This is the Accepted Manuscript version of an article published by Elsevier in the journal *Energy and Buildings* in 2023, which is available at: `https://doi.org/10.1016/j.enbuild.2023.113175`
Cite as: Dilsiz AD, Nweye KE, Wu AJ, Kämpf JH, Biljecki F, Nagy Z (2023): How spatio-temporal resolution impacts urban energy calibration. *Energy and Buildings*, 292: 113175.

How spatio-temporal resolution impacts urban energy calibration

Aysegul Demir Dilsiz[a], Kingsley E Nweye[b], Allen J Wu[b], Jérôme H. Kämpf[c], Filip Biljecki[d], Zoltan Nagy[b]

[a] *Dept. Civil and Architectural Engineering and Construction Management, University of Wyoming, Laramie WY, USA*
[b] *Intelligent Environments Lab, Dept. Civil, Architectural and Environmental Engineering, The University of Texas at Austin, Austin TX, USA*
[c] *Energy Informatics Group, Idiap Research Institute, 1920 Martigny, Switzerland*
[d] *National University of Singapore, Singapore*

Abstract

Building Energy Modeling tools help forecast the energy performance of buildings. Urban energy models (UBEMs) emerged as important instruments to analyze the energy performance of buildings aggregated at different spatial resolutions, from the building level to the district level. They heavily rely on available data on geometries and measurements to create accurately calibrated energy models. However, limited research has been conducted to understand the impact of spatial and temporal resolution on the simulation results because of the difficulty of comparing results and not having a standardized procedure to report simulation errors. We review the literature on UBEM validation compared to measured energy data and show the discrepancies in the reporting accuracy. We articulate the need for consistent reporting on model accuracy and introduce a multi-dimensional Level of Detail (LoD) specification for UBEM, including geometry, thermal zoning, and spatio-temporal resolution of the measured data used to calibrate the models. Using a university campus with 70 buildings as an extensive case study, we demonstrate the performance of Bayesian calibration from the building level to the aggregated level. Our results suggest that the accuracy of urban energy prediction with annual temporal resolution can be significantly increased if calibration is performed by using building-level data. However, whenever privacy is a concern, then the data should be provided by aggregating them based on primary use type. Additionally, using monthly data to calibrate uncertain input parameters is not improving the accuracy of the models because the obtained posterior distributions for the selected parameters are not informative for monthly data. To improve this shortcoming, we suggest seasonal calibration, which is computationally costly.

Keywords: Urban Energy Modeling, Bayesian Calibration, Temporal and Spatial Resolution, Level of Detail, Energy modeling, Sustainable urban planning

1. Introduction

There are over 7 billion people globally, and about 60% of them live in urban areas. According to United Nations [1], the share of people living in urban areas is expected to increase to over 70% by 2050, and the global population is projected to reach 10 billion. Therefore, urbanization perpetually remains an important issue, mainly because cities are responsible for 70% of global energy consumption, which will only increase in the future. Building operation and construction consume 36% of global energy use and produce 39% of energy-related CO_2 emissions [2]. Various studies have been conducted to understand the individual energy contributions of different types of buildings([3, 4, 5, 6, 7, 8]). However, it is essential to understand the relationship of each building with its surroundings, which has spurred the development of urban energy models.

Furthermore, city planners need to analyze the building energy demand for building stock and assess their future energy policies. Researchers perform building energy simulations at an urban scale as a helpful tool. With various urban building energy modeling (UBEM) software packages available, energy simulation tools are easy to employ and help researchers estimate the potential to improve building energy performance at a large scale for retrofitting. They are also helpful in discussing reconstruction options, system performance optimization, exploring options for energy supply systems, and policy assessment [9, 10, 11, 12]. Yet, due to the complexity of urban energy systems and required resources, bettering urban energy performance seems quite challenging [13, 14, 15].

Typical urban energy simulation studies use simplified models, such as archetypes for buildings, and show the impact on refurbishment only concerning typical weather data while reducing computational time at the cost of accuracy. A short computational time is especially important if the model is to be used in an optimization study or forecasting scenarios over a long period in the future, such as several years [16]. However, the existing UBEM should represent the actual behavior. This outlook can only be discussed by using measured data and calibration, which is a necessary step to reduce the uncertainties on the input parameters of UBEM. Uncertain parameters are generally chosen using statistical information and simulation results for developing a reliable and accurate UBEM [17, 18].

Calibration of uncertain parameters of UBEM can be performed at multiple temporal and spatial resolutions such as annual, monthly, or hourly on the building, neighborhood, or city. In early review papers in UBEM, it is clearly observed that collecting all the necessary data for validation or calibration processes because of privacy concerns, creating a sufficiently detailed model, and performing required simulations because of computational cost remained challenges faced regardless of the urban energy modeling approach [19, 20, 21]. Although most of these studies have been performed to overcome the stated challenges, they have not provided detailed information on the accuracy of their methodology. Some studies [22, 23, 24, 25, 26] have discussed how necessary it is to provide detailed input data, temporal resolution, or spatial resolution to obtain consistent results. Their work indicates that we cannot draw a direct relationship between increasing the resolution and complexity of the model inputs [27] from the case studies in the literature.

There is a clear need for consistent reporting in UBEM to make these existing and future studies comparable and discuss how different levels of detail in modeling and resolution of the used measured data affect our reported error on the urban scale. The objective of this paper is threefold. First, we introduce a multi-dimensional level of detail description to describe possible spatio-temporal resolutions for UBEM. Second, we review relevant work in UBEM and their validation/calibration methodology. Third, we use introduced multi-dimensional LoDs to investigate the impact of the spatio-temporal resolution on Bayesian calibration in a case study of 70 buildings. The paper is structured as follows. Section 2 reviews the levels of detail and how they are used in the literature for geometry, thermal zoning, and spatial and temporal resolution. We then introduce a general, multi-dimensional definition for LoDs. In Section 3, we used the introduced LoDs for UBEM to systematically analyze simulation errors in the literature. In Section 4, we discuss commonly used Bayesian Calibration methodology and how we implement it on different levels for the case study of the University Campus. Finally, in section 5, we provide our results, and in Section 6, we discuss our results to answer the following questions:

1. What are the relevant and frequently used levels of detail in UBEM?

2. What is the level of detail for spatial and temporal resolution necessary to validate the urban energy model with a deterministic approach? How the reported error differs for each approach?

3. How does performing Bayesian calibration with aggregated data improve the predictions of cooling demand? Which approach gives the closest results to the building level?

