Enhancing urban energy modeling

A case study of data acquisition, enrichment, and evaluation in Berlin

Rehmann, Felix; Mosteiro-Romero, Martín; Miller, Clayton; Streblow, Rita

DOI
10.1016/j.enbuild.2025.116070

Publication date
2025

Document Version
Final published version

Published in
Energy and Buildings

Energy & Buildings 346 (2025) 116070

Contents lists available at ScienceDirect

Energy & Buildings

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

Enhancing urban energy modeling: A case study of data acquisition, enrichment, and evaluation in Berlin

Felix Rehmann [a,b,*], Martín Mosteiro-Romero [c], Clayton Miller [d], Rita Streblow [a,b,e]

[a] Institute for Digital Networking of Buildings, Energy Supply Systems and Users, TU Berlin, Straße des 17. Juni 135, Berlin, 10623, Germany
[b] Einstein Center Digital Future, Wilhelmstraße 67, Berlin, 10117, Germany
[c] Department of Architectural Engineering and Technology, Delft University of Technology, Julianalaan 134, Delft, 2628 BL, the Netherlands
[d] Department of the Built Environment, National University of Singapore, 8 Architecture Drive, SDE4 #03-02, Singapore, 117356, Singapore
[e] RWTH Aachen University - E.ON Energy Research Center - Institute for Energy Efficient Buildings and Indoor Climate, Mathieustraße 10, Aachen, 52074, Germany

A R T I C L E I N F O

Keywords:
UBEM
Data pipeline
Archetype
Evaluation
Simulation
Mixed-use districts

A B S T R A C T

Urban Building Energy Modeling (UBEM) has become a critical tool for developing local heating and cooling plans, as required by the European Union. Despite growing interest, the reproducibility and reliability of UBEM studies remain limited due to data scarcity and workflow complexity. This paper presents a comprehensive framework to evaluate the data pipeline in UBEM, with a particular focus on data acquisition, enrichment, simulation, calibration, and information application. The approach applies three distinct UBEM workflows (CityEnergyAnalyst, DistrictGenerator, and SimStadt) to the Mierendorffinsel district in Berlin, Germany. We compare the based on quantitative performance metrics and qualitative framework criteria. The results highlight the influence of data sources, archetype definitions, and geometric preprocessing on simulation outcomes. The CEA models consider between 365 and 646 buildings, depending on the scenario. The study provides guidelines for practitioners to enhance model transparency, reproducibility, and accuracy in urban energy modeling. Although official data provide more accurate building functions, the geometries need extensive preprocessing. The residential archetypes are far more refined, e.g., in the status of renovation, which reflects the amount of residential buildings compared to nonresidential buildings. We show that evaluation threshold criteria for the district level are scarce and evaluate multiple metrics. Results depend on the selected evaluation method, but observed differences are generally higher in case of nonresidential buildings, with differences of more than 300 kWh/m^2 for several nonresidential building types. The heated area considered differs up to a factor of 1.9, because of different buildings, metadata, and calculation approaches.

1. Motivation and introduction

The European Union requires municipalities with more than 45,000 inhabitants to draft local heating and cooling plans [61]. Urban Building Energy Modeling (UBEM) has gained significant interest in recent years and can support the creation of such plans [1]. UBEM involves the modeling, simulation, and analysis of building energy demand for design, operation, and policymaking [2]. Consequently, a coherent understanding of the data requirements, model evaluation, and the wider ecosystem and standardization is crucial. The complexity of UBEM workflows, typically comprising five steps (data acquisition, modeling, simulation, calibration, and information used for a specific use case), is evident [3,4]. However, most studies are non-reproducible because of inaccessible models, data, and software tools, as well as insufficient descriptions of preprocessing steps and workflows [5]. Each step and interaction within a workflow is prone to individual errors and uncertainties, presenting significant challenges.

Because of the extensive political and technical significance of UBEM, it is important to understand UBEM workflows and assumptions in detail and to develop criteria for their processing. However, guidelines on data acquisition, verification, and validation are scarce and would benefit all stakeholders from the respective UBEM tasks. As Pfenninger [6] state, understanding the individual building blocks of an energy system model should be promoted, not only through open-source software but also by increasing the adequacy of guidelines that help explain the assumptions, criteria, and agendas involved.

UBEM evaluation must differentiate the many UBEM methodologies, which have their own requirements, evaluation methodologies, and

* Corresponding author.
E-mail addresses: rehmann@tu-berlin.de (F. Rehmann), m.a.mosteiroromero@tudelft.nl (M. Mosteiro-Romero), clayton@nus.edu.sg (C. Miller), rstreblow@eonerc.rwth-aachen.de (R. Streblow).

https://doi.org/10.1016/j.enbuild.2025.116070
Received 19 March 2025; Received in revised form 4 June 2025; Accepted 26 June 2025
Available online 4 July 2025

approaches to acquire, enrich, and validate data [7]. Typically, modelers distinguish between top-down and bottom-up approaches, and they further divide the latter into data-driven and physics-based methods [3]. In the following, we focus on physics and archetype-based single-zone modeling. Those approaches commonly use Reduced-Order Resistance-Capacitance (RC) Models. RC models can provide the advantage of accurate dynamic models with information about important features while requiring fewer data inputs than fully physics-based models, resulting in less computational resources [8].

As district or city-wide UBEM require significant information, detailed data at the building level are often not available. Data sources like CityGML and OpenStreetMap (OSM) are frequently used as alternative data sources for modeling [5]. These data can provide information such as geometry or usage, which is then enriched with further data to simulate the energy demand of the selected buildings. However, large scale models may lack additional data; therefore, research involving modeling and simulation frequently focuses on domains like universities [9] or residential regions [11] where information is more accessible, or even on virtual structures [12].

Focusing on residential areas makes sense for various reasons. First, most of the buildings in the world are residential [13,14]. In Germany, for example, there are approximately 19.5 million residential buildings [15] compared to 2 million nonresidential buildings [16]. However, nonresidential building modeling tends to be more complex due to heterogeneity in the category of buildings and different zoning. Hence, research on archetype descriptions has focused mainly on residential buildings [17]. This has changed in recent years because Hörner et al. [18] established a nonresidential building typology and made the typology publicly accessible within the open-source simulation tool DIBS [19]. Together with the well-established TABULA typology for residential buildings, [21] homogeneous modeling of mixed-use districts is now possible in Germany. This is one of the novelties of the presented studies: Development and evaluation of a new simulation workflow that integrates both typologies.

Another reason to focus on well-known areas such as universities is that the typical UBEM process has various factors of uncertainty [22,23], such as data, modeling, or programming errors. As one usually evaluates the data and the model together, gaining a deeper understanding of the data helps to identify errors. However, as we argue for better understanding and guidelines, we investigate the individual steps of the UBEM workflow. Especially since in large areas, as with the Energy Performance of Buildings Directive (EPBD) plans, a detailed model and understanding of the data is very difficult.

Therefore, we argue that researchers should study UBEM in a data pipeline in order to realistically evaluate the process at an urban scale under data scrutiny. A data pipeline is a set of data processing steps in which the output of an element is the input of the next, enabling the application and use of information [24,25]. Errors in the first steps can cause significant effects and propagate errors in subsequent steps, as researchers have demonstrated, for example, in the field of machine learning [26].

To examine the entire data pipeline, we conducted a case study in the Mierendorffinsel district of Berlin, Germany, to present three UBEM approaches and their evaluation. We analyze each step of UBEM data processing to provide criteria to modelers, scientists, and practitioners to build and refine validated models based on data availability. In doing so, we refine the criteria applied to UBEM and identify research gaps. Since data acquisition is the most time-consuming step in UBEM, we place special emphasis on it. As we will show, while there are various criteria for individual Building Energy Models, there is a lack of criteria for urban-scale models, especially under data scarcity. In addition, by comparing various approaches, we provide guidance on prioritizing data refinement in the UBEM process. This is a novelty, as the evaluation of tools is usually done either in literature reviews (refer to subsection 2.3), evaluation of single parameters in areas that are well-studied, such as research campuses (refer to: subsection 2.6) or Sensitivity Analysis (SA) (refer to subsection 2.4), software quality assessment (refer to subsection 2.3), or assessment of single data types (refer to: subsection 2.1). This is the case, as learning and developing modeling and simulation tools is a labor-intensive task [27].

Fig. 1 highlights the main objectives and the outline of this paper. The following section outlines the relevant literature and contributions. subsection 2 presents the literature and derives key factors that influence the different steps of the UBEM data pipeline. Additionally, it presents criteria for evaluation. Section 3 presents a framework for evaluation, based on the previously described criteria. Section 4 introduces the Mierendorffinsel, as well as the three tools and the data used. Section 5 presents the results of the case study and applies the previously introduced criteria to evaluate them. Finally, Section 6 discusses the presented approach and states recommendations.

2. Related literature

In this section, we provide an overview of studies that examine the influence of data acquisition, enrichment, modeling, simulation, and the application of UBEM in comparative approaches or data pipelines. The latter closes this section and examines the influences between different workflow steps. As explicit workflows are of importance to the results, we give examples of tools, data, and methodology within three UBEM modeling frameworks: City Energy Analyst (CEA) [28], the Tool for Energy Analysis and Simulation for Efficient Retrofit (TEASER) [29], the related DistrictGenerator [30], and SimStadt [31]. In addition, the section presents criteria considered in the literature for evaluating the different steps of the UBEM workflow.

2.1. Data acquisition

Energy models draw from a diverse set of standards and data, ranging from ontologies, geo-related information, or those related to the Building Information Modeling (BIM) method [25]. Despite the feasibility of detailed data integration for individual buildings [32], publicly accessible models and data are rare. Therefore, we argue that applying highly detailed models and related standards is not applicable at the urban scale for buildings outside exemplary areas. Hence, the city or district scale must consider substitutes [33], such as Shapefile, CityGML, and CityJSON. In UBEM applications, OSM and CityGML are often the two data sources used for geometric information [5].

In addition to geometric data, modelers must acquire non-geometric building data and weather data [34,35]. Goy et al. [36] refine these data categories and name potential data sources. They note that modeling often requires the combination of multiple data sources and that often only orientation and building geometry data are accessible. Wang et al. [35] provide a review of data acquisition for UBEM where they differentiate archetype-based enrichment from non-archetype-based approaches. They consider archetype availability and cost to be good in comparison to other approaches. However, they suggest modelers should be cautious in applying deterministic archetypes due to randomness in parameters. A useful classification by Malhotra et al. [37] distinguishes primary (direct use), secondary (classifying), and tertiary (statistical) inputs. For example, an archetype represents statistical data and, together with secondary data that classify the building as a certain category of the archetype (e.g., building type, year of construction), they substitute primary information (e.g., U-value). However, in most cases, even secondary data are rare.

The level of detail of geometric data influences simulation results. For example, the area modeled as a function of the outer surface can differ significantly from the true inner surface and has a correspondingly large influence on the simulated energy demand, as well as the gains through sunlight [38,39]. Consequently, Biljecki et al. [40] and Demir Dilsiz et al. [9] both provide refined input geometry specifications. In addition, the correct selection of weather data, which are

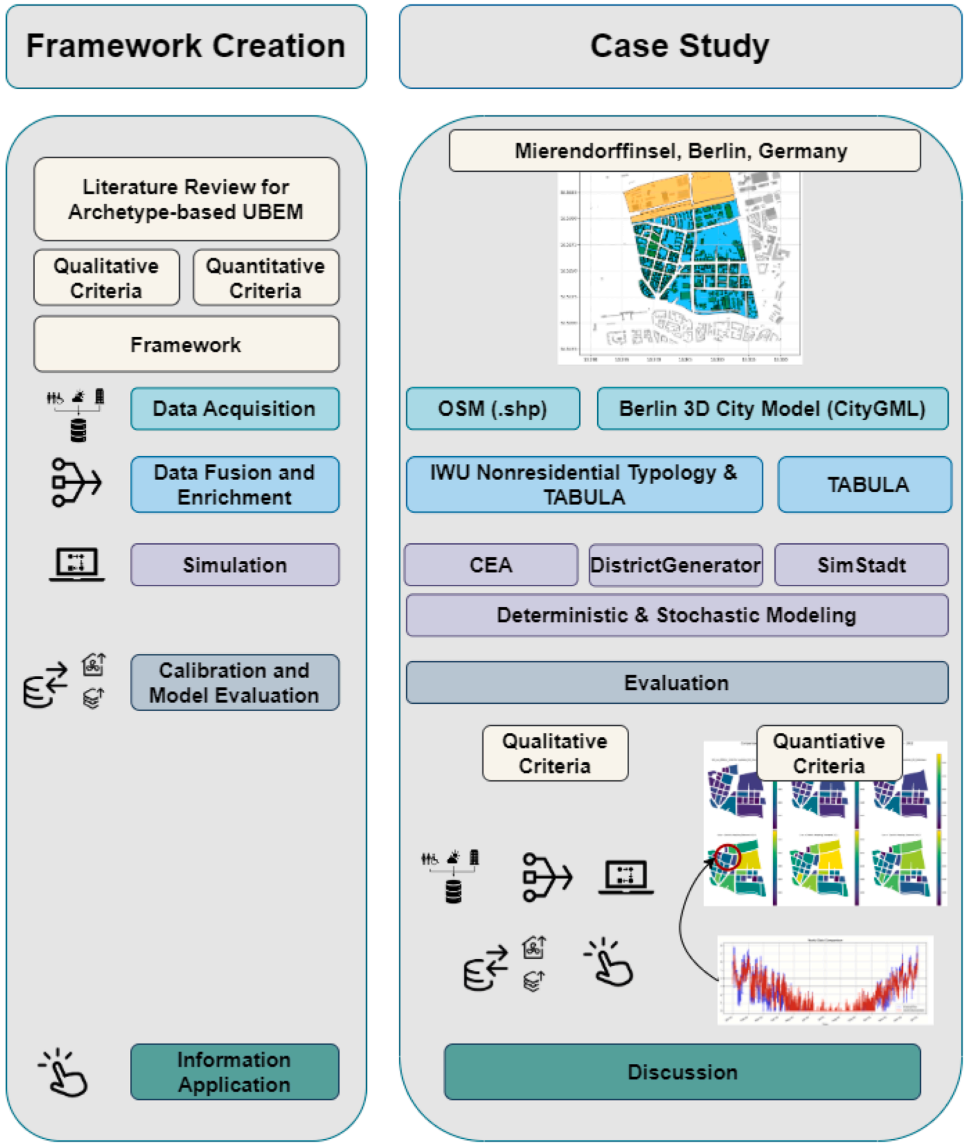


Fig. 1. The structure of the presented study. We examine the literature to collect quantitative and qualitative criteria for the steps of the UBEM process. We then implement different UBEM pipelines in a case study to analyze their application. Finally, we evaluate and discuss the case study using the framework proposed earlier.

available through test reference years (TRY) and local weather stations, affects the UBEM results, depending on parameters such as proximity or microclimate conditions [41].

Data assessment considers general criteria, such as completeness, uniqueness, timeliness, validity, accuracy, and consistency, as primary dimensions [42]. These criteria are specified for geographic data in ISO/TC 211 Geographic information/Geomatics [67], where *Completeness*, *Logical Consistency*, *Positional Accuracy*, *Thematic Quality*, *Temporal Quality*, and three *Metaqualities (Confidence, Representativity, and Homogeneity)* are stated as data quality elements. The literature also uses more general criteria, such as general availability and acquisition cost, to evaluate data acquisition [35]. Coors et al. [43] develop an application-specific approach to quality management in 3D city models. The application they focus on is SimStadt. Therefore, quality criteria are related to the calculation of monthly energy balance and related criteria, such as year of construction or valid geometry. In this case, data quality is more about valid geometry and how to ensure it in CityGML. As the authors state, assessing the geometry and (if necessary) repairing it is a complex task, that depends on the level of detail.

Biljecki et al. [44] evaluate the quality of OSM as crowd-sourced geospatial information about buildings using three criteria. First, they assess the completeness of all attributes. Second, they evaluate the consistency of attributes, because OSM provides guidelines on tagging without enforcing them. Hence, users can choose different tags that refer to the same category. By comparing Google Street View images with attributes, the authors manually assess data quality, leaving an opportunity for interpretation of the data quality due to users' perception of tags. Lei et al. [45] propose four categories of criteria for assessing and benchmarking 3D city models (such as CityGML), from the data portal to basic information about the model, to assess their thematic and attribute content. The authors benchmarked the Berlin open data set with a score of 33 % in the attribute content category. This highlights previous findings that metadata can be scarce, even in official datasets, and users must evaluate it carefully. If data about energy consumption are openly accessible, the yearly demand is a typical scale [46]. Models can use such data for evaluation and verification.

In conclusion, UBEM requires data collection, aggregation, and often substitution. Different organizations and entities can provide the data for UBEM. Official organizations provide data sources such as 3D City Models (CityGML, CityJSON) based on related standards and criteria. While official data sources often use a standardized data schema, data quality does not necessarily reflect this (e.g., timeliness or completeness). Various entities collect and provide data from energy consumption by energy providers to crowd-sourced information (e.g., OSM) or archetypes (e.g., TABULA or Department of Energy (DoE)). Hence, the evaluation of data should consider which entities provide the data. In addition, it should consider whether all the data are accessible openly or at what cost. Furthermore, general criteria such as timeliness, validity, completeness of metadata, and comprehensiveness of building coverage reflect the scope and quality of the dataset. Finally, the level of information contained in the input data are relevant to consider. These questions reflect the criteria 1C1-1C7 in Table 2.

2.2. Data fusion and enrichment

Because available data rarely meet the necessary formats, modelers must fuse, enrich, and finally model them for the simulation process. We consider there to be three relevant criteria for data fusion and enrichment: the scale at which the data enrichment process is occurring, the infrastructure and methodology used for data fusion and enrichment, and the data (archetype-based or detailed enrichment (non-archetype-based)). This section discusses these criteria and their evaluation.

Nichersu [47] discusses the advantages and disadvantages of different spatial scales, arguing that the consistency of buildings (compared to other spatial constructs with fuzzy naming such as city, town, neighborhood) enables them to be the perfect entity to bring data together. However, the author states a variety of approaches and architectures to do so. Zirak et al. [48] also reflect the scale of the enrichment process by stating that statistics from different entities (e.g., national government and municipality) affect the results when gathering information about the year of construction. Thiele et al. [49] conclude that due to differences in assumed data and actual conditions in buildings, solutions should be able to store information about individual buildings.

As the building provides a suitable scale, it is often used for data fusion. For example, Chen et al. [50] fuse multiple open datasets for the creation of a city-building dataset for San Francisco. To achieve this, they map and transform various data sources, which requires intensive data processing. The authors indicate invalid or missing data and lack of standardized tagging methodology as barriers. In this and a few other works in the literature, CityGML and EnergyADE are suggested as standards to enrich building data at an urban scale, ranging from standardized information collection [50–54] to integration into fully developed web applications [55]. Dabirian et al. [56] propose a central data model based on catalogs. They argue that such approaches provide more flexibility than standardized approaches and being able to store information, not directly required by a specify tool. However, data still need to be integrated and interfaces need to be designed, so data can be mapped to simulation engines or standards.

Monteiro et al. [57] show that the optimal number of archetypes depends on the number of parameters used to divide the subset of buildings. They note, if all archetypes use the same parameters, the peak load is overestimated. Hence, data enrichment needs some variation, at least in non-geometrical estimates (e.g., occupancy). And while parameter variety can lead to better results, the same is true for detailed enrichment at a higher level of detail than archetype-based enrichment, as Malhotra et al. [54] state. In general, the granularity in enrichment must match the supported use case (e.g., hourly schedules for dynamic simulation with hourly schedules) [56].

Because buildings typically consist of multiple zones, modelers can apply different approaches to aggregate their characteristics. Lauster [58] and Remmen [22], for example, define an approach that uses various typical zones and weights them according to their share of the net floor area, based on statistical studies. Karczewski et al. [59] propose a method that adapts and weights single-zone data from the DIN V 18599 standard [10] to align with multi-zone averages. By applying this technique, they represent the typical parameters of multi-zone buildings with a single averaged value.

To evaluate archetypes, Eisenack et al. [60] recommend quality criteria for archetype analysis in general. In addition to validity, they state the ability to combine, suitable abstraction, and obtaining fit between elements. Common approaches to derive such archetypes include data-driven, building codes-based, and hybrid methodologies [62].

Because the creation of archetypes is not the focus of this paper, we do not discuss the quality of the archetypes themselves. However, the questions in category 2C in Table 2 reflect some remarks. As Prataviera et al. [63] state, refined archetypes or assumptions during the data enrichment process might reduce model calibration efforts. Hence, one should consider the validity of the archetypes used and whether they are applicable at all to their study. During archetype enrichment, one should consider where and how the archetypes are obtained, if they reflect the metadata present in the dataset, and what additional data are used in the study for enrichment. Although national archetypes might be, in general, suitable for UBEM, they cannot reflect local particularities.

2.3. Simulation workflows and software

Several tools are available for UBEM, and various reviews have compared the metrics of these tools, e.g., [7,27,64,64,65], but rarely the process itself. In the following section, we give a brief overview of six approaches that provide data and methods for the simulation of the German building stock and discuss criteria for evaluating them.

Fonseca et al. [28] introduce CityEnergyAnalyst (CEA), a UBEM tool that has been under development and constantly improving ever since. CEA creates and simulates archetypes based on the function of the building (type of use), the year of construction and/or renovation (the CEA database for Germany requires both the year of construction and status of refurbishment) and the number of floors. The tool offers spatial and temporal distribution of a variety of outputs. At the core, CEA uses an ISO 13790 RC model [66]. CEA includes key data resources for modeling, such as options to obtain weather files or input data from OSM. Users can provide their own geometries as well. The outputs include thermal energy, domestic hot water (DHW), and electricity demands. An advantage of CEA compared to other tools is the integration of shading calculations [68], which is especially useful in the appropriate calculation of the energy demand for cooling. The software ships with two occupancy models, deterministic and Markov chain, though literature describes the implementation of other models (namely population- and agent-based) in the tool [69,70].

