Collecting Data for Urban Building Energy Modelling by Remote Sensing and Machine Learning

Philip Gorzalka, Oana M. Garbasevschi, Jacob Estevam Schmiedt, Ariane Droin
Magdalena Linkiewicz, Michael Wurm, Bernhard Hoffschmidt
German Aerospace Center (DLR), Jülich/Oberpfaffenhofen/St. Augustin/Berlin, Germany

Abstract

High-quality data on the investigated area is crucial for modelling urban building energy demands, but its availability is often insufficient. We present an approach to acquire (i) building geometries, (ii) their ages, and (iii) their retrofit states. It consists of creating a 3D model from aerial imagery, determining building ages through machine learning, generating a simulation model based on open-source tools, and assessing retrofit states by comparing simulated temperatures with infrared thermography (IRT) measurements.

The demonstration on a case study quarter in Berlin shows that heat demand results are comparable to other tools. Using machine learning is already well-suited to close knowledge gaps regarding building ages. However, retrofit state assessment using IRT was unsatisfactory due to insufficient measurement accuracy and is envisaged for improvement in future research, along with a validation of the approach.

Key Innovations

- Combined workflow using remote sensing, machine learning, and data enrichment tools to collect building stock data
- Application of an additional, Modelica-based simulation method to the case study quarter used by Dochev et al. (2020) who compare two urban building energy modelling approaches

Practical Implications

High-quality urban building energy models may be generated even in case of limited input data availability prior to the study, but some open issues exists.

Introduction

Since carbon emissions resulting from heat use in buildings significantly contribute to climate change (IPCC (2015)), public bodies at national, regional and municipal levels have set ambitious goals for reducing heat demand. For newly constructed buildings, regulations often require or incentivise their conceptualisation as low-energy houses or even net zero energy buildings. For existing buildings, current policies are less effective, and have resulted in a current weighted annual renovation rate of about 1 % in Europe (European Commission (2020)), which is far below the rate required for a pathway consistent to the goal of restricting global warming to 1.5 °C compared to pre-industrial levels (Kuramochi et al. (2018)).

Urban building energy modelling (UBEM) tools help public bodies to model the existing state of the building stock and the impact of possible future developments. They can contribute to a better understanding of the current situation and to the design of effective policies. Several UBEM tools have been recently developed. The review papers of e.g. Reinhart and Cerezo Davila (2016), Li et al. (2017), and Sola et al. (2020) give an overview of the field.

Data on the investigated area is a crucial input for modelling. Although the amount of available datasets is quickly growing around the world, their content varies and the required effort for collecting data remains an open challenge (Hong et al. (2020); Sola et al. (2020)). In Germany, a full dataset of CityGML 3D models for the whole country has the level of detail (LoD) 2, meaning it contains exterior wall, ground, and roof polygons, but the geometries of the latter are "simplified" (BKG (2020)). The digital cadastre ALKIS provides building footprints for the whole country (AdV (2015)). The years of construction, information allowing to estimate thermal properties by consulting building typologies like TABULA (Loga et al. (2016)), are only included for some regions. For the entire country, the Census 2011 (Federal Statistical Office (2011)) is the central statistical database of the residential building stock containing attributes such as building type and age, ownership, heating type and number of dwellings. For privacy reasons, the public data published consists in summaries of these characteristics at municipality level or in a 100 m or 1 km grid.

Our proposed approach is close to the bottom-up physics-based or engineering UBEM methods as defined by Reinhart and Cerezo Davila (2016), Li et al. (2017) and others. We tackle the problem of limited