4. How should we report our results so everyone can benefit from the accuracy of the methodology discussed?

2. Introducing a multi-dimensional Level of Detail (LoD)

Urban energy models are used to identify smart energy solutions for sustainable cities and policies and support energy and environmental goals. Therefore, these models provide insights to inform city decision-making on sustainability, efficiency, and resilience. However, urban energy modeling is often over-parameterized. It requires a tremendous amount of time and resources to complete accurately as it involves the calibration of simulation outputs with measured energy data for accurate urban modeling. Although the advances in sensing technologies and emerging smart city initiatives enabled the streaming of structured and unstructured data to describe buildings and their surroundings, the availability of such data to create energy models is limited at the building level because of privacy concerns. But the question is, how does this streaming data change the accuracy of our urban energy models? how should we report our results so they would be comparable to future studies with different methodologies? It is essential to understand the level of detail of model inputs to create the Urban Building Energy Modeling (UBEM) and spatial and temporal resolution of the streamed data on the model accuracy while balancing the efforts spent on model development while maintaining the reliability of the results.

An improved specification of Level of Detail (LoD) in 3D city modeling is first introduced by Biljecki et al. (2016) [28], which includes a framework that defines granularity in detail provided by the geometric model. Mathur et al. (2021) [29] also use the same LoDs for the urban energy modeling literature. Oraiopoulos and Howard (2022) [27] recent review of the literature performed in UBEM shows that the error reported in the literature is not reported consistently, and thus reporting the accuracy for simplifications is essential. Some studies have performed Bayesian calibration for buildings with simpler geometry using building-level metered data [24, 30, 31]. On the contrary, some researchers conducted it using only aggregated data while simulation results were obtained using very detailed geometry [32, 33].

Some studies have reported that oversimplification of urban data and modeling approaches might cause large discrepancies. Still, very detailed inputs and metered data are not always necessary to obtain consistent results from a UBEM [25, 34, 26]. Yet, it is hard to conclude this without comparing consistently reported errors from presented case studies. It is clear that there is a need for this subset of LoDs for the area of energy modeling to provide comparable methodologies.

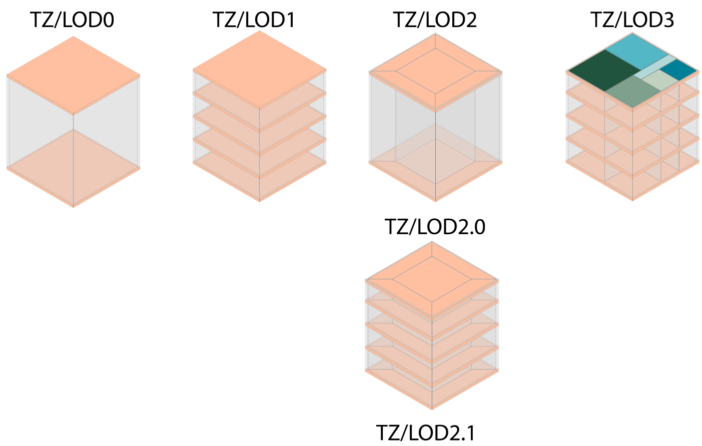

Figure 1: Levels of Detail for thermal zone layer used in urban energy modeling literature(Detailed explanations for each acronym can be found in Appendix A.).

In order to compare these case studies and discuss methodologies for different granularity and data segmentation,we included the details for four layers of Geometry, Zoning, Spatial Resolution, and Temporal Resolution defined by Mathur et al. (2021) [29] to discuss the errors for different approaches in the calibrated/validated UBEMs. Based on the provided literature review, we present a subset of LoDs for the geometry (Figure 2), thermal zone (Figure 1), temporal, and spatial resolution categories for urban energy modeling (Appendix A: Introduced LoDs). In the review, we use this developed specification to determine the LoD used in each of the reviewed studies.

3. Systematic Literature Review using Multi-dimensional LoDs

In this section, we summarize research work related to large-scale energy modeling approaches with case studies in the literature by using introduced multi-dimensional LoDs. The primary databases for the literature searches were Scopus and Google Scholar. The searches were completed based on the title, abstract, and keywords containing the following search formulas:

1. (urban OR city) AND (large OR scale) AND (building OR housing) AND (energy OR electricity)
2. (urban OR city) AND (heating OR cooling) AND (calibration OR validation)
3. (urban OR city) AND (modeling OR model) AND (calibration) AND (validation)
4. (urban OR city) AND (energy OR electricity) AND (calibration AND/OR validation)

Finally, we review 60 journal papers and include 23 articles with either validation or calibration performed using measured data on the bottom-up physics-based tools created for urban applications. Bottom-up models provide energy insights from the building level to the city level. We focus on physics-based tools, known as engineering or simulation methods based on thermodynamic simulations because they capture the entire dynamic of high-resolution building performance [20]. The presented literature papers have either applied validation or calibration to demonstrate their prediction accuracy. Thirty percent of the studies have used the validation process for the predicted energy-related data, 58% of them have applied calibration, and 55% of the calibrations are listed as Bayesian Calibration.

We organize the literature tables (Table 1 and 2 as follows: modeling tool, use of measured data, details of geometry, details of thermal zoning, details of temporal resolution, details of spatial resolution, number of buildings, primary use type, error metrics, and accuracy. The performed literature review in our study includes error and accuracy to compare the case studies with calibration or validation. The complexity of LoD is defined from 0 to n, with zero being the most straightforward and 'n' being the most complex level of detail. The following section describes the details for the given LoDs. These subset LoDs can be increased based on future studies by considering more attributes under each layer.

The accuracy of UBEM refers to the model output error when compared to measured data. Although the American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE) Guideline 14-2002 has recommendations (NMBE and CVRMSE) on energy and demand saving for error reporting, it has not been applied until recently [32, 43, 42, 36, 24]. Among these observed error measures, non-normalized measures cannot be used to compare studies. From this literature review, it is clearly seen that the current ranges for the model accuracy in the UBEM studies are not reported similarly. Therefore, it is difficult to discuss them regarding the accuracy of applied methodologies. Furthermore, the assumptions made by the modeler in UBEMs regarding all model characterizations, such as geometry and thermal zone, might have a considerable impact on the results and require more detailed documentation, as discussed in the table 1 and 2.

3.1. Geometry

According to the review, G/LOD1 is the most common approach applied for geometry, mainly for data availability. However, the impact of geometry input in urban energy modeling has hardly been discussed. To the extent of our knowledge, only Faure et al. (2022) [22] addressed the impact of G/LOD on the model accuracy for urban energy modeling by using G/LOD1 with height details (G/LOD1.2.1) and without them (G/LOD1.1.1) and concluded that for district-scale analysis, G/LOD1.2.1 might not be required. They also stated that for energy conversation measures at the building level, it would be more accurate to perform LOD1.2.1. Because of the lack of data challenge, G/LOD2 and G/LOD3 are barely discussed for UBEM. Nouvel et al. (2017) [44] compared the model accuracy and observed an increase of 15–20% in the result accuracy for buildings with pitched roofs and attics while modeling a 3D geometry, G/LOD2.2.0, as compared to 2.5D extrusion G/LOD1.2.0 geometry. Risch et al .[24] also calibrated three office buildings located in Germany with different geometries and levels of information. They provide two other metrics to put their results in the context of ASHRAE Guideline 14 [48]). Detailing the geometry increase the Coefficient of Variation of the Root Mean Squared Error (CV(RMSE)) for one building within three compared buildings.

