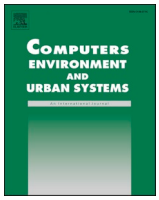
Computers, Environment and Urban Systems 88 (2021) 101637

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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



Spatial factors influencing building age prediction and implications for urban residential energy modelling

Oana M. Garbasevschi [a,b,*], Jacob Estevam Schmiedt [c], Trivik Verma [d], Iulia Lefter [d], Willem K. Korthals Altes [e], Ariane Droin [b], Björn Schiricke [a], Michael Wurm [b]

[a] Institute for Solar Research, German Aerospace Center (DLR), Cologne, Germany
[b] German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[c] Institute for the Protection of Terrestrial Infrastructures, German Aerospace Center (DLR), Saint Augustin, Germany
[d] Faculty of Technology, Policy and Management, Delft University of Technology, the Netherlands
[e] Faculty of Architecture and the Built Environment, Delft University of Technology, the Netherlands

ARTICLE INFO

Keywords:
Open data
Urban morphology
Spatial autocorrelation
Residential building age
Residential heat demand
Random forest

ABSTRACT

Urban energy consumption is expected to continuously increase alongside rapid urbanization. The building sector represents a key area for curbing the consumption trend and reducing energy-related emissions by adopting energy efficiency strategies. Building age acts as a proxy for building insulation properties and is an important parameter for energy models that facilitate decision making. The present study explores the potential of predicting residential building age at a large geographical scale from open spatial data sources in eight municipalities in the German federal state of North-Rhine Westphalia. The proposed framework combines building attributes with street and block metrics as classification features in a Random Forest model. Results show that the addition of urban fabric metrics improves the accuracy of building age prediction in specific training scenarios. Furthermore, the findings highlight the way in which the spatial disposition of training and test samples influences classification accuracy. Additionally, the paper investigates the impact of age misclassification on residential building heat demand estimation. The age classification model leads to reasonable errors in energy estimates, in various scenarios of training, which suggests that the proposed method is a promising addition to the urban energy modelling toolkit.

1. Introduction

Worldwide, urban areas account for 70% of total energy-related CO_2 emissions and two thirds of primary energy demand (International Energy Agency, 2016). It is estimated that between 2005 and 2050, due to urban population growth and economic development, end-use energy demand will triple (Creutzig, Baiocchi, Bierkandt, Pichler, & Seto, 2015). In 2018, the residential sector represented 26.1% of the final energy consumption in the EU, with the main end-use being space heating (63.6%), followed by water heating (14.5%), lighting and electrical appliances (14.1%), and other uses such as cooking or space cooling (Eurostat, 2020). Energy demand in cities could be reduced by up to 50% before 2050 by increasing the energy efficiency of buildings, appliances and distribution networks and by improving the energy-efficient behaviour of residents (Ürge-Vorsatz et al., 2012). Urban energy modelling assists this goal by providing computational techniques for estimating the energy consumption of buildings and infrastructure, and for planning and evaluating energy saving strategies (Reinhart & Cerezo Davila, 2016).

Bottom-up building energy models are based on the in-detail simulation of the energy requirements of an individual building, also called archetype, which is representative for a major class of buildings in a national building stock (Mata, Kalagasidis, & Johnsson, 2014). These models estimate the energy demand for entire groups of buildings based on the demands of the representative building archetypes and extrapolate the results from local to regional or national level (Swan & Ugursal, 2009). The factors that influence energy demand include indoor and outdoor environmental conditions, occupant behaviour, building geometry, construction techniques and building regulations existing at the time of the construction (Economidou et al., 2011). These variables constitute the input for energy calculation in complex individual building energy models. In the absence of complete building data,

* Corresponding author at: The German Remote Sensing Data Center, Oberpfaffenhofen, 82234 Weßling, Germany.
E-mail address: mihaela.garbasevschi@dlr.de (O.M. Garbasevschi).

https://doi.org/10.1016/j.compenvurbsys.2021.101637
Received 17 September 2020; Received in revised form 8 February 2021; Accepted 5 April 2021
Available online 25 April 2021

building age is used to infer model parameters, such as thermal insulation properties of construction elements (Firth, Lomas, & Wright, 2010), ventilation rate (Mata et al., 2014), storey height (Zirak, Weiler, Hein, & Eicker, 2020), and floor type or glazing ratios (Rosser, Long, Zakhary, Boyd, & Mao, 2019), thus making building age an essential variable for energy demand estimation.

Most energy modelling studies focus on the residential building stock since statistical information on the non-residential sector is sparser and modelling processes are more complex due to the large variety of building morphologies and related usages (Loga, Diefenbach, Stein, & Born, 2012). Residential building data is acquired by various methods: census data collection, formal building and dwelling registers, surveys or remote sensing (Mata et al., 2014; Van den Brom, Hansen, Gram-Hanssen, Meijer, & Visscher, 2019; Wurm et al., 2021). In Europe, the availability of building age data varies at national and regional levels. While in The Netherlands and Denmark such information is publicly available (Van den Brom, Hansen, Gram-Hanssen, Meijer, & Visscher, 2019), this is not the case for most countries due to non-uniform administrative procedures and privacy concerns. In Germany, data availability differs between regions and data sources exhibit various levels of quality and detail (Zirak et al., 2020). The need to bridge the gap between the available data and the requirements of building energy models has led to various efforts for the automatic identification of the individual age or age class of residential buildings.

Open spatial data sources have become increasingly available due to a growing number of open city data initiatives and the emergence of crowd sourced GIS maps. This provides enhanced opportunities to perform analysis of urban-relevant sustainability topics by taking into consideration various aspects of urban form, as shown in studies concerning urban pollution (Athanassiadis, 2019), smart cities (Neves, de Castro Neto, & Aparicio, 2020) or energy systems (Manfren, Nastasi, Groppi, & Astiaso Garcia, 2020). This paper integrates building age prediction in the greater scope of urban morphology analysis by investigating the effect of urban geometry structures such as street and urban block metrics in predicting building age. The study additionally explores the relevance of attributes that predict building type for building age classification. Building construction type is an important parameter in building energy models, and consitutes, along with building age, a factor for classifiying residential buildings into energy-relevant archetypes (Reinhart & Cerezo Davila, 2016). The combined analysis of building characteristics that predict age and construction type can be instrumental for the fast deployment of building energy models. Furthermore, the integration of urban energy modelling and urban morphology is an established reasearch area (Weinand, McKenna, & Fichtner, 2019) and this study is a natural addition to this research field.

Outside the context of energy applications, building age is a factor to be considered also in scenarios dealing with material stocks and flows in the built environment (Ortlepp, Gruhler, & Schiller, 2018), seismic vulnerability (Liuzzi et al., 2019), building thermal performance under climate change conditions (Nahlik et al., 2017), or real estate market valuation (Zeppelzauer, Despotovic, Sakeena, & Koch, 2018). Furthermore, investigating building age constitutes an opportunity for understanding the built environment, its spatial and temporal patterns.

Previous work in building age prediction focused on inferring the age from physical characteristics of the building, like geometry (Alexander, Lannon, & Linovski, 2009; Biljecki & Sindram, 2017; Tooke, Coops, & Webster, 2014), location (Rosser et al., 2019) or façade appearance (Li, Chen, Rajabifard, Khoshelham, & Aleksandrov, 2018; Zeppelzauer et al., 2018). Most of these studies used a combination of topographic data such as building footprints and digital surface models, with cadastre and municipal databases (Alexander et al., 2009; Biljecki & Sindram, 2017; Rosser, Boyd, et al., 2019; Tooke et al., 2014). Studies analysing building façades used building images extracted from real-estate data platforms or Google Street View. Related work in the prediction of building construction type relied on computing complex footprint shape indicators from spatial data (Droin, Wurm, & Sulzer, 2020; Wurm, Schmitt, & Taubenböck, 2016).

Studies in the broader area of urban morphology have also undertaken the task of predicting age at the level of entire neighbourhoods of buildings. The focus of these studies is either to compare or classify neighbourhoods based on construction year epoch by using metrics related to urban blocks and streets, and include analyses of both European (Gil, Beirão, Montenegro, & Duarte, 2012; Hermosilla, Palomar-Vázquez, Balaguer-Beser, Balsa-Barreiro, & Ruiz, 2014) and American cities (Lowry & Lowry, 2014). Similarly to research on individual buildings, the metrics used in neighbourhood classification are extracted from topographic and open government data.

The literature review we performed on topics concerning building age, building shape and urban morphology highlights a recurring number of machine learning methods employed: Random Forest (Biljecki & Sindram, 2017; Droin et al., 2020; Rosser, Boyd, et al., 2019; Tooke et al., 2014), clustering (Berghauser Pont et al., 2019; Gil et al., 2012), linear discriminant analysis (Wurm et al., 2016) and Convolutional Neural Networks (Li et al., 2018; Zeppelzauer et al., 2018).

The overall performance of classification is similar across studies predicting age classes, with high accuracies for specific age periods (Alexander et al., 2009; Rosser, Boyd, et al., 2019). The prediction of age as construction year leads to average errors in predicted age between 15.8 (Tooke et al., 2014) and 19.4 years (Biljecki & Sindram, 2017) for studies based on spatial data. In image-based prediction, for a selected set of residential buildings, the errors in construction year estimation are reduced, with a maximum of 10 years (Li et al., 2018). Model performance review shows that there is a need for models yielding higher accuracies over all building ages and types. Furthermore, a discussion of the power of generalization of the age prediction method is lacking, with all related work being peformed for single cities or neighbourhoods.

For this study, we developed a method to improve building age classification by using an extended set of spatial attributes and we explored the transferability of results between spatial units at different scales. First, we supplemented the classical building attributes with complex shape metrics conclusive for characterizing building types. Second, we associated buildings with block and street metrics derived from urban morphology studies that explicitly connect properties of the neighbourhood with age. Using data from different cities in Germany, we analysed the potential of identifying building age by learning from examples from similar and different geographic areas. Lastly, we applied a reduced model for space heating demand calculation in order to quantify the impact that the inferred building age has on the accuracy of energy demand estimates. The following sections describe the adopted methodology and conclude with a discussion of the strengths and limitations of the proposed approach.

2. Study area and data

The analysis is focused on residential buildings in 8 cities in the German federal state of North Rhine-Westphalia. In recent years, North Rhine-Westphalia – the state with the highest population and the fourth largest by area – has been actively developing its open data policies and is currently first in Germany in terms of the size of published open data (Open Government Germany, 2019).

The main source of information about building age is the 2011 Census, the national population and housing statistical report (Zensus, 2011). Due to data privacy concerns, individual data points are subject to statistical confidentiality (Senate Department for Urban Development and Housing, 2018) and the public data is published as aggregates in an INSPIRE-compliant grid format (INSPIRE, 2014). Building age information is presented in the Census as 10 age classes comprising unequal intervals of building construction years.

The highest level of spatial resolution of the published Census data is a grid of 100 m resolution, available from the German Federal Agency of Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie, 2019). Every grid cell has associated summaries of population and

household statistics. The building age distribution in a grid cell is represented by the number of buildings in each age class. By selecting cells that contain a single age class, we mapped age classes to individual buildings through the spatial overlay of grid cells and building footprints, as illustrated in Fig. 1. The buildings contained in the selected grid cells constitute the input data for model training. Thus, for a set of selected 7 cities, we obtained a sample of individual buildings comprising of approximately a third of the residential building stock. The German residential building stock at the end of 2009 was formed in proportion of 34% of buildings built after 1979, 26% of buildings built before 1948, and 40% of buildings built between 1919 and 1978 (Loga et al., 2012). With the Census age-extraction method we have obtained a sample containing almost twice as many buildings from the 1949–1978 category, and a considerably reduced number of buildings built after 1979. For validation purposes, we analysed an auxiliary building dataset. The city of Wuppertal is the only city in the state for which the age of individual residential buildings is publicly available (Open NRW, 2019) therefore it served as test site for developing our method. The distribution of samples per city and age class is presented in Table 1.