Remmen et al. [29] introduce TEASER as an UBEM tool, providing a variety of archetypes, from TABULA data for Germany to limited other archetypes such as institute buildings on campus or the KFW40 (a type of building that meets all the energy requirements for state subsidies). In addition to typical data for the archetypes (e.g., U-values, wall-to-window-ratio), TEASER provides detailed information about the wall structures, obtained by reviewing and testing material combinations with experts. The minimum input data required are usage, year of construction, gross floor area, and cubature (obtained from footprint, number, and height of floors). Various approaches describe the integration with CityGML [38,71]. However, we did not use the code for this in the presented study as it is outdated or was inaccessible. Although TEASER itself provides Modelica models, other approaches integrate the tool as a data source, for example, Schiefelbein et al. [72]. The latter not only provides thermal loads but also electricity and DHW using a Markov Chain for occupancy modeling, based on the Python package *richardsonpy* [73,74].

Nouvel et al. [31] introduce SimStadt, which has been extensively updated since then, e.g., in Coors et al. [75]. The tool offers a monthly simulation according to DIN 18599 [10] and an hourly calculation according to VDI 4710 [76]. The input must be a valid CityGML file with information on the function of the building and its age. The tool offers the benefit of repairing invalid geometry within CityGML data. The tool integrates the IWU residential building typology from TABULA. The output data are the thermal (monthly and hourly), DHW (monthly) and electricity (hourly) demands of the buildings.

Hillen et al. [77] presents an integration of the full European TABULA database, a 5RC model in the Open Energy Modeling Framework (oemof). In comparison to the previous tools and approaches, oemof originates in energy system analysis, highlighting the importance of understanding modeling at a spatial scale. In addition to building type, the tool requires the construction year, location, and status of the renovation, internal heat gains, and demand time series for components. The tool currently does not integrate spatial relations and hence we excluded it from further analysis.

Bischof et al. [19] presents the development and evaluation of a simplified hourly model for energy demand calculation of nonresidential buildings. At the core, it uses an ISO 13790 RC model. The input data is more detailed compared to other tools, including factors such as workplace lighting, with a total of 36 different inputs. The outputs are hourly time series for heating and cooling demand. The tool integrates the nonresidential typology by Hörner et al. [18]. As this tool can only model a minority of the building stock, we excluded it from further analysis.

Maile et al. [78] present a tool called GenSim that automatically generates input profiles for urban simulation. In contrast to the previously mentioned tools, it combines German standards and archetypes with data and information provided by the DoE. At the core, it uses EnergyPlus to provide heating, cooling, and electricity profiles. The interface enables modeling of cubic geometries; more complex geometries require SketchUp. We excluded the tool for two reasons. Firstly, GenSim lacks support for simulating DHW demand. Secondly, there is an absence of an accessible workflow to automatically model geometries from OSM or CityGML data and incorporate them into GenSim.

The literature reports multiple criteria for evaluating and choosing simulation software. Review articles select and evaluate tools based on criteria such as spatial and temporal scale, type of energy (e.g., heating, cooling, electricity) or license [7,27,64,65]. ISO/IEC 250002 is a relevant standard for evaluating software [79]. According to this standard, product and service quality, as well as quality-in-use, in addition to the previously discussed data quality, affect the software. Further tests refine and adapt these criteria for energy-related use cases such as the Building Energy Simulation Test (BESTEST) [80], the Building Optimization Testing Framework (BOPTEST) [81], and the German research project SIMQUALITY [82]. The District Energy Simulation Test (DESTEST) provides measures to adapt these tests for District Energy Systems [83,84]. The BESTEST, BOPTEST and SIMQUALITY criteria include reference cases, data exchange standards, and KPIs. The next section discusses the latter in more detail. BOPTEST explicitly states that open-source software is a criterion for validation. All presented cases compare different software, similar to this study. Therefore, when validation with real data or verified software is not possible, comparing different approaches seems to be a suitable way to overcome this limitation.

Considering these tools and the required inputs and outputs, some questions should be asked when choosing the right tool for modeling and simulation. In addition to the approach to modeling occupancy and time-varying parameters (deterministic or stochastic), the outputs and their time scale, as well as the additional input data required, are crucial factors. Most importantly, the tool must support the functionality to model and simulate the required use case. Criteria 3C1–3C5 reflect the questions raised.

2.4. Calibration and model evaluation

Researchers often use model calibration and sensitivity analysis (SA) synonymously with data quality, e.g., in Nouvel et al. [85]. Before calibration, a SA investigates the influence of parameters on the model [22]. Calibration can be done on various spatial scales (such as zone, building, urban) and temporal scales (such as sub-hourly, hourly, and yearly). In addition to this heterogeneity, there is a certain fuzziness depending on the modeling approach and the related terminology.

Nouvel et al. [85] use SimStadt as a platform to assess data quality in 3D city models. They rank various input parameters and their impact on the results of the simulation process using the annual specific heating demand. The data requirements are categorized into three categories: “Must-have” (building year of construction, building function, refurbishment information, residence type), “Relevant-to-have” (number of stories, information on basement, occupant profile, locally measured weather data), and “Nice-to-have” (LoD2 instead of LoD1, information on attic/roof, window-to-wall-ratio, number of occupants, heated/energy reference area). The authors describe the air change ratio, heating schedule, and room air temperature as parameters that have a significant impact but deem it unrealistic to collect such data at the urban scale. They list several reasons to approach the results of the study with caution, such as not considering the mutual influence between parameters. However, they emphasize one universal prerequisite: adopting intelligent data collection strategies that prioritize relevant parameters to ensure accurate results.

In Mosteiro-Romero et al. [86], SA in CEA considers the heating and cooling demand and the impact of parameters on yearly space heating and cooling demand. The authors differentiate spatial, occupancy, and building shape effects. Smaller and non-compact buildings are more easily influenced by the properties of their thermal envelope, while larger buildings are more affected by air exchange rates and set point temperatures. Spatial properties did not appear to have a measurable effect within the study. The same author conducted a study in Mosteiro-Romero et al. [87], where they calibrated 34 buildings on a university campus and investigated cooling and electricity demand under flexible occupancy using CEA. For each building, they state the best set of parameters considering the set point, cooling setback, window-to-wall ratio, infiltration rate, and U-value, as well as g-value. The study uses hourly metered data for electricity and cooling data for calibration. Both studies highlight the interdependence of SA.

Demir Dilsiz et al. [88] produce similar findings, by investigating various climates and building forms using CitySim. The minimum temperature has the greatest impact on the heating demand, regardless of the weather conditions or the form of the building. Demir Dilsiz et al. [9] extend this work, by examining how different spatio-temporal resolutions affect UBEM calibration. The authors emphasize that errors at each level of the study should be highlighted. In addition, calibration should reflect the level of detail at which the data are used.

Remmen [22] proposes a framework for automated calibration of UBEM models, evaluating various levels of information, using TEASER, non-intrusive thermal load monitoring, and the CityGML standards. However, the approach requires hourly measured data supplied by heat meters as well as detailed information about the buildings (e.g., material layers). This approach requires plenty of a priori information. Hence, the case study presented examines a well-known area, a research campus. The authors confirm the hypothesis that with an increased level of information, diverging from archetypes, the quality of the model and calibration increases, underlying the importance of reliable input quality in the first stage. In addition, the sensitivity ranking of the parameters depends on the building function and age.

Chen and Hong [39] investigate another important parameter in UBEM, the zoning configuration. The authors state the significant impact on the final energy demand but do not compare them to approaches considering another impact. Not only operational parameters and building characteristics, but geometric uncertainties affect UBEM (e.g., wall thickness, building volume) [63]. The latter are often neglected in SA because, in single-building studies, fewer geometrical uncertainties exist because of known geometries.

The quality of the SA and the calibration depend on the availability of the data. Additionally, the temporal resolution and the model itself,

Table 1
Input Data for UBEM and its importance on the results based on sensitivity analysis studies. We calculated the input parameters' relative impact calculated for single impact studies and [85]. For each study, we considered the parameters in the first third of the ranking high impact, the second third medium impact, and the final third low impact.

Input Data	Description	Impact on heating demand (Relative ranking)	Impact Discussed by
Number of floors	Number of floors in the building	Medium	[85]
Geometric zoning	How many zones are in the building and their configuration.	–	[39]
Reference area	Area that is heated or cooled	Low	[85]
CityGML LoD2	CityGML LoD2 instead of LoD1	Low	[85]
Building proportions	Information about windows, roof, etc.	Low (roof and window-to-wall ratio); High (Window Area)	[22,85]
Building characteristics	Information about material in walls, windows, etc.	High (Energy transmittance through windows); High, (heating and window U-value); High (Wall thermal resistance)	[22,36,86]
Building age	Year of Construction of the building.	High	[85]
Building use	Typical use of the building	–	[22]
Retrofit status	Retrofit Status of the building	High	[85]
Building function	Main use of the building (e.g., residential)	High	[85]
Control strategy	Control strategy, e.g., setpoints	High; High	[22,36]
Appliances	Information about appliances (e.g., number of computers) and type of lighting within a building.	Medium to Low	[22]
Occupancy data	Information about whether a building is occupied and at what times	High (Type of vacancy), Medium (Occupant profile), Low (Number of occupants)	[85]
Ventilation/ Infiltration	Infiltration or ventilation rate (air leakage)	High to Medium; High Impact	[22,86]
Local weather data	Weather data, containing information about local temperature, e.g.	Medium	[85]

including factors such as the archetypes, significantly influence the ranking of parameters. These aspects are frequently examined under varying conditions, which leads to differences in the presentation of the findings. Consequently, it is not feasible to provide an absolute assessment of the order of the parameters. For Table 1, we extracted the relative ranking of the cited articles. We consider the parameters that were in the first third of the ranking high impact, the second third medium impact, and the final third low impact. It is questionable whether the calibrated model reflects the actual building properties or merely matches the data investigated. Data are often implicitly contained within other information. For example, occupancy profiles are frequently derived from the assumed building archetype. Still, they may be an independent variable or the conclusion of different assumptions, such as vacancy rate, maximum occupancy, and occupancy profile. For example, regarding the occupancy profile and the cause of energy demand, a study considers the set point as a proxy or as an assumption that it is fixed. Additionally, the kind of data available may vary between regions and the access of research groups to non-publicly available data.

UBEM calibration is closely linked to the use of key performance indicators (KPIs) and thresholds for their evaluation. In the following, we discuss both. Since the first generation of UBEM, a variety of KPIs such as yearly demand per square meter [89] or time series energy demand output data [23] have been used to evaluate the results of the simulation. The often-used KPI of energy consumption per square meter is applicable to UBEM use cases such as building stock assessment or the provision of retrofitting strategies. As Saad and Eicker [90] note, this approach standardizes the comparison between various geometric calculation assumptions by using different reference areas (floor, envelope, or space). However, because of the increased flexibility of energy consumption and greater potential insights, the analysis of time series has become more relevant.

Typical indicators to evaluate building simulation time series can be described as absolute, relative, and forecasting evaluation [91]. The absolute evaluation indicators used for the simulation performance analysis are the Mean Bias Error (MBE), the Mean Absolute Error (MAE), and the Root-Mean-Square Error (RMSE). The suggested relative Indicators are the Coefficient of Variation of the Root-Mean-Square Error (CV(RMSE)) and the Normalized Root-Mean-Square Error (NRMSE). Relative indicators have the advantage and disadvantage of relating the absolute value to a reference, usually the measured value, which is affected by the boundary conditions. The authors of Johari et al. [92] suggest using MAE and Mean Absolute Percentage Error (MAPE) at the district level. The authors note that the accuracy of simplified methodologies generally increases with lower spatial and temporal resolutions.

Oraiopoulos and Howard [93] conclude that UBEM studies should strive for transparent applied evaluation metrics and a case-based use context (energy efficiency retrofit analysis, energy demand quantification, energy systems integration, and climate resilience). They consider CV(RMSE), RMSE, MBE, NMBE, R^2 and % as a difference between the measured and simulated energy demand. They compare a variety of studies with the ASHRAE Guideline 14–2002 [94]. Oraiopoulos and Howard [93] note, that these metrics often neglect important factors, such as hourly shifting demand. Other commonly applied benchmark thresholds [90,95] include the Federal Energy Management Program Monitoring and Verification Guide (FEMP) [96] and the International Performance Measurement and Verification Protocol (IPMVP) [97]. All three criteria reflect the level of the building, and the urban level is currently neglected in the verification and criteria for this evaluation [9,93]. A further problem in this study is that these thresholds apply monthly and hourly criteria. Data at this level are not publicly available in Germany on an urban scale.

In addition to these thresholds, model comparison between models is a common approach. For example, in the DESTEST [83,84]. For evaluation, the authors suggest using NMBE, CV(RMSE), and CV(RMSE) of daily amplitude (to evaluate the building dynamics). To evaluate yearly demand, modelers can evaluate their results against reference values, e.g., those provided by TABULA [11]. Hence, comparing a new model with a verified model can be a valid approach to measuring and evaluating a model's performance.

The SA, calibration and evaluation of UBEM significantly depend on model/tool selection and spatial and temporal resolution. As Table 1 shows, the exact ranking of the parameters affecting UBEM parameters varies according to the case study context, climate conditions, spatial scales, and the granularity of available data. Due to this variability, it is not possible to establish an universal hierarchy of parameters [22]. Instead, the study should prioritize the parameters based on the specific objectives, available data, and spatial and temporal scale. The key parameters for examination can rank from more abstract levels (such as retrofit level) to setpoints or occupancy presence. UBEM can use the metrics and thresholds stated in this section to conduct the evaluation, from the building level to the comparison between models or defined

Table 2
Criteria for the assessment of UBEM. For each of the five key steps, we propose criteria for evaluation.

UBEM pipeline category	Criterion
Data Acquisition	1C1 – Do the data have a valid format? 1C2 – Are the data recent (within the last 5 years)? 1C3 – Are the data obtained from verified sources and provided by recognized entities? 1C4 – Does the dataset cover all relevant buildings? 1C5 – Does the dataset provide complete metadata? 1C6 – Does the data include detailed geometry information? 1C7 – Are all the data in the study open to access?
Data Fusion and Enrichment	2C1 – Are the archetypes obtained from a single source? 2C2 – Are multiple archetypes used in this study? 2C3 – Do the archetypes reflect all building functions and age groups? 2C4 – Is the spatial scale of the data adequate for the study's objective? 2C5 – Is the temporal scale of the data adequate for the study's objective? 2C6 – Do the buildings have consistent identifiers? 2C7 – Is standardized terminology used for mapping the attributes during enrichment? 2C8 – Is additional non-geometric used for enrichment? 2C9 – Are zone simplifications applied in the study? 2C10 – Is there noticeable variation within the data for a single archetype? 2C11 – Are the data used for enrichment openly accessible, including the archetypes? 2C12 - Can other studies apply these archetypes effectively?
Simulation	3C1 – Is the simulation done using open-source software? 3C2 – Is a uniform approach applied for the modeling of non-building, time-dependent parameters? 3C3 – What are the minimum required input data for simulation? 3C4 – Is every building included in the data acquisition simulated? 3C5 – Is the code openly accessible?
Calibration and Model Evaluation	4C1 – Is sensitivity analysis conducted on key model parameters? 4C2 – Is the calibration performed at an appropriate temporal resolution? 4C3 – Is the calibration performed at an appropriate spatial resolution? 4C4 – Are standard verification metrics used to validate the model? 4C5 – Are predefined thresholds applied for model verification?
Information Application	5C1 – Is the Urban Energy Model applied to a practical use case? 5C2 – Do the results have a temporal resolution that aligns with the intended application? 5C3 – Does the approach allow the infrastructure to be reusable or updated? 5C4 – Are the results openly available?

test cases (see also subsection 2.3. The questions 3C1–3C5 in Table 2 reflect these considerations.

2.5. UBEM spplication

A model should be evaluated against only the background of its intended use. Hence, the quality term needs to reflect this intended use, which is an application or use case of the model. Biljecki et al. [98] identify 29 use cases for 3D City Models. The main differentiation stated is visualization-based and non-visualization-based use cases. However, due to fuzziness in terminology, classification provides ambiguity, and often use cases are related to a certain extent. The authors list estimation of solar irradiation and energy demand, semantic enrichment, and cadasters as exemplary use cases that are related to UBEM.

Modeling of building energy demand at the urban scale has seen an increasing number of applications in recent years. Ang et al. [4] state four applications for UBEM: urban planning and new neighborhood design, stock-level carbon reduction strategies, building-level recommendation, and building-to-grid integration. The review by Hong et al. [2] has a different terminology, however their applications are almost the same. Nevertheless, they state energy auditing and benchmarking, as well as climate resilience, as additional tasks. The authors of Schildt et al. [99] propose using the energy model and the related material for life cycle assessment at the district scale, embedding carbon emissions from Ökobaudat within the material in UBEM.

The information contained in the model should reflect the use cases. The higher the required spatial and temporal resolution, the finer the calibration and input data should be. If building-level recommendations are the chosen level of application, modelers must consider building-level data and calibration. For tasks such as life cycle assessment, U-values are not sufficient and models must embody materials. In addition, developers and users of tools should consider the re-usability of the results, and they can increase it by using standardized terminology. Finally, openly sharing the results and methods applied can increase understanding. The criteria 5C1-5C4 reflect these considerations.

2.6. Data pipelines and full workflows

In this section, we provide an overview of studies that examine the influence of data acquisition, enrichment, modeling, simulation, and the application of UBEM in comparative approaches or data pipelines. In contrast to the previous sections, the presented studies evaluate various ways of processing data or tools. Hence, the studies are comparable to the presented study.

Saad and Eicker [90] investigate data pipelines for energy models, comparing various data sources for geometry (i.e., input data, model complexities). To our knowledge, this is the most advanced investigation of data pipelines so far. However, the authors compare the criteria for a single building and do not provide criteria for urban scale. They find that different pipelines lead to different geometric errors and suggest standardized evaluation (demand per conditioned floor area, building envelope area, and conditioned space volume) results.

Dochev et al. [100] calculate the energy demand in a mixed-use area in Berlin using SimStadt and compare it to the so-called specific heat demand (SHD) approach. The authors compare a variety of parameters (heated volume, building typology, U-values, surface temperature, window area, and heat demand) between both approaches and additional data, such as data gained through window recognition. Due to the lack of publicly available data at the time, the authors do not give a final statement on which approach they considered superior. Because these data are available now, it is possible to continue with the evaluation.

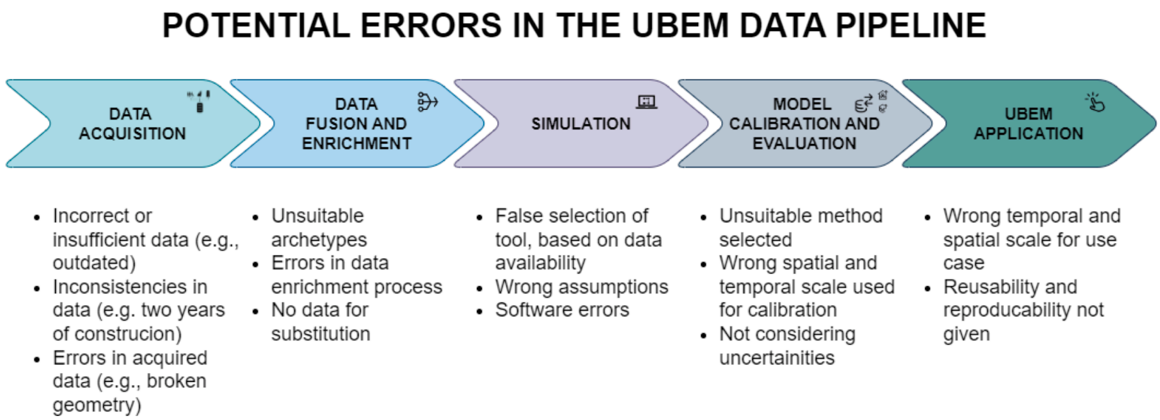


Fig. 2. Potential errors in the UBEM workflow or data pipeline. Each step has potential errors that can affect the result and applicability to a specific use case.

Malhotra et al. [11] investigate two workflows using CityGML models, TEASER, and information on the age and usage of the building. The authors simulate residential buildings and highlight inconsistencies in GML-based and *Amtliche Liegenschaftskatasterinformationssystem* (ALKIS)-based building type assessment. In addition, they state the importance of information on remodeling on an urban scale. To deduce the refurbishment state at a larger scale, the authors suggest simulating various refurbishment states and comparing them against measured data. The authors also recommend conducting further comparisons with other simulation approaches. Building on this recommendation, we conduct a comparative analysis of three simulation tools.

Johari et al. [92] examine various zoning configurations for three types of residential buildings in IDA ICE, TRNSYS, and EnergyPlus. At the building level, the complexity of the building shape, the simulation tool, and the zoning configuration affect the results. At the district level, regardless of the zoning configuration used, evaluation metrics improve.

Prataviera et al. [63] perform a sensitivity analysis using Monte Carlo sampling for two different RC models (ISO 13790 and VDI6007). The authors concluded that the peak power ratio is better after sampling and closer to reality, for two reasons: First, a temporal variation in the assumed operation times leads to a more realistic pattern by flattening the peak; Second, classical archetypes tend to systematically overestimate parameters such as building envelopes and indoor temperature patterns. Therefore, refining assumptions can enhance UBEM by yielding more realistic results.

Chen and Hong [39] investigate the impacts of geometry preprocessing, such as zoning, in the context of UBEM using EnergyPlus. The authors state that the selection of zones and floors can significantly impact equipment sizing. Similar, Dogan and Reinhart [101] present and investigate an algorithm for abstracted multi-zone urban building energy model generation called Shoeboxer. They find that in well-insulated construction (e.g., retrofit, new construction), geometric errors tend to have higher impacts on the simulated energy demand.

Wang et al. [35] review acquisition approaches for geometric and non-geometric data in UBEM and evaluate their performance, availability, and cost. The authors reflect data quality and derivations in geometric and non-geometric data. The authors consider probabilistic archetype approaches to have the potential to increase simulation accuracy at the urban scale while increasing operation time. The study reflects on criteria but does not apply them to a use case.