3.2. Thermal Zone

The general approach for thermal zoning is TZ/LOD1, thermal zone per floor, followed by TZ/LOD0 single zoning. Chen and Hong (2018) [25] created UBEM with 940 buildings and

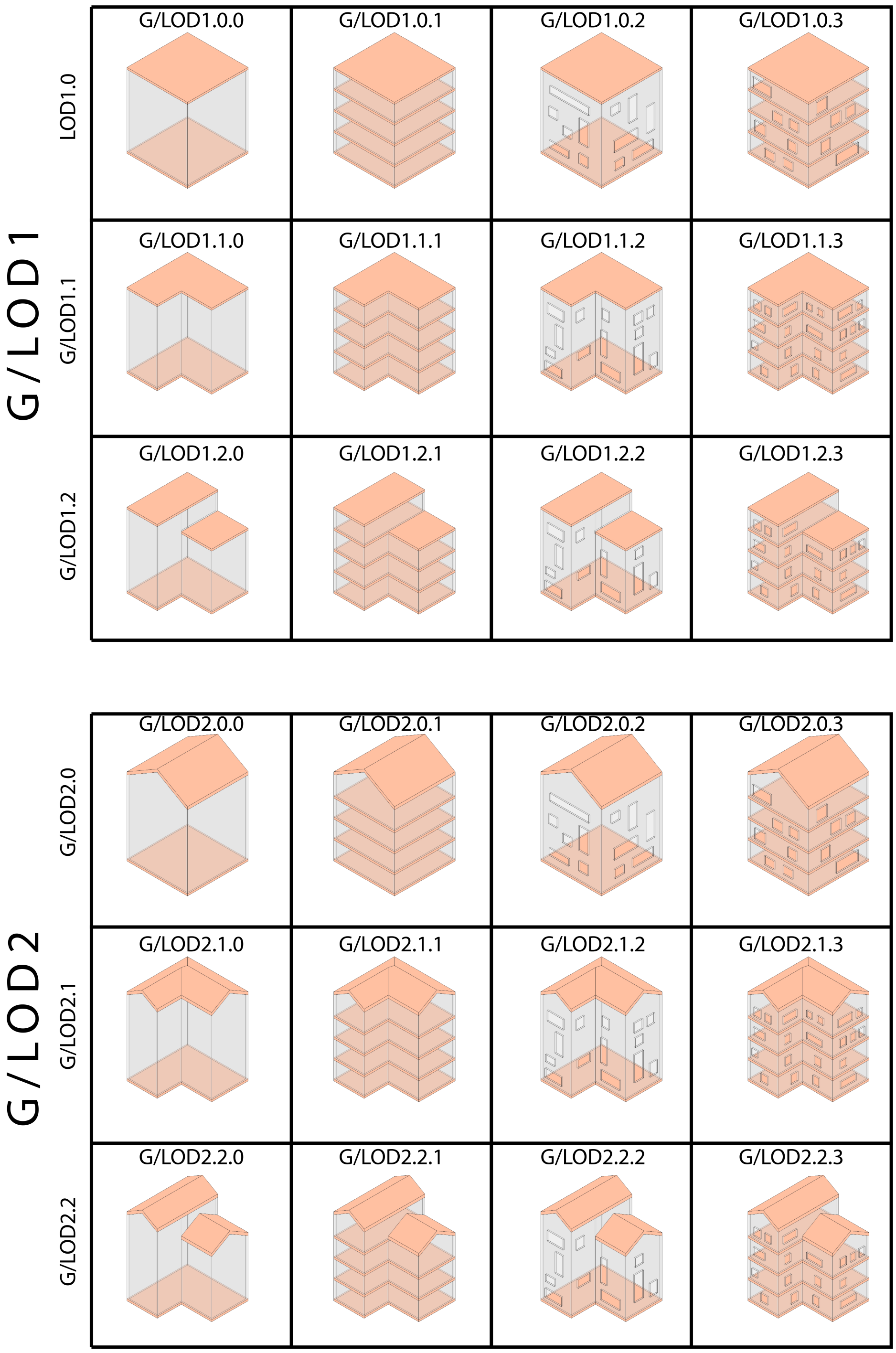

Figure 2: Levels of Detail for geometries used in urban energy modeling literature (Detailed explanations for each acronym can be found in Appendix A.)

Table 1: Details of UBEM studies- Input and Metered Data used for Validation and Calibration.