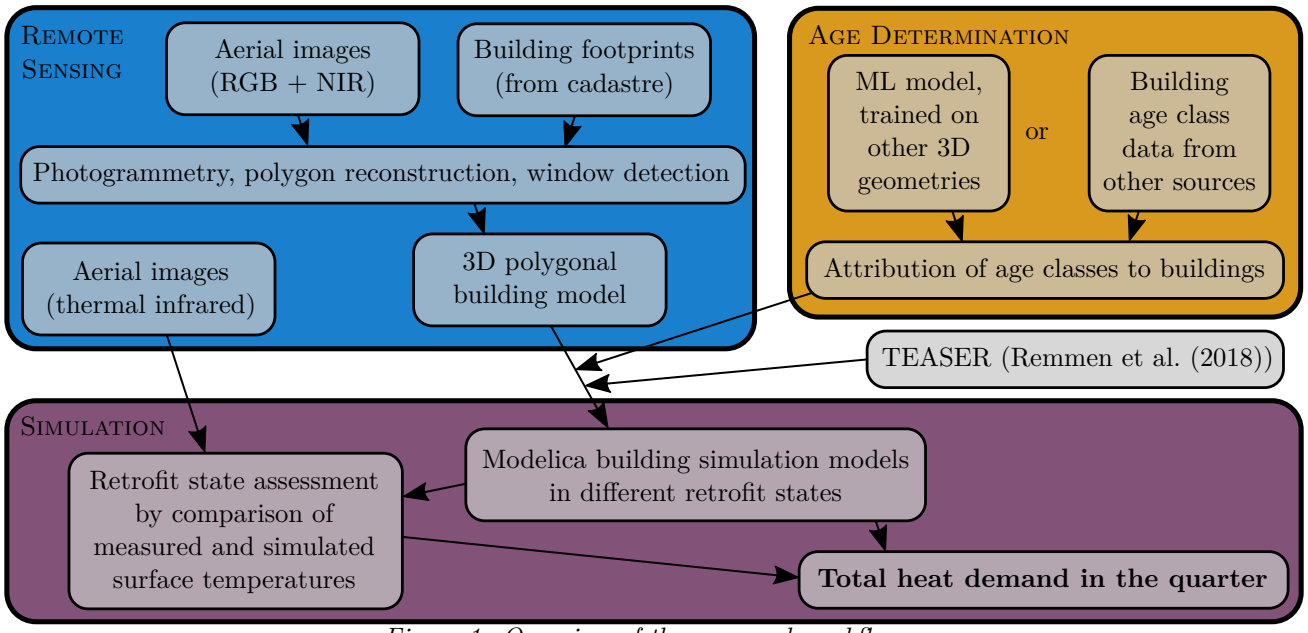

Figure 1: Overview of the proposed workflow.

data availability by a combination of measurement, machine learning (ML), and data enrichment. As a result, the only prerequisites are the geometries of individual buildings and streets, building types and uses (as available from ALKIS in Germany), a building typology (like TABULA), and age classes of the buildings or an appropriate set of training data. In contrast to Wurm et al. (2021) who apply similar tools, we use ML only for building age determination to maximise the share of deterministic calculations. The thereby increased computational effort currently restricts our approach to district rather than city scale. However, the primary aim, shared by both their and our approach, is to provide a highly automatable method to generate UBEM-ready datasets for residential buildings.

As a first step of the procedure, we process aerial imagery to obtain more elaborate 3D geometries. Second, a ML workflow provides building ages. Finally, we make use of aerial infrared thermography (IRT) to assess retrofit states of the residential building stock.

Methods

This section elaborates our approach and the methods it consists of. The steps are also illustrated in the overview of the workflow in Figure 1.

Case Study Quarter

The case study is located in the urban neighbourhood of Berlin-Moabit. The area is a good example of a district for which fuel consumption data is unavailable. This is caused by the heterogeneity of heating energy sources as well as by data protection regulations and is a strong motivation for applying UBEM. In a previous study on the quarter, Dochev et al. (2020) created a dataset to compare different UBEM methods. From the 387 buildings and building parts in the investigated area, 279 are classified as "residential" in the cadastre. We focus on this subset of the building stock in this paper. Regarding years of construction, a large share of the quarter was built before 1900. Figure 2 shows the distribution into age classes according to Berlin's "Geoportal" (Senatsverwaltung für Stadtentwicklung und Wohnen (1993)) for buildings built before 1993 and to historic aerial imagery for the rest. To achieve correspondence with the TABULA typology age classes, relevant for heat demand estimation, the middle of each age class in the Geoportal was chosen as the building year of construction as in Dochev et al. (2020).

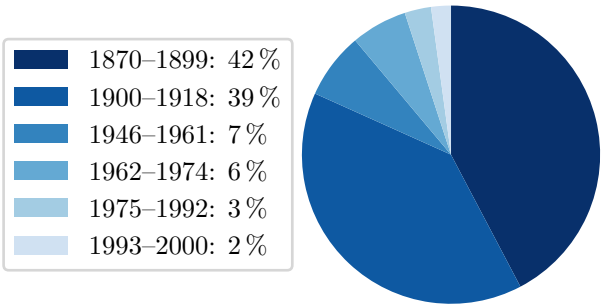

Figure 2: Distribution of the residential buildings in the investigated quarter into age classes.

3D Geometry Generation

Our approach relies on high-quality 3D models of the investigated quarter. In this case, aerial RGB imagery served for creating them. From an airplane, the area was recorded with three RGB cameras in both nadir and oblique orientation and a near-infrared (NIR) camera. On the resulting imagery, featuring a ground resolution of 8–14 cm, we applied the photogrammetric approach of Frommholz et al. (2017) with its six main steps:

1. Generation of a 3D point cloud, a digital surface model (DSM) and a digital terrain model (DTM) from the RGB images after having filtered out

vegetation using the NIR orthoimage

2. Projection of the point cloud to the ground plane, enabling the extraction of walls by local linear regression
3. Roof reconstruction by performing plane local linear regression within the building footprints defined by the walls
4. Intersection of the DTM (as ground plane), the walls (assumed vertical) and the roof plane patches to create polygons in 3D space
5. Texturing of wall and roof polygons from the original RGB imagery, avoiding occlusions
6. Window recognition on the textured wall polygons, leading to a separation of walls and windows