Demir Dilsiz et al. [9] investigate various levels of spatiotemporal aggregation of data and their impact on the UBEM process for the calculation of cooling demand. They observe that while a calibration process can improve results at the annual and aggregated levels, it might worsen accuracy at the building level. Hence, the authors conclude that there should be more detailed reporting metrics, proposing four levels of detail that modelers should consider for each of the following five dimensions: *Building geometry layer*; *Building-thermal zoning layer*, *Model calibration/validation-temporal resolution layer*, *Model calibration/validation-spatial resolution layer*, and *Model accuracy layer*.

The findings of the literature review highlight the impacts on the UBEM pipeline, depending on building characteristics, and the importance of detailed and reliable data. Various studies highlight the necessity of choosing appropriate tools and approaches for specific use cases, reflecting the intended application, whether it is urban planning, energy auditing, or life cycle assessment. Errors at each step of the UBEM pipeline can affect the whole process; for example, the acquisition of incorrect data or unsuitable model calibration can have negative impacts on subsequent steps. Fig. 2 highlights several potential sources of errors identified in the UBEM workflow in Sections 1 and 2. Although calibration can solve potential errors in acquisition or enrichment, the errors can also be pervasive throughout the workflow. We consider the figure a motivation for the evaluation presented later. Although this section reviews existing literature on the interdependence present in various pipelines, there remains a lack of research dedicated to the comparative analysis of multiple approaches and the enhancement of accuracy in urban-scale modeling.

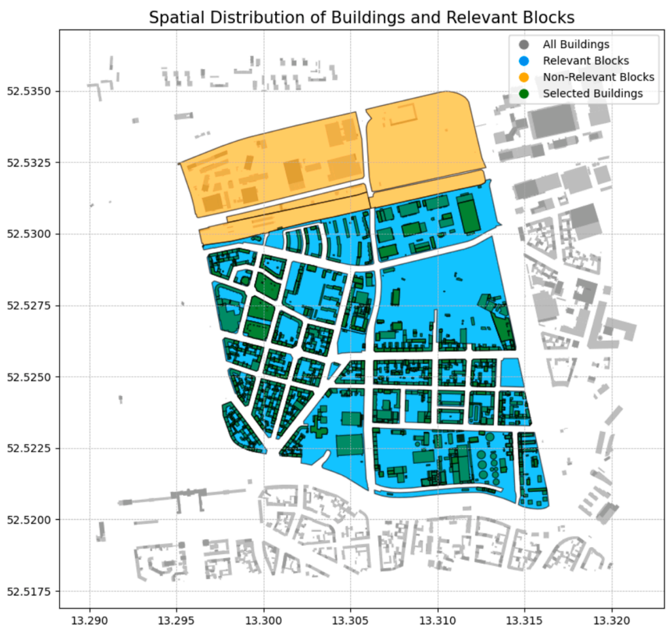
3. Framework creation

Subsection 2 identifies relevant data and evaluation criteria for UBEM. Based on these findings, this section introduces a framework and questions for evaluation of the UBEM process. As discussed in other studies, the terminology is sometimes fuzzy and open to interpretation [44,45]. The idea of the proposed criteria and framework is to help modelers, developers, and users to reflect on the interdependence of the different steps of the UBEM process and provide further criteria to evaluate the impact of choices along the UBEM pipeline. We have intentionally omitted weighting or prioritization for the checklist. Weighting would imply universal priorities, yet the literature shows that priorities need to be tool- and application-dependent, as summarized in Table 1.

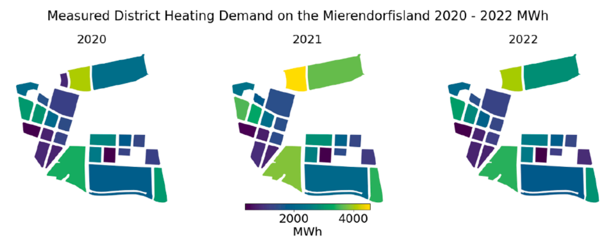
4. Methodology – use case

4.1. Mierendorffinsel

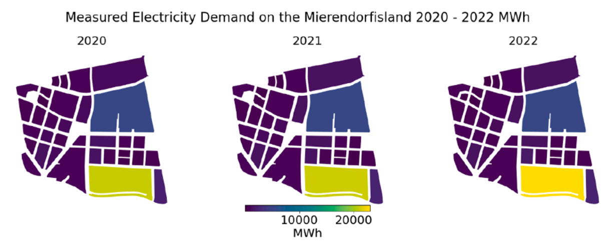
As a representative district, we consider the Mierendorffinsel in Berlin, Germany, home to approximately 15,000 residents. This district is of significant relevance due to its infrastructure, which exemplifies the heterogeneity and key challenges of the German energy transition within urban areas. The diversity in ownership structures, building usage, and building ages present on the island highlight these complexities. Furthermore, Mierendorffinsel is undergoing substantial change,



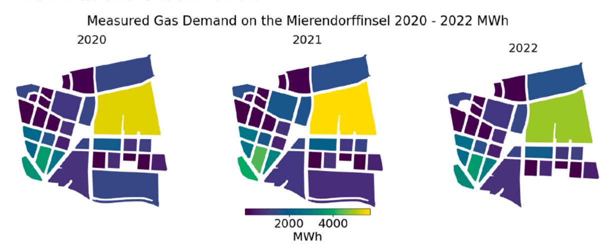
(a) The area selected for the case study. The buildings shown in the figure originate from the GML files. We investigated the areas highlighted in blue.



(b) Aggregated district heating for the years 2020 - 2022 at the block level.



(c) Aggregated electricity consumption for the years 2020 - 2022 at the block level.



(d) Aggregated gas consumption for the years 2020 - 2022 at the block level.

Fig. 3. The area for the case study. Fig. 3a highlights the levels of statistically reported data. The study examines the blue areas. Highlighted in green are the buildings considered for the study, based on a footprint calculation of the Berlin open 3D City Model. The three figures on the right display the reported energy demand at the block level. In blocks with fewer than three potential users of the energy form, for privacy reasons no data are available. It is also visible, that in some cases data availability changed in between years at block level due to unknown reasons. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with its population expected to grow to 22,000 inhabitants. Various research projects investigated this dynamic environment [102,103].

In modeling the district, we rely exclusively on publicly available information supplemented by archetypes described in our study. For CEA, we use the integrated OSM data extraction tool within the software. We sourced CityGML data from the Berlin Open Data Portal, provided in 1x1 km tiles [104]. The relevant information for Mierendorffinsel spans four such tiles. We obtained the energy consumption data for 2020, 2021, and 2022 from the Energy Atlas of Berlin at the block level [105]. In this context, a block represents the smallest unit within Berlin's regional statistical reporting system. Mierendorffinsel comprises 44 blocks, 40 of which are included in our study, covering approximately 1.47 km^2. In addition, we used the available data on the age of residential buildings per block in this study. Fig. 3a illustrates the study area and the energy consumption data used. As shown in Fig. 3b to d, the availability of energy consumption is subject to change and is only available in blocks with more than three buildings, due to privacy concerns.

4.2. Archetype enrichment

Since nonresidential buildings are completely absent in SimStadt and individual assumptions are missing in CEA, this section provides a brief overview of the data and assumptions used. Furthermore, the residential typology lacks information about apartment blocks from 1979 onward. It is challenging to estimate the number of buildings in Germany that fall into this group, as official reports use a different age group system. According to the German census, buildings that potentially belong to this age group represent less than 0.71–2.17 % of the total German building stock [106].[1] Consequently, these buildings generally do not play a particularly relevant role in a country-specific national context, but are significant in urban areas.

The research project Data:NWG provides an overview of the German nonresidential building stock [18]. As stated previously, the DIBS simulation engine provides a software side and a database implementation of this [19]. This database provides information on typical U-values, building sizes, and window and wall areas for thirteen building categories and three age groups. In contrast to Loga et al. [21], only average U-values are present, and there is no information on the structure of the wall or possible retrofits. Ceruti et al. [107] describe the implementation, which provides the archetypes for Germany for CEA. The implementation follows a similar approach by using typical U-values for both residential and nonresidential buildings. DistrictGenerator implemented these data in an object-oriented manner, similar to the original

[1] Depending on the calculation, between 141,925 and 433,585 out of a total of 19,957,268 residential buildings could belong to this building group. The first number only counts buildings with more than 13 apartments, and the second number also includes all kinds of *other* buildings constructed since 1970, according to census data from 2022.

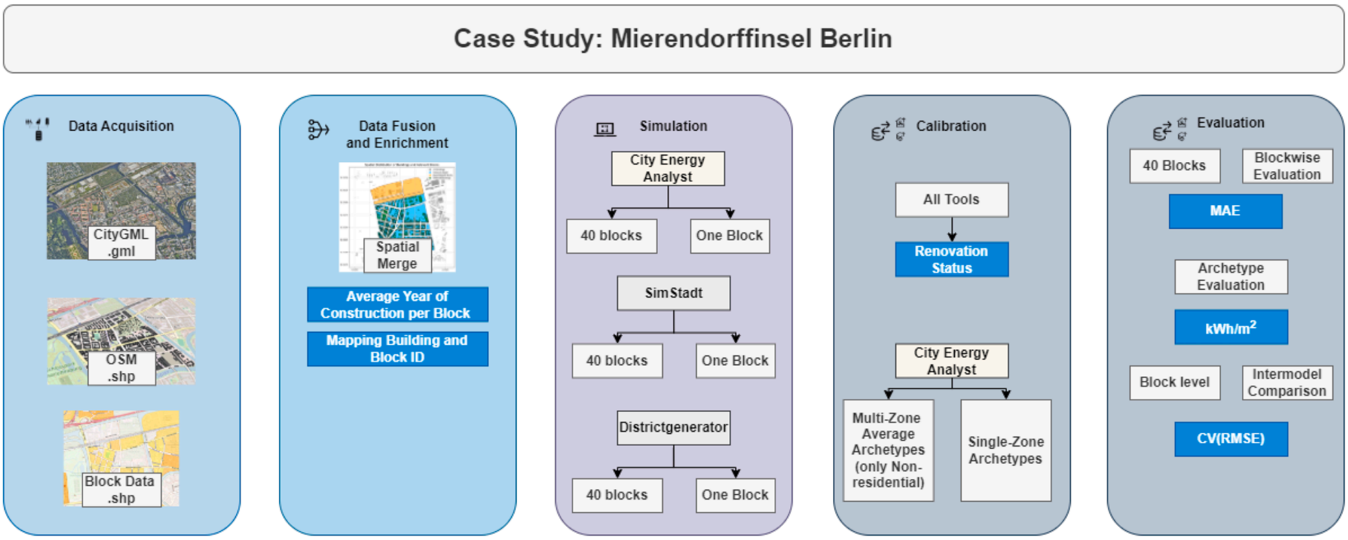


Fig. 4. The different steps for the workflow for the case study. Output of the steps is displayed in blue boxes. Grey boxes are different scenarios. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

approach from TEASER [29]. Both SimStadt and DistrictGenerator aggregate the typical material and properties for wall structures to match TABULA values. For nonresidential buildings DistrictGenerator uses the same envelope data as the CEA database [18,107].

Two recent investigations report the electricity consumption of nonresidential buildings in Germany and provide information on the configuration (low, medium, high) of appliances [108,109]. DistrictGenerator randomly assigns one configuration to each building. We obtain data on maximum occupancy by considering fire protection guidelines as maximum capacity [110]. For occupant presence we consider the Swiss Society of Engineers and Architects (SIA) profiles originally used in CEA [111]. In the absence of similar German standards, these profiles are expected to be comparable to typical German occupancy profiles or extend them [19]. In addition, we obtain data for comfort data and control strategy settings from Karczewski et al. [59]. Both DistrictGenerator and Scenario E5 implement all data on set points, maximum number of occupants, and average electrical consumption. Where multiple configurations are available, the medium setting corresponds to the CEA setting.

4.3. UBEM workflow

Each of the workflows requires its own specific steps, which are explained in the respective sections. However, there are a few general data preparation steps. We prepared the data using Python and GeoPandas. We spatially joined the data at the level of the extracted ground surfaces. To meet the 2D requirement of GeoPandas we transformed the CityGML dataset from 3D to 2D. For this, we developed a Python library, called *TECDEM - Tool for Efficient CityGML District Energy Modeling*. The library integrates with the DistrictGenerator and enables the extraction of information from CityGML and EnergyADE files. We used this library to calculate the area and shared walls and map the type of building for simulation. Subsection 4.3.1 gives a complete mapping of use types based on building function. This process is fuzzy, as for example, the CEA zone helper uses the OSM tag *Building* as the default source for building types, leading to labels that do not match any building use type, such as *Yes*.

The data merging process involves aligning attributes with the largest overlapping block shape. We estimated the age of each building based on the average age of the predominant construction period of residential buildings within its block. This approach assumes that nearby buildings generally share a similar construction time period. For example, if ten buildings were built between 1911 and 1920, and four were built between 1951 and 1960, the approach assumed that all buildings in the block date from around 1915. Fig. 4 shows the workflows and the general approach. Further limitations included invalid shapes in the CityGML, where building footprint polygons were not closed or had overlaps.

We ran a single simulation for each year (2020, 2021, and 2022) to reflect the possible variety in stochastic simulation. For residential buildings we set the indoor temperature to 20 °C, for nonresidential it varies depending on the use type.

4.3.1. Mapping of building functions

Archetype enrichment requires mapping the attributes of the extracted building function with those of similar or matching characteristics. Mapping archetypes is fuzzy, as different typologies and different standards lack a common schema or standardization. Appendix A presents the proposed mappings for the study. All approaches use the metadata of the respective geometric input data. To explore the limitations of the existing archetypes, in the first scenario for CEA (E1), we only simulated existing construction archetypes. In all other CEA scenarios, we modeled the apartment blocks built after 1979 as multifamily houses, by modifying the mapping accordingly.

4.3.2. CityEnergyAnalyst

Subsection 2.3 introduced CEA. This study used version 3.39. CEA is the first tool to integrate both the German residential and nonresidential typologies; hence we consider this database and approach as benchmarks. Ceruti et al. [107] describe the integration and verification that the benchmarks use. Because the authors describe the lack of data for nonresidential buildings, we develop an updated database which we evaluate against the original database. Subsection 4.2 describes the updated database. We use the CEA graphical interface to acquire the input OSM data and conduct all other steps (e.g., data enrichment) using Python. For each simulation, a Python function maps the archetype data from the respective database and a Python script runs all simulation steps. For each scenario, the simulation script copies and cleans the original OSM by removing buildings' shapes that are invalid archetypes.

4.3.3. DistrictGenerator

The DistrictGenerator is based on the work by Henn et al. [30] and integrates a variety of aspects previously described in work such as Remmen et al. [29], Schiefelbein et al. [72] to generate load profiles for residential districts with low-level data input used in UBEM (Year of

Table 3
Overview of the experiments conducted at district level.

Experiment	Description	Comparability
E1: CEA Baseline, original building functions, and no retrofit	Examination if there are no archetypes for apartment blocks after 1979.	E2
E2: CEA Benchmark, updated building functions and no retrofit	Same as E1, but modeling apartment blocks with the year of construction after 1979 as MFH	E1 for the number of simulated buildings E6, E7, E10 for archetypes
E3: CEA Benchmark normal retrofit, updated building functions, and normal retrofit	Same as E2, but all apartment buildings set to normal refurbishment.	E8, E11
E4: CEA Benchmark advanced retrofit, updated building functions, and advanced retrofit	Same as E2, but all apartment buildings set to advanced refurbishment	E9, E12
E5: CEA updated database, multi-zone average database	Same geometry as E2, but multi-zone average database for nonresidential buildings and not setbacks.	E7
E6: CEA stochastic, stochastic occupancy	Same as E2, but stochastic occupancy	E2, E7, E10
E7: DistrictGenerator no retrofit	DistrictGenerator 13790 model, no retrofit	E2, E7, E10
E8: DistrictGenerator normal retrofit	DistrictGenerator 13790 model, normal retrofit for residential buildings	E3, E11
E9: DistrictGenerator advanced retrofit	DistrictGenerator 13790 model, advanced retrofit for residential buildings	E4, E12
E10: SimStadt no retrofit	SimStadt model, considering no retrofit for residential buildings. DIN 18599 for yearly demand,	E2
E11: SimStadt normal retrofit	Same es E10, but normal retrofit for residential buildings	E3, E8
E12: SimStadt advanced retrofit	Same es E10, but advanced retrofit for residential buildings.	E4, E9

Construction, Renovation Status –Original, Retrofit, Advanced Retrofit–, Building Type –Single Family Home, Terraced House, Multi Family Home and Apartment Block–, and Floor Area) for typical years. This approach presents an enhanced version that considers the nonresidential typology of the IWU and related tertiary data. DistrictGenerator uses the *Teilenergiekennwerte* (TEK) method for calculating DHW load profiles, closely resembling the DIBS methodology presented by Bischof et al. [19]. TEK is a methodology that uses comparative metrics for a specific energy parameter, which simplifies the estimation of nonresidential and area-related energy demand.

In addition, we developed a tool (TECDEM)that reads and writes GML files, enabling visualization and further refinement of data at the building level in Python. This tool extracts the building metadata (e.g., function or renovation status), as well as the building area, from the GML data and creates the input simulation data for DistrictGenerator. For the calculation of geometry and shared walls, TECDEM uses CityDPC [112]. TECDEM provides a simulation scenario file, which DistrictGenerator requires to run the simulation. To calculate the heated area, we introduce a factor based on [113] that scales the outer area to the heated area. SimStadt and CEA use a similar approach. If the CityGML files contain no information about floors, we assume an average floor height of 3 m, as is in CEA. Because DistrictGenerator normally generates simple load profiles, the basic version considers no interaction between buildings, for example, shared walls. In this presented version, TECDEM extracts the geometry of buildings from CityGML and adjusts the simulation input accordingly. However, we must integrate shading from surrounding buildings into future versions.

4.3.4. SimStadt

For SimStadt, we made no further adjustments regarding programming or archetype enrichment. We prepared the CityGML files as XML, merging four tiles into one and adding the average year of construction per block. For this and further preprocessing we used Python. For the simulation, SimStadt preferred GML building parts over entire buildings. In the latter case, SimStadt only detected 131 buildings detected for the simulation. SimStadt also requires the weather data to follow the TMY3 format, for which we had to transform the EPW files. For each of the years, a single SimStadt workflow was performed. SimStadt models five categories of archetypes based on the TABULA typology: Single Family Home (*Einfamilienhaus*, EFH), Terraced House (*Reihenhaus*, RH), Multi-Family Home (*Mehrfamilienhaus*, MFH), Apartment Block (*Großes Mehrfamilienhaus*, GMH), and High-Rise (*Hochhaus*, HH). We selected a 100 % retrofit quote to generate retrofit scenarios. Although SimStadt considers multiple building functions for the simulation of heating and DHW demand, it considers only residential buildings for electricity and envelope data. To simulate residential electricity demand, SimStadt uses a function that generates random load profiles using an auto-regressive stochastic model based on measured data [114]. SimStadt can generate hourly DHW profiles, by soft-linking its output to DHW calc [75]. The presented work does not use this feature.

4.4. Overview of experiments

This section lists all the experiments we modeled and simulated in the study. We consider multiple scenarios for CEA, SimStadt, and DistrictGenerator. For all of the tools we consider no retrofit, normal retrofit, and advanced retrofit scenarios. The strategies model the envelope according to the respective TABULA retrofit standards. For CEA, we consider three additional scenarios. First, in scenario E1 we do not model apartment blocks (AB) if they are constructed after 1979, because no envelope archetype data are available for this. In all other scenarios, the archetype data (e.g., U-value for the walls) is adapted from the age group of MFH. In both cases, we select the occupancy profile of the *Multi Residential* home. Second, in scenario E5, we consider the multi-zone average data from Karczewski et al. [59], evaluating the zoning configurations for nonresidential buildings. This results in adapted set points, occupancy, and electrical configurations, matching the DistrictGenerator settings. In addition, the scenario does not consider nightly setbacks. Finally, in scenario E6, we consider stochastic modeling of occupancy to investigate the effect of simultaneity in the evaluation of time series and compare it against the electricity modeling in SimStadt and the occupancy modeling in DistrictGenerator. Table 3 lists an overview of all experiments conducted at the district level, their description, and comparability.

5. Results

5.1. Data acquisition evaluation

This section evaluates the CityGML and OSM input data according to the framework in Section 3. Fig. 5 gives an overview of the shape file data, separated between CityGML and OSM. In contrast to CityGML the OSM contains newer buildings and also very small buildings, such as sheds. The OSM data obtained has a total of 791 buildings, and the CityGML data 673.