Reference	Modeling Tool	Use of Measured Data	Details of Geometry	Details of Thermal Zoning	Details of Temporal Resolution	Details of Spatial Resolution	Number of Buildings	Primary Use Type	Error Metric	Accuracy
Katal et al. 2022 [35]	CityBEM	Validation	G/LOD1.1.0	TZ/LOD0	TR/LOD2	NS	255	Mixed	NRMSE	22
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.1.1	TZ/LOD0	TR/LOD0	SR/LOD2.0	33	Residential	TEDI	-0.5
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.1.1	TZ/LOD0	TR/LOD0	SR/LOD3	33	Residential	TEDI	Between -1.5 and 0.2
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.1.1	TZ/LOD2.0	TR/LOD0	SR/LOD2.0	33	Residential	TEDI	+1.7
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.1.1	TZ/LOD2.0	TR/LOD0	SR/LOD3	33	Residential	TEDI	Between 1.6 and 2.5
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.1.1	TZ/LOD2.1	TR/LOD0	SR/LOD2.0	33	Residential	TEDI	+2.1
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.1.1	TZ/LOD2.1	TR/LOD0	SR/LOD3	33	Residential	TEDI	Between 0.2 and 2.5
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.2.1	TZ/LOD0	TR/LOD0	SR/LOD2.0	33	Residential	TEDI	-0.3
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.2.1	TZ/LOD0	TR/LOD0	SR/LOD3	33	Residential	TEDI	Between -2 and 0.4
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.2.1	TZ/LOD2.0	TR/LOD0	SR/LOD2.0	33	Residential	TEDI	+1.7
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.2.1	TZ/LOD2.0	TR/LOD0	SR/LOD3	33	Residential	TEDI	Between 1.8 and 2.1
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.2.1	TZ/LOD2.1	TR/LOD0	SR/LOD2.0	33	Residential	TEDI	+2.0
Faure et al. 2022 [22]	MUBES/EnergyPlus	Validation	G/LOD1.2.1	TZ/LOD2.1	TR/LOD0	SR/LOD3	33	Residential	TEDI	Between 0.2 and 2.4
Todeschi et al. 2021 [23]	CitySIM	Calibarion	G/LOD1.2.1	TZ/LOD1	TR/LOD0	SR/LOD3	200	Mixed	MAPE	19.3
Risch et al. 2021 [24]	TEASER	Bayesian Calibration	G/LOD1.0.1	TZ/LOD3	TR/LOD0	SR/LOD3	3	Residential	R-squared (R^2)	63 and 80 and 86
Risch et al. 2021 [24]	TEASER	Bayesian Calibration	G/LOD1.0.3	TZ/LOD3	TR/LOD0	SR/LOD3	3	Residential	R^2	63 and 80 and 78
Risch et al. 2021 [24]	TEASER	Bayesian Calibration	G/LOD1.0.1	TZ/LOD3	TR/LOD0	SR/LOD3	3	Residential	CVRMSE	64 and 44 and 26
Risch et al. 2021 [24]	TEASER	Bayesian Calibration	G/LOD1.0.3	TZ/LOD3	TR/LOD0	SR/LOD3	3	Residential	CVRMSE	63 and 43 and 33
Rashidfarokhi 2021 [30]	NS	Bayesian Calibration	NS	NS	TR/LOD0	SR/LOD3	2	Residential	%	15 and 19
Mathur et al. 2021 [29]	EnergyPlus	Calibration	G/LOD1.0.3	TZ/LOD1	TR/LOD1	SR/LOD3	250	Mixed	N/A	N/A
Gholami et al.2021 [36]	EnergyPlus	Bayesian Calibration	G/LOD1.1.3	TZ/LOD2.1	TR/LOD0	SR/LOD1	1,156	Mixed	CVRMSE	Between 0.3-0.6
Gholami et al.2021 [36]	EnergyPlus	Bayesian Calibration	G/LOD1.1.3	TZ/LOD2.1	TR/LOD0	SR/LOD1	1,156	Mixed	MAPE	Between 2.5-5.1
Ledesma et al. 2021 [37]	Design Builder	Manual Calibration	G/LOD2.2.3	TZ/LOD3	TR/LOD0	SR/LOD2.0	51	Educational	%	below 1
Tardioli et al. 2020 [33]	EnergyPlus	Bayesian Calibration	G/LOD2.2.3	TZ/LOD1	TR/LOD0	SR/LOD2.0	2,646	Mixed	%	5
Tardioli et al. 2020 [33]	EnergyPlus	Bayesian Calibration	G/LOD2.2.3	TZ/LOD1	TR/LOD0	SR/LOD3	326	Mixed	%	+-20
Wang et al. 2020 [38]	CitySim	Bayesian Calibration	G/LOD1.2.0	TZ/LOD0	TR/LOD0	SR/LOD2.0	2,178	Residential	%	Between 7.7-8.3
Berthou et al. 2019 [39]	Smart-E	N/A	N/A	TZ/LOD1	TR/LOD0	SR/LOD2.0	4,000	Residential	%	above 10
Krayem et al. 2019 [40]	EnergyPlus	Clustering Calibration	G/LOD1.1.3	TZ/LOD1	TR/LOD1	SR/LOD1	2311	Mixed	%	Between 2-262

Table 2: Details of UBEM studies- Input and Metered Data used for Validation and Calibration-Continue

Reference	Modeling Tool	Use of Measured Data	Details of Geometry	Details of Thermal Zoning	Details of Temporal Resolution	Details of Spatial Resolution	Number of Buildings	Primary Use Type	Error Metric	Accuracy
Haneef et al. 2019 [41]	CitySim	Calibration	G/LOD1.1.0	TZ/LOD0	TR/LOD0	N/A	11	Residential	%	within 15
Nutkiewicz et al. 2018 [42]	DUE-S/EnergyPlus	Validation	G/LOD1.2.3	TZ/LOD2.1	TR/LOD1	SR/LOD2.0	22	Mixed	CVRMSE	11.4
Nutkiewicz et al. 2018 [42]	DUE-S/EnergyPlus	Validation	G/LOD1.2.3	TZ/LOD2.1	TR/LOD2	SR/LOD2.0	22	Mixed	CVRMSE	14.4
Nutkiewicz et al. 2018 [42]	DUE-S/EnergyPlus	Validation	G/LOD1.2.3	TZ/LOD2.1	TR/LOD3	SR/LOD2.0	22	Mixed	CVRMSE	25.6
Nutkiewicz et al. 2018 [42]	DUE-S/EnergyPlus	Validation	G/LOD1.2.3	TZ/LOD2.1	TR/LOD1	SR/LOD3	22	Mixed	CVRMSE	27.9
Nutkiewicz et al. 2018 [42]	DUE-S/EnergyPlus	Validation	G/LOD1.2.3	TZ/LOD2.1	TR/LOD2	SR/LOD3	22	Mixed	CVRMSE	31.3
Nutkiewicz et al. 2018 [42]	DUE-S/EnergyPlus	Validation	G/LOD1.2.3	TZ/LOD2.1	TR/LOD3	SR/LOD3	22	Mixed	CVRMSE	46
Nageler et al. 2018 [43]	IDA ICE	Validation	G/LOD2.2.3	TZ/LOD3	TR/LOD1	SR/LOD2.0	34	Mixed Use	R^2	92
Nageler et al. 2018 [43]	IDA ICE	Validation	G/LOD2.2.3	TZ/LOD3	TR/LOD1	SR/LOD2.0	34	Mixed Use	CVRMSE	21.4
Nageler et al. 2018 [43]	IDA ICE	Validation	G/LOD2.2.3	TZ/LOD3	TR/LOD1	SR/LOD3	2	Mixed Use	R^2	68 and 92
Nageler et al. 2018 [43]	IDA ICE	Validation	G/LOD2.2.3	TZ/LOD3	TR/LOD1	SR/LOD3	2	Mixed Use	CVRMSE	40.2 and 24.9
Chen et al. 2018 [25]	EnergyPlus	Validation	G/LOD1.1.3	TZ/LOD1	TR/LOD0	SR/LOD3	940	Office and Retail	N/A	Varies
Chen et al. 2018 [25]	EnergyPlus	Validation	G/LOD1.1.3	TZ/LOD2.1	TR/LOD0	SR/LOD3	940	Office and Retail	N/A	Varies
Nouvel et al. 2017 [44]	SimStadt	Validation	G/LOD1.2.0	TZ/LOD0	TR/LOD0	SR/LOD2.0	14,000	Residential	MAPE	Between 32-55
Nouvel et al. 2017 [44]	SimStadt	Validation	G/LOD1.2.2	TZ/LOD0	TR/LOD0	SR/LOD2.0	14,000	Residential	MAPE	Between 20-34
Nouvel et al. 2017 [44]	SimStadt	Validation	G/LOD2.2.0	TZ/LOD0	TR/LOD0	SR/LOD2.0	14,000	Residential	MAPE	Between 2-15
Nouvel et al. 2017 [44]	SimStadt	Validation	G/LOD2.2.2	TZ/LOD0	TR/LOD0	SR/LOD2.0	14,000	Residential	MAPE	Between 3-15
Dogan et al. 2017 [45]	ShoeBoxer/EnergyPlus	Calibration	G/LOD1.2.1	TZ/LOD2.1	N/A	N/A	121	Mixed Use	RMSE	Between 11% and 20%
Chen et al. 2017 [46]	CityBES/EnergyPlus	Validation	G/LOD1.1.3	TZ/LOD1	N/A	N/A	540	Office and Retail	N/A	N/A
Sokol et al. 2017 [32]	EnergyPlus	Bayesian Calibration	G/LOD1.1.1	TZ/LOD1	TR/LOD0	SR/LOD0	399	Residential	%	87
Sokol et al. 2017 [32]	EnergyPlus	Bayesian Calibration	G/LOD1.1.1	TZ/LOD1	TR/LOD0	SR/LOD1	399	Residential	CVRMSE	66
Sokol et al. 2017 [32]	EnergyPlus	Bayesian Calibration	G/LOD1.1.1	TZ/LOD1	TR/LOD0	SR/LOD1	399	Residential	CVRMSE	58
Davila et al. 2016 [31]	UMI/EnergyPlus	Bayesian Calibration	G/LOD1.1.3	TZ/LOD1	TR/LOD0	SR/LOD1	83,541	Mixed	%	between 5-20
Davila et al. 2016 [31]	UMI/EnergyPlus	Bayesian Calibration	G/LOD1.1.3	TZ/LOD1	TR/LOD0	SR/LOD2.0	83,541	Mixed	%	40%
Nouvel et al. 2015 [26]	SimStadt	Validation	G/LOD2.2.0	TZ/LOD0	TR/LOD0	SR/LOD2.0	1,000	Mixed	MAPE	49
Ascione et al. 2013 [47]	EnergyPlus	Validation	G/LOD1.1.1	TZ/LOD1	TR/LOD0	SR/LOD3	7	Mixed	%	between 5-44