It must be noted that the original method is designed to use cadastral building footprints in case of incomplete wall shapes in step 2. In our case, difficulties in receiving the flight permission due to the proximity to the government district delayed the image recording until July. Since many façades were covered by vegetation by then, wall shapes were often incomplete. Therefore, we inverted the method, refining cadastral footprints using the point cloud data. The success of step 6 was also affected by tree coverage: many windows could not be detected. Still, the recognised windows fit well to reality on many façades and the results outmatch usual window-wall ratio assumptions (Dochev et al. (2020)). The outcome of the process is the CityGML model visualised in Figure 3.

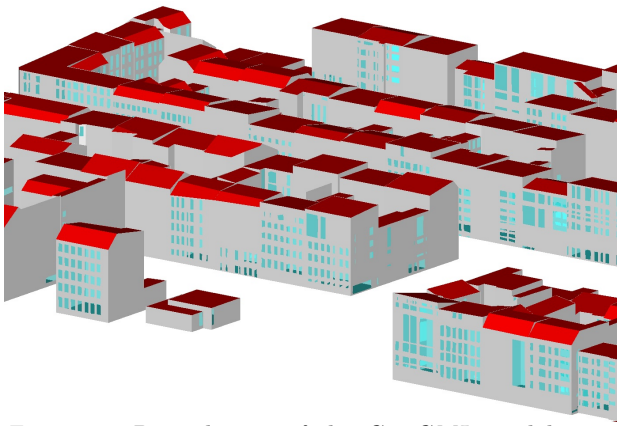

Figure 3: Partial view of the CityGML model visualised in FZKViewer.

Building Age Determination

Building age and construction type are two essential attributes for estimating buiding energy demand (Reinhart and Cerezo Davila (2016)), and a growing number of recent studies have dealt with the prediction of these attributes in countries where the information is not publicly available (Tooke et al. (2014); Wurm et al. (2016); Rosser et al. (2019); Droin et al. (2020)). Relying on modern machine learning models and spatial data sources that include topographic data as well as aerial and satellite images, these studies aim at rendering available information that should support energy-efficiency policies and decision making.

We apply a Random Forest classification method (Breiman (2001)) for labelling the residential buildings in our study with age information, by using morphological features of the building and its surroundings (Garbasevschi et al. (2021)). The building attributes determined relevant for age prediction include shape properties of the 3D building model and of the 2D building footprint, attributes describing the position of the building with respect to neighbouring buildings, and attributes of the surrounding streets and urban block. These features are extracted exclusively from open data sources, like municipality (Berlin Business Location Center (2015)) or state open data repositories (Open.NRW (2019)) or OpenStreetMap.

Statistical learning models require labelled instances as training classes. The building age classification model was trained using two separated datasets of buildings with associated construction year or age class, extracted from two locations, Wuppertal and Berlin. For Wuppertal, residential buildings' construction years are available on the open data portal of the state of North Rhine-Westphalia (Open.NRW (2019)). Ages in the Berlin dataset were extracted from Census 2011 (Federal Statistical Office (2011)). With this method of extraction, a fraction of the residential building stock of a city can be assigned one of ten Census age classes. The years of construction of the Berlin-Moabit and Wuppertal buildings were mapped to Census age classes, to allow comparison and have full correspondence between training and validation datasets.

Buildings newer than 1978 are under-represented in our sample. Problems of statistical learning with imbalanced label classes in the training lead to biased results, favouring the identification of the majority class. One technique proven successful in minimizing this effect is minority oversampling, or artificially adding new samples from the minority classes. We applied the Synthetic Minority Oversampling (SMOTE) method on the samples of buildings newer than 1978, in varying proportions (Chawla et al. (2002)). Table 1 shows the numbers of available samples and their distribution into age classes, originally and after resampling.

Simulation

To interface our 3D model with a building simulation environment, we make use of the open-source software tools TEASER (Remmen et al. (2018)) and AixLib (Müller et al. (2016)). TEASER allows the enrichment of the 3D geometry with essential input data: default use conditions, interior thermal masses, and common building part materials as collected by the German TABULA typology (Loga et al. (2016)) for different age and size classes. Loga et al. (2015) men-

Table 1: Residential building samples per age class in the training data from Wuppertal (W) and Berlin (B), originally acquired and after resampling.