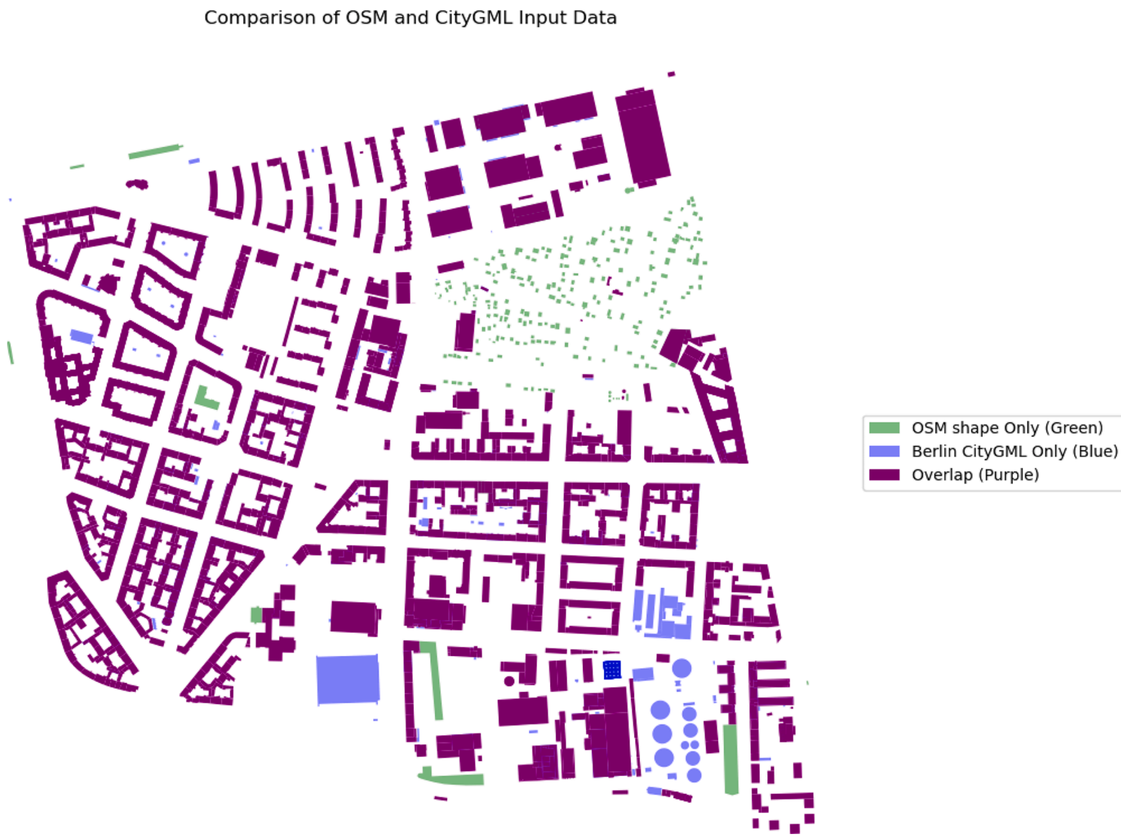


Fig. 5. Comparison of the geometric input data for CityGML and OSM. Although both input formats contain most buildings, OSM contains more up-to-date data as well as smaller, non-permanent buildings (sheds).

Table 1 compares the criteria. As DistrictGenerator and SimStadt both use CityGML as input data, the results are given in a single column. Although the metadata are complete in all scenarios, the validity of the data is sometimes unrealistic. For example, the height of buildings can be as low as 0.424 m. Additionally, the CityGML data provided by Berlin tends to have unrealistic or invalid geometries (e.g., a wall being only 6 cm in height). Although the OSM dataset provides comprehensive details about floors above and below ground level, a manual review reveals inaccuracies; for instance, a multi-story AB building is incorrectly listed as a single-story. SimStadt has numerous functions to repair such invalid CityGML geometries, which can lead to errors, see examples of this in the comparison of Fig. 6b and c. In the case of Berlin, courtyards are a typical feature of the local construction method. The repair functionalities of SimStadt repair those geometries both correctly and incorrectly.

Arguably, the most significant aspect derived from the obtained geometry is the heated area. The heated area differs between 1,152,177 m^2 in SimStadt (factor 5.58 to the ground floor area), 1,626,205.9 m^2 in DistrictGenerator (factor 5.97 relative to the ground floor area), and 855,103 m^2 in CEA scenarios E2–E6 (factor 2.68 relative to the floor area). Consequently, the heated area calculated by the DistrictGenerator model is 1.9 times larger than that in the CEA model. Given that the average building in the area comprises between four and six floors, and noting that buildings are only partially heated, it is probable that both DistrictGenerator and SimStadt overestimate the heated area (Table 4).

5.2. Data fusion and enrichment evaluation

This section discusses the data fusion and enrichment process and presents the results in Table 5. As discussed previously, there are two data sources for archetypes that are publicly available in the related case. All three approaches partially implement the typology of the residential buildings of TABULA. CEA and DistrictGenerator additionally implement the nonresidential typology. All three tools support the import of new archetypes. CEA supports creations within CSV databases. SimStadt offers two tools (Building Physics Library Editor and Building Usage Library) that enable the addition of new archetypes. Finally, the DistrictGenerator requires knowledge of Python to add new archetype data.

In the case study, SimStadt assigns eleven different usage categories. SimStadt has enabled mapping of archetypes and use types, similar to the pipeline from DistrictGenerator. CEA has a catalog of 26 use types. DistrictGenerator has richardsonpy for residential buildings and nine SIA use types for nonresidential buildings. To estimate the total number of archetypes for envelope properties we calculate each year, retrofit variant and building category. We exclude archetypes that are not fully implemented (e.g., the KFW40 is included in TEASER but not in DistrictGenerator) or that are not designed for Germany (e.g., in CEA, the Swiss standard values). In addition, we exclude potential variations (e.g., changing shading types). In SimStadt, 85 SFH, 76 TH (Row House), 86 MFH, 61 AB, and 29 High-Rise are supported out of the box. Only one building from the former East Germany (GDR) is implemented (HR, 1979–1983), while for the other categories full and partial *Energieeinsparverordnung* (EnEv) retrofit options are included. In DistrictGenerator, 39 versions for SFH and MFH are included, as well as 15 apartment blocks and 33 types of terraced houses. In addition, DistrictGenerator supports 33 nonresidential building categories. Currently, CEA is the only tool that supports the full TABULA typology, including the former GDR, out of the box. In total, CEA supports 190 archetypes, including an additional category of nonresidential buildings, a reference building by the Federal Ministry of Transport, Building and Urban Affairs. As previously stated, DistrictGenerator varies the assumptions in between buildings of the same archetype, while CEA and SimStadt apply deterministic modeling as default.

(a) The buildings considered in the CEA simulations E2-E6, based on OSM. The circle highlights an error, where sheds are considered as buildings.

(b) The buildings considered in the DistrictGenerator simulation, based on CityGML. The circle highlights an error, where different building parts overlap each other.

(c) The buildings considered in the SimStadt simulation, based on repaired CityGML. The circle highlights an error where SimStadt does not detect a courtyard or wrongfully repairs it.

Fig. 6. Comparison of the input data of the different modeling approaches. The images display the buildings considered in the heat demand simulation after data enrichment and processing. Different approaches present different archetypes based on input data. The circles highlight potential errors that occurred during the processing.

Table 4
Evaluation of criteria 1C.

Criterion	CEA model	DistrictGenerator and SimStadt models
1C1 – Do the data have a valid format? Which?	Yes, OSM shapefile	Yes, CityGML
1C2 – Are the data recent (within the last 5 years)?	Yes	Yes (January 2023)
1C3 – Are the data obtained from verified sources and provided by recog- nized entities?	Yes, OSM, crowdsourced data	3D City Model Berlin, Senate Department for Urban Development, Building and Housing
1C4 – Does the dataset cover all relevant buildings?	Yes, all buildings (791) in the area included	Yes, all buildings (673 DistrictGenerator/ 1304 SimStadt) in the area included
1C5 – Does the dataset provide complete metadata?	Yes, Building Function (100 %), Building height (100 %), Floors above and below ground (100 %)	Yes, Rooftype (100 %), Building function (100 %), Measured height (100 %), Actuality of floor plan (100 %)
1C6 – Does the data include detailed geometry information?	LoD0	LoD2
1C7 – Are all the data in the study open to access?	Yes	Yes

Table 5
Evaluation of criteria 2C.

	CEA	DistrictGenerator	SimStadt
2C1 – Are the archetypes obtained from a single source?	No	No	Yes
2C2 – Are multiple archetypes used in this study?	Yes, 19 (E1), (22) (E2–E6)	Yes, 42	Yes, 19
2C3 – Do the archetypes reflect all building functions and age groups?	No, only a subset selected	No, only a subset selected	No, only a subset selected
2C4 – Is the spatial scale of the data adequate for the study's objective?	No, data at block level	No, data at block level	No, data at block level
2C5 – Is the temporal scale of the data adequate for the study's objective?	Yes, yearly	Yes, yearly	Yes, yearly
2C6 – Do the buildings have consistent identifiers?	No (different scenarios can assign the same ID to different buildings)	Yes	Yes
2C7 – Is there standardized terminology used for mapping the attributes in enrichment?	No (because of study design)	Yes	Yes
2C8 – Is additional non-geometric used for enrichment?	Average building age at block level	Average building age at block level	Average building age at block level
2C9 – Are zone simplifications applied in the study?	Yes, one zone building (E1–E6); Multi-zone averages for nonresidential buildings (E7)	One zone per residential building and multi-zone averages for nonresidential buildings	One zone per building
2C10 – Is there a noticeable variation within the data for a single archetype?	No (E1–E5); Yes, (occupancy - E6 -)	Yes, occupancy, equipment (low, medium, high)	No, (heating); Yes (occupancy for electricity load profiles)
2C11 – Are the data used for enrichment openly accessible, including the archetypes?	Yes	Yes	Yes
2C12 - Can other studies apply these archetypes effectively?	Yes	Yes	Yes

None of the tools implements one archetype by itself; instead, they often combine multiple data sources (e.g., SIA and DIN 18599) and offer the user to adjust these combinations. We determine whether the data can be exported and used in other studies by two factors: First, whether the data are accessible outside the program, e.g., in the form of a CSV file; this is true (*Yes*) for all tools. The second factor is, whether archetypes can be combined, e.g., by using features from one archetype (e.g., U-values or mapped occupancy) with another; this is also true (*Yes*) for all tools as they implement the archetypes in categories (e.g., envelope separated from setpoints).

Based on the selected buildings and the data enrichment process, we generate the final data set considered for simulation. Fig. 6 shows these data which are the result of the spatial aggregation of the average residential age and the function of the building. We do not consider areas in the simulation where the open data does not state information on the residential age. This restriction reduces the amount of considered blocks from 40 to 30. As shown in the comparison of Fig. 6b and c, the geometric processing of the CityGML data is relevant but leads to errors in itself. For example, the building encircled in the DistrictGenerator geometry correctly contains six courtyards but also has overlaps. SimStadt correctly repairs such errors, but also incorrectly repairs some courtyards, as highlighted in the graphic. In the CEA case, errors in the OSM data lead to the false inclusion of buildings.

Regarding the temporal and spatial scale, we considered the yearly energy demand adequate but not the spatial aggregation (block-level). In this study, we obtained the OSM geometry via the CEA interface. During this process, CEA assigns an ID to each building (e.g., B1013), which is the same for all scenarios with the same input. This leads to two limitations in this workflow: First, in comparing two scenarios with an overlap, those buildings might have different IDs. Second, the OSM terminology is not standardized to map with the archetypes in the German CEA. Hence, we consider SimStadt and the presented version of DistrictGenerator a *Yes* for questions 2C6 and 2C7, while we consider CEA a *No*.

5.3. Simulation evaluation

Both CEA and SimStadt provide an interface for simulation. All three approaches provide an API. Although CEA and DistrictGenerator are open-source, SimStadt is distributed as free software, but the source code is not publicly available. In all three examined pieces of software, stochastic simulation is possible to at least some extent. In CEA, stochastic simulation is optional, in DistrictGenerator, it is the default for residential buildings, and in SimStadt, it is included for electricity.

The evaluation of question 3C4 highlights the importance of having archetypes for all building age classes while considering the availability of the various tags. The comparison of scenarios E1 and E2 shows that without modeling AB after 1979 as MFH, 35 % of the acquired buildings are not considered within the further UBEM pipeline (Table 6).

5.4. Calibration and model evaluation

This section presents the evaluation criteria and thresholds for assessing the modeling and simulation approaches. Initially, we compare the results of the entire district with the reported data, followed by an inter-model comparison at the block level. We aggregate DHW and heating demand for each simulated building, comparing this data with publicly available gas and district heating information at the block scale. We consider this approach due to the absence of data on building equipment and measured data. As measured data at the building level are unavailable, we refrain from applying thresholds to verify models at this scale. Instead, we employ MAE and MAPE to evaluate the district comprehensively at the block level. Furthermore, we compare and discuss results at the archetype level. Finally, we evaluate the time series comparisons using CV(RMSE). MAE and MAPE clearly illustrate the interpretability of results. Assessing archetype demand enables us to avoid the influence of differences in geometric processing. Using CV(RMSE) allows us to evaluate how closely the simulated energy demands align at the block level.

5.4.1. Full district

To evaluate the entire district, we calculate the MAE and MAPE at the block level by comparing the results of the annual simulations of heating and DHW to the annual sum of the gas and district heating demand. Fig. 7 displays the results. As expected, the MAE and MAPE change with the retrofit level. The MAPE exceeds 300 % in multiple cases, all of which have in common that the measured energy demand is low compared to other blocks. For example, block 20,029 has a measured energy demand of 74.7 MWh for 2020 while the neighboring and comparable block 20,030 has a measured district heating and gas demand of 1568 MWh. In Block 19742, multiple buildings are present that lack an archetype in scenario E1. Hence, the baseline model in CEA is significantly worse. In most cases, for all three tools, the MAE is higher for no retrofits, indicating that the buildings might have undergone some kind of renovation or that the non-retrofit scenarios lead to larger errors. The worst block in all CEA simulations is a block with industrial buildings and a former power plant; some of these buildings lack geometries in the underlying OSM files. The block 21,061 is characterized by industrial buildings and a former power plant, leading to a large error in all simulations due to the untypical building.

The MAE is relatively high for the three DistrictGenerator scenarios, which is in proportion to the simulated energy demand, which is higher than in SimStadt and CEA. In most cases, the DistrictGenerator No Retrofit model performed the worst, as the model's results exceed the measured data. The large heated area, combined with a comparatively high energy demand per archetype, causes this overestimation. SimStadt tends to perform better, for the likely reason that SimStadt repairs the geometries, and thus the building areas are lower for many buildings. This is especially the case in block 20036. The block contains two large AB, which are a single building in the GML file, with a respective ground

Table 6
Evaluation of criteria 3C.

Criteria	CEA	DistrictGenerator	SimStadt
3C1 – Is the simulation done using open-source software?	Yes	Yes	No (but software is free to use)
3C2 – Is a uniform approach applied for the modeling of non-building, time-dependent parameters?	Yes, Deterministic (E1–E5); No, stochastic (E6)	No, Stochastic (richardsonpy) for Residential, Stochastic Modeling of occupants and electrical settings for Nonresidential	Stochastic for Electricity, Deterministic for Heating
3C3 – What are the minimum required input data for simulation?	Building function, year of construction, refurbishment status	Area (extracted from CityGML file), building function, year of construction, refurbishment status	CityGML file, ALKIS building function and year of construction
3C4 – Is every building included in the data acquisition simulated?	No, (365 / 791 - E1 - , 646 / 791 - E2–E6)	No, (520 / 673)	No, (794 / 1204)
3C5 – Is the code openly accessible?	Yes	Yes	Yes

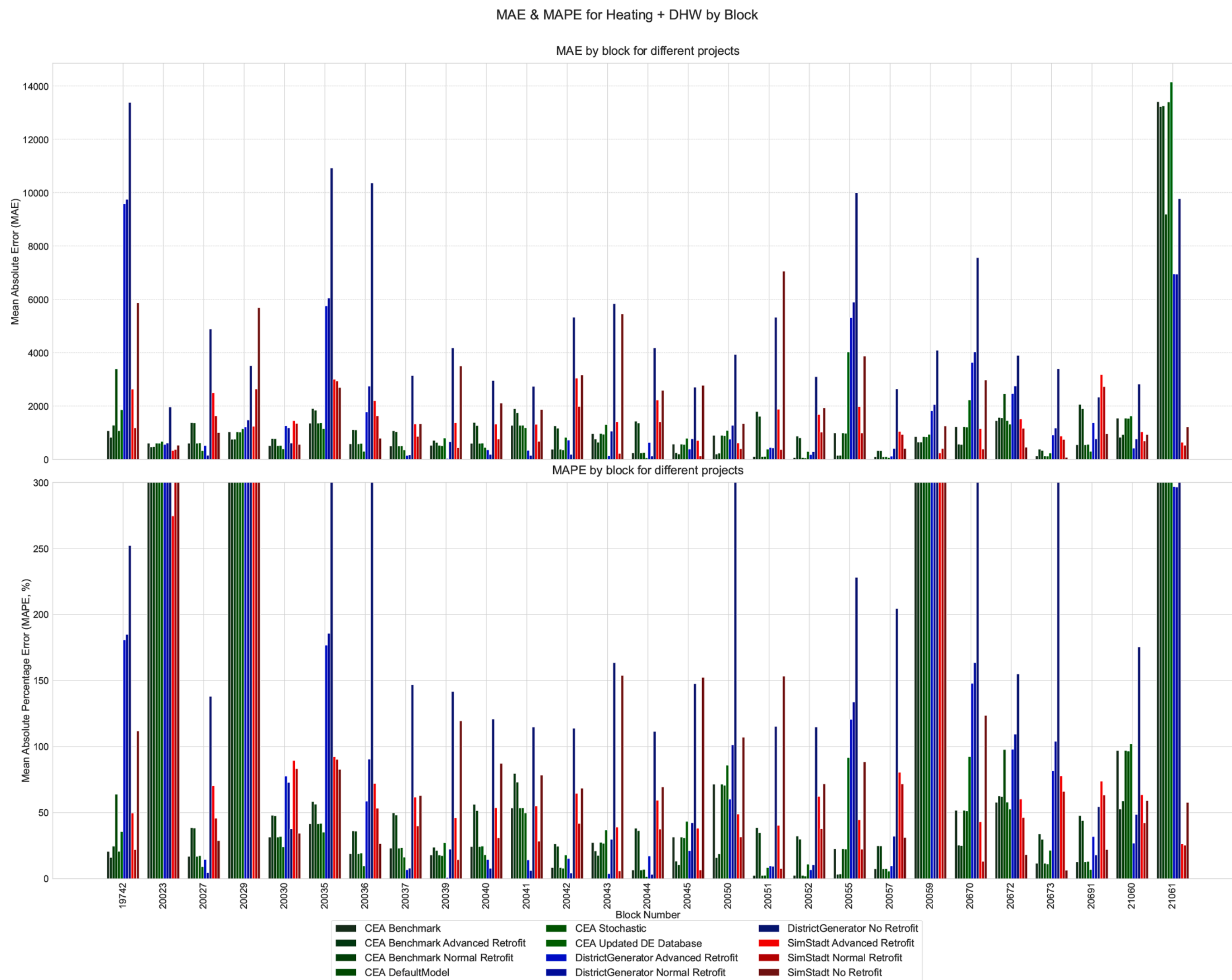


Fig. 7. Comparison of the MAE and MAPE for the different simulation approaches, regarding the simulated DHW and heat demand, as well the open gas and district heating demand. The MAPE y-axis is cut off at 300. The results show the comparison between simulated heating and DHW demand and measured gas and district heating demand. All metrics refer to the period 2020–2022.

floor area of 6789 m^2 and 5507 m^2 in DistrictGenerator. The SimStadt geometry processor calculates the areas as 3006.1 m^2 and 3006 m^2.

Fig. 8 displays the MAE and MAPE for the electricity demand. For most blocks, the MAE is reduced in comparison to the previous analysis. One potential explanation for this phenomenon is the presence of a uniform and centralized electricity demand reporting. This is attributable to the fact that the district in question has fewer decentralized technologies in electricity. Similarly to the previous analysis, block 21061, characterized by industry and the power plant, is among those blocks with the highest MAE. Since SimStadt models the building stock deterministically (but not the occupancy), the distributions do not change, as the figure shows. In DistrictGenerator, the stochastic modeling of electrical demand has a significant impact, so the MAE changes in between scenarios. In CEA, the same configurations produce identical results, as expected. In CEA, the stochastic modeling of occupancy affects the energy demand and, consequently, the MAE, as demonstrated by the comparison between scenarios E2 and E6. Scenario E1 has an increase in MAE compared to scenarios E2–E4, highlighting the relevance of the missing archetype.

Fig. 9 displays the annual energy demand per square meter per residential archetype for all the pipelines considered. For this, we calculated the average of all three years and divided it by heated area. The y-axis is cut off at 350 kWh/m^2 for better visibility. Both DistrictGenerator (SFH B, no retrofit) and SimStadt (SFH D, no retrofit) have a single outlier building that exceeds that energy demand.

This comparison shows that the different data pipelines use different mapping approaches, which leads to different archetypes and the number of them, visualized by the size of the dot. This visual inspection also makes it evident that SimStadt implements archetypes that do not exist in TABULA, apartment blocks (1984–2013 and after 2014). The SimStadt analysis also includes more buildings overall because the model uses building parts instead of buildings. As expected, the energy demand decreases with increasing levels of retrofit, and as the building construction year increases, the differences in energy demand between the retrofit classes also diminish. The maximum energy demand varies for the three tools depending on the archetype. As expected in scenario E5, the waiver of the nightly setback leads to an increased energy demand.

To contextualize the results displayed, the energy demand for older buildings (e.g., SFH C in original state has an end energy demand of 3796 kWh/m^2 per year) seems to be to low compared to the results suggested by TABULA [21]. For higher retrofit levels and increased year of construction this gap between simulated and reported energy demand narrows.