used annual measured data on the building level to validate their model. They also compared the model accuracy on two different zoning approaches, TZ/LOD1 and TZ/LOD2.1. With the zoning approach, their results demonstrated improved accuracy of predicted cooling and heating loads by 7.5% and 16.9%, respectively. Faure et al. (2022) [22] also discussed the impact of thermal zoning on the model accuracy by normalizing results per heated area to compare them easily for urban energy modeling and found that a single zone option, (TZ/LOD0) for heated and non-heated volumes should be avoided while having one zone per floor (TZ/LOD1) is still acceptable. They presented the results as a change in thermal energy demand intensity (TEDI) 1 and discussed change in the level of geometry detailed led up to 20% TEDI on the building level, but it remained below 1% for the district level. Thus, they recommend using more detailed geometry when the impact of energy conversation measures is discussed on the building level.

3.3. Spatial and Temporal Resolution

Urban building energy modeling tools can be used to predict energy demand on different spatial and temporal resolutions (SR: building level, block level, neighborhood level, city level, and TR: hourly, daily, weekly, monthly, annual), and literature shows that buildings in urban energy modeling are calibrated and validated against measured energy use data at different spatio-temporal resolutions, as well. However, the definition of spatial resolution in the late literature is inconsistent. While Oraiopoulos et al. (2022) [27] explain it as an error of the output, which is calculated whether for a cluster of buildings (aggregated) or on a per building basis, Mathur et al.(2021) [29] discuss the spatial resolution on three granularity as archetype level, aggregate level, and building level. Many models employ coarser resolutions (archetype or aggregated) than desired when appropriate spatial data is unavailable. Therefore, we presented different spatial resolutions for measured data used in the validation and calibration processes, as seen in the section Appendix A.1.3. This definition can be enlarged depending on the attributes added to the aggregation approaches.

The most common temporal resolution used to report model accuracy is the TR/LOD0 reporting error for annual temporal resolution. Nutkiewicz et al. (2018) [42] performed validation by using TR/LOD1, TR/LOD2, and TR/LOD3 and found that the model accuracy decreases when validation is performed for TR/LOD3 with hourly data, yet daily and monthly data does not show considerable differences in terms of model accuracy. Sokol et al. [32] propose a probabilistic approach to define archetypes by defining the most uncertain parameters as prior probability distributions and discussing the model accuracy on two different temporal resolutions. Their findings show that the model's accuracy is higher when monthly data is used instead of annual data to perform Bayesian calibration.

Wate and Coors (2015) [49] have shown the spatial resolution as aggregated to disaggregated data (Building Level, Urban Level, and Regional Level) in their early urban energy modeling study. Aggregated demand data (SR/LOD2) is the most frequently used in the absence of granular data. Typically, sample buildings' annual metered energy bills are aggregated based on location, zip code, block level, or urban level and used for calibration and validation depending on the spatial resolution of the model. Cerezo et al. (2016) used total metered energy consumption for archetype use and age, aggregated data at zip code level from all buildings, and reported average errors of 5–20% for SR/LOD1 archetype level and average 40% for SR/LOD2.0 aggregate level. Archetype level is the second most common method used in the literature. This review observes that as the spatial resolution changes from SR/LOD1 to SR/LOD3, the reported error increases [22, 33, 43, 32].

Nine out of 23 case studies discussed building-level temporal resolution(highest granularity), and only four provided results for each building [43, 47, 24, 22]. However, these studies have a limited number of buildings(less than five buildings), and only Faure et al. (2022) [22] discuss 33 buildings at once.

Our study discussed the Bayesian methodology applied in many UBEM studies while we evaluated the performance of the methodology from the building level to aggregated level. Reports with aggregated data tend to average the error when analyzing aggregate demands at any annual scale. As seen in the provided literature Table 1 and 2, these reports are relatively low errors in the 1% and 15% range. Provided platforms are generally validated with archetype or aggregated level data; however, it is never discussed how much error in predicting demands these simplifications introduce on the building level.