Age class	Training data distribution [%]			
	Original		After res.	
	W	B	W	B
Pre-1919	19.3	23.3	16.0	16.8
1919–1948	13.2	31.8	11.0	22.9
1949–1978	33.6	30.6	28.0	22.1
1979–1986	10.2	4.0	8.5	8.6
1987–1990	7.0	1.5	8.8	5.4
1991–1995	6.9	1.3	8.6	4.6
1996–2000	4.2	5.4	7.0	11.6
2001–2004	4.4	1.1	7.4	3.8
2005–2008	0.8	1.0	3.4	3.6
Post-2019	0.2	0.1	1.3	0.5
No. of samples	47,361	77,227	56,981	107,021

tion that representative buildings are classified into the different size classes mainly based on the number of apartments. As our model does not contain this parameter, we use a threshold building volume based on assumptions for the apartment size and the floor-to-floor height. Since the TABULA archetype buildings have either flat or tilted roofs, we replace the archetype roofs by those from other types in the same age class if necessary. For an improved comparability of the results with Dochev et al. (2020), we change the default heating setpoint temperatures to 20 °C with a night setback of 4 K, both according to the German standard DIN V 18599 (DIN (2016)).

We export the TEASER data to a Modelica simulation model using the interface to the AixLib "ReducedOrder" package. To make the simulation meet the requirements of our study, we made several changes to the code. The original export process lumps the building parts to at most one single resistance-capacitance module per building element type (roof, wall, window, ground floor). We introduced the possibility to avoid lumping and simulate each model surface individually. Additionally, we adjusted the calculation of the heat exchange with the exterior such that the specific heat gain from long-wave radiation $q_{\mathrm{r,lw}}$ for a surface is calculated as

$$\begin{aligned} q_{\mathrm{r,lw}} = \varepsilon_{\mathrm{s}} \cdot F_{\mathrm{s,sky}} \cdot \left(E_{\mathrm{sky}} - \sigma \cdot T_{\mathrm{s}}^4\right) \\ + \varepsilon_{\mathrm{s}} \cdot F_{\mathrm{s,g}} \cdot \left(E_{\mathrm{g}} - \sigma \cdot T_{\mathrm{s}}^4\right), \end{aligned} \tag{1}$$

where ε_{s} is the long-wave emissivity of the surface, $F_{\mathrm{s,sky}}$ and $F_{\mathrm{s,g}}$ are the view factors of the surface to the sky and the ground respectively, E_{sky} is the sky long-wave radiation, E_{g} is the terrestrial long-wave radiation, σ is the Stefan-Boltzmann constant, and T_{s} is the thermodynamic temperature of the surface. The view factors are derived from the building 3D model. Regarding their sum, the equality

$$F_{\mathrm{s,sky}} + F_{\mathrm{s,g}} + \sum_i F_{\mathrm{s,i}} = 1, \tag{2}$$

holds. The view factors to other building's surfaces $F_{\mathrm{s,i}}$ do not appear in Equation (1) because long-wave radiative exchange with these surfaces is cancelled out by the assumption that they have the same temperature and emissivity. Thereby we avoid the computing effort of linking the individual building simulation models. E_{g} is available from the test reference years provided by DWD and BBSR (2017). These data also cover local (urban) effects on the ambient temperature, which is why we added an interface for them to AixLib.

For each building in the quarter, two Modelica models are created: using the building parts of the TABULA archetype buildings first without retrofit and second in the standard retrofit scenario.

Retrofit State Assessment

Aerial infrared images are a well-known tool to assess the thermal quality of building envelopes. The dataset we evaluated resulted from recording the investigated quarter using nadir and oblique IRT in March 2019 and subsequently texturing the 3D model with the imagery. The approach applied to derive surface temperatures (one average temperature for each surface) is fully described by Dochev et al. (2020). It largely follows Byrnes and Schott (1986), adjusts emissivities in case of oblique view according to Monien et al. (2016) and accounts for radiation from the surroundings when correcting for reflected radiation.

By simulating the buildings using actual weather data for a few days before the recording, we can calculate the surface temperatures $T_{\mathrm{s,ret}}$ and $T_{\mathrm{s,std}}$ that the building parts *should* show, given the conditions defined through TEASER and the thermal properties of the archetype building with and without retrofit respectively. Comparing them with the surface temperature measured through IRT $T_{\mathrm{s,meas}}$ using

$$\vartheta_{\mathrm{s}} = \frac{T_{\mathrm{s,meas}} - T_{\mathrm{s,ret}}}{T_{\mathrm{s,std}} - T_{\mathrm{s,ret}}} \tag{3}$$

yields a relative temperature ϑ_{s} for each surface in the quarter. As the influence of small changes in the U-value on surface temperatures is not significant enough, we do not compare more than two simulated envelope qualities. To assess the retrofit state of a particular building b, we calculate its relative surface temperature ϑ_{b} as the area-weighted average over all its surfaces with area shares a_{s}. If

$$\vartheta_{\mathrm{b}} = \sum_{\mathrm{s}} a_{\mathrm{s}} \cdot \vartheta_{\mathrm{s}} < 0.5 \tag{4}$$

holds, the building is considered retrofitted. As we cannot determine the thermal quality of insulation behind unheated attic spaces, we manually filtered for buildings with signs of an apartment in the uppermost storey, like large windows and roof terraces. Relative roof surface temperatures contributed to ϑ_{b} only in that subset.