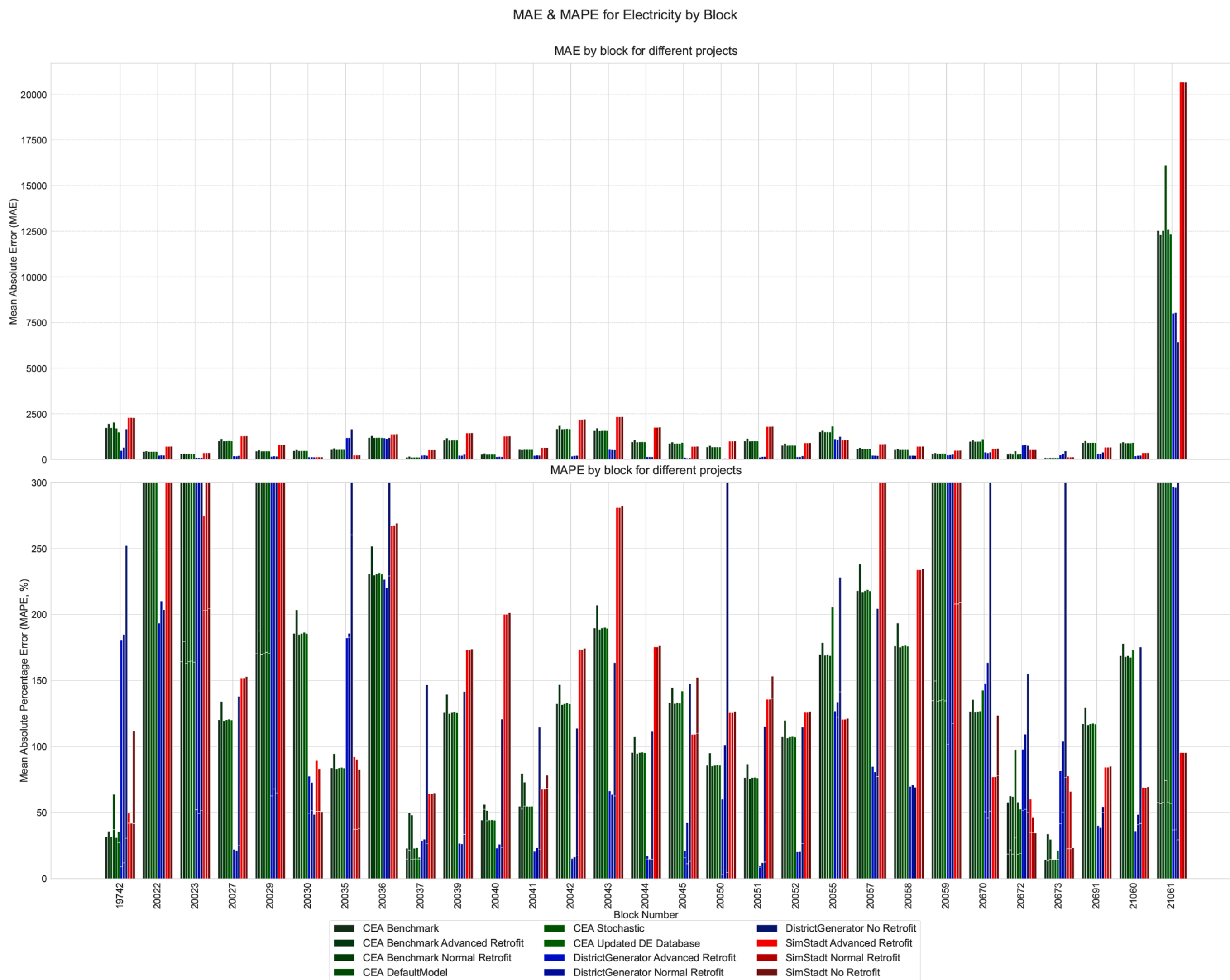


Fig. 8. Comparison of the MAE and MAPE for the different simulations approaches regarding the electricity consumption. The MAPE y-axis is cut of at 300. All metrics refer to the period 2020–2022.

Fig. 10 presents the annual energy demand per square meter for the nonresidential archetypes as simulated by both CEA and DistrictGenerator. For readability, the y-axis is cut off at 500 kWh/m^2. Four CEA archetypes exceed that threshold: IWU Generalized (2) Production Buildings A (all scenarios), IWU Production, Workshop, Warehouse or Operations (all scenarios), and Sports Facilities A (only scenario E5). The selected pipeline approaches again yield different distributions of archetypes, a result that stems from differences in the underlying metadata. DistrictGenerator simulations reflect the stochastic variability in building equipment, while CEA simulations, except for the updated database approach (E5), consistently produce similar outcomes across scenarios.

Compared to the residential buildings, a more pronounced divergence between the approaches is noticeable. Several factors may explain this: First, due to the smaller number of buildings, outliers exert a greater impact on the average. Second, while the typology establishes common aspects, such as U-values for the walls, which both approaches implement, CEA and DistrictGenerator have for example, differences in the calculation of the DHW requirements. Third, the typology establishes features that both tools neglect, such as rooftop windows. Furthermore, certain buildings in the CEA simulations exhibit a very low energy demand. We investigated this anomaly, but could not identify irregularities in the underlying databases. We further describe this phenomenon in subsubsection 5.4.2.

Although the nonresidential building typology only specifies the average specific useful heating energy requirement [18], we can approximate comparisons with the total heat demand for water and space heating. For example, supply and disposal buildings in age group B report an energy demand of 330 kWh/m^2, suggesting that DistrictGenerator underestimates the energy demand. In contrast, office, administrative, and government buildings in age group A report an average specific heating demand of 108 kWh/m^2, which aligns closely with DistrictGenerator's estimates.

5.4.2. Block inter-model analysis

For detailed inter-model analysis, we evaluate the CV(RMSE) of all blocks. As SimStadt does not support the export of hourly DHW demand, this part of the study only considers the demand for heating and electricity. The calculation of CV(RMSE) involves determining the mean across all scenarios and years that share similar retrofit assumptions, which is then used as a reference to compare the mean of each individual scenario. For example, we calculate the mean of all three advanced retrofit scenarios and then compare the three advanced retrofit scenarios against it to calculate CV(RMSE).

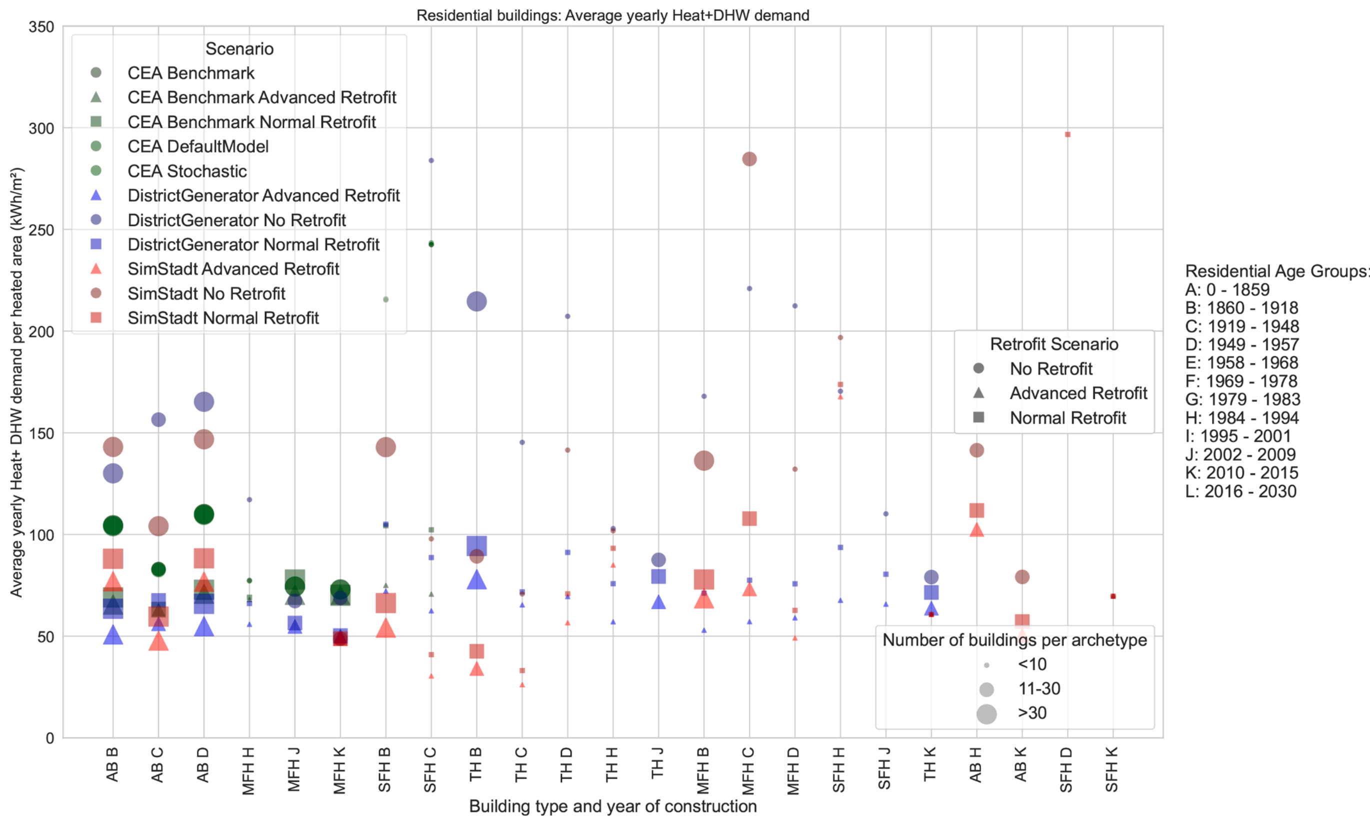


Fig. 9. Yearly energy demand in kWh per square meter for each residential archetype and all pipelines considered in the study. The size of the marker indicates the amount of archetypes in the respective simulation. As the graphic shows, different approaches lead to different archetypes and quantities of archetypes.

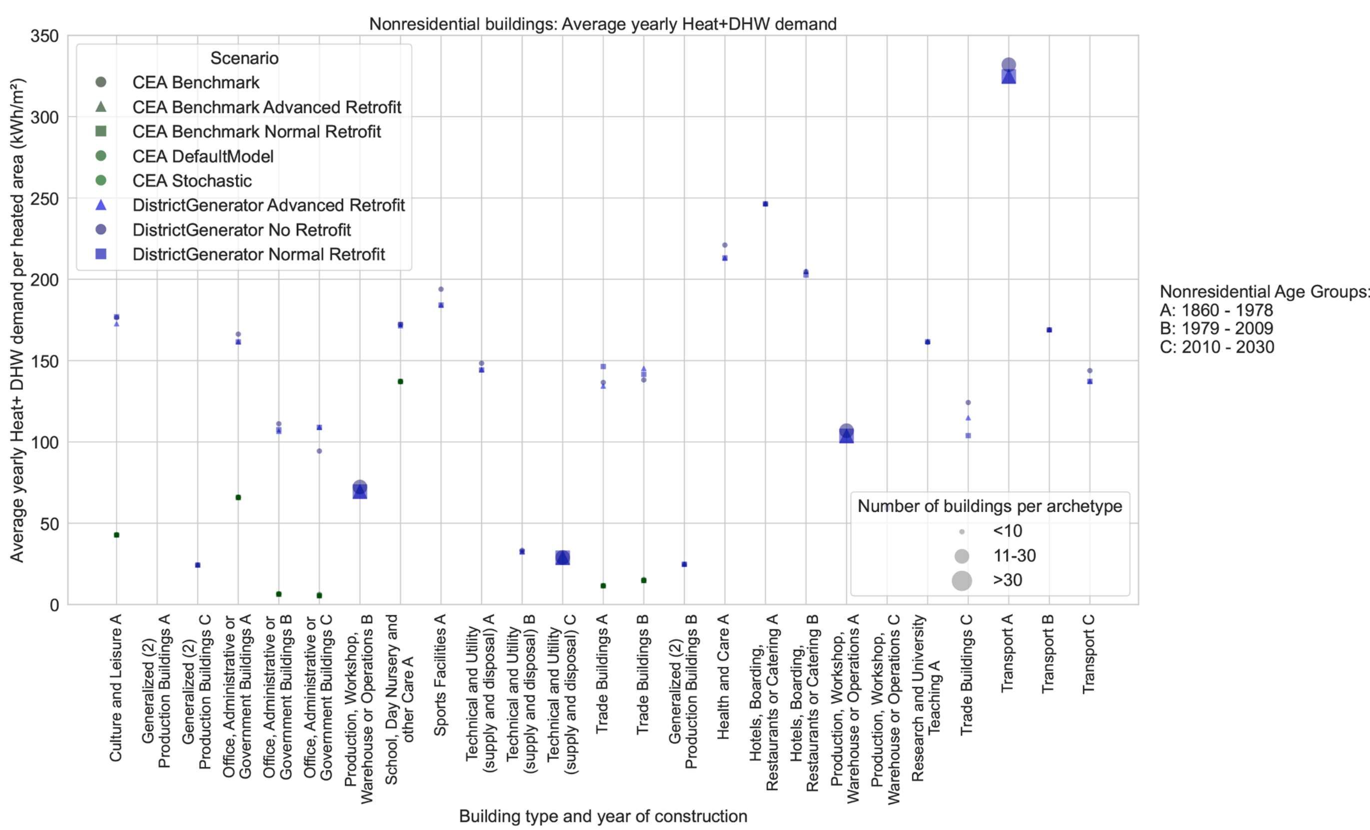


Fig. 10. Yearly energy demand in kWh per square meter, for each nonresidential archetype and all pipelines considered in the study. The size of the marker indicates the amount of archetypes in the respective simulation. As the graphic shows, different approaches lead to different archetypes and quantities of archetypes.

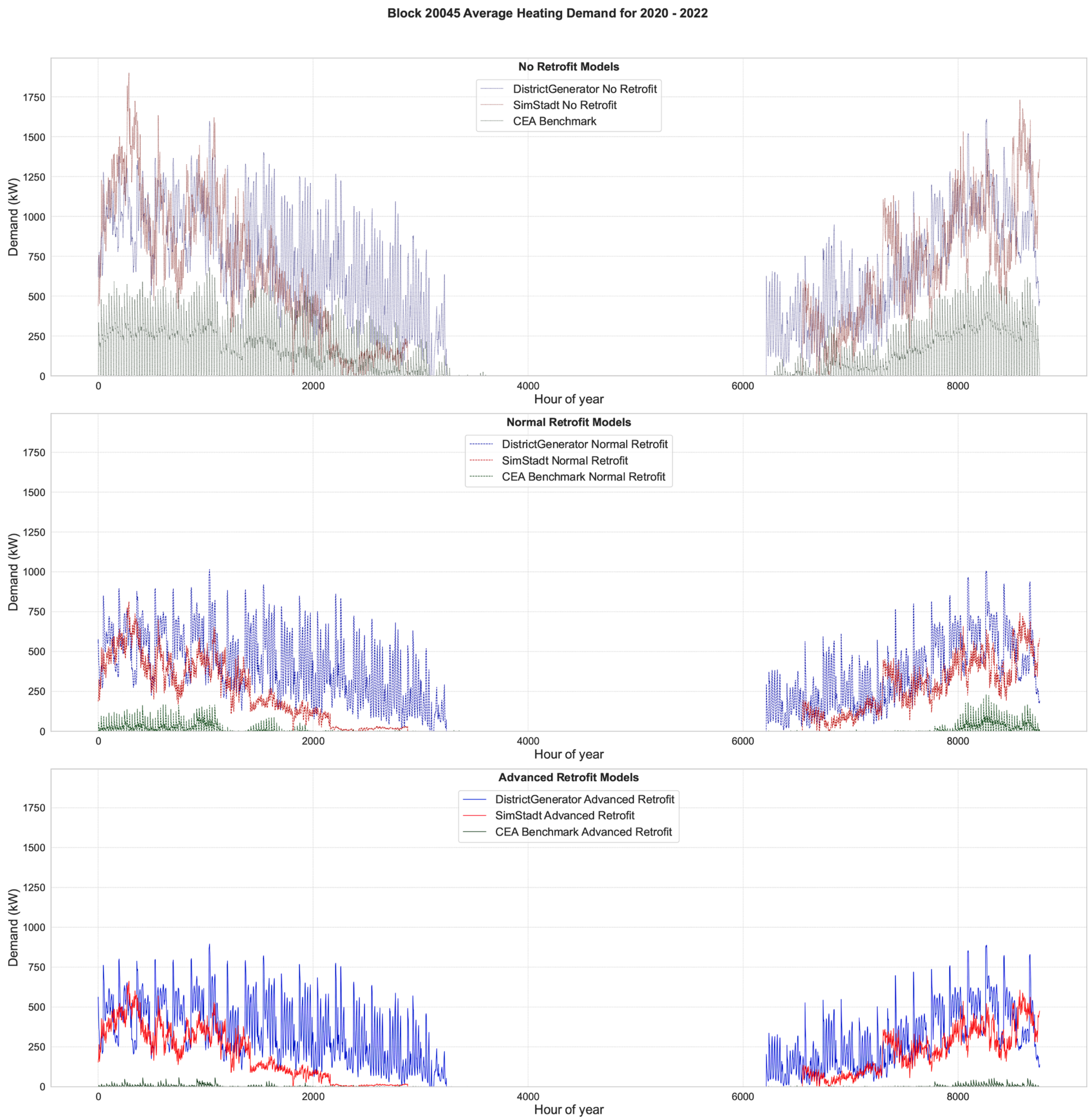


Fig. 11. Comparison of the average hourly simulation demand for the years 2020–2022 for the block 20045. To aid clarity, we only visualize the retrofit scenarios with the aim to highlight the effects of different modeling assumptions of the non-envelope parameters. For example, in case of CEA the setback is visible.

To improve the understanding of the tools and the simulated energy demand, Fig. 11 presents the average hourly heating demand. To aid clarity, the figure visualizes and compares only the retrofit scenarios. These scenarios provide similar assumptions in the envelope, and hence the comparison enables the interpretation of effects related to other modeling aspects. The figure displays block 20045, for which we assumed that the year of construction is 1905. In CityGML, the block comprises 17 buildings (14 residential, two churches, and one commercial building), whereas in OSM it comprises eleven buildings (nine residential, one church, and one commercial building). When inspecting the graphs, two effects become apparent: First, the energy demand in CEA is significantly lower and regularly drops to zero, especially at night, which reflects the setback in the default model. Furthermore, in the CEA and DistrictGenerator scenarios, the buildings have a longer heating period compared to SimStadt. The CEA advanced retrofit demand is close to zero, which is probably an error. In that case, residential buildings have no heating demand. However, a thorough inspection of U-values and g-values showed no deviation from the TABULA typology. In addition, we inspected the advanced retrofit archetypes by selecting normal retrofit modeling parameters (e.g., ventilation), which had no effect on the energy demand of residential buildings. Hence, the energy demand of the CEA advanced

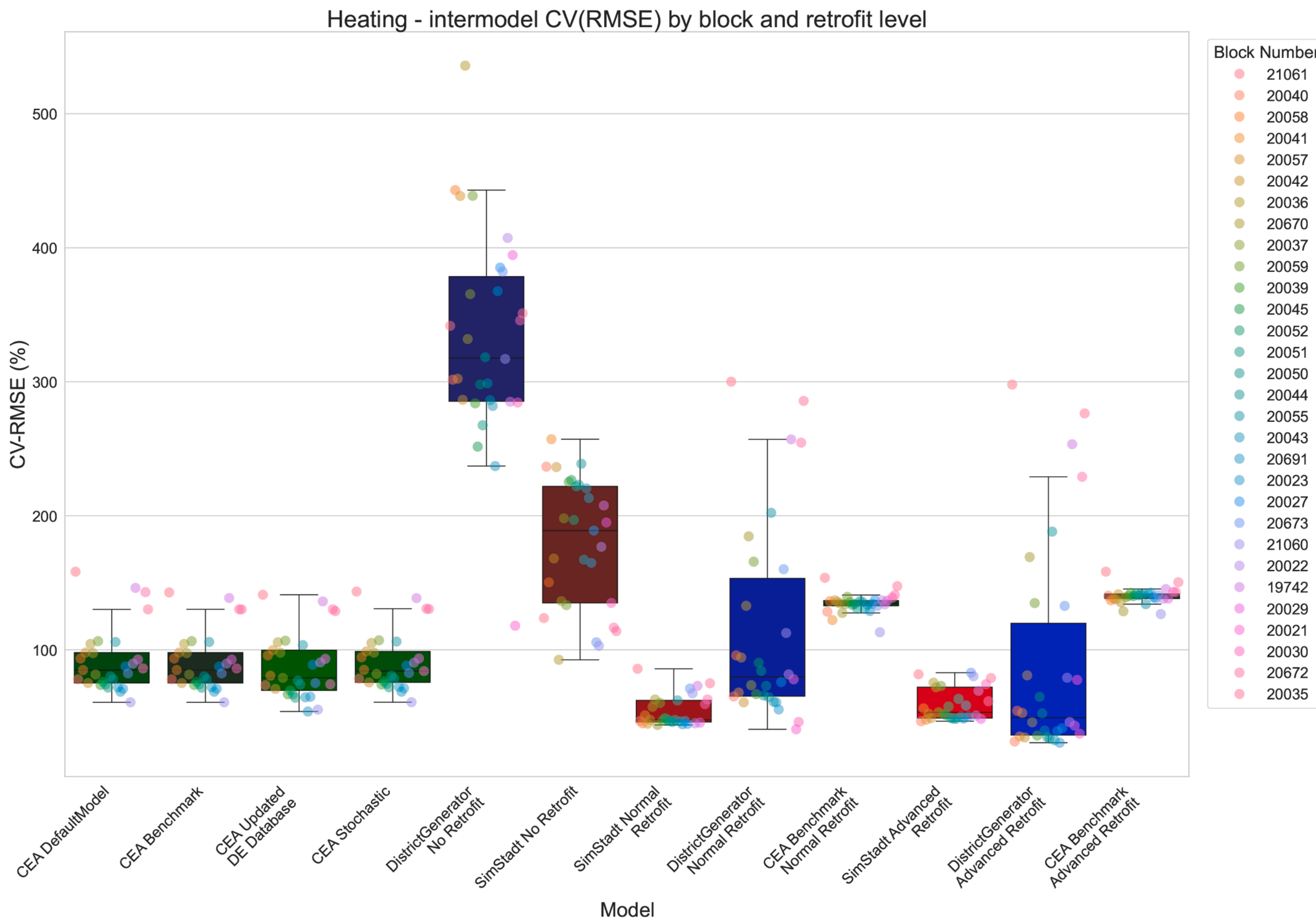


Fig. 12. Comparison of the CV(RMSE) of the average hourly heating demand at block level for the years 2020–2022. We determine the CV(RMSE) by evaluating similar retrofit scenarios. For instance, we compute the average for all three years across all three advanced retrofit scenarios at the block level. Subsequently, we determine the CV(RMSE) for each scenario.

retrofit archetypes for heating is too low and mostly driven by DHW demand.

Fig. 12 presents a block wise comparison of the CV(RMSE) of the heating demand. The plot highlights the relationship between the simulated energy demand and the large spread between the various scenarios. The average of the no retrofit scenarios follows the CEA results, as these are four out of six no retrofit scenarios. In addition, the graphic highlights previous findings. For example, the no retrofit DistrictGenerator model is above average, due to effects such as the larger heated area. The figure shows that the stochastic modeling of presence affects the outliers by comparing the CEA No Retrofit and CEA Stochastic models. However, this effect is random and can vary in strength depending on the respective simulation.