4. Bayesian Calibration

We first discuss the level of details for geometry, thermal zoning, and spatial and temporal resolution in the literature and provide a sub-level of details for each layer. Then we discuss the modeling errors to predict cooling demand before and after Bayesian calibration is performed for the model by using a different spatial and temporal resolution of the metered data. The case study is located in Austin, Texas, where cooling demand has a higher portion of total energy consumption. Therefore, we only calibrate the model for the cooling demand. We investigate the impact of the choices made at the UBEM calibration stage, namely around the LoD of spatial and temporal resolution. We quantify the simulation error for aggregated and building-level spatial resolutions by using multiple spatio-temporal measured data to calibrate the model inputs.

The data collection and modeling process is broken into the following steps in this study, as seen in the schematic (Figure 3). First, we collected building and cooling demand data for each building on the campus. Then we defined baseline parameters for the buildings in the energy model. Next, using measured data with different temporal resolutions, we calibrate each building with Bayesian methodology with two different temporal resolutions. Finally, we compare results with the calibration performed by aggregating the data and calibrating the buildings with aggregated data instead of building-level data to understand how the CV(RMSE) differs from higher resolution to lower resolution.

4.1. Data Preparation- Urban-Scale Building Energy Simulation-CitySim

The inputs provided for urban energy modeling tools primarily depend on the adapted simulation engine and modeling purpose. Yet, they can be grouped into the following data categories: weather, geometry, construction, energy system, operation, and energy consumption. The engine we use in this study is a bottom-up physics-based urban-scale building energy simulation platform, CitySim, a C++ based command-line integrated solver initially developed at the Solar Energy and Building Physics Laboratory (LESO-PB) of Ecole Polytechnique Fédérale de Lausanne (EPFL). The CitySim solver has been released open-source since 2020 under BSD-3 Clause license on GitHub[1]. In addition, we took a university campus located in downtown Austin (UT's Main Campus) as a case study to discuss the accuracy of the proposed Bayesian Methodology with different spatial and temporal resolutions.

[1]`http://www.github.com/kaemco/CitySim-Solver`

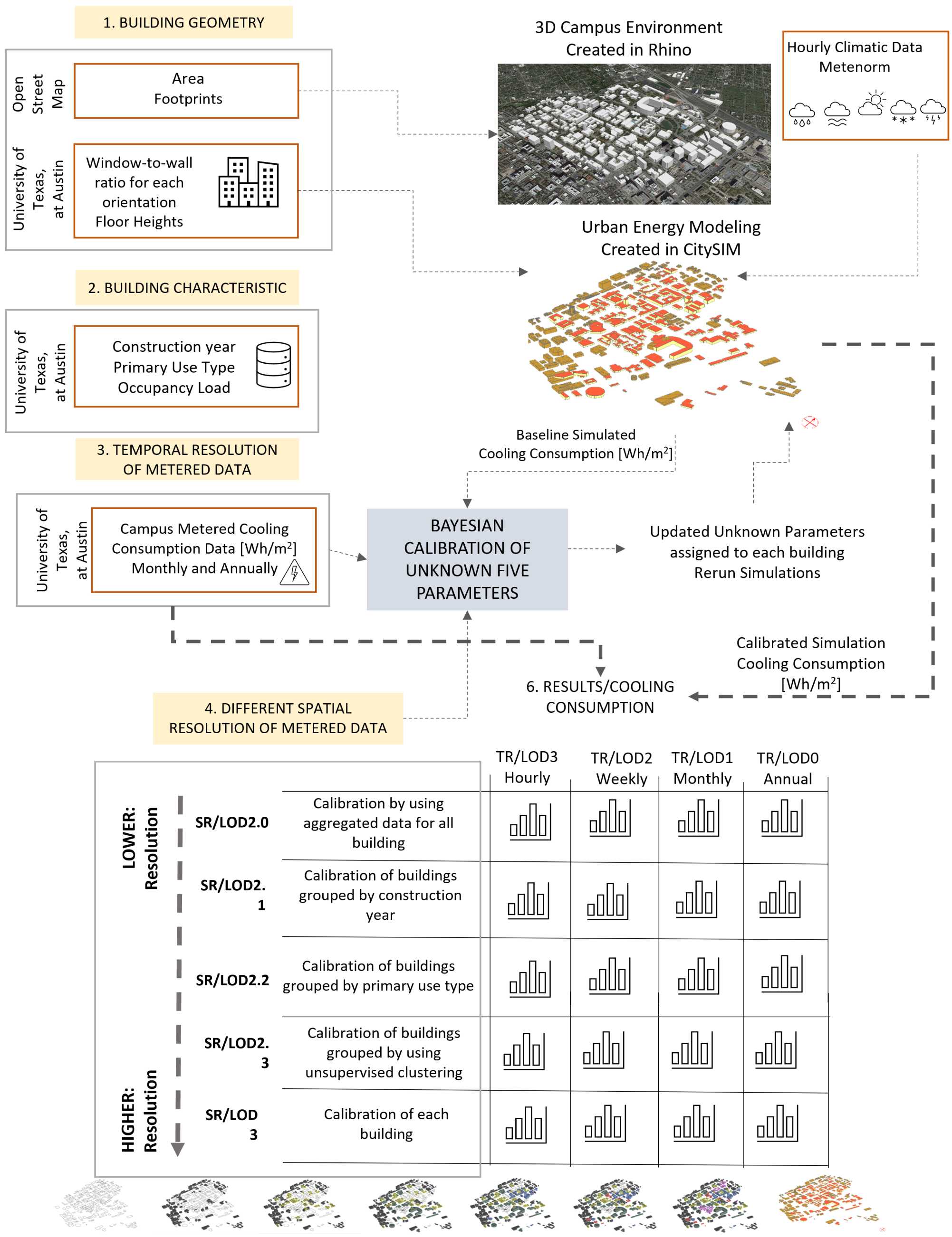

Figure 3: Schematic of the Bayesian Calibration and Modeling Process.

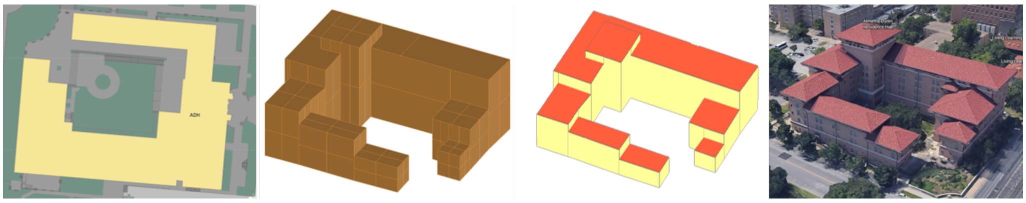

Figure 4: Sample campus building represented with (a) Building Footprint and (b) 3D Model in Rhino, (c) CitySim Energy Model (G/LOD1.2.2 and TZ/LOD0) in comparison to (d) a Google Earth view.