By simulating the fitting model (with or without retrofit) of each building for the time of a full year, we calculate the annual heat demand for space heating (without hot water) in the quarter.

Results

In this section, we present the results of applying our approach, separately for the components of age class determination, retrofit state assessment, and heat demand estimation. The implications of the findings are covered in the Discussion.

Building Age Classes

For learning models where training and test instances are in close spatial vicinity, the spatial autocorrelation of features leads to better prediction results than for instances farther apart (Meyer et al. (2019)). Street and block attributes are a naturally common attribute for all buildings that share the same street or are situated in the same block. Building features such as height and the area containing a building's direct neighbors have also been shown to exhibit spatial autocorrelation (Garbasevschi et al. (2021)). Classification results confirmed this hypothesis. As shown in Table 2, the Berlin model has an overall accuracy of 84.2 % compared with 41.9 % for the Wuppertal model. In comparison, the sensitivity of both models is low, with the majority of buildings built after 1948 being classified as belonging to the "Pre-1919" age class by the Berlin model and vice versa by the Wuppertal model. A comparison between the actual and the predicted age classes of buildings can be observed in Figure 4 for the Berlin model.

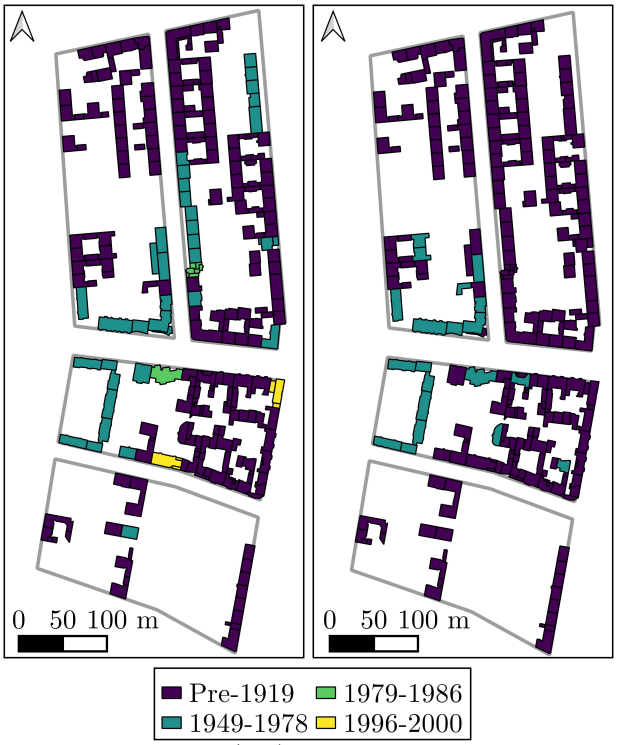

Figure 4: Actual (left) and predicted building age (right) using the Berlin dataset as training data.

Table 2: Indicators of classification performance, per training model.

Training model		Wuppertal	Berlin
Overall accuracy		41.9 %	84.2 %
Sensitivity	**Pre-1919**	35.1 %	97.4 %
	1949–1978	100 %	35.1 %
	1979–1986	0 %	0 %
	1996–2000	0 %	0 %
	average	33.8 %	33.1 %

Retrofit States

Using the decision criterion of Equation (4), only 4 (or 1.4 %) of the buildings were classified as retrofitted. When simulating $T_{\mathrm{s,ret}}$ and $T_{\mathrm{s,std}}$ with the models generated using the Berlin and Wuppertal age classification models, the number changes to 2 and 1 respectively.

Given their definition (see Equations (3) and (4)), we expected the values for ϑ_{s} and ϑ_{b} to be distributed between 0 and 1 or slightly outside this interval. The histogram of ϑ_{b} for all buildings in Figure 5 illustrates that our data does not support that expectation. Measured temperatures are systematically higher than expected. Furthermore, they are clearly more widely spread than the difference between simulated temperatures with and without retrofit, which was 1.2 K on average when using the actual years of constructions to obtain archetype building parts.

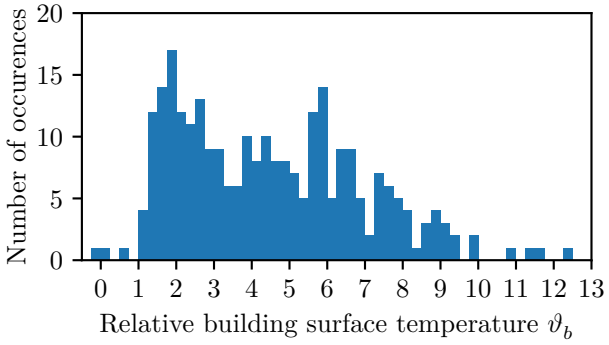

Figure 5: Histogram of the average relative surface temperature of the buildings calculated with actual years of construction, omitting two outliers at very low and six at very high relative temperatures.