The source of information for building geometries are building models in LoD2 format, available on the data portal of North Rhine-Westphalia (Open NRW, 2019). LoD2 is the simplest 3D building representation that includes ground surfaces, walls and roof shapes (Kolbe, Gröger, & Plümer, 2005). Building addresses were extracted from the real estate cadastre data of North Rhine-Westphalia (Geobasis, 2019). City blocks are administrative areas enclosed by streets and have been mapped at various spatial resolutions for the territory of Germany in the Official Topographic-Cartographic Information System. Block data at a scale of 1:10,000 was extracted from the data portal of the state (Open NRW, 2019). We retrieved the street information from Open Street Map and processed it through the APIs provided by the *OSMnx* Python library (Boeing, 2017). Remote sensing images were extracted from Sentinel-2 data . We used the information from the 10 m resolution red and infrared spectral bands to compute a normalized difference vegetation index (NDVI) and derive from it texture attributes of blocks. Fig. 2 illustrates the different types of data used.

3. Building age classification method

3.1. Classification features

The most relevant attributes for predicting building age have been selected from previous studies concerning building age and building type prediction and from studies concerning neighbourhood age prediction. The first type of studies deals mostly with the analysis of building geometry, while the second type also includes attributes of surrounding streets and urban block.

In the majority of studies reviewed, building characteristics concerning height, roof shape, number of storeys, volume, surface and shape of building footprint are a common denominator. The building footprint is described by simple metrics, such as area and perimeter, and also by more complex statistics, such as shape smoothness and complexity (McGarigal & Marks, 1995), normalized perimeter index (Angel, Parent, & Civco, 2010), shape compactness and fractal dimension (Wurm et al., 2016). The building's 3D shape is characterized by smoothness and compactness (Tooke et al., 2014; Wurm et al., 2016). In addition to shape features, the building type was used as a classification feature in our approach. In a prior study the building type of German residential buildings was predicted using geometric features of the building footprint and the connections between a building and the neighbouring buildings (Droin et al., 2020). The different types of building features are described in Table 2.

For lack of available data, some features that were determined important for building age prediction in related studies have not been added to our analysis. These features include ceiling height, a building attribute judged relevant for identifying older buildings, which tend to have higher ceilings (Biljecki & Sindram, 2017), and the area of the lot allocated to a building and building-to-lot area, attributes relevant for identifying newer buildings, which maximize the space available on the lot (Tooke et al., 2014). Since these features were highlighted in connection to the Dutch and respectively, the British residential building stock, their relevance for German buildings is still to be determined. The most promising of these features is the ceiling height, with various sources indicating that German buildings built before 1949 have higher

Fig. 1. Selection of samples for analysis by overlaying building footprints with INSPIRE grid cells. The statistical information of the selected cells indicates that the area of the cell contains only residential buildings in the same age class. Illustration of a neighbourhood in Duisburg.

Table 1
Proportion of building samples acquired per city and the age class distribution of samples. For the first 7 cities, samples contain only residential buildings that can be accurately labelled with age class information extracted from Census data. For Wuppertal, samples represent the entire residential building stock.

City	Age class distribution (%)										Total samples
	Pre–1919	1919–1948	1949–1978	1979–1986	1987–1990	1991–1995	1996–2000	2001–2004	2005–2008	Post–2009	
Bielefeld	4.4	9.2	73.6	5.2	0.6	1.1	2.8	1.5	1.2	0.4	22,499
Cologne	1.5	8.1	80.8	2.7	0.5	1.4	1.5	1.0	1.3	1.2	44,274
Dortmund	6.4	7.4	77.1	3.1	0.4	1.2	1.9	1.1	0.9	0.5	36,617
Duisburg	5.0	13.2	72.7	2.1	1.0	1.4	2.3	1.1	0.9	0.3	22,313
Dusseldorf	3.3	10.6	76.3	3.3	1.2	1.8	1.2	0.9	0.8	0.7	25,588
Essen	5.4	10.6	76.3	2.2	0.4	0.8	1.2	1.2	1.4	0.4	28,903
Munster	0.4	4.3	78.7	4.4	0.7	1.2	3.3	1.5	3.1	2.3	17,935
Wuppertal	19.3	13.2	33.6	10.2	7.0	6.9	4.2	4.4	0.8	0.2	47,361

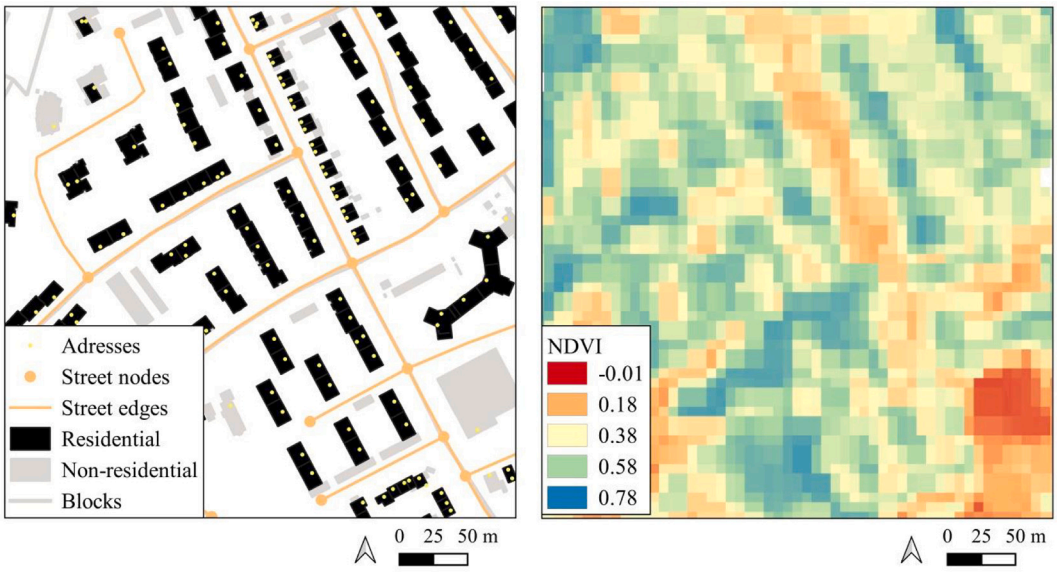


Fig. 2. Illustration of types of data used: urban structures in vector format (left) and NDVI information in raster format (right).

ceilings than newer buildings (Zirak et al., 2020).

Urban form metrics are an established technique to quantify and understand the structure of the built environment (Lowry & Lowry, 2014). Streets represent city connectivity and are increasingly used to characterize the urban form. Connectivity can be expressed by street length, number of intersections or ratio of intersections to street length (Berghauser Pont et al., 2019). Complex centrality metrics describe the position of a street node in the overall city street network (Boeing, 2017). We extracted centrality metrics for the street nodes closest by Euclidean distance to the vertices of a building footprint and assigned the maximum value over all nodes as a building attribute.

Blocks are described by area, perimeter and shape related metrics, such as compactness, fractal dimension and perimeter index. A classical notion in urban research is the density of the built environment. Density can be one, two or three dimensional, when considering the number, area or the volume of the buildings in a specific area, respectively. Another type of density refers to the richness of functions inside a neighbourhood. Lowry and Lowry (2014) have applied an ecological metric, Simpson's index, to compute the diversity of land uses in a block area. We adapted this metric by computing the diversity of building functions both in terms of number of buildings and area occupied by buildings having the same function. Building function is defined as the usage of a building outside residential purposes and includes broad classes, such as commercial, industrial, office and administration, sport and entertainment etc. The availability of open space is a feature that differentiates old from new neighbourhoods (Hermosilla et al., 2014) and can be expressed either as the non-built or the vegetation area within a block. Street and block metrics are shown in Table 3.

3.2. Classification model

The classification method used in this study is *Random Forest*, an ensemble-based supervised learning algorithm (Breiman, 2001). Its main features include robustness to noise, computational efficiency, in-built importance estimation and treatment of both categorical and continuous data. Additionally, the method handles high data dimensionality and multicollinearity of features well and is widely used in classification tasks based on remote sensing and geographic data (Belgiu & Dragut, 2016).

The performance of a classification algorithm is generally evaluated by its *accuracy* or *success rate*, defined as the ratio of correctly labelled observations to the size of input data. For problems of multi-class classification, the overall accuracy can be misleading when the representation of classes in the sample is unequal. This being the case with most of the building datasets under analysis, we also report the *sensitivity* or *recall*, defined as the ratio of correct positive predictions to the total number of actual observations per class. The average of sensitivity scores over all age classes was our chosen metric for evaluating and optimizing classification results. Additionally, the *kappa coefficient* is listed, to illustrate the agreement between predicted and actual classes, corrected for chance (Cohen, 1960).

Table 2
Building features.

Name	Description	Ref.[a]
Height and volumetric features		
Building height	Height measured from highest point of the building.	1,2,3
Number of storeys	$\frac{height}{3}$ (storey height assumed to be 3 m).	
Volume	Building volume.	
Floor area	Sum of all floor areas.	
Shape compactness	Exposed building per unit volume: ratio of surface area to $V^{2/3}$.	3
3D Shape index	Indicator of 3D smoothness, expressed as the ratio of perimeter to perimeter of the equal volume cube.	4
Height-area ratio	Ratio of building height to footprint area.	5
Roof and wall features		
Wall height	Difference between building height and roof height.	
Roof slope	Angle between the ground surface and the roof surface with the greatest area.	2,3
Roof angle (avg, sd)	Average and standard deviation of all angles formed between roof surfaces and the ground surface.	
Roof area	Sum of all roof surfaces.	
Number of walls	Approximated by the number of vertices that make up the footprint.	2,3
Wall area	Sum of wall surfaces.	3
Surface area	Sum of all surfaces.	
Simple footprint shape features		
Area, perimeter	Area (a) and length (p) of building footprint.	1,2,3
Perimeter index	Indicator of building footprint compactness, expressed as the ratio of perimeter of the equal area circle (EAC) to perimeter $\frac{2\sqrt{a\pi}}{p}$.	2,4,6
2D Shape index	Indicator of shape smoothness, expressed as the ratio of perimeter to perimeter of the equal area square $\frac{p}{4*\sqrt{a}}$.	4,7
Direct neighbours	Number of buildings that are connected (the footprints have a common edge or a fraction of a full edge).	5
Area neighbours	Area of the minimum-sized polygon containing all direct neighbours.	
Perimeter neighbours	Perimeter of the minimum-sized polygon containing all direct neighbours.	
Relative area neighbours	The relative area of the building footprint compared to the polygon containing all neighbours.	
Building type	Detached, semi-detached, terraced, multi-family.	
Complex footprint shape features		
Fractal	Indicator of shape complexity and fragmentation, expressed as a function of perimeter and area $\frac{ln\left(\frac{p}{4}\right)^2}{lna}$.	3,4,7
Building parts	Number of building parts.	
Detour	Length of the convex hull of the footprint polygon. The convex hull is the shortest connection of vertexes that fully contains the polygon.	5,6
Detour index	Ratio of perimeter of EAC to detour length.	
Range	Distance between the furthest most vertex points of the building footprint.	
Range index	Ratio of diameter of EAC to range.	
Exchange	Area of the polygon within EAC considering they have the same centroid.	
Exchange index	Ratio of exchange to footprint area.	
Cohesion	Average Euclidean distance between 30 randomly selected interior points.	
Cohesion index	Ratio of radius of EAC to cohesion.	
Proximity	Average Euclidean distance from all interior points to the footprint centroid.	
Proximity index	Ratio of two thirds of EAC radius to proximity.	
Spin	The average of the square of the Euclidean distances between all interior points and the footprint centroid.	
Spin index	Ratio of spin to the half of the squared radius of EAC.	

[a] 1 – Biljecki & Sindram, 2017; 2 – Rosser, Boyd, et al., 2019; 3 – Tooke et al., 2014; 4 – Wurm et al., 2016; 5 – Droin et al., 2020; 6 – Angel et al., 2010; 7 – McGarigal & Marks, 1995.