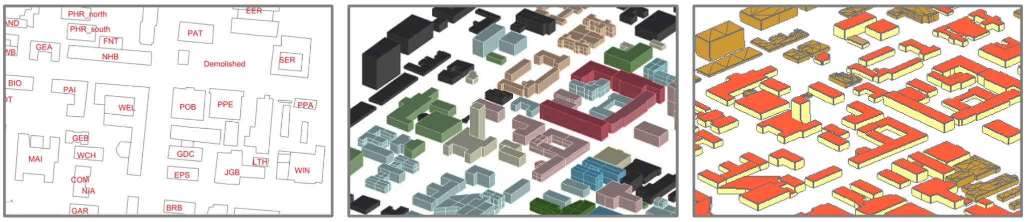

Figure 5: Example of creating UBEM (from left to right, the whole process of creating CitySim model, openings such as window to wall ratio have been assigned in the xml file; thus they are not visible on the graphics).

4.1.1. Building Geometry

The complete data integration and 3D city model enrichment process involve several steps. It is accomplished by combining multiple tools, such as OpenStreetMap (OSM), Geometry processing in Rhinoceros 3D software with Grasshopper plugin, and Python script as seen in Figure 3. Accessible geographical information for University of Texas at Austin (UT)'s Main campus buildings in the form of 2.5D data OSM is pre-processed through Rhino in order to define the geographical coordinates of the floor plan vertices. Building heights for 123 buildings on campus have been collected from available UT's Architectural drawings, e.g., elevations. When the data was not available, the height of the buildings was calculated by assuming that the height of each floor was 3 meters (10 feet) and multiplied by the number of floors available, a standard approach in the field [50]. Based on the building floor plan extracted from the .osm file and the building height, a single 3D thermal zone for each building is created, taking all floors.

2.5D data of buildings define building size, shape, absolute geographical location, and orientation of each building. Figure 5 shows the 2D footprint of an example building and its neighboring buildings (shading objects colored black) and the resulting XML file generated by Rhino for CitySim, a bottom-up physics-based UBEM tool (right). All buildings have been modeled as a level of detail 1 (G/LOD1.1.1) block model, the coarsest volumetric representation defined in the Open Geospatial Consortium CityGML standard [51] with height differences and opening details as illustrated in the provided sample building (Figure 4). To reduce simulation complexity, each building in this work is modeled as a single thermal zone to simplify geometric processing complexity (TZ/LOD0). It is important to state that CitySim provides an option to create multiple zones [52].

4.1.2. Building Characteristics

Non-geometric characteristics are also required as input to create an energy model of the existing buildings, and the information on the building construction year, primary use type, and

Table 3: Baseline U-values in different construction periods for residential and non-residential constructions. (* when a minimum requirement changes for non-residential buildings).

Parameters	Pre 1980	1980-2004	2004-2007	2007-2013	Post 2013
Wall U-Value [W/m^{2}K]	1.35	0.85	0.44	0.30/0.44*	0.30/0.35*
Floor U-Value [W/m^{2}K]	1.21/0.96*	1.21/0.96*	0.79/0.60*	0.60/0.50*	0.60/0.50*
Roof U-Value [W/m^{2}K]	0.57	0.38	0.30	0.30	0.23
Window U-Value [W/m^{2}K]	1.22	1.22	1.22	0.72	0.60
Window SHGC	0.54	0.25	0.25	0.25	0.25

Table 4: Defined baseline values and uncertainty ranges of simulation inputs [53, 54, 55, 56, 57, 58].

Variables	Units	Baseline	Uncertainty	Selected Values
Average walls U-value	W/m^2K	Table 3	U(0.1-3)	Calibrated
Floor U-value	W/m^2K	Table 3	U(0.1-3)	Table 3
Roof U-value	W/m^2K	Table 3	U(0.1-3)	Calibrated
Average windows U-value	W/m^2K	Table 3	U(0.5-5)	Calibrated
Windows average g-value	-	Table 3	U(0.1-0.8)	Table 3
Min. Temperature set-point	°C	18	U(16-20)	18
Max. Temperature set-point	°C	22	U(20-24)	Calibrated
Infiltration rate	h^{-1}	0.5	U(0.1-1)	Calibrated
Average walls short-wave ref.	-	0.3	U(0.1-0.6)	0.3
Window to Wall Ratio	-	Measured	Measured	Measured
Average roofs short-wave ref.	-	0.3	U(0.1-0.6)	0.3
Average ground short-wave ref.	-	0.3	U(0.1-0.6)	0.3

glazing ratios were available at the building level. Therefore, input information for construction, type and usage, and glazing ratios have been collected for each building. The detailed information on the Window to Wall Ratio (WWR) is collected for each orientation (East, West, South, North) from building elevations when they are available; otherwise, Google Earth is used for WWR prediction. Only 13 buildings out of 123 buildings do not have drawings available. For buildings with adjacent buildings, the WWR has been defined as zero for the orientation with the adjacent building. Construction methods and thus the resulting heat transfer coefficients (U-values) of building elements have changed considerably over time. Therefore, a categorization is adopted in the model based on ASHRAE 90.1 Standards release years. Thus, we create construction data referring to the thermal transmittance coefficient of roofs, walls, floors, and windows; the solar energy transmittance of window glazing (g-value of the window). It is impractical to collect specific construction parameters for each building by collecting minimum requirements for given years based on ASHRAE 90.1 standards [48] for five different construction periods.

In our model, we also defined the cooling source as a heat pump with minimum and maximum temperatures of 5 and 20, respectively to cover the maximum load in the buildings related to the indoor temperature setting. Pmax, representing the electrical capacity of the Heat Pump, is chosen to cover the largest load and the same for each building.

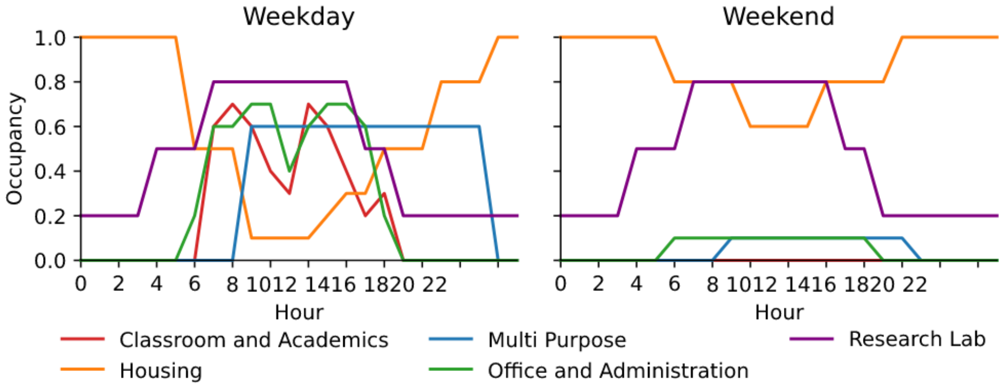

Figure 6: Occupancy Schedules [60].