Annual Heat Demand

We obtained a total heat demand for space heating of 35.6 GWh/a with the model containing the actual age class information and retrofit states as detected through IRT. As the four buildings with detected retrofit are small and relatively young, the difference to the quarter's heat demand without any retrofit is minimal. Figure 6 shows the influence of the retrofit states and the source of information on building age classes. It also contains a comparison to the heat demand calculated by the software SimStadt (Nouvel et al. (2015)) for the same buildings by Dochev et al. (2020). It must be noted that the SimStadt retrofit scenario differs from the TABULA retrofit scenario that we used for our model.

Retrofits compensate for the differences in thermal qualities between different years of constructions. Therefore, the symmetric mean absolute percentage errors (SMAPE) of individual building heat demands between the model based on actual age classes and the models based on age classes sourced from Berlin and Wuppertal decrease from 4.7 % and 13.6 % respectively without retrofits to 1.7 % and 5.4 % with full retrofits.

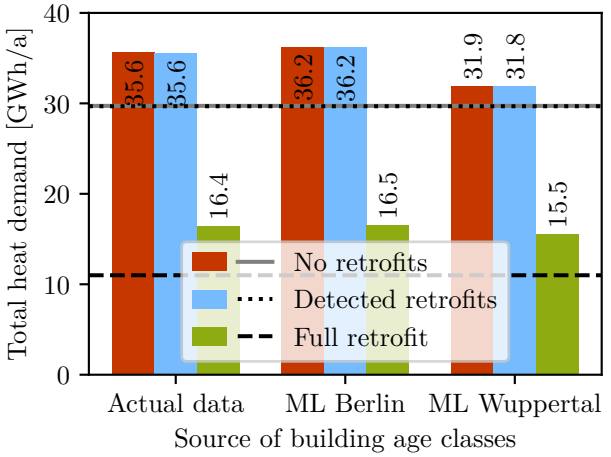

Figure 6: Simulated total heat demand of residential buildings in the investigated quarter depending on different retrofit states and the source of building age classes (bars), compared to results from SimStadt (lines, Nouvel et al. (2015); Dochev et al. (2020)).

Discussion

Based on the Results, we discuss the the suitability of our approach in the following.

Reliability of Heat Demand Results

Our results show that connecting 3D building geometries generated from aerial images to a Modelica building energy simulation gives yearly heating energy demands comparable to the established SimStadt software. The difference of about 20 % is in the same range as for the comparison of SimStadt with a specific heat demand approach by Dochev et al. (2020) for the same district. Due to missing data, a validation with actual energy use is impossible.

Already executed retrofits reduce the heat demand significantly. Therefore, detecting them is essential for an accurate representation of the building stock. Unfortunately, our method of comparing building envelope surface temperatures measured by IRT and simulated in a non-retrofit and a retrofit scenario did not yield satisfying results. The uncertainty of measured temperatures was higher than the difference between simulated temperatures with and without retrofit. Furthermore, the distribution of relative building surface temperatures in Figure 5 leads to the conclusion of a systematic error in the measurements.

This goes along with the findings of Dochev et al. (2020) that a better accuracy of IRT is required, possibly including the uncertainty of the camera itself as well as the modelling of atmospheric conditions, the radiation properties of surface materials, and radiance from surrounding objects and the sky. There are also relevant factors of uncertainty regarding simulated temperatures, e.g. locally different ambient temperatures and convection coefficients, as well as the influences of vegetation and differing interior temperatures. Most of theses issues are envisaged to be improved by future studies.

Applicability of Machine Learning for Building Age Determination

Our results show that the Berlin model classifies buildings built before 1919 correctly, but mislabels the majority of the buildings built after 1949 as older. For a study sample including only a compact neighbourhood comprising a few blocks of buildings, the rationale behind building age misclassification is probably not representative for larger samples. In our study, it can be presumed that newer buildings have been built in gaps in a predominantly pre-1919 neighbourhood and their appearance is conforming to the style. For example, a closer examination of the characteristics of the buildings built after 1979 reveals that they share attributes with the neighbouring buildings, such as height, footprint area, and roof inclination. Additionally, these similar-looking buildings are in close vicinity spatially, share the same street and block, which are all attributes of importance for the classification model. The impact on heat demand estimation depends on the extent of misclassifications. As Figure 6 shows, the Wuppertal model classifies many buildings built before 1919 as “1949–1978” and results in a lower heat demand than the demand estimated using actual age classes. For the Berlin model, the misclassification of post-1945 buildings as “Pre-1919” results in higher heat demand estimates. Nevertheless, we consider both classification models to be suited for an initial estimation of the heat demand, as the differences are within a 10 % range.