Table 3
Street and block metrics.

Metric	Description	Ref.[a]
Street metrics		
Betweenness centrality	For node v, sum of the fraction of all-pairs shortest paths that pass through v.	1,2
Load centrality	For node v, fraction of shortest paths that pass through v.	2
Closeness centrality	For node v, reciprocal of the sum of the shortest path distances from v to all n-1 other nodes.	
Degree centrality	For node v, fraction of nodes v is connected to.	
Neighbourhood degree	The average degree of the neighbourhood of each node.	
Distance to road	Distance from the centroid of the building footprint to closest street segment.	3
Length	Street length.	4
Width	Street width.	
Connections	The number of connecting street segments to a node v; minimum and maximum values over all street nodes close to a building.	
Connectivity	Ratio of street length to number of intersections.	1
Intersections	Street intersections.	
Block metrics		
Area, perimeter	Area (a) and length of block footprint (p).	4,5
Shape index	$\frac{p}{4\sqrt{a}}$	5
Fractal dimension	$2*\frac{log p/4}{log a}$	
Block compactness	The degree to which the block shape is close to a circle $\frac{4\pi a}{p}$.	
Number of buildings	Total number of buildings in a block.	
Built-up area	Ratio of built-up area to block area.	
Gross floor area	Sum of floor areas over all buildings.	1
Floor space index	Ratio between gross floor area and block area.	
Open space ratio	Ratio between non-built area and gross floor area.	
Maximum building height	The maximum height of buildings in a block.	5
All storeys per block	Sum of all storeys for all buildings in a block.	
Maximum number of storeys per block	Maximum number of storeys for all buildings in a block.	
Built-up volume	Sum of building volumes in a block.	
Built-up volume mean	Mean building volume per block.	
Built-up volume normalized	Ratio of mean building volume to block area.	
Function richness	Number of different function classes of buildings in a block.	6
Simpson diversity index	Function of number of classes and buildings per class: $1-\frac{\sum_i n_i(n_i-1)}{N*(N-1)}$ where N is the total number of buildings and n_i the buildings of function i.	
NDVI average	Normalized vegetation index averaged over the area of the block.	
Vegetation area	Area covered with vegetation (as estimated using NDVI).	5
Vegetation ratio	Ratio of vegetation area to block area.	

[a] 1 – Berghauser Pont et al., 2019, 2 – Boeing, 2017; 3 – Alexander et al., 2009; 4 – Gil et al., 2012; 5 – Hermosilla et al., 2014; 6 – Lowry & Lowry, 2014.

3.3. Imbalanced learning

The building datasets extracted using Census data are characterized by an unequal distribution of age classes, as illustrated in Table 1. Buildings constructed between 1949 and 1978 constitute the majority class and represent more than 70% of buildings in each city sample. The Random Forest algorithm attempts to minimize the overall error rate and is prone to overemphasizing the accuracy of prediction of the majority class in imbalanced learning problems (Chen, Liaw, & Breiman, 2004; Wurm, Taubenböck, Weigand, & Schmitt, 2017).

A popular method to counteract this tendency is to use a combination of majority class undersampling and minority class oversampling with

the Synthetic Minority Oversampling (SMOTE) technique (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). The tests we performed have shown that under- and oversampling increases sensitivity by 10% without any other loss in classification performance. As recommended in various studies, resampling was performed inside the cross-validation loop (Santos, Soares, Abreu, Araujo, & Santos, 2018) using SMOTE's implementation from the *imbalanced-learn* Python package (Lemaitre, Nogueira, & Aridas, 2017).

3.4. Cross-validation strategy and spatial autocorrelation

Urban structures, though artificially built, exhibit patterns of spatial autocorrelation similar to those encountered with natural phenomena and obey the first law of geography that states that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). Similarities in the appearances of neighbouring residential building can stem from urban planning regulations, building codes legislation, design patterns introduced by construction companies and safety or aesthetic considerations.

Intuitively, there are two sources of spatial correlation in the datasets under analysis. The first is artificially introduced by the way in which metrics concerning street and blocks are defined. Buildings positioned across and along the same street are very likely to share street metrics, and all buildings in an urban block are associated with the same block metrics. The second is an inherent property of the buildings' physical features. We tested this hypothesis by exploring the spatial autocorrelation of selected building features, using Moran's I statistic of autocorrelation. Moran's I is a classical measure of interpreting correlated structures in ecological data (Moran, 1950). Values range between −1 and 1 with strong spatial dispersion at −1, strong spatial clustering at 1, and a random pattern of no correlation for values close to 0 (Cliff & Ord, 1981). The variation of spatial autocorrelation values as a function of the distance between observations is visually represented by a correlogram (Oden, 1984). The correlograms in our analysis were computed and plotted using the R package *ncf* (Bjornstad, 2020).

An analysis of a neighbourhood in Wuppertal showed that the building height and the area containing a building's direct neighbours are highly correlated up to a distance of 50 m between buildings (a correlation of 0.45, p value = .002, at 47 m distance for building height, and a correlation of 0.57, p value = .002, at the same distance for area of direct neighbours, see Fig. 3). Other features like roof angle and footprint area are correlated only for neighbouring buildings (a correlation of 0.49, p value = .002, at 7 m for footprint area, and a correlation of 0.41, p value = .002, at 12 m for roof angle). These findings confirm the intuition that spatial autocorrelation is a characteristic of building features in our analysis.

Spatial autocorrelation leads to optimistically biased prediction results (Pohjankukka, Pahikkala, Nevalainen, & Heikkonen, 2017). Classification and regression models can reproduce training data but fail to output similar performance for new data at different locations (Meyer, Reudenbach, Wöllauer, & Nauss, 2019). For reducing the classification bias, alternative spatial cross-validation methods have been proposed. These methods ensure spatial division between training and test samples either by separating the test and training sets using a space or time buffer (Pohjankukka et al., 2017) or by partitioning the spatial extent of observations into equal-sized non-overlapping sets, a technique called block cross-validation (Meyer et al., 2019; Roberts et al., 2016; Wurm, Stark, Zhu, Weigand, & Taubenböck, 2019).

In order to evaluate the effect of spatial autocorrelation in building age prediction, two methods of model cross-validation were tested. The first is a normal *random* stratified 10-fold cross-validation and the second a *spatial* 10-fold cross-validation built on the "block" principle, where the units of separation are the areas of the urban blocks (see Fig. 4). The buildings' dataset was split by block into 10 sets of blocks containing an approximately equal number of buildings with the added constraint that each set contains buildings of all age classes. While the condition of equal distribution of classes between training and test sample can no longer be reinforced as in the case of a random stratified draw, our tests showed that dividing by block produces similar class distributions between training and test sets for the majority of validation folds.

3.5. Variable importance

Predictor values that are correlated with the underlying spatial structure lead to model overfitting with non-causal predictors (Roberts et al., 2016). When judging their ability to predict new data, building features should be interpreted as independently as possible from the importance derived with spatial autocorrelation. While the performance of feature selection algorithms for Random Forest is well-debated (Degenhardt, Seifert, & Silke Szymczak, 2019; Speiser, Miller, Tooze, & Ip, 2019) the issue of features that include a spatial component is addressed infrequently (Meyer et al., 2019). Feature selection methods that rely on Random Forest's out-of-bag error generated internally during training do not capture the variability in new test data generated by spatial cross-validation (Meyer et al., 2019).

To account for the spatial autocorrelation of building features, we chose the permutation importance method for evaluating feature

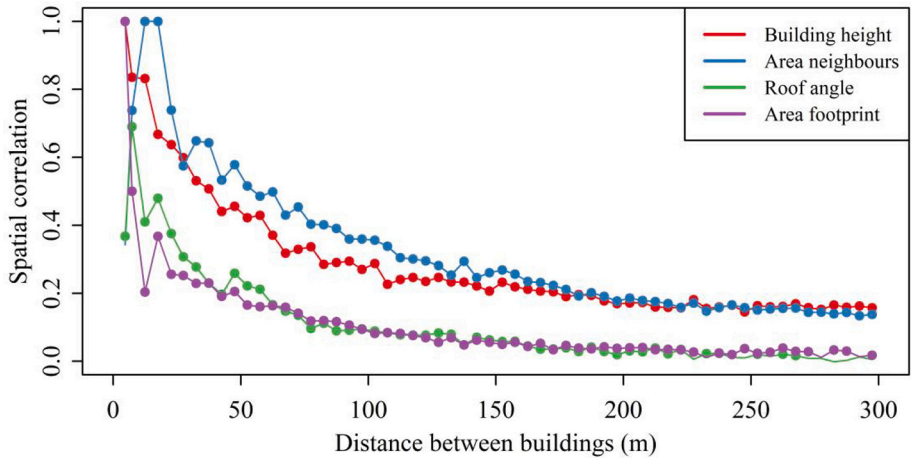


Fig. 3. Correlogram illustrating the difference in spatial autocorrelation values for selected building features. Spatial autocorrelation is measured by Moran's I statistic. Distances between pairs of distinct buildings have been grouped in 5 m intervals, and within each interval the spatial correlation significance is assessed by a randomization test with 500 permutations.

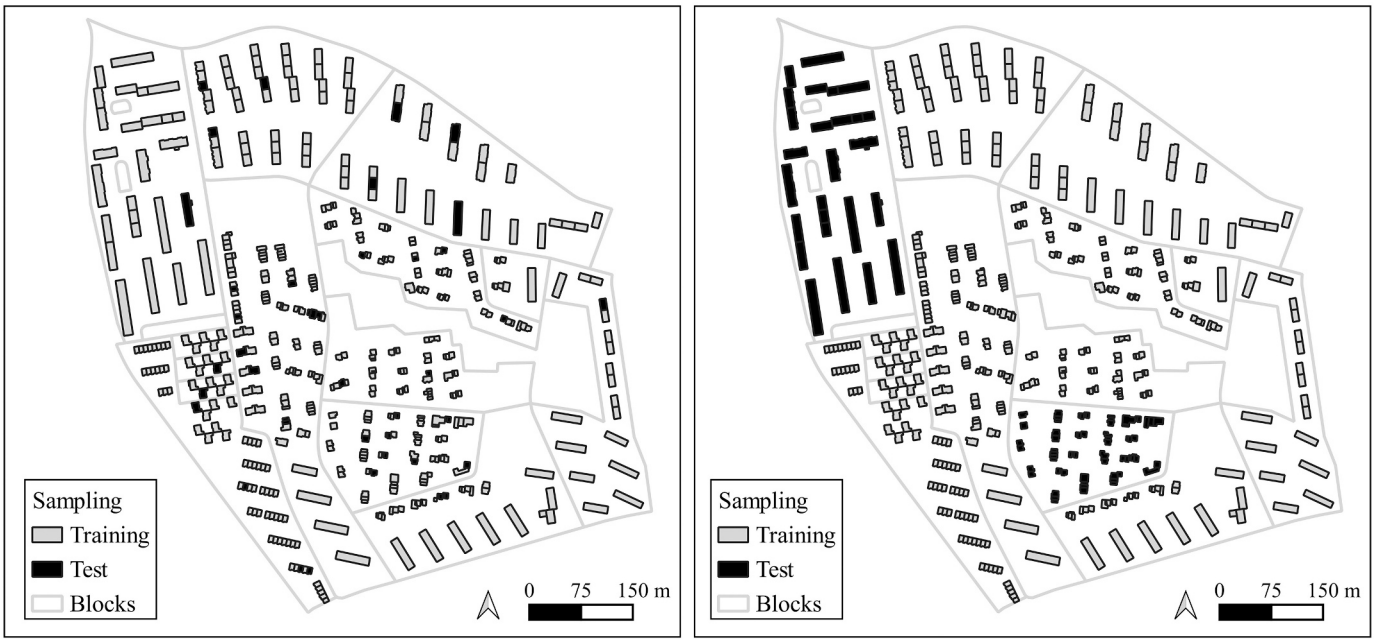


Fig. 4. Strategies for sampling training and test data in 10-fold cross-validation: random (left) and spatial (right).

importance (Breiman, 2001). The algorithm randomly permutes the values of a feature and evaluates feature importance based on the decrease in model performance. The method can be applied on either the training or test set which makes it relevant for tests in spatial cross-validation setup. We verified the results obtained by permutation importance with the fast-forward selection procedure proposed by Meyer et al. (2019), an algorithm that exhaustively searches the feature space by incrementally enlarging the set of relevant features with attributes that increase model performance. We used the permutation importance method from the Python package *sklearn* (Pedregosa et al., 2011) and the fast forward selection method from the R package *CAST* (Meyer, 2020).

