Planning B: Planning and Design*, 18, 277–299.

Davies, M. and Oreszczyn, T., 2012. The unintended consequences of decarbonising the built environment: A UK case study. *Energy and Buildings*, 46, 80–85.

Davila, C.C., Reinhart, C., and Bemis, J., 2016. Modeling Boston: A workflow for the generation of complete urban building energy demand models from existing urban geospatial datasets. *Energy*, 117, 237–250.

DCLG, 2015. *Deparment for Communities and Local Government, English Housing Survey: Housing Stock Report, 2014-2015*.

Fan, H., Zipf, A., Fu, Q., and Neis, P., 2014. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28 (4), 700–719.

Firth, S.K., Lomas, K.J., and Wright, A.J., 2010. Targeting household energy-efficiency measures using sensitivity analysis. *Building Research and Information*, 38 (1), 24–41.

Fulkerson, B., Vedaldi, A., and Soatto, S., 2009. Class Segmentation and Object Localization with Superpixel Neighborhoods. *In*: *Proceedings of the International Conference on Computer Vision*.

Hamilton, I.G., Steadman, P.J., Bruhns, H., Summerfield, A.J., and Lowe, R., 2013. Energy efficiency in the British housing stock: Energy demand and the Homes Energy Efficiency Database. *Energy Policy*, 60, 462–480.

Hastie, T., Tibshirani, R., and Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer New York.

Hermosilla, T., Palomar-Vázquez, J., Balaguer-Beser, Á., Balsa-Barreiro, J., and Ruiz, L.A., 2014. Using street based metrics to characterize urban typologies. *Computers, Environment and Urban Systems*, 44, 68–79.

Hermosilla, T., Ruiz, L.A., Recio, J.A., and Cambra-López, M., 2012. Assessing contextual descriptive features for plot-based classification of urban areas. *Landscape and Urban Planning*, 106 (1), 124–137.

Hernandez-Vela, A., Zlateva, N., Marinov, A., Reyes, M., Radeva, P., Dimov, D., and Escalera, S., 2012. Graph cuts optimization for multi-limb human segmentation in depth maps. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 726–732.

Van Hoesen, J. and Letendre, S., 2013. Characterizing the spatiotemporal evolution of building-stock age in Poultney, Vermont: A GIS-based approach to improve thermal efficiency in historical buildings. *Environment and Planning B: Planning and Design*, 40 (4), 630–643.

Huang, H., Kieler, B., and Sester, M., 2013. Urban building usage labeling by geometric and context analyses of the footprint data. *In*: *Proceeding of 26th international cartographic conference (ICC)*.

James, G., Witen, D., Hastie, T., and Tibshirani, R., 2007. *An Introduction to Statistical Learning with Applications in R*. Performance Evaluation.

Jensen, F. V, 1996. *Introduction to Bayesian Networks*. 1st ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Kolmogorov, V. and Zabih, R., 2004. What Energy Functions can be Minimized via Graph Cuts? *IEEE transactions on Pattern Analysis and Machine Intelligence*, 26 (2), 147–159.

Landis, J.R. and Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33 (1), 159–174.

Long, G., Alwany, M., and Robinson, D., 2015. Building typologies simulation report, Nottingham. [online]. Available from: http://www.insmartenergy.com/wp-content/uploads/2014/12/D.2.1.-Simulation_Report_for_building_Typologies-Nottingham.pdf [Accessed 12 Jul 2017].

Lucon, O., Ürge-Vorsatz, D., Zain Ahmed, A., Akbari, H., Bertoldi, P., Cabeza, L.F., Eyre, N., Gadgil, A., Harvey, L.D.D., Jiang, Y., Liphoto, E., Mirasgedis, S., Murakami, S., Parikh, J., Pyke, C., and Vilariño, M.V., 2014. Buildings. *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 33, 1–66.

Mavrogianni, A., Wilkinson, P., Davies, M., Biddulph, P., and Oikonomou, E., 2012. Building characteristics as determinants of propensity to high indoor summer temperatures in London dwellings. *Building and Environment*, 55, 117–130.