Another important source of uncertainty, for both age prediction and heat demand estimation, is the mapping between the three age class definitions of the Geoportal, TABULA and the Census. For example, a building labelled as class “1975 or after” in the Geoportal could be of the Census class “1949–1978” or “1979–1986” or a more recent age class. Similarly, it could be either of multiple TABULA classes. Ensuring uniformity between the formats, terminology and resolutions of different public datasets would be an important step forward for UBEM and, more generally, for the open urban data science.

Conclusion

We presented an approach to tackle the problem of limited data availability for UBEM by a combination of measurement, machine learning, and data

enrichment. For demonstration, we used a dataset of Berlin-Moabit (Dochev et al. (2020)). The approach consists of generating a 3D model from aerial imagery and optionally determining building ages through a classification model trained on existing data by machine learning. For generating a simulation model, we applied the complementary UBEM tools TEASER (Remmen et al. (2018)) and AixLib (Müller et al. (2016)) with minor custom variations. Retrofit states were assessed by comparing simulated and IRT-measured temperatures. Altogether, we achieved low requirements for previously known data and a highly automatable process at the same time.

The determination of age classes using a trained model confirmed other studies regarding the correlation of distance between training and test instance and prediction accuracy (Meyer et al. (2019)). Although the sensitivity of the models was low, the difference in the modelled overall heat demand of the residential buildings in the investigated quarter was adequate and lower than the difference to the demand modelled using the alternative UBEM tool SimStadt (Nouvel et al. (2015)).

Assessing retrofit states using IRT did not yield satisfying results. Some starting points for the necessary improvement of IRT surface temperature measurement were already mentioned by Dochev et al. (2020), including camera uncertainty, modelling atmospheric conditions as well as reflected radiance, and the assessment of surface radiation properties.

All in all, we conclude that using machine learning is already a viable option to close knowledge gaps. Regarding the other components of our proposed approach, prospective studies envisage the improvement of retrofit state assessment through remote sensing and a validation of the overall approach with actual energy use or individual building energy certificates.

Acknowledgements

The authors would like to thank Verena Weiler (HFT Stuttgart) and Ivan Dochev (HCU Hamburg) for providing intermediate data and final results from their heat demand calculations in the case study quarter.

References

AdV Working Committee of the Surveying Authorities of the Laender of the Federal Republic of Germany (2015). *Authoritative Real Estate Cadastre Information System (ALKIS).* http://www.adv-online.de/Products/Real-Estate-Cadastre/ALKIS/. Accessed 15 January 2021.

Berlin Business Location Center (2015). Berlin 3D - Download Portal. https://www.businesslocationcenter.de/en/economic-atlas/download-portal/. Accessed 8 June 2020.

BKG Bundesamt für Kartographie und Geodäsie (2020). *Dokumentation 3D Gebäudemodell Deutschland - LoD2 (LoD2-DE). Produktstand 2020.* https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/LoD2-DE.pdf. Accessed 15 January 2021.

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

Byrnes, A. E. and J. R. Schott (1986). Correction of thermal imagery for atmospheric effects using aircraft measurement and atmospheric modeling techniques. *Applied optics 25*(15), 2563.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research 16*, 321–357.

Deutsches Institut für Normung (2016). *Energetische Bewertung von Gebäuden - Berechnung des Nutz-, End- und Primärenergiebedarfs für Heizung, Kühlung, Lüftung, Trinkwasser und Beleuchtung - Teil 10: Nutzungsrandbedingungen, Klimadaten (DIN V 18599-10).*

Dochev, I., P. Gorzalka, V. Weiler, J. Estevam Schmiedt, M. Linkiewicz, U. Eicker, B. Hoffschmidt, I. Peters, and B. Schröter (2020). Calculating urban heat demands: An analysis of two modelling approaches and remote sensing for input data and validation. *Energy and Buildings 226*, 110378.

Droin, A., M. Wurm, and W. Sulzer (2020). Semantic labelling of building types. A comparison of two approaches using Random Forest and Deep Learning. In *40. Wissenschaftlich-Technische Jahrestagung der DGPF*. Stuttgart (Germany), 4-6 March 2020.

DWD and BBSR (2017). Ortsgenaue Testreferenzjahre von Deutschland für mittlere, extreme und zukünftige Witterungsverhältnisse: Handbuch. https://www.bbsr.bund.de/BBSR/DE/forschung/programme/zb/Auftragsforschung/5EnergieKlimaBauen/2013/testreferenzjahre/try-handbuch.pdf. Accessed 11 January 2021.