4. Heat demand estimation method

For assessing the effect of age misclassification errors on heat demand estimation, we used a simplified building energy model to compute the relative difference between heat demand based on real building age class and heat demand based on predicted building age class. Results for individual building demand were aggregated for various-sized groups of buildings to show the variation in heat estimates at different spatial scales. Our aim is not to provide exact heat demand values per building, but to offer a comparative perspective.

The methodology for calculating building heating demand was adapted from (Dochev, Gorzalka, et al., 2020; Dochev, Seller, & Peters, 2020; Wurm et al., 2021). First, the method estimates the total heated area of a building based on ground area, roof shape, and number of floors. Second, total heat demand is computed by associating the heated area with reference values for heat demand per square meter based on building age class, type and function. For residential buildings, reference heat demand values are extracted from energy estimates for the Institut Wohnen und Umwelt (Institute for Housing and Environment, IWU) typology (Loga et al., 2012). For mixed-residential buildings, reference heat demand values are found in the VDI 3807 report (Verein Deutscher Ingenieure, 2014) based on the non-residential function of the building, e.g. administrative, sales or services. Additional information on the energy estimation methodology applied can be consulted in the Supplementary Material.

Aggregating the heat demand of individual buildings should ideally respect the actual spatial structure of neighbourhoods. Urban development throughout the years leads to an uneven spatial distribution of construction epochs. In many European cities, as is the case for the cities under analysis, the majority of buildings built after 1995 are situated more likely in the outer regions of the city than in the central areas and form age-homogeneous neighbourhoods (see Fig. 5). Given that new buildings have on average higher energy efficiency levels than older buildings, it is expected that the energy demand will be unequal across neighbourhoods. We were able to account for the spatial dispersion of building ages across the city for Wuppertal in our energy calculation scenario, since the available dataset represents the entirety of residential buildings in the city. In other words, for this dataset, aggregated heat demands are estimated for increasingly large groups of neighbouring buildings chosen randomly from different locations across the area of the city. For the other cities under analysis, the sparser, incomplete datasets render such initiative less conclusive. In this case, heat demand is aggregated for groups of buildings randomly selected from the available building dataset.

5. Results

We explored the performance of three classification models defined as combinations of different types of building features. *Model 1* includes building features related to overall building appearance, shape of building footprint and relationship with neighbouring buildings. *Model 2* extends the first model by including street features. *Model 3* further extends the second model with attributes that describe the urban block where the building is situated. The spatial relation between data used for learning and data used for age prediction determines a parallel division of models into: *local models*, where the age class is inferred using data from buildings in the same city, and *regional models*, where the age class of buildings in one city is inferred from data gathered from other cities. *Local* models are further on defined by the cross-validation strategy used and are divided into: local models using a *random* 10-fold cross-validation and local models using a *spatial* 10-fold cross-validation.

5.1. Local classification

Classification performance is summarised by accuracy and sensitivity, where the accuracy is the ratio of correctly labelled observations to the size of the data and the sensitivity is the ratio of correct positive

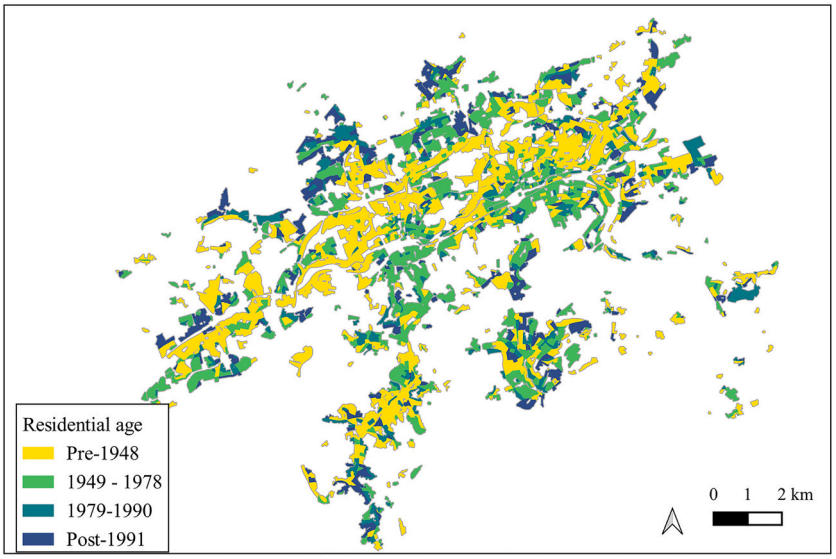


Fig. 5. Distribution of residential buildings across the city of Wuppertal, grouped according to broad age classes.

predictions to the total number of actual observations per class. In a random cross-validation scenario, accuracy and sensitivity increase with the addition of new features, irrespective of study area. Conversely, the differences in classification performance between cities decrease with higher number of features. The sensitivity of classification with *Model 3* for all cities and all age classes is greater than 70%, as shown in Table 4.

Common patterns can be observed in the errors of classification across all cities (Fig. 6). For all cities and all models, the class of buildings built between 1949 and 1978 is correctly classified in more than 85% of the cases. There is minimal misclassification between buildings built before 1949 and buildings built after 1978. For Bielefeld and Münster, where the predictive ability is lower than for the other cities, we observe a greater misclassification between the age classes of buildings built after 1990. The classification errors obtained with *Models 2* and *3* are illustrated in the Supplementary Material.

From random stratified cross-validation to spatial cross-validation, the sensitivity of the models decreases significantly, as shown in Table 4. The impact on overall accuracy is less severe due to unequal distribution of age classes in the sample. The difference of classification performance between the two strategies is especially significant for buildings built after 1979. The power of identification of buildings built after 1979 is weak, with an average of less than 20% per class. Within this period buildings built after 1996 have higher classification accuracies for most cities. Buildings built in the 1919–1948 period have, over all cities, a sensitivity score of 45%. The sensitivity of classification for buildings built before 1919 ranges between 1% and 40%. The results for this class are correlated with the size of the age sample. These observations are illustrated in the results obtained for the city of Dortmund, shown in Table 5. Sensitivity scores per age class for the other cities under analysis can be consulted in the Supplementary Material.

To validate our method, the classification is also tested on the complete building dataset of the city of Wuppertal. Results are shown in Table 6. For random stratified 10-fold cross-validation the sensitivities of classification for *Model 1* and *Model 2* are within the ranges of values obtained for the other cities while the score for *Model 3* decreases by more than 10% compared with the sensitivity of classification for the other city datasets. This is due to the fact that the new dataset presents an increased spatial density of heterogeneous building ages. Labelling buildings of different ages with identical attributes, either due to the street along which they are situated or the block that encloses them, renders this attribute less relevant in the classification process. Fig. 7 illustrates the results obtained for age prediction using for classification *Model 2* with random cross-validation.

5.2. Regional classification

Testing the transferability of learning between geographical regions is accomplished by training the classification model on the combined data of 6 of the 7 cities under analysis and testing it on the remaining city. The process has been repeated with each city as a test region. The sensitivity of prediction of each model decreases significantly as the results in Table 7 show. Breaking down the results by age class, we have observed that buildings constructed between 1949 and 1978 are identified accurately in more than 80% of cases. The time period between 1919 and 1948 is labelled with an average success rate of 40% while the oldest buildings, built before 1919 have the lowest sensitivity scores (between 7 and 19%). The greatest variation in results is observed for new buildings, built after 1995 (accuracies between 28 and 61%). Applying a regional classification model on the city of Wuppertal yields a sensitivity score within the range observed for the other cities and a lower overall accuracy of 43%.

5.3. Variable importance

The theoretical work of Gregorutti, Michel, and Saint-Pierre (2017) shows that the permutation importance of a variable is strongly dependent on the degree of correlation with other variables, and that, in a regression framework, the higher the number of correlated features,

Table 4
Range of results over 7 cities for local models with random and spatial 10-fold cross-validation.

Cross-validation strategy	Random			Spatial		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Mean sensitivity (%)	43–65	55–73	78–86	21–35	24–38	14–24
Mean Kappa	0.44–0.64	0.55–0.73	0.77–0.88	0.28–0.47	0.34–0.52	0.11–0.45
Overall accuracy (%)	73–87	81–90	90–96	68–81	74–84	74–82

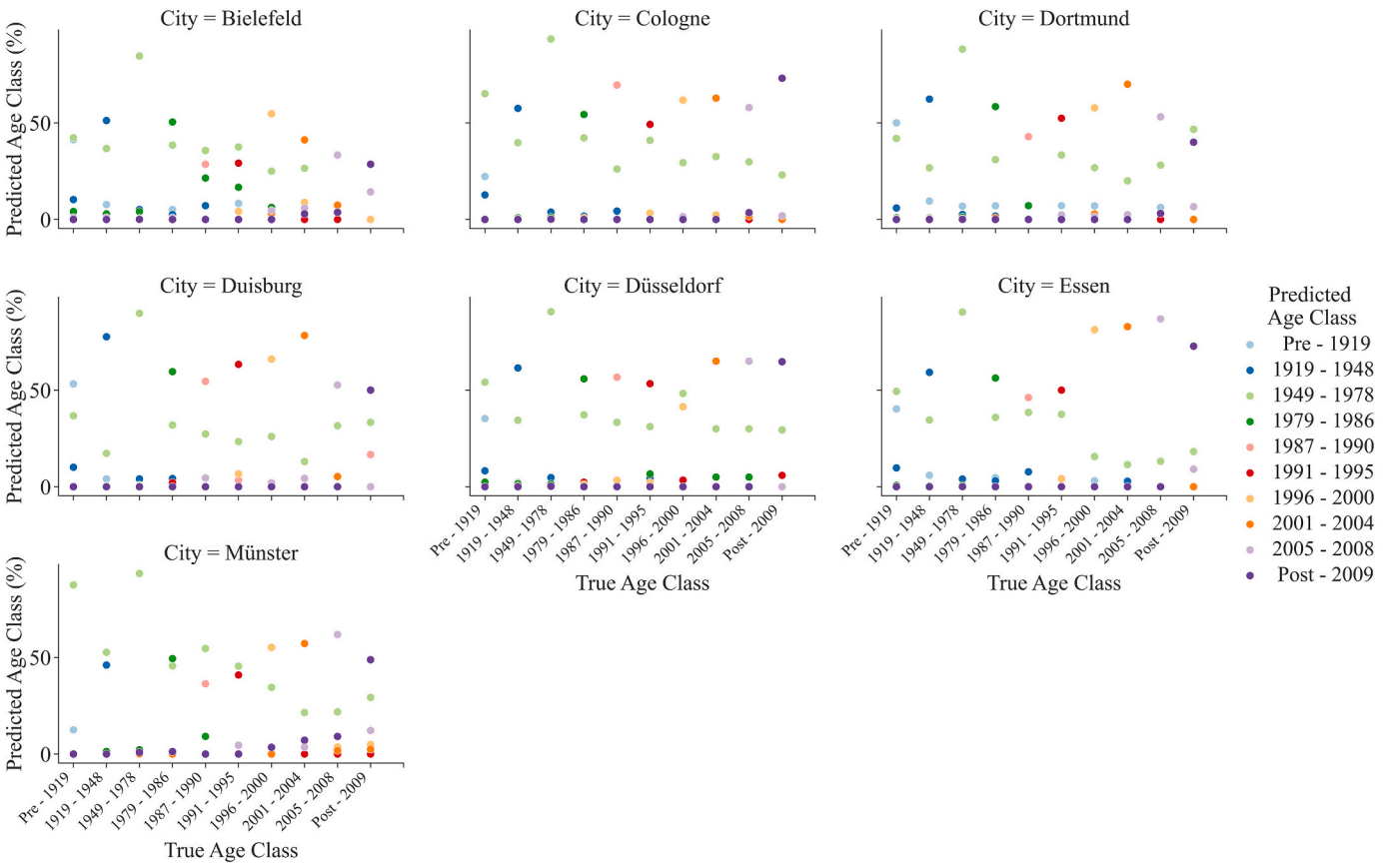


Fig. 6. Actual and predicted building age with *Model 1* and random 10-fold cross-validation.