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

Nouvel, R., Zirak, M., Coors, V., and Eicker, U., 2017. The influence of data quality on urban heating demand modeling using 3D city models. *Computers, Environment and Urban Systems*, 64, 68–80.

ONS Geography, 2011. Output Areas (December 2011) Boundaries [online]. Available from: http://geoportal.statistics.gov.uk/datasets/output-area-december-2011-full-extent-boundaries-in-england-and-wales [Accessed 6 Sep 2017].

Ordnance Survey, 2016. AddressBase Plus [online]. Available from: https://www.ordnancesurvey.co.uk/business-and-government/products/addressbase-plus.html [Accessed 9 Dec 2016].

Ordnance Survey, 2017. OS MasterMap Topography Layer [online]. Available from: https://www.ordnancesurvey.co.uk/docs/technical-specifications/os-mastermap-topography-layer-technical-specification.pdf [Accessed 4 Jan 2018].

Oshiro, T.M. and Perez, P.S., 2012. Machine Learning and Data Mining in Pattern Recognition, 7376 (July 2012).

RIBA, 2015. *Space Standards For Homes*. Royal Institute of British Architects (RIBA) technical report. London.

Schindler, K., 2012. An overview and comparison of smooth labeling methods for land-cover classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50

(11 PART1), 4534–4545.

Shorrock, L.D. and Dunster, J.E., 1997. The physically-based model BREHOMES and its use in deriving scenarios for the energy use and carbon dioxide emissions of the UK housing stock. *Energy Policy*, 25 (12), 1027–1037.

Shorrock, L.D., Henderson, J., and Utley, J.I., 2005. Reducing carbon emissions from the UK housing stock. *BRE Press.*, 80.

Sousa, G., Jones, B.M., Mirzaei, P.A., and Robinson, D., 2017. A review and critique of UK housing stock energy models, modelling approaches and data sources. *Energy and Buildings*, 151, 66–80.

Steimen, S., Fäh, D., Giardini, D., Bertogg, M., and Tschudi, S., 2004. Reliability of building inventories in seismic prone regions. *Bulletin of Earthquake Engineering*, 2 (3), 361–388.

Strobl, C., Boulesteix, A.L., Zeileis, A., and Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8.

Tooke, T.R., Coops, N.C., and Webster, J., 2014. Predicting building ages from LiDAR data with random forests for building energy modeling. *Energy and Buildings*, 68 (PARTA), 603–610.

Uzielli, M., Catani, F., Tofani, V., and Casagli, N., 2015. Risk analysis for the Ancona landslide—II: estimation of risk to buildings. *Landslides*, 12 (1), 83–100.

Wang, Y., Zhang, L., Takis Mathiopoulos, P., and Deng, H., 2015. A Gestalt rules and graph-cut-based simplification framework forurban building models. *International Journal of Applied Earth Observation and Geoinformation*, 35 (PB), 247–258.

Wieland, M., Pittore, M., Parolai, S., Zschau, J., Moldobekov, B., and Begaliev, U., 2012. Estimating building inventory for rapid seismic vulnerability assessment: Towards an integrated approach based on multi-source imaging. *Soil Dynamics and Earthquake Engineering*, 36, 70–83.

Wilson, A. and Boehland, J., 2005. Small is Beautiful U.S. House Size, Resource Use, and the Environment. *Journal of Industrial Ecology*, 9 (1–2), 277–287.

Wurm, M., Schmitt, A., and Taubenböck, H., 2016. Building Types' Classification Using Shape-Based Features and Linear Discriminant Functions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9 (5), 1901–1912.

Yu, B., Liu, H., Wu, J., Hu, Y., and Zhang, L., 2010. Automated derivation of urban building density information using airborne LiDAR data and object-based method. *Landscape and Urban Planning*, 98 (3–4), 210–219.

Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S., and Stoica, I., 2010. Spark: Cluster Computing with Working Sets. *HotCloud'10 Proceedings of the 2nd USENIX conference on Hot topics in Cloud Computing*, 10.