European Commission (2020). *A Renovation Wave for Europe - greening our buildings, creating jobs, improving lives*, Volume COM/2020/662 final. Brussels (Belgium): Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions.

Federal Statistical Office (2011). Zensus2011. https://www.zensus2011.de/EN/Home/home_node.html. Accessed 17 October 2019.

Frommholz, D., M. Linkiewicz, H. Meissner, and D. Dahlke (2017). Reconstructing Buildings with Discontinuities and Roof Overhangs from Oblique Aerial Imagery. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-1/W1*, 465–471.

Garbasevschi, O. M., J. Estevam Schmiedt, T. Verma, I. Lefter, W. K. Korthals Altes, A. Droin, B. Schiricke, and M. Wurm (2021). Spatial factors influencing building age prediction and implications for urban residential energy modelling. *Computers, Environment and Urban Systems 88*, 101637.

Hong, T., Y. Chen, X. Luo, N. Luo, and S. H. Lee (2020). Ten questions on urban building energy modeling. *Building and Environment 168*, 106508.

IPCC (2015). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva (Switzerland): IPCC.

Kuramochi, T., N. Höhne, M. Schaeffer, J. Cantzler, B. Hare, Y. Deng, S. Sterl, M. Hagemann, M. Rocha, P. A. Yanguas-Parra, G.-U.-R. Mir, L. Wong, T. El-Laboudy, K. Wouters, D. Deryng, and K. Blok (2018). Ten key short-term sectoral benchmarks to limit warming to 1.5°C. *Climate Policy 18*(3), 287–305.

Li, W., Y. Zhou, K. Cetin, J. Eom, Y. Wang, and G. Chen (2017). Modeling urban building energy use : A review of modeling approaches and procedures. *Energy 141*, 2445–2457.

Loga, T., B. Stein, and N. Diefenbach (2016). TABULA building typologies in 20 European countries—Making energy-related features of residential building stocks comparable. *Energy and Buildings 132*, 4–12.

Loga, T., B. Stein, N. Diefenbach, and R. Born (2015). *Deutsche Wohngebäudetypologie*. Darmstadt (Germany): IWU.

Meyer, H., C. Reudenbach, S. Wöllauer, and T. Nauss (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling 411*, 108815.

Monien, D., R. Wilting, E. Casper, M. Brennenstuhl, and V. Coors (2016). WeBest – Automatisierte Korrektur und Mapping von Fassadenthermographien auf 3D-Gebäudemodelle. *Photogrammetrie - Fernerkundung - Geoinformation 2016*(4), 246–257.

Müller, D., M. Lauster, A. Constantin, M. Fuchs, and P. Remmen (2016). AixLib - An Open-Source Modelica Library within the IEA-EBC Annex 60 Framework. In *BauSIM 2016*, pp. 3–9. Dresden (Germany), 14-16 September 2016.

Nouvel, R., K.-H. Brassel, M. Bruse, E. Duminil, V. Coors, U. Eicker, and D. Robinson (2015). SimStadt, a new workflow-driven urban energy simulation platform for CityGML city models. In *Proceedings of CISBAT 2015 International Conference*, pp. 889–894. Lausanne (Switzerland), 9-11 September 2015.

Open.NRW (2019). Open.NRW – Das Open-Government-Portal in NRW. `https://open.nrw/`. Accessed 17 October 2019.

Reinhart, C. F. and C. Cerezo Davila (2016). Urban building energy modeling – A review of a nascent field. *Building and Environment 97*, 196–202.

Remmen, P., M. Lauster, M. Mans, M. Fuchs, T. Osterhage, and D. Müller (2018). TEASER: An open tool for urban energy modelling of building stocks. *Journal of Building Performance Simulation 11*(1), 84–98.

Rosser, J., D. Boyd, G. Long, S. Zakhary, Y. Mao, and D. Robinson (2019). Predicting residential building age from map data. *Computers, Environment and Urban Systems 73*, 56–67.

Senatsverwaltung für Stadtentwicklung und Wohnen (1993). Geoportal Berlin / Gebäudealter 1992/93. `https://fbinter.stadt-berlin.de/fb/index.jsp?loginkey=showMap&mapId=gebaeudealter@senstadt`. Licensed according to dl-de/by-2-0. Accessed 6 December 2019.

Sola, A., C. Corchero, J. Salom, and M. Sanmarti (2020). Multi-domain urban-scale energy modelling tools: A review. *Sustainable Cities and Society 54*, 101872.

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

Wurm, M., A. Droin, T. Stark, C. Geiß, W. Sulzer, and H. Taubenböck (2021). Deep Learning-Based Generation of Building Stock Data from Remote Sensing for Urban Heat Demand Modeling. *ISPRS International Journal of Geo-Information 10*(1), 23.

Wurm, M., A. Schmitt, and H. Taubenböck (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*, 1901–1912.