Table 5
Sensitivity results for the classification of Dortmund data.

Classification model	Cross-validation	Pre–1919	1919–1948	1949–1978	1979–1986	1987–1990	1991–1995	1996–2000	2001–2004	2005–2008	Post–2009
Model 1	Random	50.2	62.3	88.2	58.2	40.7	51.9	58.3	68.2	53.5	39.5
	Spatial	37.8	47.8	85.8	19.6	12.4	11.0	13.6	13.1	10.5	5.1
Model 2	Random	57.0	64.9	92.5	62.3	54.0	58.9	62.9	74.0	62.1	51.1
	Spatial	41.8	47.4	90.4	18.2	18.4	13.1	16.4	17.5	9.8	6.8
Model 3	Random	83.2	81.4	96.4	82.8	83.3	81.5	82.8	85.3	84.4	86.1
	Spatial	35.9	33.2	95.6	3.1	0.0	1.2	4.6	6.3	2.9	17.5

Table 6
Results for Wuppertal for random and spatial10-fold cross-validation.

Cross-validation	Random			Spatial		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Mean sensitivity (%)	53	58	65	30	32	26
Mean Kappa	0.50	0.54	0.61	0.35	0.37	0.32
Overall accuracy (%)	61	65	70	52	55	53

the lower their permutation importance score. We have chosen a simple way in which to counteract this effect and assess as accurately as possible the importance of variables by filtering out predictor variables which have a coefficient of correlation greater than 0.90. On account of computational time, we proceed with the analysis of continuous features only, since the singular categorical feature in our analysis, building type, was found in our initial tests to be the least relevant attribute.

We found the roof angle to be the most important feature for classification, at all spatial scales. Building height and volume are the other two attributes related to building appearance that are important throughout all tests. Shape exchange and shape detour index are important features related to the building footprint shape. The results concerning the importance of these variables are consistent between spatial and random cross-validation strategies. The area including neighbouring buildings is an attribute that is judged important in random cross-validation and that loses its importance in a spatial setting.

Street features such as street closeness centrality and, to a lesser extent, street width and length are important with local models of classification, both for random and spatial cross-validation scenarios (see Fig. 8). Their importance is however diminished in region models. Street closeness centrality was found to be the single most important street feature for classification with spatial cross-validation in Wuppertal. The average building height per block is the most important block feature in local models. The importance of block metrics decreases significantly in regional and local spatial cross-validation models. The comparative importance of all variables between local and regional

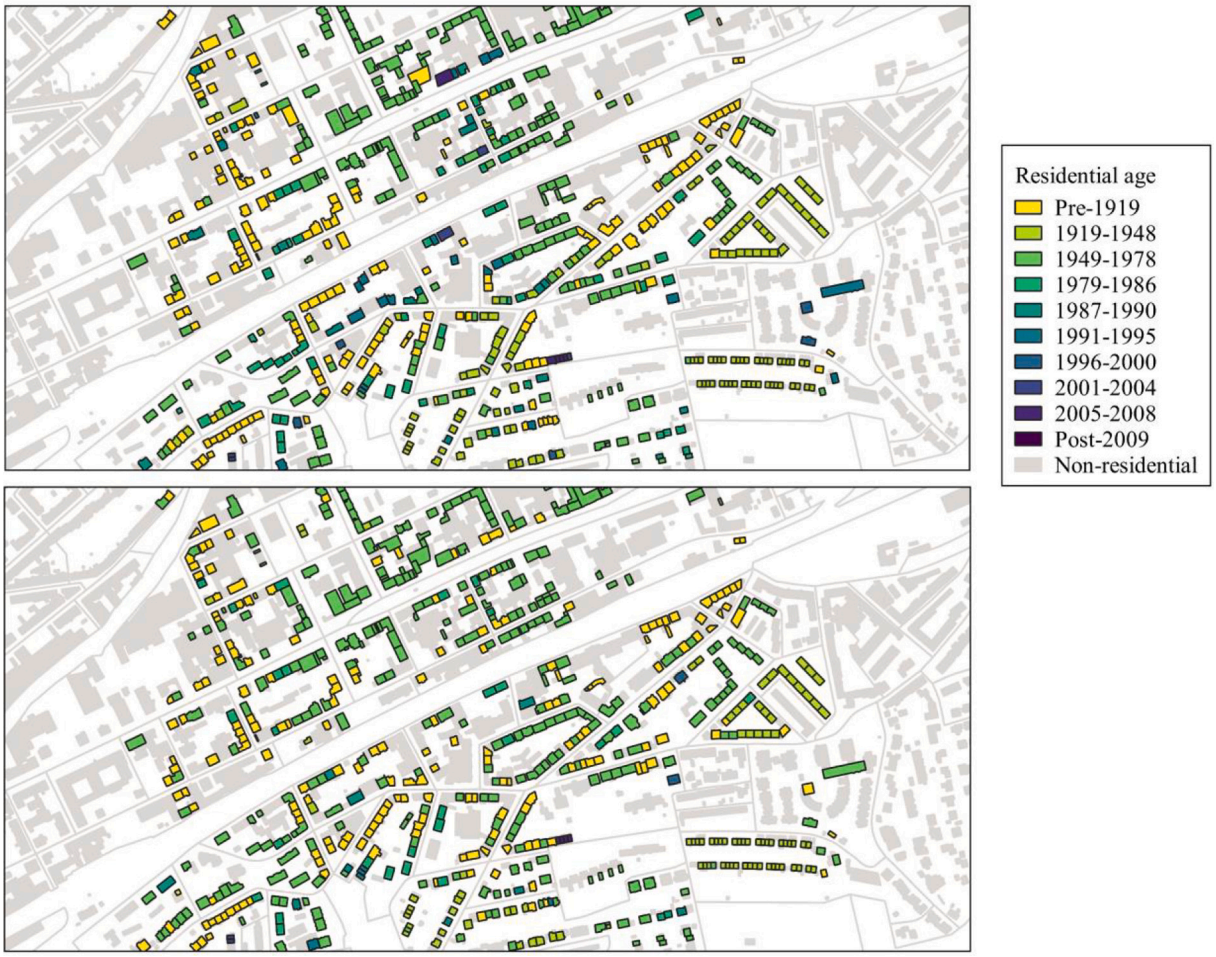


Fig. 7. Actual (top) and predicted (bottom) building age. Classification performed with random cross-validation, using *Model 2*. Illustration of a neighbourhood in Wuppertal.

Table 7
Range of results over selected cities for region models.

Classification model	Model 1	Model 2	Model 3
Mean sensitivity (%)	18–22	17–21	13–16
Mean Kappa	0.21–0.33	0.2–0.34	0.13–0.27
Overall accuracy (%)	69–74	70–77	73–80

models can be consulted in the Supplementary Material.

Reducing the set of features to a limited set of the most important variables yields higher sensitivity scores for all cities, when using a local random classification model. For example, using the ten most important building and street features identified across all cities, classification sensitivity is improved by 5% to 13% for all cities except Wuppertal, where the improvement is only of 3%. However, applying the same process for local spatial models, the differences in sensitivity scores ranges from −5% to 2%. To analyse further the effect of reducing the size of the input feature set we recommend more complex methods of feature selection (Degenhardt et al., 2019; Speiser et al., 2019).

5.4. Heat demand

Heat demand estimates were derived for buildings for which age has been predicted with different models and cross-validation strategies with the aim of testing the impact of a wide range of classification performances: *Model 1* with spatial 10-fold cross-validation (52% accuracy), *Model 1* with random 10-fold cross-validation (61% accuracy) and *Model 3* with random 10-fold cross-validation (70% accuracy). As expected, the percentage difference between the reference and predicted heat demand is greater when age class is predicted with lower accuracies. For the Wuppertal dataset, the difference can reach 48% for individual buildings. By aggregating heat demand for an increasing number of buildings, results converge to stable results. For groups of more than 500 buildings, the average heat demand difference over 500 iterations ranges between 4% and 7.6%, depending on the classification model used. Fig. 9 illustrates these observations. Concerning the spatial distribution of errors, it can be observed that the errors are greater for areas comprising of newer buildings (see Fig. 10 and Fig. 11).

The results are reproduced for the other 7 cities, as shown in Fig. 12. The differences in energy errors between classification models correlate with the models' accuracy scores, the ratio of correctly labelled observations to the size of input data. The difference in energy estimates is smaller than for Wuppertal since classification models exhibit better accuracy on average for the other cities and the distribution of building ages is different. Compared with Wuppertal, these datasets include a smaller number of buildings built after 1979. These buildings are in general harder to identify, and when misclassified as older buildings, lead to high heat demand estimates, which in turn increases the aggregated heat demand. An example of such a dataset is the city of Münster, where the difference in heat demand is higher than for other cities with comparable accuracies.

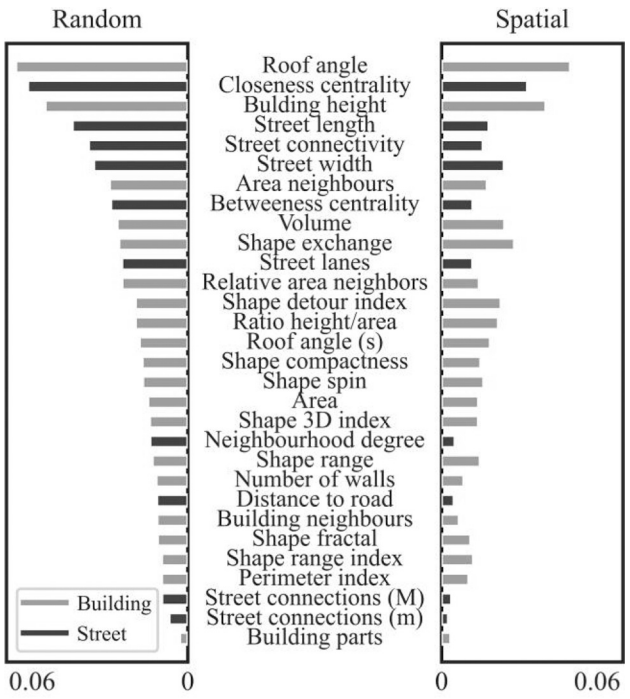


Fig. 8. Building and street feature importance in local models with random (left) and spatial (right) 10-fold cross-validation. Importance scores were averaged over all cities. Annotations: m – minimum, M – maximum, s – standard deviation.

6. Discussion

Results have shown that the combination of classical building features and street and block features can improve the accuracy of classification of building age, in specific model training scenarios. The observed improvement is due to underlying spatial autocorrelation trends in the chosen urban form metrics. Overall, we observed that classification improves for datasets where the spatial distribution of age classes tends to be homogeneous. An important outcome of our study is that attributes describing the building geometry are spatially autocorrelated, similarly to environmental or ecological data, and that spatial autocorrelation should be a factor of analysis in studies dealing with buildings and urban morphology research, similarly to other urban science studies that focus on air pollution (Bertazzon, Johnson, Eccles, & Kaplan, 2015), surface temperature (Yin, Yuan, Lu, Huang, & Liu, 2018), land use (Fan & Myint, 2014) or energy use (Tian, Song, & Li, 2014).