The CV(RMSE) is large compared to the threshold of 15 % from ASHRAE [94]. There are several reasons for this: First of all, the CV(RMSE) compares the average of different models against each other. The considerable large CV(RMSE) is therefore a result of the assumptions previously discussed. The assessment examines the simulated energy demand, influenced by varied factors in present archetypes (e.g., nightly setback), calculation of heated area, and different buildings in input data. In summary, the notably large CV(RMSE) underscores the impact of varying modeling pipelines on the CV(RMSE), as evidenced by its considerable large size.

Fig. 13 displays a block wise comparison of the CV(RMSE) of the electricity demand. Overall, it can also be observed that the electrical simulations diverge less from one another than for the thermal simulations. Again, SimStadt and DistrictGenerator have higher CV(RMSE) scores, indicating the relevance of the calculated area. As expected, the box-plots highlight the variation in modeling of CEA scenarios.

Based on the findings, Table 7 presents the answers to the question 4C1-4C5 of the framework. Due to the lack of measured data, we consider no calibration or threshold at building level to verify the models.

5.5. Information application evaluation

Table 8 presents the evaluation of the information application. All three applications simulate the energy demand for heating, DHW, and electricity. However, SimStadt does not support the hourly simulation of DHW without soft linking it to DHW calc. Since the results and code are openly available, the answers to questions 5C1, 5C2, and 5C4 do not differ. Whether the approaches offer infrastructure to make the results reusable and update them depends on the definition of *reusable*. SimStadt best illustrates integration into further applications, for example in the work by Eicker et al. [115]. Additionally, using standardized data formats and consistent identifiers, both provided by CityGML, can increase reusability. However, users currently need to model information about the buildings themselves (such as function and year of construction) with Python in the GML files.

CEA can enable users to update individual parameters through both APIs and the provided interface. However, if the automated zone generator based on OSM is used, CEA automatically generates building names (e.g., B1000), which hinders the comparison of different pipelines. DistrictGenerator, as the newest tool, so far offers the least integration into other infrastructures and currently does not support updating

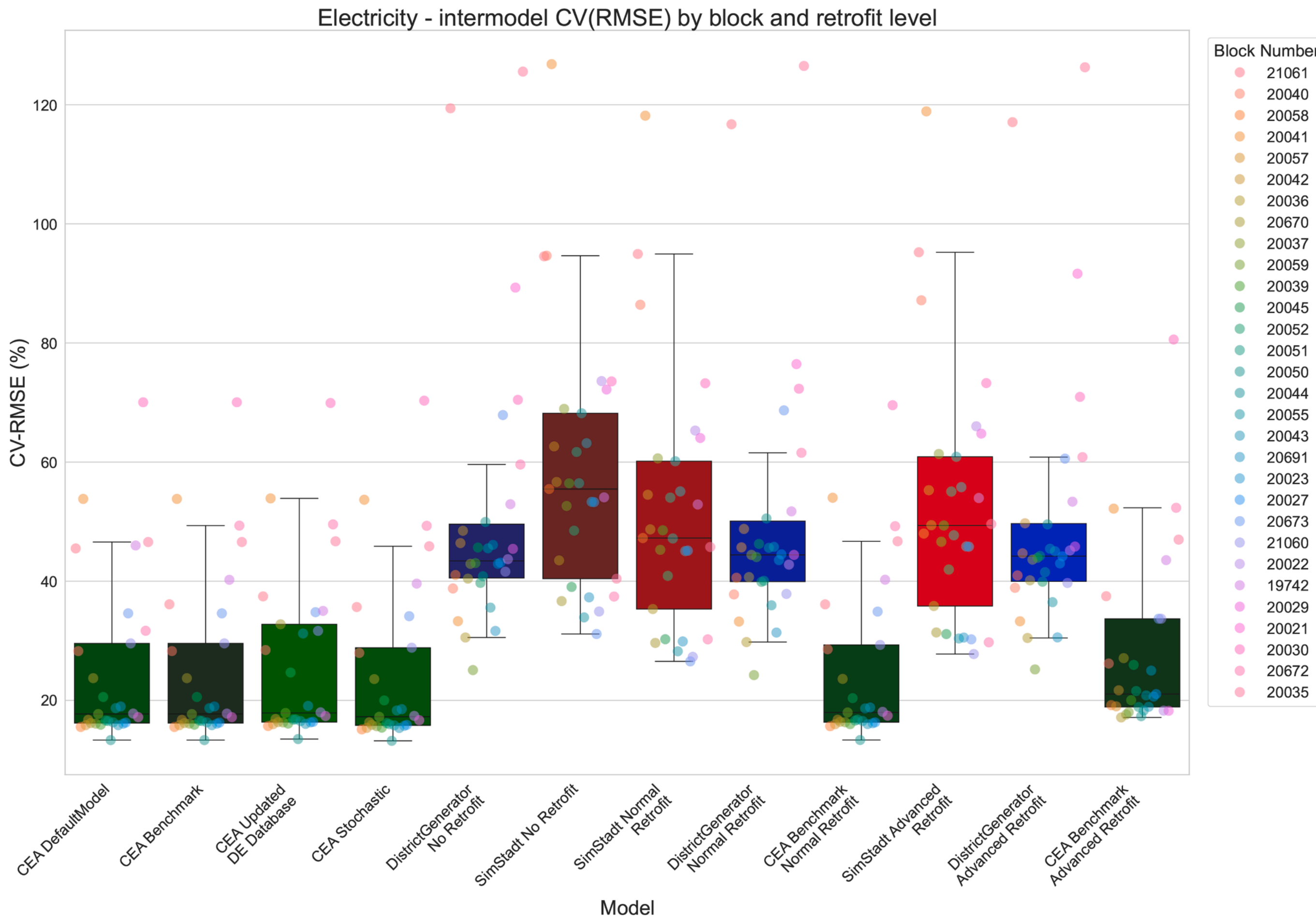


Fig. 13. Comparison of the CV(RMSE) of the average hourly electricity demand at block level for the years 2020–2022. We determine the CV(RMSE) by evaluating similar retrofit scenarios. For instance, we calculate the mean for all three years across all three advanced retrofit scenarios at the block level. Subsequently, the CV(RMSE) for each scenario is determined.

Table 7
Evaluation of criteria 4C.

Criteria	CEA	DistrictGenerator	SimStadt
4C1 – Is sensitivity analysis conducted on key model parameters?	Retrofit, Comparison of single zone vs. multi-zone average profiles	Retrofit	Retrofit
4C2 – Is the calibration performed at an appropriate temporal resolution?	None	None	None
4C3 – Is the calibration performed at an appropriate spatial resolution?	None	None	None
4C4 – Are standard verification metrics used to validate the model?	Yes, MAE, kWh/m2	Yes, MAE, kWh/m2	Yes, MAE, kWh/m2
4C5 – Are predefined thresholds applied for model verification?	No	No	No

Table 8
Evaluation of criteria 5C.

Criteria	CEA	DistrictGenerator	SimStadt
5C1 – Is the Urban Energy Model applied to a practical use case?	Yes, comparison of UBEM pipelines	Yes, comparison of UBEM pipelines	Yes, comparison of UBEM pipelines
5C2 – Do the results have a temporal resolution that aligns with the intended application	Yes, yearly and hourly	Yes, yearly and hourly	Yes, yearly (DHW, heating, and electricity and hourly (heating and electricity)
5C3 – Does the approach allow the infrastructure to be reusable or updated?	Yes, by using the API or interface	No	Yes, by sharing EnergyADE models
5C4 – Are the results openly available?	Yes	Yes	Yes

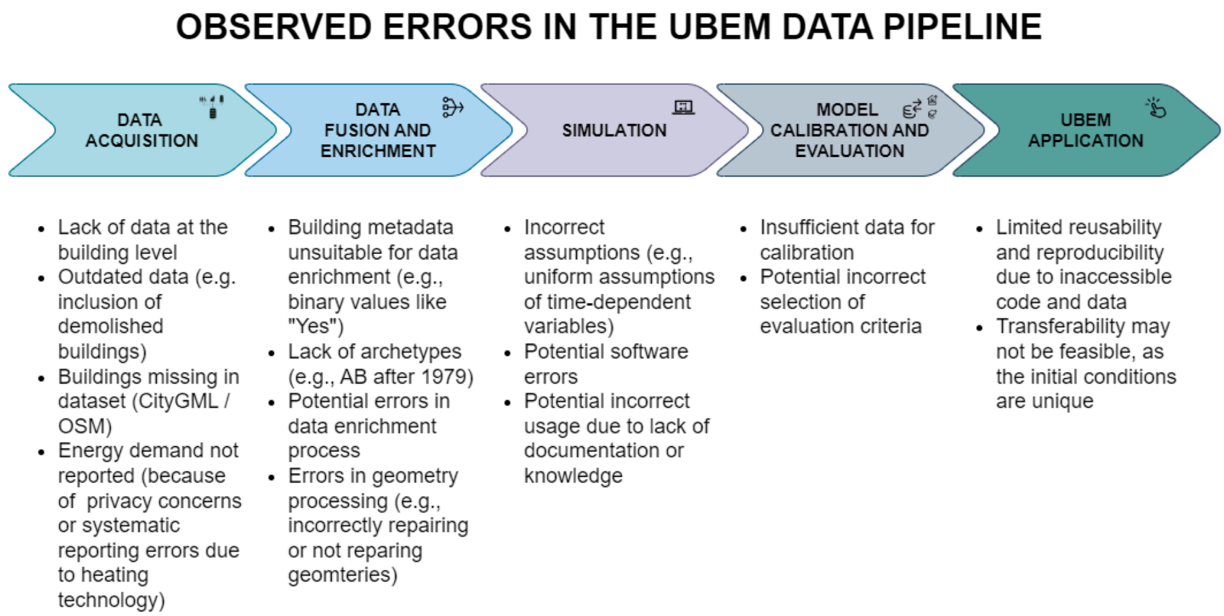


Fig. 14. The figure summarizes the (potential) observed errors identified in the study. Each step in the UBEM data pipeline influences the subsequent steps, and these errors can either accumulate or cancel each other out.

parameters other than the archetype identifiers without programming knowledge. Based on this assessment, we consider SimStadt and CEA *Yes* in rating, while we give DistrictGenerator a *No* in category 5C3.

6. Discussion

6.1. Contribution

This study introduces a new workflow for modeling mixed-use districts. For this purpose, we integrate CityGML and a nonresidential building typology by Hörner et al. [18] into DistrictGenerator by Henn et al. [30]. To highlight interdependence in large-scale modeling, we evaluate the tool together with CEA and SimStadt in a district in Berlin using open data. To conduct the evaluation, we apply an integrated approach based on data pipelines. For each step, we formulate several questions that aim to help modelers evaluate and reflect on their decisions. We derive the questions from an extensive literature discussion. This section critically discusses those contributions. By developing new UBEM pipelines, such as DistrictGenerator, and comparing them against existing pipelines (e.g., CEA, SimStadt), we provide a comparative analysis for each step and its respective interdependence within the data pipeline. In this section, we link the observed errors to the initial UBEM data pipeline presented in Fig. 14. The respective sections present and discuss the observed errors.

6.2. General discussion

This paper combines both quantitative and qualitative approaches for UBEM, evaluating these methods under varying assumptions. The methodology highlights various approaches that yield different results and provides examples for their comparison and evaluation. Although there are more detailed approaches for individual buildings with high data availability, both time and data availability can represent significant obstacles to generating calibrated models at the building scale. The initially discussed heating and cooling plans require innovative solutions in large urban areas, where suitable standards are still lacking, as discussed in subsection 2.4. Such standards should consider the complexities inherent in the complete UBEM pipeline, as discussed in this paper.

6.2.1. Data acquisition

In this paper, we rely on publicly available information for data acquisition using crowd-sourced data (OSM) and official data (CityGML, block-level data). Although open data portals using standards like CityGML or platforms such as OSM offer building information, data are often invalid or have inconsistent quality. For example, file errors are common, information may be incomplete, and mapping standards are inconsistent. In our study, the OSM data tended to be more current; however, the quality of the *Building* tag was variable and often subjective (for example, binary values such as *Yes*). However, the format exhibited fewer geometric errors, such as missing building footprints and inconsistent methods for representing surface areas.

A key data shortage in this study is the lack of building-level information. Because of privacy concerns, there are currently no data about building energy end use and building age that are publicly available. As discussed in subsection 5.4, not only the geometric data, but also the energy data have potential errors. The data have unrealistic values, such as very high or low demand or missing values for some years.

In addition, simulating energy use over multiple years reflects changes in the built environment during the observation period. For example, in the OSM files, different buildings are present than in CityGML. If building-level energy data are not available that match the measured energy with the simulated energy demand, this can introduce errors. The *Data Acquisition* category of Fig. 14 lists the observed errors.

6.2.2. Data fusion and enrichment

As demonstrated in Figs. 9 and 10, different pipelines produce different archetypes. Although this variance may be manageable in small-scale scenarios (where detailed knowledge about individual buildings or manual inspections are feasible), in urban-scale scenarios, it becomes a significant factor. For example, where SimStadt models buildings as single-family homes (TH homes), major discrepancies become apparent across archetype categories. This is a potential error in the data enrichment process, as the area of investigation has limited single-family homes. In addition, the absence of metadata (CityGML) regarding the number of stories necessitates estimating the heated area. As demonstrated in subsection 5.1, these missing metadata or errors (OSM) influence the aggregate energy demand on the block scale, along with the general threshold in buildings analyzed by the methods studied.

As the comparison between the geometries extracted by the presented approach and SimStadt highlights, fewer data processing and enrichment steps are necessary to process the OSM data. Hence, the OSM CEA integration simplifies the data utilization process. However, CityGML, as an official data source, provides more detailed information on building usage. This is especially relevant in nonresidential buildings, where the typology is less dependent on other factors, such as the size of the building. In contrast, as Fig. 9 highlights, highlights, the process of identifying and classifying residential buildings has more ambiguity.

Moreover, the reviewed tools only partially support regional specificity. Although data for former GDR buildings are largely missing, additional local variations are also absent. SimStadt includes modern apartment blocks, yet other necessary data are lacking. And while the total AB constructed in Germany after 1979 are few, a lack of such archetypes in urban areas can lead to a corresponding error, as found in this study. By comparing the approaches E1 and E2, this work highlights the need for archetype substitution in the case study examined. Although all tools support the integration and modification of archetype data, this task can be complex. The integration of the database within CEA or the built-in tools in SimStadt appear to be viable solutions to facilitate this process.

A further limitation is the estimation of the year of construction. Currently, there is only open data about the distribution of the year of construction of residential buildings openly available. To reduce complexity, we assumed the full block to be the majority age. This neglects buildings of younger or older construction and affects the simulated energy demand. While this impacts block level MAE, MAPE and CV(RMSE), the simulated energy demand for the archetypes and general effects remain valid. As the aim of the study is to highlight how different pipelines affect the result, we consider this as an example of how one modeling choice to reduce complexity has manifold impacts on the output. The respective section of Fig. 14 summarizes the discussion of data fusion and enrichment.

6.2.3. Simulation

Although all three tools provide documentation, there is always the possibility of errors in handling tools or data processing. For example, CEA requires the initialization of new databases, which can be a potential error in scenario E6. Additionally, the vast assumption of internal modeling criteria might not be suitable for one use case. For example, the data availability for the residential archetypes is higher (e.g., information about material layers), and the reference simulation for the nonresidential archetypes [19] has higher data requirements (e.g., lighting configurations at the workplace). These data are hard to implement at the archetype level that matches the TABULA typology, because such data is currently not systematically collected. In addition, there might be potential simulation errors. For example, as visible in Fig. 9, there are buildings that have a very high energy demand per square meter in SimStadt. To us, it is unclear where this error originates. We also emphasize that using multiple parameters for evaluation allows us to identify such errors and deepens the understanding of the evaluation.

The investigated area also contains buildings that the metadata identifies as production buildings, or other more technical buildings (e.g., the power plant). Although it is already difficult to estimate the energy demand of normal equipment in nonresidential buildings [109], evaluating heating needs is even more dependent on the respective business and the process needs.

Subsection 5.3 describes and evaluates methods for software verification linked to *Calibration and Model Evaluation*. This study does not explore the origins of errors that are from data, archetypes, or the software itself. This represents a limitation that this study addresses through a comparison of various tools, comparable archetypes, and different KPIs. Yet, an independent analysis is not within the scope of this study. Achieving such a task would require comprehensive test cases and uniform data, both of which are currently unavailable for the scenario presented. Instead, the study provides an explicit and detailed discussion of potential error sources. Further, expert users (in best case the developers) are suggested to avoid pitfalls of incorrect usage. Category *Simulation* of Fig. 14 summarizes the potential errors outlined in this section. In addition, we also omit the analysis of computational effort. Generally, the computational burden increases with stochastic modeling, potentially imposing constraints in larger case study contexts.

6.2.4. Calibration and model evaluation

Because of the lack of building-level information, we did not conduct a calibration at the building level. This lack of calibration is one of the main limitations of this study. To address this limitation Table 1 highlights the impact of different parameters on the presented UBEM tools. It is important to understand, which impact the chosen model parameters have on the results and if they are similar in all chosen engines. This can impact for example, the ranking of suggested measures in building-level recommendations.

For the evaluation, we selected and discussed several metrics. Various uncertainties influence the modeling and evaluation of energy end-use beyond the model itself and the assumed input data. In the presented case study, we faced a limitation in the evaluation due to the absence of certain technologies in the available statistics. In Fig. 15, we display the Energy Performance Certificate (EPC) for a building in Mierendorffinsel, which we obtained by scraping a real estate advertisement. The certificate identifies *oil* as the energy carrier. However, municipal authorities in Germany do not systematically collect such data of decentral technologies, which impacts our evaluation process, as we do not categorize it as *measured* data within this study.

This study addresses the limitations by evaluating archetypes, block level simulation, and inter-model data. The use of multiple metrics helps to circumvent potential disadvantages associated with relying on a single evaluation criterion. The intermodel comparison uses CV(RMSE). The altercated scores far exceed typical verification thresholds. However, as this study aims to investigate heterogeneity in modeling, we calculate the CV(RMSE) by comparing the models. As expected, the scenarios have different energy demand at block level, resulting from various factors. The large CV(RMSE) highlights the impact of intelligent data collection and verification strategies, in cases of large scale UBEM models.

Bausubstanz & Energieausweis

Baujahr:	1910	Wesentliche Energieträger:	Öl
Objektzustand:	Gepflegt	Energieausweis:	liegt zur Besichtigung vor
Ausstattung:	Normale Qualität	Baujahr laut Energieausweis ⓘ:	1998
Heizungsart:	Zentralheizung		

Fig. 15. Energy Performance Certificate (EPC) of a building on the Mierendorffinsel obtained from a real estate advertisement webpage. The EPC highlights oil (Öl) as the energy carrier, which the publicly available energy consumption does not report. This can result in an error in the comparison of the reported measured and simulated energy demand.

There are other strategies to verify the modeling and simulation results, that we have not discussed in detail. For example, comparison with other studies (e.g., Brandt et al. [102]), verification with experts or EPC data. We do not consider them in more detail, as they are unlikely to scale to large scale applications or are inaccessible to us. A scalable method applicable in the modeling of grid-scale energy systems, as demonstrated by Booshehri et al. [116], could involve the use of ontology-based comparisons of input data and modeling techniques to further investigate potential discrepancies.

6.2.5. Information application

As discussed in subsection 5.5, there are a variety of applications for UBEM. This study evaluates the resulting energy demand of various UBEM workflows. Accordingly, any scale that meets this objective in all compared workflows is sufficient.

However, the overarching goal of this study is to illustrate how the different steps of the UBEM pipeline influence the results and how these effects can impact real-world decisions, such as the energy plans required by the EPBD. These decisions should be transparent and, consequently, reproducible. As demonstrated in this study, reproducibility remains a significant challenge. To address this, we strive to ensure reproducibility by relying on publicly available data and established approaches and by providing access to the source code. However, a potential challenge arises from the different development environments (e.g., varying Python versions) used by the three approaches, which can complicate reproducibility and increase installation efforts.

6.2.6. Framework

The objective of the proposed framework is to provide guidance for researchers, practitioners, and other stakeholders to enhance their understanding by highlighting the trade-offs they must consider when simulating at a spatial level other than the building level. To ensure practicality, the framework must make several compromises, such as determining the number of questions or the required level of detail. The design of the questions aims to achieve an appropriate balance in this trade-off. Consequently, the selection of questions remains somewhat subjective and may vary depending on the applied methodology. The questions are formulated to allow for simple *Yes* or *No* responses for a quick assessment, while still enabling more detailed answers if necessary. This approach aims to facilitate a quick and accessible introduction to the framework. By applying the framework to this case study, the framework itself is effectiveness is evaluated and iteratively refined; resulting in the presented set of questions. Because archetype-based UBEM originally informed this framework, it does not necessarily guarantee full transferability to other regions or contexts. In the examined region, we observed reduced data availability due to General Data Protection Regulation (GDPR) regulations. Therefore, using an archetype-based approach and evaluating areas above the building level seem appropriate in this context. However, the situation may differ in other regions of the world, where alternative approaches might be more suitable, because of better data availability. However, the concept of integrated analysis in form of data pipelines and the conclusion of smart data management strategies is generally transferable.

6.3. Outlook

Future studies should investigate whether the proposed qualitative criteria enhance both the understanding and comparability of simulation workflows across broader application contexts. Researchers may adopt or adapt the developed framework to assess a wider range of studies from the existing body of literature. Further, we intend to strengthen the integration of future versions of DistrictGenerator with the CityGML ecosystem to facilitate standardized data exchange and improve interoperability.

Subsequent research should also evaluate which methodological approaches are most suitable for different data enrichment tasks. Incorporating innovative enrichment techniques may improve the interpretation of uncertainty factors. For instance, the data-driven methodology proposed by Blanco et al. [117] for estimating building age introduces distinct types of uncertainty compared to those encountered in our own framework. In this context, the *Smart City* paradigm and associated data governance principles provide promising foundations for handling such factors. They may support the reproducibility of simulation studies by enabling continuous data updates, promoting data quality assurance, and facilitating systematic integration across platforms and stakeholders.