The large-scale data available allowed us also to analyse how classification accuracy varies across different scenarios of classification, where the model learns age relevant characteristics from buildings in close vicinity or buildings in different neighbourhoods or even different cities. The transferability of learning results between geographical regions is a topic that has not yet been explored in previous similar studies. Our results indicate that in scenarios where training and test data are spatially separated, classification accuracy deteriorates significantly, irrespective of the type of metrics used for classification. For predicting building construction type, the geographical transfer of learning between cities produced only a slight decrease in performance according to Wurm et al. (2016). For building age prediction, we report a decrease of more than 10% in accuracy and more than 20% in sensitivity. Further applications of the building age classification approach should consider this aspect during training design. In what concerns the differences in the accuracy of classification between building type and building age, it could be presumed that the building type is a building property inherently defined through notions such as size, shape, connections with other buildings, while the building age is a feature that refers at the same time at construction methods and materials, external appearance and internal planning of the building (Senate Department for Urban Development and the Environment, 2018). Therefore, beyond geometrical properties of the building appearance and its spatial relationships with the environment, the inclusion of other data sources, such as façade images of high-resolution rooftop images, could be beneficial for improved classification accuracy of building age.

The availability of data from different locations highlighted similarities and differences between cities. The variables found most important for prediction are similar across regions. The disagreement consists in different prediction accuracies for the oldest and newest buildings between cities. We estimate that this is due to different distribution of ages in the available building sample and furthermore, to differences in age-typical building appearances between regions. The latter perspective, the morphological similarities of German buildings per epoch and geographical location, is an avenue of research which deserves closer attention in the context of understanding and improving the accuracy of age prediction for the national building stock

In previous studies, buildings built after 1980 yield the lowest accuracy over all age classes (Alexander et al., 2009; Rosser, Boyd, et al., 2019). These results have been reproduced in this paper. The power of identification of buildings built after 1979 is weaker than for buildings built before this year. We observed however that buildings built after 1996 have higher classification accuracies, as a single age class, than buildings built between 1979 and 1996. Since new buildings are underrepresented in most age prediction studies, we recommend further tests on larger samples of this class of buildings. Our findings diverge from the aforementioned studies in what concerns the category of buildings built before 1919. In the UK-based studies, these buildings are classified with the highest accuracy, while our results indicate an ability of

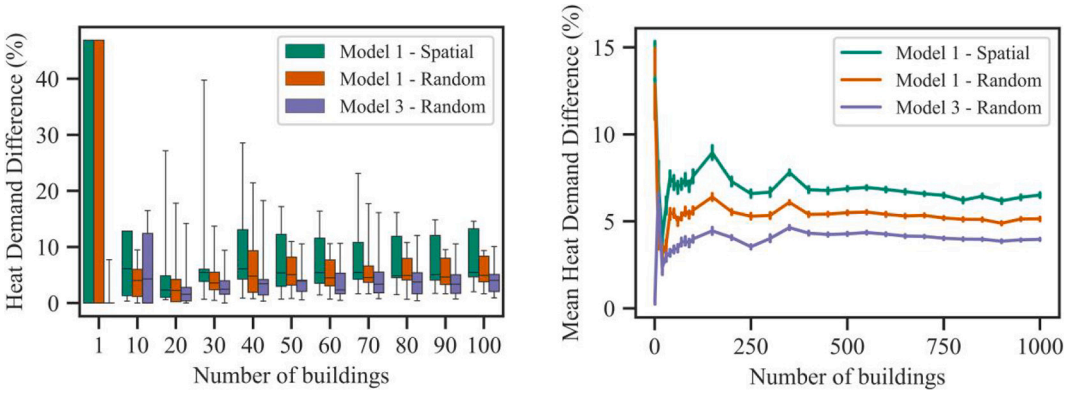


Fig. 9. Difference between actual and model estimated heat demand, per number of buildings, for Wuppertal. Left, the distribution of results for groups of less than 100 buildings is presented. Left, the mean difference for groups of 1 up to 1000 buildings is displayed as a 95% confidence interval over 500 iterations.

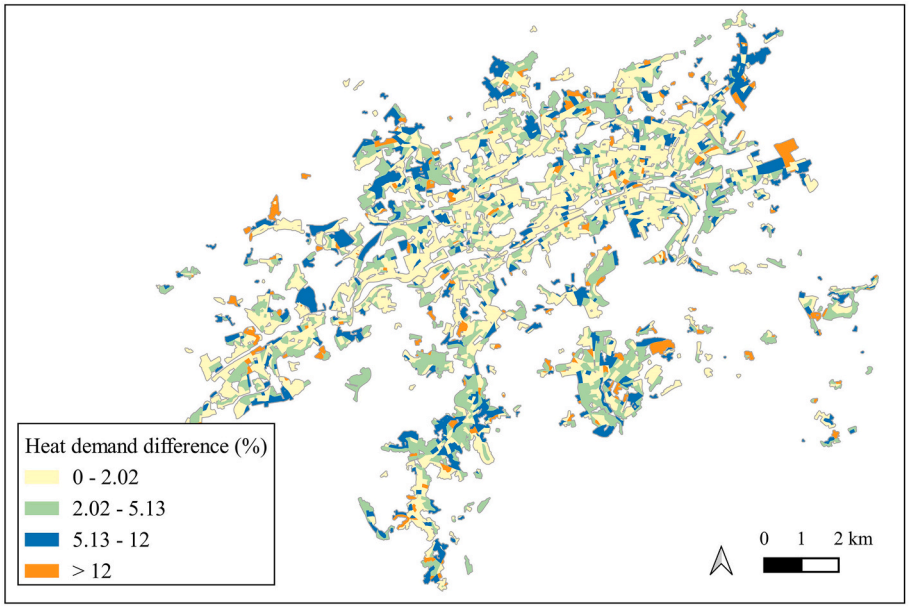


Fig. 10. Differences between actual and model estimated heat demand (*Model 1* with random 10-fold cross-validation), for heat demand aggregated per block, color-coded according to the percentiles of the distribution of differences (50%, 75%, 92.5%). Illustration of the city of Wuppertal.

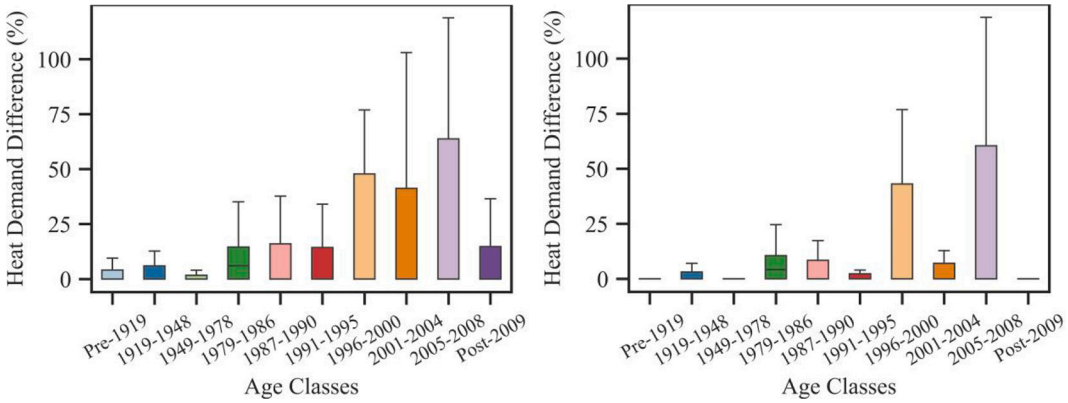


Fig. 11. Distribution of absolute differences between actual and model estimated heat demand for individual buildings in Wuppertal, grouped by the age class of buildings, for classification using *Model 1* with random cross-validation (left) and *Model 3* with random cross-validation (right). Outlier values are not shown, where outliers are values outside the 1.5 IQR (interquartile range) interval.

identification inferior to buildings built between 1919 and 1978. The difference could be due to the fact that in UK buildings built before 1919 are distinguishable from interwar buildings and more recent residential constructions due to the particularities of architectural styles and plan forms (Alexander et al., 2009).

The buildings and infrastructure in both UK and Germany have been massively affected by World War II. In Germany, World War II brought about the large-scale destruction of cities and at the same time, a large influx of refugees, which lead to a long reconstruction process (Pahl-Weber & Henckel, 2008). Immediately after the war, the necessity to satisfy basic housing needs resulted in simple three- or four-storey rows of apartments, aligned perpendicularly to the street and without courtyards (Gieselmann, 2011). This phase was followed in the 60s and 70s by the construction of large-scale housing estates and developments in the urban fringes (Pahl-Weber & Henckel, 2008), with a higher accent on open and green space surrounding buildings. The reconstruction was followed by a phase of urban renewal in the 80s, which also included the construction of buildings to fill in gaps left by the war, and a preference for smaller dwellings, like row houses and urban villas (Gieselmann, 2011). This trend continued in the 90s, when individual detached houses were developed in peri-urban areas, mostly for families and retired persons seeking open spaces outside the inner city (Caruso, 2001). Finally, contemporary German architectural style can be characterized as "straight-forward, low key and cautious" (Hackel, 2004). This evolution of the German building stock seems to be mostly characterized by its functional aspect, and can be associated with considerations of size, available space and urban development, as well as with particularities of appearance. For this reason, identifying building age is a difficult process, and it should be integrated into the analysis of the urban environment.

The model identifies most accurately buildings built before 1978, which form the majority of the building stock in Germany. The ability to correctly identify older buildings is an important factor for estimating energy demands, since this category introduces a high degree of uncertainty in energy demand calculation (Firth et al., 2010). Recent studies show for example that building occupants adjust their energy behaviour to building characteristics, and energy models that do not account for the relationship between building age and occupant behaviour overestimate the energy use of older buildings and underestimate that of newer, more energy-efficient, buildings (Guerra Santin,

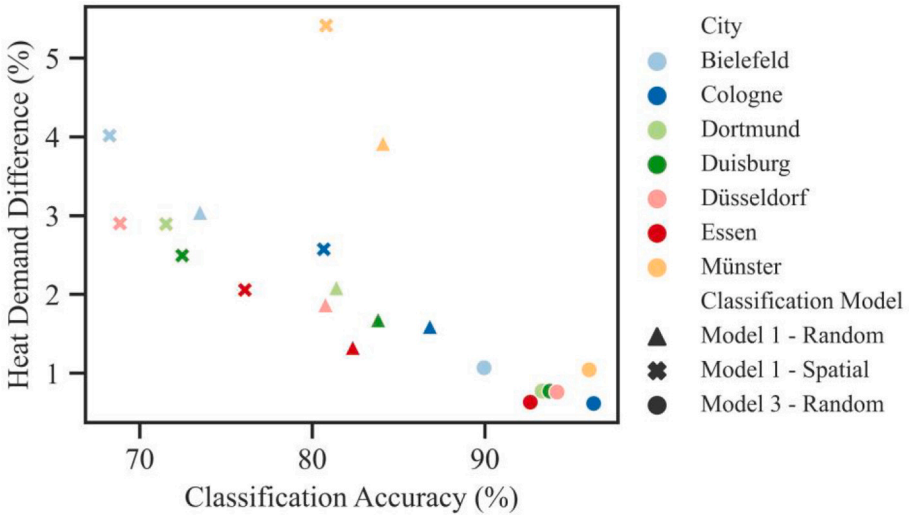


Fig. 12. Difference between actual and model estimated heat demand for groups of 100 buildings, in relationship with the accuracy of classification, per city and classification model. The heat demand difference values are the mean values over 1000 iterations.

Itard, & Visscher, 2009; Van den Brom, Hansen, Gram-Hanssen, Meijer, & Visscher, 2019). Additionally, our results indicate that the classification confusion between the oldest, most energy-consuming buildings, and the newest buildings is limited, which diminishes the potential error when using the age inferred from classification in energy demand estimation. Errors in energy estimation are expected to be higher for groups of new buildings, an important factor to consider when applying the age classification model to specific study areas for energy modelling purposes.

The energy demand calculation exercise we reported provides insight into how the accuracy of building age classification impacts estimation of residential heat demand. Previous age prediction studies suggest that despite age classification errors, the relative difference between energy demand inferred from real age and demand inferred from predicted age is small at a city level (Alexander et al., 2009; Tooke et al., 2014). The energy modelling results we obtained confirm and extend these observations by showing that the difference remains low also at block level. For the city of Wuppertal, for groups of 100 buildings or more, model-predicted energy estimates are within 4% to 7.6% of the actual energy demand value.

The proposed approach of associating age classes to individual buildings has the potential to reduce the uncertainty in modelling energy consumption compared with methods of associating age classes to groups of buildings based on aggregated statistic data. Zirak et al. (2020) have used the aggregated Census age information to compute a complex model of energy demand and concluded that heating demand for an entire city is acceptable, yielding a difference of 4.5% compared with municipality reference data. Our heat demand estimation produces similar results and at a finer spatial scale. Precise localisation of above-average energy consumption can inform local authorities in the preparation of energy-efficiency information campaigns, can pinpoint barriers in the process of retrofit decision-making when coupling this information with building type and tenancy data, e.g. joint tenancy in multi-family houses (Meijer, Itard, & Sunikka-Blank, 2009), and can direct the implementation of extensions of central heating networks, which are considered more energy-efficient than individual heating solutions. All these benefits constitute a strong argument for using age prediction models, but further tests on heat demand computation are advised. Dochev, Gorzalka, et al. (2020) reported that the heat demand estimation method we adopted in this work results in lower values than methods that employ a physical building energy model. Further work in estimating the gap between energy demand based on actual building age and the energy demand based on the classified building age could include a more complex energy modelling method coupled with a sensitivity analysis.

Despite the increased efficiency of newly constructed buildings, which are regulated by increasingly stringent building regulations, the identification of old buildings is still an important factor in decision-making. In EU, 90% of the building stock is constructed before 1990 and 50% before 1970, the decade when energy efficiency restrictions were beginning to be introduced into building code regulations (Filippidou & Jimenez Navarro, 2019). The yearly rate of renovation of European buildings is between 0.4 and 1.2%, with most of the measures being minor adjustments and with different rates across EU countries (Filippidou & Jimenez Navarro, 2019). The status of renovations in European countries makes it clear that the potential of energy efficiency of the building stock is not reached, and the identification of buildings ages, in cases where the information is missing, can help the implementation of targeted retrofit measures. Furthermore, the retrofit of these buildings can have an impact on other aspects of energy consumption, not directly related to space heating. An example would be the improvement of natural light inside a dwelling, and a consequent reduction in lightning energy consumption. Beyond energy efficiency, renovations can bring about other co-benefits such as increased comfort and aesthetics, improved health due to higher levels of thermal comfort, structural safety and noise insulation (Almeida & Ferreira, 2017), and even more so for occupants of older building than newer ones.

One of the limitations of this study is that the sampling method we employed for deducing individual building age classes using uniform grid cells from the census has some important drawbacks and results in a building sample that is incomplete and arguably not representative of the entire building stock. First, the different age classes are disproportionately represented in the building sample. Second, it introduces a bias towards identifying age-homogeneous neighbourhoods. Results obtained for the incomplete building datasets have been validated on an additional, complete, building stock dataset but further tests are recommended. Another limitation, though not inherent to our study, is the lack of data concerning building renovation. In addition, since not all renovation activities are notified to building authorities, and there is no sustained national monitoring of renovation effects in European countries, the data can be unreliable (Meijer et al., 2009). Considering that renovations include not only changes in building appearance and insulation properties, but also component replacement, the present method would not be able to distinguish between retrofitted and non-retrofitted buildings, which is an important aspect of energy consumption, with simulations showing that renovation impacts energy performance

substantially (Sandberg et al., 2016).

Another source of uncertainty related to the age extraction method concerns heat demand estimation. As mentioned in section 4 and fully explained in the Supplementary Material, we have employed a method based on IWU age classes and made an approximate transition from the available Census classes. Considering that the period 1949–1978 comprises of three IWU classes (Loga et al., 2012), with decreasing reference values for heat demand, the misclassification between buildings built after 1979 and building built between 1949 and 1978 could result in smaller differences in heat demand estimates than what our analysis showed. Training a classification model with information labelled with IWU age classes could provide a better perspective of the accuracy of heat demand estimation, provided that such training data is available.

7. Conclusion

This paper shows the potential of improving the accuracy of age class prediction for residential buildings by associating buildings with metrics describing complex footprint shape, street and block characteristics. A selected set of these attributes has been found relevant for classification at different spatial scales. Due to the underlying spatial autocorrelation of predictors, a higher accuracy of classification is obtained when training and test samples are located in close vicinity. Conversely, classification accuracy decreases for predictions in new locations. The findings suggest that the advantages of transferring building age classification learning between geographical regions are limited.

Buildings built before 1978 are better classified than newer buildings irrespective of the spatial scale of training design. New buildings are however underrepresented in this study and are divided into age classes with a fine granularity. Imbalanced classification problems are notoriously difficult and we recommend further tests on extended datasets for clarifying the effect of sample size on variations in classification accuracy between age classes. Alternatively, broader or different age classes could be defined to fit the purposes of specific applications.

The theoretical exercise of energy calculation demonstrates that using model-classified building age classes for inferring heat demand produces estimates that are close to values inferred from actual building age. Moreover, heat demand predictions can be achieved with acceptable accuracy from block to city level. The relevance for energy applications of the proposed age classification approach relies in the ability to identify and study variations in energy demand at fine spatial resolution with reduced efforts in the acquisition of building age information.

Financial disclosure/acknowledgements

This work was supported by the German Federal Ministry for Economic Affairs and Energy within the 6th Energy Research Programme of the Federal Government/Applied Energy Research (grant code O20E-SY 48,192 ANSWER-Kommunal).

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest associated with this publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compenvurbsys.2021.101637.

References

Alexander, D., Lannon, S., & Linovski, O. (2009). *The identification of analysis of regional building stock characteristics using map based data. Building Simulation.* Glasgow: Eleventh International IBPSA Conference. Retrieved from http://www.ibpsa.org/proceedings/BS2009/BS09_1421_1428.pdf.

Almeida, M., & Ferreira, M. (2017). Cost effective energy and carbon emissions optimization in building renovation (annex 56). *Energy and Buildings, 152*, 718–738. https://doi.org/10.1016/j.enbuild.2017.07.050.

Angel, S., Parent, J., & Civco, D. L. (2010). Ten compactness properties of circles: Measuring shape in geography. *The Canadian Geographer, 54*(4), 441–461. https://doi.org/10.1111/j.1541-0064.2009.00304.x.

Athanassiadis, A. (2019). Urban metabolism and Open data: Opportunities and challenges for urban resource efficiency. In S. Hawken, H. Han, & C. Pettit (Eds.), *Open Cities | Open Data.* Singapore: Palgrave Macmillan. https://doi.org/10.1007/978-981-13-6605-5_8.

Belgiu, M., & Dragut, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Berghauser Pont, M., Stavroulaki, G., Bobkova, E., Gil, J., Marcus, L., Olsson, J., … Legeby, A. (2019). The spatial distribution and frequency of street, plot and building types across five European cities. *Environment and Planning B: Urban Analytics and City Science, 46*(7). https://doi.org/10.1177/2399808319857450.

Bertazzon, S., Johnson, M., Eccles, K., & Kaplan, G. G. (2015). Accounting for spatial effects in land use regression for urban air pollution modeling. *Spatial and Spatio-Temporal Epidemiology, 14-15*, 9–21. https://doi.org/10.1016/j.sste.2015.06.002.

Biljecki, F., & Sindram, M. (2017). Estimating building age with 3D GIS. In , *4. ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*. https://doi.org/10.5194/isprs-annals-IV-4-W5-17-2017. Melbourne.

Bjornstad, O. N. (2020). Ncf: Spatial covariance functions. R package version 1.2–9. Retrieved from https://CRAN.R-project.org/package=ncf.

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems, 65*, 126–139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324.

Bundesamt für Kartographie und Geodäsie. (2019). *Geographische Gitter für Deutschland in Lambert-Projektion (GeoGitter Inspire). © GeoBasis-DE / BKG (2019)*. Retrieved from Bundesamt für Kartographie und Geodäsie https://gdz.bkg.bund.de/index.php/default/open-data/geographische-gitter-fur-deutschland-in-lambert-projektion-geogitter-inspire.html.

Caruso, G. (2001). Peri-urbanisation : The situation in Europe. In *A bibliographical note and survey of studies in the Netherlands, Belgium, Great Britain, Germany, Italy and the Nordic countries.* Retrieved from https://orbilu.uni.lu/bitstream/10993/10153/1/Caruso_PeriUrbanEuropeDATAR.pdf.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953.

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. Retrieved from https://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf.

Cliff, A., & Ord, J. (1981). *Spatial processes : Models & applications. London: Pion*.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104.

Creutzig, F., Baiocchi, G., Bierkandt, R., Pichler, P.-P., & Seto, K. C. (2015). Global typology of urban energy use and potentials for an urbanization mitigation wedge. *Proceedings of the National Academy of Science, 112*, 6283–6288. https://doi.org/10.1073/pnas.1315545112.

Degenhardt, F., Seifert, S., & Silke Szymczak, S. (2019). Evaluation of variable selection methods for random forests and omics data sets. *Briefings in Bioinformatics, 20*(2), 492–503. https://doi.org/10.1093/bib/bbx124.

Dochev, I., Gorzalka, P., Weiler, V., Estevam Schmiedt, J., Linkiewicz, M., Eicker, U., … Schröter, B. (2020). Calculating urban heat demands: An analysis of two modelling approaches and remote sensing for input data and validation. *Energy and Buildings*. https://doi.org/10.1016/j.enbuild.2020.110378.

Dochev, I., Seller, H., & Peters, I. (2020). Assigning energetic archetypes to a digital cadastre and estimating building heat demand. an example from Hamburg, Germany. *Environmental and Climate Technologies*, 233–253. https://doi.org/10.2478/rtuect-2020-0014.

Droin, A., Wurm, M., & Sulzer, W. (2020). Semantic labelling of building types. A comparison of two approaches using random Forest and deep learning. In *Wissenschaftlich-Technische Jahrestagung der DGPF, Band 29*.

Economidou, M., Atanasiu, B., Staniaszek, D., Maio, J., Nolte, I., Rapf, O., … Zinetti, S. (2011). Europe's buildings under the microscope. In *A country-by-country review of the energy performance of buildings*. Buildings Performance Institute Europe (BPIE).

Eurostat. (2020, June 6). Energy consumption in households. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_consumption_in_households#Energy_products_used_in_the_residential_sector.

Fan, C., & Myint, S. (2014). A comparison of spatial autocorrelation indices and landscape metrics in measuring urban landscape fragmentation. *Landscape and Urban Planning, 121*, 117–128. https://doi.org/10.1016/j.landurbplan.2013.10.002.

Filippidou, F., & Jimenez Navarro, J. P. (2019). *Achieving the cost-effective energy transformation of Europe's buildings*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/278207.

Firth, S. K., Lomas, K., & Wright, A. (2010). Targeting household energy efficiency measures using sensitivity analysis. *Building Research and Information, 38*(1), 25–41. https://doi.org/10.1080/09613210903236706.

Geobasis, N. R. W. (2019). Bestandsdatenauszug ohne Eigentümer. Retrieved from Open Geodata NRW https://www.opengeodata.nrw.de/produkte/geobasis/lk/bda_oe_xml/.

Gieselmann, R. (2011). Historical development of housing plans. In O. Heckmann, & F. Schneider (Eds.), *Floor plan manual* (pp. S. 14–25). Basel: Birkhauser Verlag AG. https://doi.org/10.1515/9783034610407.