6.4. Recommendations

This section provides modeling recommendations, complementing general guidelines for modeling energy or power systems, such as those by Hirth [118] and Pfenninger et al. [119]. As noted above, the EPDB requires municipalities to prepare energy plans. While data availability often presents a major challenge [120], researchers and modelers must also consider the interdependence among data sources and systems, as emphasized in this study. As presented, large scale simulations must consider multiple effects, from geometric input data, to metadata, or data for enrichment.

Hence, energy reports should document not only the models but also data used as input and used for enrichment, as well as potential factors of uncertainty. The TABULA typology, for example, remains a strong resource because it encompasses most European countries [21]. However, it lacks coverage for specific building types, such as modern large multi-family housing or nonresidential structures. Researchers can address and replace these gaps, by extending the database as done by Hörner et al. [18] or as demonstrated in this study. However, as discussed in this study, it is virtually impossible to the vast amount of explicit and implicit assumptions and the resulting calculations. The choice of well documented and established tools helps to avoid such erroneous assumptions.

Modelers must select tools based on the specific requirements of their applications. The appropriate level of spatial or temporal resolution should guide this decision, as also highlighted by Höffner and Glombik [64]. However, additional factors matter. This study illustrates two such cases. First, limited archetype data constrains tools like CEA and DistrictGenerator. Second, although TABULA includes regional archetypes such as GDR-era buildings, SimStadt and DistrictGenerator do not implement them. While we strongly recommends using open-source tools and codebases to promote broader adoption, facilitate validation, and disclose implicit and explicit assumptions, users must verify whether the tools have already integrated the database components relevant to their context. Tools such as SimStadt and CEA include graphical interfaces that enable non-programmers to adopt and exchange them more easily.

A key consideration concerns whether modelers should invest time and resources in collecting data at a higher level of detail or focus first on broader data collection. Based on this study, we propose several conclusions.

First, it is essential to clearly identify the factors that influence the results. Based on our findings, we recommend using average energy consumption per square meter per year as a benchmark to assess the validity of the archetypes. However, as we illustrate in subsection 5.1, significant variations in input data can strongly affect outcomes, so modelers should examine them as a priority. Given the dynamic conditions in urban environments, the lack of publicly available data, and the insufficient documentation of existing datasets, many actors still struggle to obtain coherent and consistent data. Nevertheless, modelers must make efforts to ensure the accuracy of the most critical data inputs-specifically, building floor area, energy demand, and the set of buildings included in the analysis. Our comparison of the two CityGML approaches shows that correctly processing geometric input data has a greater overall impact than improving the accuracy of archetypes. Furthermore, our comparison of SimStadt and DistrictGenerator demonstrates that selecting an appropriate building scale significantly influences the results. Solutions such as Smart City data platforms could improve data quality by, for example, repairing broken geometries or updating changes in building functions [20]. We agree with Nouvel et al. [85], for example, that modelers should treat the data quality of the models informing urban planning as a relevant criterion and report it transparently. The research of Biljecki et al. [44] and Lei et al. [45] can inform the development of such metrics. Our analysis shows that multiple factors influence the quality of acquired energy demand data. In regions where these factors play a significant role and become apparent during equipment acquisition, metrics such as energy demand should reflect and highlight these potential inaccuracies.

A diverse set of usage-type profiles influences the peak ratio at larger aggregation levels beyond individual blocks. When actual data at the single-building level is unavailable, modelers can apply this approach to generate more realistic peak demand estimates. However, as the presented use cases demonstrate, they must account for the randomness of the effect.

CRediT authorship contribution statement

Felix Rehmann: Writing – review & editing, Writing – original draft, Validation, Software, Project administration, Methodology, Investigation, Data curation, Conceptualization; **Martín Mosteiro-Romero:** Methodology, Writing – review & editing, Supervision; **Clayton Miller:** Writing – review & editing, Supervision; **Rita Streblow:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Funding

The authors gratefully acknowledge the financial support by the Federal Ministry for Economic Affairs and Climate Action (BMWK), promotional reference(s): 03EWB004A (Felix Rehmann, Rita Streblow).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used cursor, gpt-4o, claude-sonnet 3.5 in order to help with coding. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Data availability

All data and code is open available.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Mapping of building functions

Table A.9
The mapped archetypes across different standards and data sources. A '–' indicates that this category is not present in the standard or respective data sources. ALKIS (used in the Berlin CityGML data) provides more data types, but for mapping, we only used those present in our case study. Ten types have been excluded from the mapping, such as 31001_2612 for public toilets.

IWU / DistrictGenerator	CEA	OSM	18599	TEK	SIA	ALKIS CityGML	SimStadt
AB	MULTI_RES	yes, residential, apartments	Wohnen (EFH)	–	Wohnen (MFH)	31001_1000, 31001_1221, 31001_1231,	1000, 1010
MFH	MULTI_RES	yes, residential, apartments	Wohnen (MFH)	–	Wohnen (MFH)	31001_1000, 31001_1221, 31001_1231, 31001_1130	1000, 1010
SFH	SINGLE_RES	house	Wohnen (EFH)	–	Wohnen (EFH)	31001_1000, 31001_1331	1000
TH	SINGLE_RES	–	Wohnen (MFH)	–	Wohnen (EFH)	31001_1000, 31001_1321,	1000
IWU Hotels, Boarding, Restaurants or Catering	HOTEL		Hotelzimmer	Hotels / Pensionen	Hotelzimmer	31001_2071, 31001_2083	
IWU Office, Administrative, or Official building	OFFICE	office	Gruppenbüro	Bürogebäude	Einzel-, Gruppenbüro	31001_2020, 31001_2030, 31001_2050,31001_3010, 31001_3015	
	OFFICE_1P		Einzelbüro	–	–	–	
IWU Trade Buildings	RETAIL		Einzelhandel	Verkaufs-stätten	Fachgeschäft	31001_2000, 31001_2010, 31001_2054, 31001_2055, 31001_2210,	
-	FOODSTORE		–	–	–	–	
-	RESTAURANT		–	–	–	–	
IWU Technical and Utility (supply and disposal)	INDUSTRIAL	industrial	Industriehallen – schwer	–	Produktion	31001_2500, 31001_2520 - 31001_2523	
IWU School, Day Nursery and other Care	SCHOOL	school	Klassen-zimmer	Schulen	Schulzimmer	31001_3021, 31001_3065	
IWU Health and Care	HOSPITAL		Bettenzimmer	Beherbergungs-stätten	Bettenzimmer	31001_3060, 31001_1022	
IWU Sports Facilities	GYM		Turnhalle	Sporthallen	Fitnessraum	31001_3211, 51006_1440	
-	SWIMMING		–	–	–	–	
-	SERVERROOM		–	–	–	–	
IWU Transport	PARKING		Parkhaus	Verkehrs-gebäude	Parkhaus	31001_2460, 31001_2461, 31001_2462, 31001_2463	
-	COOLROOM		–	–	–	–	
-	LAB		–	–	–	–	
IWU Culture and Leisure	MUSEUM		Ausstellung-sräume	Ausstellungs-gebäude	Ausstellungs-halle	31001_3041, 31001_3044	
-	LIBRARY		–	–	–	–	
IWU Research and University Teaching	UNIVERSITY		Hörsaal	Hochschule	Hörsaal	31001_3020, 31001_3023	
IWU Generalized (1) Services building	OFFICE	office	–	Verwaltungs-gebäude	Einzel-, Gruppenbüro	–	
IWU Production, Workshop, Warehouse or Operations	INDUSTRIAL	industrial, warehouse	Industriehallen – mittel	Gewerbliche Gebäude	Produktion	31001_2100, 31001_2111, 31001_2120	
IWU Generalized (2) Production buildings	INDUSTRIAL	–	Industriehallen – mittel	Gewerbliche Gebäude	Lagerhalle	–	

References

[1] C.F. Reinhart, C. Cerezo Davila, Urban building energy modeling—A review of a nascent field, Build. Environ. 97 (2016) 196–202. https://doi.org/10.1016/j.buildenv.2015.12.001
[2] T. Hong, Y. Chen, X. Luo, N. Luo, S.H. Lee, Ten questions on urban building energy modeling, Build. Environ. 168 (2020) 106508. https://doi.org/10.1016/j.buildenv.2019.106508
[3] M. Ferrando, F. Causone, T. Hong, Y. Chen, Urban building energy modeling (UBEM) tools: a state-of-the-art review of bottom-up physics-based approaches, Sustain. Cities Soc. 62 (2020) 102408. https://doi.org/10.1016/j.scs.2020.102408
[4] Y.Q. Ang, Z.M. Berzolla, C.F. Reinhart, From concept to application: a review of use cases in urban building energy modeling, Appl. Energy 279 (2020) 115738. https://doi.org/10.1016/j.apenergy.2020.115738
[5] A. Malhotra, J. Bischof, A. Nichersu, K.-H. Häfele, J. Exenberger, D. Sood, J. Allan, J. Frisch, C. van Treeck, J. O'Donnell, G. Schweiger, Information modelling for urban building energy simulation—A taxonomic review, Build. Environ. (2021) 108552. https://doi.org/10.1016/j.buildenv.2021.108552
[6] S. Pfenninger, Open code and data are not enough: understandability as design goal for energy system models, Prog. Energy 6 (3) (2024) 033002. https://doi.org/10.1088/2516-1083/ad371e
[7] U. Ali, M.H. Shamsi, C. Hoare, E. Mangina, J. O'Donnell, Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis, Energy Build. 246 (2021) 111073. https://doi.org/10.1016/j.enbuild.2021.111073

[8] E. Prataviera, P. Romano, L. Carnieletto, F. Pirotti, J. Vivian, A. Zarrella, EUReCA: an open-source urban building energy modelling tool for the efficient evaluation of cities energy demand, Renew. Energy 173 (2021) 544–560. https://doi.org/10.1016/j.renene.2021.03.144

[9] A. Demir Dilsiz, K. Nweye, A.-M. Wu, J.H. Kämpf, F. Biljecki, Z. Nagy, How spatio-temporal resolution impacts urban energy calibration, Energy Build. 292 (2023) 113175–113175. https://doi.org/10.1016/j.enbuild.2023.113175

[10] DIN-Normenaussschuss Bauwesen (NABau), DIN-Normenausschuss Heiz- und Raumlufttechnik sowie deren Sicherheit (NHRS), DINNormenausschuss Lichttechnik (FNL), 2018. DIN V 18599-10:2018-09: Energetische Bewertung von Gebäuden - Berechnung des Nutz-, End- und Primürenergiebedarfs für Heizung, Kühlung, Lüftung, Trinkwarmwasser und Beleuchtung - Teil 10: Nutzungsrandbedingungen, Klimadaten https://doi.org/10.31030/2874436

[11] A. Malhotra, M. Shamovich, J. Frisch, C. van Treeck, Urban energy simulations using open CityGML models: a comparative analysis, Energy Build. 255 (2022) 111658. https://doi.org/10.1016/j.enbuild.2021.111658

[12] C. Wang, X. Wang, F. Causone, Y. Yang, N. Gao, Y. Ye, P. Li, X. Shi, Addressing uncertainty to achieve stability in urban building energy modeling: a comparative study of four possible approaches, Build. Environ. 267 (2025) 112197. https://doi.org/10.1016/j.buildenv.2024.112197

[13] L.D.D. Harvey, K. Korytarova, O. Lucon, V. Roshchanka, Construction of a global disaggregated dataset of building energy use and floor area in 2010, Energy Build. 76 (2014) 488–496. https://doi.org/10.1016/j.enbuild.2014.03.011

[14] S. Zhang, M. Ma, N. Zhou, J. Yan, W. Feng, R. Yan, K. You, J. Zhang, J. Ke, Estimation of global building stocks by 2070: unlocking renovation potential, Nexus 1 (3) (2024) 100019. https://doi.org/10.1016/j.ynexs.2024.100019

[15] S. Becker, J. Hagen, J. Saikiran, R. Krüger, S. de la Serna, DENA GEBÄUDEREPORT 2024. Zahlen, Daten, Fakten zum Klimaschutz im Gebäudebestand, Technical Report, Berlin, 2023. https://www.dena.de/fileadmin/dena/Publikationen/PDFs/2023/dena-Gebaedereport_2024.pdf.

[16] M. Hörner, J. Bischof, Typologie der Nichtwohngebäude in Deutschland – Methodik, Anwendung und Ausblick, Technical Report, 2022. https://www.iwu.de/fileadmin/publikationen/gebaeudebestand/2022_IWU_HoernerEtBischof_WorkingPaper_Typologie-der-Nichtwohngebaeude-Deutschlands.pdf.

[17] M. Buschka, J. Bischof, C. Meier-Dotzler, W. Lang, Developing non-residential building stock archetypes for LCI—A German case study of office and administration buildings, Int. J. Life Cycle Assess. 26 (9) (2021) 1735–1752. https://doi.org/10.1007/s11367-021-01963-5

[18] M. Hörner, H. Cischinsky, M. Behnisch, R. Busch, J. Bischof, M. Rodenfels, A. Hartmann, R. Hecht, G. Meinel, M. Schorcht, S. Schwarz, G. Spars, A.-K. Tigges, Exploring an unknown: representative sample survey on structure and energy-related quality of the non-residential building stock in Germany, Build. Environ. 255 (2024) 111407. https://doi.org/10.1016/j.buildenv.2024.111407

[19] J. Bischof, S. Knoll, A. Duffy, Development of a Python-based simplified hourly building model for non-domestic building stock operational energy simulations, in: Proceedings of BauSim Conference 2022: 9th Conference of IBPSA-Germany and Austria, Weimar, 2022. https://doi.org/10.26868/29761662.2022.5

[20] Leibniz-Institut für ökologische Raumentwicklung, Colouring Cities Research Colouring (Dresden), https://colouring.dresden.ioer.info/

[21] T. Loga, B. Stein, N. Diefenbach, TABULA building typologies in 20 European countries—Making energy-related features of residential building stocks comparable, Energy Build. 132 (2016) 4–12. https://doi.org/10.1016/j.enbuild.2016.06.094

[22] P. Remmen, Automated calibration of non-residential urban building energy modeling, Ph.D. thesis, E.ON Energy Research Center, RWTH Aachen University, Aachen, 2022. OCLC: 1343893830, https://publications.rwth-aachen.de/record/843586/files/843586.pdf.

[23] D. Robinson, F. Haldi, J. Kämpf, P. Leroux, D. Perez, A. Rasheed, U. Wilke, CITYSIM: comprehensive micro-simulation of resource flows for sustainable urban planning, in: Proceedings of the Eleventh International IBPSA Conference, 2009. https://doi.org/10.26868/25222708.2009.1083-1090

[24] M. Pau, P. Kapsalis, Z. Pan, G. Korbakis, D. Pellegrino, A. Monti, MATRYCS—A big data architecture for advanced services in the building domain, Energies 15 (7) (2022) 2568. https://doi.org/10.3390/en15072568

[25] F. Rehmann, F. Cudok, J. Schölzel, T. Schreiber, S. Henn, R. Streblow, District energy management systems: key data points for system integration and related challenges: lessons learned from experts in Germany, Energy Technol. (2024) 2400297. https://doi.org/10.1002/ente.202400297

[26] N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, L.M. Aroyo, "Everyone wants to do the model work, not the data work": data cascades in high-stakes AI, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, ACM, Yokohama Japan, 2021, pp. 1–15. https://doi.org/10.1145/3411764.3445518

[27] C. Klemm, P. Vennemann, Modeling and optimization of multi-energy systems in mixed-use districts: a review of existing methods and approaches, Renew. Sustain. Energy Rev. 135 (2021) 110206. https://doi.org/10.1016/j.rser.2020.110206

[28] J.A. Fonseca, T.-A. Nguyen, A. Schlueter, F. Marechal, City energy analyst (CEA): integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts, Energy Build. 113 (2016) 202–226. https://doi.org/10.1016/j.enbuild.2015.11.055

[29] P. Remmen, M. Lauster, M. Mans, M. Fuchs, T. Osterhage, D. Müller, TEASER: an open tool for urban energy modelling of building stocks, J. Build. Perform. Simul. 11 (1) (2018) 84–98. https://doi.org/10.1080/19401493.2017.1283539

[30] S. Henn, M. Brunjes, J. Scholzel, T. Beckholter, D. Muller, Districtgenerator: generating building-specific load profiles for residential districts, 2023. https://www.researchgate.net/publication/375448112_Districtgenerator_Generating_building-specific_load_profiles_for_residential_districts.

[31] R. Nouvel, K.-H. Brassel, M. Bruse, E. Duminil, V. Coors, U. Eicker, D. Robinson, SimStadt, a new workflow-driven urban energy simulation platform for CityGML city models, 2015. https://doi.org/10.5075/EPFL-CISBAT2015-889-894

[32] M. Wetter, C. Van Treeck, L. Helsen, A. Maccarini, D. Saelens, D. Robinson, G. Schweiger, IBPSA Project 1: BIM/GIS and modelica framework for building and community energy system design and operation—ongoing developments, lessons learned and challenges, IOP Conf. Ser. 323 (1) (2019) 012114. https://doi.org/10.1088/1755-1315/323/1/012114

[33] V. Todeschi, R. Boghetti, J.H. Kämpf, G. Mutani, Evaluation of urban-scale building energy-use models and tools—Application for the city of Fribourg, Switzerland, Sustainability 13 (4) (2021) 1595. MAG ID: 3129102633. https://doi.org/10.3390/su13041595

[34] C.S. Monteiro, C. Costa, A. Pina, M.Y. Santos, P. Ferrão, An urban building database (UBD) supporting a smart city information system, Energy Build. 158 (2018) 244–260. https://doi.org/10.1016/j.enbuild.2017.10.009

[35] C. Wang, M. Ferrando, F. Causone, X. Jin, X. Zhou, X. Shi, Data acquisition for urban building energy modeling: a review, Build. Environ. 217 (2022) 109056. https://doi.org/10.1016/j.buildenv.2022.109056

[36] S. Goy, F. Maréchal, D. Finn, Data for urban scale building energy modelling: assessing impacts and overcoming availability challenges, Energies 13 (16) (2020) 4244. https://doi.org/10.3390/en13164244

[37] A. Malhotra, J. Bischof, J. Allan, J. O'Donnel, T. Schwenglers, J. Benner, G. Schweiger, A review on country specific data availability and acquisition techniques for city quarter information modelling for building energy analysis, in: BauSIM 2020 : 8th Conference of IBPSA Germany and Austria, Graz, 2020.

[38] A. Malhotra, M. Shamovich, J. Frisch, C. Van Treeck, Parametric study of the different level of detail of CityGML and energy-ADE information for energy performance simulations, in: 16th International Conference of the International Building Performance Simulation Association (Building Simulation 2019), Rome, Italy, 2019, pp. 3429–3436. https://doi.org/10.26868/25222708.2019.210607

[39] Y. Chen, T. Hong, Impacts of building geometry modeling methods on the simulation results of urban building energy models, Appl. Energy 215 (2018) 717–735. MAG ID: 2789313714. https://doi.org/10.1016/j.apenergy.2018.02.073

[40] F. Biljecki, H. Ledoux, J. Stoter, An improved LOD specification for 3D building models, Comput., Environ. Urban Syst. 59 (2016) 25–37. https://doi.org/10.1016/j.compenvurbsys.2016.04.005

[41] T.R. Dougherty, R.K. Jain, TOM.D: taking advantage of microclimate data for urban building energy modeling, Adv. Appl. Energy 10 (2023) 100138. https://doi.org/10.1016/j.adapen.2023.100138

[42] N. Askham, D. Cook, M. Doyle, H. Fereday, M. Gibson, U. Landbeck, R. Lee, C. Maynard, G. Palmer, J. Schwarzenbach, The Six Primary Dimensions for Data Quality Assessment - Defining Data Quality Dimensions, Technical Report, 2013. https://www.sbctc.edu/resources/documents/colleges-staff/commissions-councils/dgc/data-quality-deminsions.pdf.

[43] V. Coors, M. Betz, E. Duminil, A concept of quality management of 3D city models supporting application-specific requirements, PFG—J. Photogramm., Remote Sens. Geoinf. Sci. 88 (1) (2020) 3–14. https://doi.org/10.1007/s41064-020-00094-0

[44] F. Biljecki, Y.S. Chow, K. Lee, Quality of crowdsourced geospatial building information: a global assessment of OpenStreetMap attributes, Build. Environ. 237 (2023) 110295. https://doi.org/10.1016/j.buildenv.2023.110295

[45] B. Lei, R. Stouffs, F. Biljecki, Assessing and benchmarking 3D city models, Int. J. Geogr. Inf. Sci. 37 (4) (2023) 788–809. https://doi.org/10.1080/13658816.2022.2140808

[46] X. Jin, C. Zhang, F. Xiao, A. Li, C. Miller, A review and reflection on open datasets of city-level building energy use and their applications, Energy Build. 285 (2023) 112911. https://doi.org/10.1016/j.enbuild.2023.112911

[47] A. Nichersu, Scale aware modeling and monitoring of the urban energy chain, Ph.D. thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, 2022. https://publikationen.bibliothek.kit.edu/1000149384.

[48] M. Zirak, V. Weiler, M. Hein, U. Eicker, Urban models enrichment for energy applications: challenges in energy simulation using different data sources for building age information, Energy 190 (2020) 116292. https://doi.org/10.1016/j.energy.2019.116292

[49] C.-D. Thiele, P.A. Zadeh, N. Hashempour, S. Staub-French, U. Rüppel, A Multi-stage approach to understand GIS model enrichment used for decision-making support when developing energy retrofit strategies on a neighborhood level, in: Advances in Information Technology in Civil and Building Engineering Proceedings of ICCCBE 2022 - Volume 1, Cape Town, 2023, pp. 367–381. MAG ID: 4387165215. https://doi.org/10.1007/978-3-031-35399w-4_28

[50] Y. Chen, T. Hong, X. Luo, B. Hooper, Development of city buildings dataset for urban building energy modeling, Energy Build. 183 (2019) 252–265. https://doi.org/10.1016/j.enbuild.2018.11.008

[51] G. Agugiaro, CityGML 3DCityDB-Loader plugin for QGIS A quick overview, 2022. https://3d.bk.tudelft.nl/gagugiaro/pdf/2022_10_3DCityDB-Loader_for_QGIS_OGC.pdf.

[52] J.F. Rosser, G. Long, S. Zakhary, D.S. Boyd, Y. Mao, D. Robinson, Modelling urban housing stocks for building energy simulation using CityGML EnergyADE, ISPRS Int. J. Geoinf. 8 (4) (2019) 163. https://doi.org/10.3390/ijgi8040163

[53] A. Geiger, A. Nichersu, K.-H. Häfele, V. Hagenmeyer, Usage profile enrichment of citygml models for urban building energy modeling, 2022. https://doi.org/10.26868/29761662.2022.28

[54] A. Malhotra, M. Shamovich, S. Raming, J. Frisch, C. Van Treeck, An open-source citygml enrichment tool (CityEnrich), 2022. https://doi.org/10.26868/29761662.2022.56

[55] T. Santhanavanich, V. Coors, CityThings: an integration of the dynamic sensor data to the 3D city model, Environ. Plan. B 48 (3) (2021) 417–432. https://doi.org/10.1177/2399808320983000

[56] S. Dabirian, M.M. Saad, S. Hussain, S. Peyman, N. Rahimi, P. Monsalvete Alvarez U, P. Yefi, U. Eicker, Structuring heterogeneous urban data: a framework to develop the data model for energy simulation of cities, Energy Build. 296 (2023) 113376. https://doi.org/10.1016/j.enbuild.2023.113376

[57] C.S. Monteiro, A. Pina, C. Cerezo, C. Reinhart, P. Ferrão, The use of multi-detail building archetypes in urban energy modelling, Energy Procedia 111 (2017) 817–825. https://doi.org/10.1016/j.egypro.2017.03.244

[58] M. Lauster, Parametrierbare Gebäudemodelle für dynamische Energiebedarfsrechnungen von Stadtquartieren, Ph.D. thesis, RWTH Aachen, Aachen, 2018. OCLC: 1189024187, https://publications.rwth-aachen.de/record/749705/files/749705.pdf.

[59] C. Karczewski, J. Bischof, M. Hörner, Evaluating non-domestic building stock simulation based on single-zone models with multi-zone average usage profiles, in: ECEEE Summerstudy 2024 Proceeding, Center Parcs Lac d'Ailette, 2024.

[60] K. Eisenack, S. Villamayor-Tomas, G. Epstein, C. Kimmich, N. Magliocca, D. Manuel-Navarrete, C. Oberlack, M. Roggero, D. Sietz, Design and quality criteria for archetype analysis, Ecol. Soc. 24 (3) (2019) art6. https://doi.org/10.5751/ES-10855-240306

[61] European Parliament, EU Council, 2023. Directive (EU) 2023/955 of the European Parliament and of the Council of 13 September 2023 on energy efficiency and amending Regulation (EU) 2023/955 (RECAST), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AJOL_2023_231_R_0001&qid=1695186598766.

[62] P. Shen, H. Wang, Archetype building energy modeling approaches and applications: a review, Renew. Sustain. Energy Rev. 199 (2024) 114478. https://doi.org/10.1016/j.rser.2024.114478

[63] E. Prataviera, J. Vivian, G. Lombardo, A. Zarrella, Evaluation of the impact of input uncertainty on urban building energy simulations using uncertainty and sensitivity analysis, Appl. Energy 311 (2022) 118691. https://doi.org/10.1016/j.apenergy.2022.118691

[64] D. Höffner, S. Glombik, Energy system planning and analysis software—A comprehensive meta-review with special attention to urban energy systems and district heating, Energy (2024) 132542. https://doi.org/10.1016/j.energy.2024.132542

[65] Y. Xu, J. Litardo, C. Del Pero, F. Leonforte, P. Caputo, District energy models: a comparative assessment of features and criteria for tools selection, Energy Build. 314 (2024) 114291. https://doi.org/10.1016/j.enbuild.2024.114291

[66] ISO/TC 163/SC, ISO 13790:2008 - Energy performance of buildings – Calculation of energy use for space heating and cooling, 2008. https://www.iso.org/standard/41974.html.

[67] ISO/TC 211 Geographic information/Geomatics, CEN/TC 287 Geoinformation, 2023. Geographic information Data quality Part 1: General requirements (ISO 19157-1:2023)

[68] J. McCarty, C. Waibel, A. Galimshina, A. Hollberg, A. Schlueter, Do we need a saw? Carbon-based analysis of facade BIPV performance under partial shading from nearby trees, J. Phys. 2600 (4) (2023) 042002. https://doi.org/10.1088/1742-6596/2600/4/042002

[69] M. Mosteiro-Romero, I. Hischier, J.A. Fonseca, A. Schlueter, A novel population-based occupancy modeling approach for district-scale simulations compared to standard-based methods, Build. Environ. 181 (2020) 107084. https://doi.org/10.1016/j.buildenv.2020.107084

[70] M. Mosteiro-Romero, M. Quintana, R. Stouffs, C. Miller, A data-driven agent-based model of occupants' thermal comfort behaviors for the planning of district-scale flexible work arrangements, Build. Environ. 257 (2024) 111479. https://doi.org/10.1016/j.buildenv.2024.111479

[71] P. Remmen, M. Lauster, M. Mans, T. Osterhage, D. Müller, CityGML import and export for dynamic building performance simulation in modelica, in: Proceedings of BSO Conference 2016: Third Conference of IBPSA-England, Newcastle, UK, 2016, pp. 330–337. http://www.ibpsa.org/proceedings/BSO2016/p1047.pdf.

[72] J. Schiefelbein, J. Rudnick, A. Scholl, P. Remmen, M. Fuchs, D. Müller, Automated urban energy system modeling and thermal building simulation based on OpenStreetMap data sets, Build. Environ. 149 (2019) 630–639. https://doi.org/10.1016/j.buildenv.2018.12.025

[73] I. Richardson, M. Thomson, D. Infield, A high-resolution domestic building occupancy model for energy demand simulations, Energy Build. 40 (8) (2008) 1560–1566. https://doi.org/10.1016/j.enbuild.2008.02.006

[74] Institute for Energy Efficient Buildings and Indoor Climate, RWTH-EBC/richardsonpy, 2024. original-date: 2017-05-30T19:21:03Z, https://github.com/RWTH-EBC/richardsonpy.

[75] V. Coors, P. Rodrigues, V. Weiler, E. Duminil, A. Klöber, D. Holweg, T. Brüggemann, K. Bohn, L. Groll, B. Balbach, F. Spath, EnEff:Stadt - SimStadt 2.0 : Schlussbericht: Laufzeit des Vorhabens: 01.07.2017-31.12.2020, Berichtszeitraum: 01.07.2017-31.12.2020, Technical Report, Hochschule für Technik Stuttgart, 2021. Artwork Size: 1 Online-Ressource (155 Seiten, 7,74 MB) Medium: application/pdf Version Number: 1.0, https://doi.org/10.2314/KXP:1787325431

[76] V.-F.T. Gebäudeausrüstung, VDI 4710 -Meteorologische Daten in der technischen Gebäudeausrüstung Gradtage, 2013. https://www.vdi.de/richtlinien/details/vdi-4710-blatt-1-meteorologische-grundlagen-fuer-die-technische-gebaeudeausruestung-aussereuropaeische- klimadaten.

[77] M. Hillen, P. Schönfeldt, P. Groesdonk, B. Hoffschmidt, Integration of a Europe-wide public building database with retrofit strategies and a thermal inertia model into an open-source optimization framework, IOP Conf. Ser. 1363 (1) (2024) 012013. https://doi.org/10.1088/1755-1315/1363/1/012013

[78] T. Maile, H. Steinacker, M.W. Stickel, E. Ott, C. Kley, Automated generation of energy profiles for urban simulations, Energies 16 (17) (2023) 6115. https://doi.org/10.3390/en16176115

[79] ISO/IEC JTC 1/SC, ISO/IEC 25002:2024 Systems and software engineering—Systems and software Quality Requirements and Evaluation (SQuaRE)—Quality model overview and usage, 2024. https://www.iso.org/standard/78175.html.

[80] J. Neymark, R. Judkoff, I. Beausoleil-Morrison, A. Ben-Nakhi, M. Crowley, M. Deru, R. Henninger, H. Ribberink, J. Thornton, A. Wijsman, M. Witte, International Energy Agency Building Energy Simulation Test and Diagnostic Method (IEA BESTEST): In-Depth Diagnostic Cases for Ground Coupled Heat Transfer Related to Slab-on-Grade Construction, Technical Report NREL/TP-550-43388, 937333, 2008. https://doi.org/10.2172/937333

[81] D. Blum, F. Jorissen, S. Huang, Y. Chen, J. Arroyo, K. Benne, Y. Li, V. Gavan, L. Rivalin, L. Helsen, D. Vrabie, M. Wetter, M. Sofos, Prototyping the BOPTEST framework for simulation-based testing of advanced control strategies in Buildings, Rome, Italy, 2019, pp. 2737–2744. https://doi.org/10.26868/25222708.2019.211276

[82] J. Grunewald, A. Nicolai, S. Hirth, D. Weiß, C. van Treeck, J. Frisch, A. Nouri, C. Emunds, M. Madjidi, J. Agudelo, I. Reichenbach, R. Rolffs, R. Strobel, R. Tang, H. Falko, ENOB: SimQuality - Entwicklung von Qualitätsstandards für die energetische Gebäude- und Quartierssimulation als Planungswerkzeug, Technical Report, TU Dresden, 2022. https://doi.org/10.2314/KXP:1859246915.

[83] H. Johra, M. Mans, K. Filonenko, I.D. Jaeger, D. Saelens, Evaluating different metrics for inter-model comparison of urban-scale building energy simulation time series, in: Building Simulation Conference Proceedings, Bruges, Belgium, 2021. S2ID: 5d0668cf4d7f66460d5f95f7dbc84a048c63bc2e. https://doi.org/10.26868/25222708.2021.30410

[84] D. Saelens, I. De Jaeger, F. Bünning, M. Mans, A. Maccarini, E. Garreau, O. Rønneseth, I. Sartori, A. Vandermeulen, B. van der Heijde, L. Helsen, Towards a DESTEST: a district energy simulation test developed in IBPSA project 1, in: Proceedings of Building Simulation 2019: 16th Conference of IBPSA, Rome, Italy, 2019, pp. 3569–3577. https://doi.org/10.26868/25222708.2019.210806

[85] R. Nouvel, M. Zirak, V. Coors, U. Eicker, The influence of data quality on urban heating demand modeling using 3D city models, Comput., Environ. Urban Syst. 64 (2017) 68–80. https://doi.org/10.1016/j.compenvurbsys.2016.12.005

[86] M. Mosteiro-Romero, J.A. Fonseca, A. Schlueter, Seasonal effects of input parameters in urban-scale building energy simulation, Energy Procedia 122 (2017) 433–438. https://doi.org/10.1016/j.egypro.2017.07.459

[87] M. Mosteiro-Romero, C. Miller, A. Chong, R. Stouffs, Elastic buildings: calibrated district-scale simulation of occupant-flexible campus operation for hybrid work optimization, Build. Environ. 237 (2023) 110318. https://doi.org/10.1016/j.buildenv.2023.110318

[88] A. Demir Dilsiz, K. Ng, J. Kämpf, Z. Nagy, Ranking parameters in urban energy models for various building forms and climates using sensitivity analysis, Build. Simul. 16 (9) (2023) 1587–1600. https://doi.org/10.1007/s12273-022-0961-5

[89] A. Krüger, T.H. Kolbe, Building analysis for urban energy planning using key indicators on virtual 3D city models - the energy atlas of Berlin, in: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Melbourne, Australia, XXXIX-B2, 2012, pp. 145–150. https://doi.org/10.5194/isprsarchives-XXXIX-B2-145-2012

[90] M.M. Saad, U. Eicker, Investigating the reliability of building energy models: comparative analysis of the impact of data pipelines and model complexities, J. Build. Eng. 71 (2023) 106511. https://doi.org/10.1016/j.jobe.2023.106511

[91] M. Lauster, D. Müller, Methods of time series analysis for simulation-based urban scale evaluations, in: Proceedings of BauSim Conference 2018: 7th Conference of IBPSA-Germany and Austria, 2018, pp. 463–470.

[92] F. Johari, J. Munkhammar, F. Shadram, J. Widń, Evaluation of simplified building energy models for urban-scale energy analysis of buildings, Build. Environ. 211 (2022) 108684. https://doi.org/10.1016/j.buildenv.2021.108684

[93] A. Oraiopoulos, B. Howard, On the accuracy of urban building energy modelling, Renew. Sustain. Energy Rev. 158 (2022) 111976. https://doi.org/10.1016/j.rser.2021.111976

[94] T.E. Watson, C.G. Arnold, V.D. Baxter, D.S. Borges, P.W. Cabot, W.S. Clements, C.W. Coward, H.L. Crowder, B.P. Dougherty, R.A. Evans, A.D. Hallstrom, R.D. Hermans, J.F. Hogan, F.H. Kohloss, W.J. Landman, R.H. Lewis, R.D. Montgomery, D. Novosel, D.A. Stanke, M. Tavares, S.T. Taylor, J.R. Wright, L.W. Burgett, G.V.R. Holness, C.B. Ramspeck, ASHRAE Guideline 14–2002: Measurement of Energy and Demand Savings, 2002. http://www.eeperformance.org/uploads/8/6/5/0/8650231/ashrae_guideline_14-2002_measurement_of_energy_and_demand_saving.pdf.

[95] G. Ruiz, C. Bandera, Validation of calibrated energy models: common errors, Energies 10 (10) (2017) 1587. https://doi.org/10.3390/en10101587

[96] U.S. Department of Energy, M&V Guidelines: Measurement and Verification for Performance-Based Contracts Version 4.0, 2015. https://www.energy.gov/sites/default/files/2016/01/f28/mv_guide_4_0.pdf.

[97] Efficiency Valuation Organization, International Performance Measurement and Verification Protocol - Concepts and Options for Determining Energy and Water Savings Volume 1, 2012. https://www.eeperformance.org/uploads/8/6/5/0/8650231/ipmvp_volume_i_2012.pdf.

[98] F. Biljecki, J. Stoter, H. Ledoux, S. Zlatanova, A. Çöltekin, Applications of 3D city models: state of the art review, ISPRS Int. J. Geoinf. 4 (4) (2015) 2842–2889. https://doi.org/10.3390/ijgi4042842

[99] M. Schildt, J. Cuypers, M. Shamovich, S. Herzogenrath, A. Malhotra, C. Van Treeck, J. Frisch, On the potential of district-scale life cycle assessments of buildings, Energies 16 (15) (2023) 5639. https://doi.org/10.3390/en16155639

[100] I. Dochev, P. Gorzalka, V. Weiler, J. Estevam Schmiedt, M. Linkiewicz, U. Eicker, B. Hoffschmidt, I. Peters, B. Schröter, Calculating urban heat demands: an analysis of two modelling approaches and remote sensing for input data and validation, Energy Build. 226 (2020) 110378. https://doi.org/10.1016/j.enbuild.2020.110378

[101] T. Dogan, C. Reinhart, Shoeboxer: an algorithm for abstracted rapid multi-zone urban building energy model generation and simulation, Energy Build. 140 (2017) 140–153. https://doi.org/10.1016/j.enbuild.2017.01.030

[102] D. Brandt, S. Brandt, D. Kreulitscht, M. Kriegel, C. Nytsch-Geusen, O. Tcvetkova, G. Wang, Schlussbericht Energetische Simulation der Mierendorff-Insel, 2020. https://www.berlin.de/ba-charlottenburg-wilmersdorf/verwaltung/aemter/umwelt-und-naturschutz/klimaschutz/energetische-quartierskonzepte/energetische-simulation-mierendorffinsel_schlussbericht.pdf.
[103] F. Hewelt, H. Rogall, A.M. Welz, K. Gapp-Schmeling, KoWa - Wärmewende im Quartier Berlin Mierendorff-Insel. Erfahrungsbericht: Clusteranalyse und Konzeptionierung, 2022. https://www.kowa-projekt.de/wp-content/uploads_kowa/2022/05/M4_Erfahrungsbericht_MiDoI_WEB.pdf.
[104] Senatsverwaltung für Stadtentwicklung, Bauen und Wohnen, 3D-Gebäudemodelle im Level of Detail 2 (LoD 2). https://fbinter.stadt-berlin.de/fb/berlin/service_intern.jsp?id = a_lod2@senstadt&type = FEED.
[105] Senatsverwaltung für Wirtschaft, Energie und Betriebe, Energieatlas Berlin, 2025. https://energieatlas.berlin.de/.
[106] Statistische Ämter des Bundes und der Länder, Deutschland, Gebäude: Baujahr (Jahrzwanzigste) - Gebäudetyp (Größe), 2024. https://ergebnisse.zensus2022.de/datenbank/online/statistic/3000G/table/3000G-2005.
[107] A. Ceruti, M. Geske, U. Hartmann, H. Spliethoff, C. Voelker, From building to district: accelerating urban building energy modeling with an open-source database for Germany, Vienna, 2024. https://doi.org/10.26868/29761662.2024.22
[108] S. Henning, K. Jagnow, Statistische Untersuchung der Flächen- und Nutzstromanateile von Zonen in Nichtwohngebäuden (Fortführung), Technical Report 51/2023, 2023. https://www.bbsr.bund.de/BBSR/DE/veroeffentlichungen/bbsr-online/2023/bbsr-online-51-2023-dl.pdf?__blob = publicationFile&v = 3.
[109] K. Jagnow, S. Henning, Statistische Untersuchung der Flächen- und Nutzstromanteile von Zonen in Nichtwohngebäuden, 2020. https://www.h2.de/fileadmin/user_upload/Fachbereiche/Bauwesen/Forschung/Forschungsberichte/Endbericht_SWD-10.08.18.7-18.29.pdf.
[110] vfdb - Vereinigung zur Förderung des Deutschen Brandschutzes e.V., Leitfaden Ingenieurmethoden des Brandschutz, Technical Report vfdb TB 04-01, Technisch-Wissenschaftlicher Beirat (TWB) der Vereinigung zur Förderung des Deutschen Brandschutzes e.V. (vfdb), Münster; Braunschweig, 2020. https://www.vfdb.de/media/doc/technischeberichte/TB_04_01_Leitfaden_IngMethoden_4Auflage_2020-03-26.pdf.
[111] Schweizerischer Ingenieur- und Architektenverein, Raumnutzungsdaten für die Energie- und Gebäudetechnik, 2021. https://shop.sia.ch/normenwerk/architekt/2024_2021_d/D/Product.
[112] M. Shamovich, S. Raming, A. Malhotra, C. Van Treeck, J. Frisch, CityDPC: a Python library for handling 3D city model datasets, Bauphysik 46 (6) (2024) 340–347. https://doi.org/10.1002/bapi.202400038
[113] R. Kaden, Berechnung der Energiebedarfe von Wohngebäuden und Modellierung energiebezogener Kennwerte auf der Basis semantischer 3D-Stadtmodelle, Ph.D. thesis, Technische Universität München, München, 2014. https://mediatum.ub.tum.de/doc/1210304/1210304.pdf.
[114] S. Köhler, M. Betz, U. Eicker, Stochastic generation of household electricity load profiles in 15-min resolution on building level for whole city quarters, Ljubljana, 2019. https://iaee2019ljubljana.oyco.eu/download/contribution/presentation/214/214_presentation_20190828_130329.pdf.
[115] U. Eicker, V. Weiler, J. Schumacher, R. Braun, On the design of an urban data and modeling platform and its application to urban district analyses, Energy Build. 217 (2020) 109954. https://doi.org/10.1016/j.enbuild.2020.109954
[116] M. Booshehri, L. Emele, S. Flügel, H. Förster, J. Frey, U. Frey, M. Glauer, J. Hastings, C. Hofmann, C. Hoyer-Klick, L. Hülk, A. Kleinau, K. Knosala, L. Kotzur, P. Kuckertz, T. Mossakowski, C. Muschner, F. Neuhaus, M. Pehl, M. Robinius, V. Sehn, M. Stappel, Introducing the open energy ontology: enhancing data interpretation and interfacing in energy systems analysis, Energy AI 5 (2021) 100074. https://doi.org/10.1016/j.egyai.2021.100074
[117] L. Blanco, A. Alhamwi, B. Schiricke, B. Hoffschmidt, Data-driven classification of urban energy units for district-level heating and electricity demand analysis, Sustain. Cities Soc. 101 (2024) 105075. https://doi.org/10.1016/j.scs.2023.105075
[118] L. Hirth, Open data for electricity modeling: legal aspects, Energy Strategy Rev. 27 (2020) 100433. https://doi.org/10.1016/j.esr.2019.100433
[119] S. Pfenninger, L. Hirth, I. Schlecht, E. Schmid, F. Wiese, T. Brown, C. Davis, M. Gidden, H. Heinrichs, C. Heuberger, S. Hilpert, U. Krien, C. Matke, A. Nebel, R. Morrison, B. Müller, G. Pleßmann, M. Reeg, J.C. Richstein, A. Shivakumar, I. Staffell, T. Tröndle, C. Wingenbach, Opening the black box of energy modelling: strategies and lessons learned, Energy Strategy Rev. 19 (2018) 63–71. https://doi.org/10.1016/j.esr.2017.12.002
[120] G. Schweiger, Johannes Exenberger, A. Malhotra, Thomas Schranz, T. Boiger, C.v. Treeck, J. O'Donnell, Data shortage for urban energy simulations? An empirical survey on data availability and enrichment methods using machine Learning, in: EG-ICE 2021 Proceedings: Workshop on Intelligent Computing in Engineering, Universitätsverlag der TU Berlin, Berlin, 2021, pp. 301–309. https://doi.org/10.14279/DEPOSITONCE-12021