Gil, J., Beirão, J. N., Montenegro, N., & Duarte, J. P. (2012). On the discovery of urban typologies: Data mining the many dimensions of urban form. *Urban Morphology, 16*, 27–40.

Gregorutti, B., Michel, B., & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Statistics and Computing*, S. 659–678. https://doi.org/10.1007/s11222-016-9646-1.

Guerra Santin, O., Itard, L., & Visscher, H. (2009). The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. *Energy and Buildings, 41*(11), 1223–1232. https://doi.org/10.1016/j.enbuild.2009.07.002.

Hackel, M. (2004). *Identity and German architecture: Views of a German architect*. Retrieved from Faculty of Architecture. Chulalongkorn University https://www.arch.chula.ac.th/nakhara/files/article/vUrcHQFIFkSun15040.pdf.

Hermosilla, T., Palomar-Vázquez, J., Balaguer-Beser, Á., Balsa-Barreiro, J., & Ruiz, L. A. (2014). Using street based metrics to characterize urban typologies. *Computers, Environment and Urban Systems, 44*, 68–79. https://doi.org/10.1016/j.compenvurbsys.2013.12.002. March.

INSPIRE. (2014). *D2.8.I.2 Data Specification on Geographical Grid Systems – Technical Guidelines*. INSPIRE Thematic Working Group Coordinate Reference Systems & Geographical. Retrieved from https://inspire.ec.europa.eu/id/document/tg/gg.

International Energy Agency. (2016). Energy technologyperspectives 2016: Towards sustainable urban energy systems. Retrieved from https://webstore.iea.org/download/summary/1057.

Kolbe, T. H., Gröger, G., & Plümer, L. (2005). CityGML – Interoperable access to 3D city models. In Z. S. V. O. P. Hrsg (Ed.), *International symposium on geo-information for disaster management* (pp. S. 883–899). Delft: Springer. https://doi.org/10.1007/3-540-27468-5_63.

Lemaitre, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research, 18*(17), 1–5. Retrieved from http://jmlr.org/papers/v18/16-365.html.

Li, Y., Chen, Y., Rajabifard, A., Khoshelham, K., & Aleksandrov, M. (2018). Estimating building age from Google street view images using deep learning. In S. Winter, A. Griffin, & M. Sester (Eds.), *114. 10th International Conference on Geographic Information Science* (pp. S. 1–7). Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik https://doi.org/10.4230/LIPIcs.GISCIENCE.2018.40.

Liuzzi, M., Pelizari, P., Geiß, C., Masi, A., Valerio, T., & Taubenböck, H. (2019). A transferable remote sensing approach to classify building structural types for seismic risk analyses: The case of Val d'Agri area (Italy). *Bulletin of Earthquake Engineering, 17*. https://doi.org/10.1007/s10518-019-00648-7.

Loga, T., Diefenbach, N., Stein, B., & Born, R. (2012). *Tabula: Further development of the German residential building typology*. Darmstadt: Institut Wohnen und Umwelt GmbH. Retrieved from https://www.building-typology.eu/downloads/public/docs/scientific/DE_TABULA_ScientificReport_IWU.pdf.

Lowry, J. H., & Lowry, M. B. (2014). Comparing spatial metrics that quantify urban form. *Computers, Environment and Urban Systems, 44*, 59–67. https://doi.org/10.1016/j.compenvurbsys.2013.11.005.

Manfren, M., Nastasi, B., Groppi, D., & Astiaso Garcia, D. (2020). Open data and energy analytics - an analysis of essential information for energy system planning, design and operation. *Energy, 213*. https://doi.org/10.1016/j.energy.2020.118803.

Mata, E., Kalagasidis, A. S., & Johnsson, F. (2014). Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. *Building and Environment, 81*, 270–282. https://doi.org/10.1016/j.buildenv.2014.06.013.

McGarigal, K., & Marks, B. J. (1995). *Fragstats: spatial pattern analysis program for quantifying landscape structure*. Portland, OR, U.S: F. S. Department of Agriculture, Hrsg. Gen. Tech. Rep. PNW-GTR-351.

Meijer, F., Itard, L., & Sunikka-Blank, M. (2009). Comparing European residential building stocks: Performance, renovation and policy opportunities. *Building Research and Information, 37*, 533–551. https://doi.org/10.1080/09613210903189376.

Meyer, H. (2020). *CAST: Caret applications for spatial-temporal models. R package version 0.4.1*. Retrieved from https://CRAN.R-project.org/package=CAST.

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*. https://doi.org/10.1016/j.ecolmodel.2019.108815.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*(1–2), 17–23. https://doi.org/10.1093/biomet/37.1-2.17.

Nahlik, M. J., Chester, M. V., Pincetl, S. S., Eisenman, D., Sivaraman, D., & English, P. (2017). Building thermal performance, extreme heat, and climate change. *Journal of Infrastructure Systems, 23*. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000349.

Neves, F. T., de Castro Neto, M., & Aparicio, M. (2020). The impacts of open data initiatives on smart cities: A framework for evaluation and monitoring. *Cities, 106*.

Oden, N. L. (1984). Assessing the significance of a spatial Correlogram. *Geographycal Analysis, 16*(1), 1–16. https://doi.org/10.1111/j.1538-4632.1984.tb00796.x.

Open Government Germany. (2019). *Second National Action Plan (NAP) 2019–2021 in the Framework of Germany's Participation in the Open Government Partnership (OGP)*. Berlin: Federal Chancellery. Retrieved from https://www.opengovpartnership.org/wp-content/uploads/2019/09/Germany_Action-Plan_2019-2021_EN.pdf.

Open NRW. (2019). Retrieved from https://open.nrw/.

Ortlepp, R., Gruhler, K., & Schiller, G. (2018). Materials in Germany's domestic building stock: Calculation model and uncertainties. *Building Research and Information, 46*, 164–187. https://doi.org/10.1080/09613218.2016.1264121.

Pahl-Weber, E., & Henckel, D. (2008). The planning system and planning terms in Germany: A glossary. Retrieved from Studies in Spatial Development http://www.special-eu.org/assets/uploads/ARL-German_Planning_System.pdf.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. https://doi.org/10.5555/1953048.2078195.

Pohjankukka, J., Pahikkala, T., Nevalainen, P., & Heikkonen, J. (2017). Estimating the Prediction Performance of Spatial Models via Spatial k-Fold Cross Validation. *International Journal of Geographical Information Science*, 1–18.

Reinhart, C. F., & Cerezo Davila, C. (2016). Urban building energy modeling: A review of a nascent field. *Building and Environment, 97*, 196–202. https://doi.org/10.1016/j.buildenv.2015.12.001.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., & Guillera-Arroita, G. (2016). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 913–929. https://doi.org/10.1111/ecog.02881.

Rosser, J., Boyd, D., Long, G., Zakhary, S., Mao, Y., & Robinson, D. (2019). Predicting residential building age from map data. *Computers, Environment and Urban Systems, 73*, 56–67. https://doi.org/10.1016/j.compenvurbsys.2018.08.004.

Rosser, J. F., Long, G., Zakhary, S., Boyd, D. S., & Mao, Y. R. (2019). Modelling urban housing stocks for building energy simulation using CityGML EnergyADE. *International Journal of Geo-Information, 8*(163). https://doi.org/10.3390/ijgi8040163.

Sandberg, N., Sartori, I., Heidrich, O., Dawson, R., Dascalaki, E., Dimitriou, S., … Brattebø, H. (2016). Dynamic building stock modelling: Application to 11 European countries to support the energy efficiency and retrofit ambitions of the EU. *Energy and Buildings, 132*, 26–38. https://doi.org/10.1016/j.enbuild.2016.05.100.

Santos, M. S., Soares, J. P., Abreu, P. H., Araujo, H., & Santos, J. (2018). Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches. *IEEE Computational Intelligence Magazine, 13*(4), 59–76. https://doi.org/10.1109/MCI.2018.2866730. November.

Senate Department for Urban Development and Housing. (2018). Building age in residential development. Retrieved from Senate Department for Urban Development and Housing https://www.stadtentwicklung.berlin.de/umwelt/umweltatlas/e_text/ek612.pdf.

Senate Department for Urban Development and the Environment. (2018). Building age in residential development. Retrieved from The Official Websit of Berlin https://www.stadtentwicklung.berlin.de/umwelt/umweltatlas/e_text/ek612.pdf.

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028.

Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews, 13*, 1819–1835.

Tian, W., Song, J., & Li, Z. (2014). Spatial regression analysis of domestic energy in urban areas. *Energy, 76*, 629–640. https://doi.org/10.1016/j.energy.2014.08.057.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*, 234–240. https://doi.org/10.2307/143141.

Tooke, T. R., Coops, N. C., & Webster, J. (2014). Predicting building ages from LiDAR data with random forests for building energy modeling. *Energy and Buildings, 68*, 603–610. https://doi.org/10.1016/j.enbuild.2013.10.004.

Van den Brom, P., Hansen, A. R., Gram-Hanssen, K., Meijer, A., & Visscher, H. (2019). Variances in residential heating consumption – Importance of building characteristics and occupants analysed by movers and stayers. *Applied Energy, 250*, 713–728. https://doi.org/10.1016/j.apenergy.2019.05.078.

Ürge-Vorsatz, D., Eyre, N., Graham, P., Harvey, D., Hertwich, E., Jiang, Y., … Jochem, E. (2012). Energy end-use: Buildings. In G. E. Team (Ed.), *Global energy assessment: Toward a sustainable future* (pp. S. 649–760). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511793677.016.

Verein Deutscher Ingenieure. (2014). *VDI 3807 - Characteristic consumption values for buildings: Characteristic heating-energy, electrical-energy and water consumption values*.

Weinand, J., McKenna, R., & Fichtner, W. (2019). Developing a municipality typology for modelling decentralised energy systems. *Utilities Policy*, 75–96. https://doi.org/10.1016/j.jup.2019.02.003.

Wurm, M., Droin, A., Stark, T., Geiß, C., Sulzer, W., & Taubenböck, H. (2021). Deep learning-based generation of building stock data from remote sensing for urban heat demand modeling. *ISPRS International Journal of Geo-Information, 10*(23). https://doi.org/10.3390/ijgi10010023.

Wurm, M., Schmitt, A., & Taubenböck, H. (2016). Building types' classification using shape-based features and linear discriminant functions. *IEEE Journal Of Selected Topics Inieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9*(5), 1901–1912. https://doi.org/10.1109/JSTARS.2015.2465131. May.

Wurm, M., Stark, T., Zhu, X. X., Weigand, M., & Taubenböck, H. (2019). Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing, 150*, 59–69. https://doi.org/10.1016/j.isprsjprs.2019.02.006.

Wurm, M., Taubenböck, H., Weigand, M., & Schmitt, A. (2017). Slum mapping in polarimetric SAR data using spatial features. *Remote Sensing of Environment, 194*, 190–204. https://doi.org/10.1016/j.rse.2017.03.030.

Yin, C., Yuan, M., Lu, Y., Huang, Y., & Liu, Y. (2018). Effects of urban form on the urban heat island effect based on spatial regression model. *Science of the Total Environment, 634*, 696–704. https://doi.org/10.1016/j.scitotenv.2018.03.350.

Zensus. (2011). *Retrieved November 8, 2019, from Zensus*. October 15 https://www.zensus2011.de/EN/Home/home_node.html.

Zeppelzauer, M., Despotovic, M., Sakeena, M., & Koch, D. (2018). Automatic prediction of building age from photographs. In *ICMR '18: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval* (pp. 126–134). https://doi.org/10.1145/3206025.3206060.

Zirak, M., Weiler, V., Hein, M., & Eicker, U. (2020). Urban models enrichment for energy applications: Challenges in energy simulation using different data sources for building age information. *Energy, 190*. https://doi.org/10.1016/j.energy.2019.116292.