4.1.3. Occupancy Loads

The mappings between space type and ASHRAE standard 60.2 [48], the resulting occupancy density values for individual space types listed in UT space use are presented in Table 5. Space used for each building is obtained from UT-interactive maps that are available to UT affiliates [59]. In order to reflect building diversity in terms of occupant densities, we assigned a number of occupancy to each building which has been calculated using standards for a person/ft^2 defined in ASHRAE 62.1-2013 Occupancy density calculation Table 6.2.2.1 and the spaces for different purposes by multiplying by the number of people/per floor area of 100 m^2(1070 ft^2) (seen in table 5 [48]).

Table 5 shows the library as the sum of the area of the library stack room, open Stack reading room, reading/study room, and library processing room. The Laboratory area has been calculated as the sum of the class laboratory and research laboratory. Finally, the Athletic Service and Recreation facility area are calculated by the gym area, sports arena, and play area. We also created occupancy schedules for five different primary use types to reflect the stochastic occupant presence and activity patterns 6 based on Ahmed et al. (2016) study [60]. It shows the occupancy presence variation between 0 (absence) and 1 (presence). For the non-operational hours, such as after working hours and on the weekend, the value is given as 0. The profiles are given on an hourly basis, and for some hours, it shows a fractional value. For instance, the occupancy rate from 10 am to 11 am is 0.7, indicating that the building reaches 70% of full occupancy between 10:00–11:00.We assume that the mechanical rooms and circulation areas (stairways and elevators) would be empty most of the time and assigned them occupancy density values of zero, and the internal gains from occupants were calculated assuming 80 W of heat per person of sensible heat for all buildings [48]. The heated area that is used to compare simulation to metered data is found by summing the total size of each space, except mechanical rooms, circulation, and alteration areas.

4.2. Temporal Resolution of Campus Metered Data

Utilities and Energy Management of UT, at Austin, has been recording energy consumption data for most of the buildings on campus. In this study, we included three years of energy consumption records (hourly) from 2017 to 2019 at the building level and used them to vali-

Table 5: Occupancy density used to calculate occupancy for each building [48].

List of Space Definitions Used at UT campus	Person / 100 m^2	List of Space Definitions Used at UT campus	Person / 100 m^2
Office Space	5	Public Restroom	1
Conference Room	50	General-Purpose Classroom	65
Library	10	Laboratories	25
Study Sleep/Dormitory Bedroom	10	Athletic Service and Recreation facility	7
Daycare	25	Assembly Facilities (Conference, dining, gym)	100
Multi-use Cafeteria/Dining	100	Lounge-Public Assembly Lobbies	150
Exhibition Facilities (museum)	40	Storage	-
Locker room	2	Radio	25
Data Processing	1	Shop Facilities	15
Nurses Station, Surgery, Healthcare	200		

date simulation results on hourly, weekly, monthly, and annual levels for cooling consumption. Calibration was only performed by using annual data.

4.3. Different Spatial Resolution of Campus Buildings

In this study, we predict the cooling consumption for 70 buildings by using CitySim. We perform the Bayesian calibration process for five different spatial resolutions by using the annual temporal resolution of metered data, as seen in Figure 3. Spatial resolution is explained as the error of the output has been calculated for a cluster of buildings based on building properties (archetype), or for a cluster of buildings based on their location (neighborhood, city), or at a building scale, [27, 29]. For each spatial resolution, groups have been created based on available data, and we excluded some groups when there were less than three buildings in each group. Finally, we perform Bayesian calibration by using two different levels of spatial resolution for the measured data. The scenarios presented in this study can be listed as follows:

- SR/LOD2.0 Aggregated Level. We use aggregated annual and monthly data of buildings on the location, UT' Campus while calibrating each building's unknown input parameters to predict cooling demand realistically. The error has been discussed on building levels for different temporal resolutions. **This aggregation methodology is commonly used to validate the UBEMs.**
- SR/LOD2.1 Aggregated Level. We use aggregated annual and monthly data of buildings on the location, UT' Campus, grouped by their construction year on the location UT' Campus while calibrating each building's unknown input parameters to predict cooling

demand realistically. The error has been discussed on building levels for different temporal resolutions. (Construction year groups: Before 1980, 1980-2004, 2004-2007, 2007-2013).

- SR/LOD2.2 Aggregated Level. We use aggregated annual and monthly data of buildings on the location, UT' Campus, grouped by the definition of their primary use type, while calibrating each building's unknown input parameters to predict cooling demand realistically, and the error has been discussed in building levels for different temporal resolutions (R: Research Laboratory, H: Housing, OA: Office and Administration, CA: Classroom and Academics, PA: Public Assembly and Multi-Purpose). Grouping is performed based on UT's Portal Database [59]. Classroom and Academic buildings contain teaching and community space for faculty and students. These buildings often have longer operational hours due to students gathering for classes, group study, or organizational meetings. These buildings' Heating, Ventilation and Air Conditioning (HVAC) systems are typically modified to save energy during unoccupied periods. Research Laboratories are the most energy-intensive buildings on campus due to their high ventilation needs. Air delivered to laboratory spaces is often 100% outside air that is conditioned, delivered to the space, and exhausted from the building. Therefore they are the most energy-intensive buildings. Housing facilities provide student housing as well as dining, community space, exercise facilities, and other student services. Office and Administrative buildings contain faculty and staff offices primarily and maintain fairly regular hours. Public Assembly and Multipurpose buildings are those that house museums, libraries, sports facilities, and other community gathering spaces. These buildings often have periods of high occupancy or specific indoor environmental requirements that can cause them to be very energy intensive. Based on these definitions, the occupancy schedule in the model is created accordingly.
- SR/LOD2.3 Aggregated Level. We use aggregated annual and monthly data of buildings on the location, UT' Campus, clustered by using unsupervised learning to discover data-driven building classes from the buildings' chilled water energy profiles. The hypothesis is that each building has a dominant energy profile, and a group of buildings that have the same dominant profile belong to the same data-driven class. We use the k-means algorithm to cluster the conditioned area-normalized daily chilled water profiles (Wh/m^2) for summer weekdays, i.e., Monday - Friday in June, July, and August, for the 70 buildings. Using the elbow method on the sum of square error, We infer that the area-normalized daily chilled water profiles are distributed across three distinct clusters where 18.3%, 40.1%, and 39.5% of profiles belong to Cluster 1, 2, and 3 respectively. The dominant profile in each building is identified such that a building's dominant profile is the cluster to which greater than 50% of the building's profiles belong. If there is no such cluster for a building, the building is assigned to the "Unidentified" class. We conclude that 18.3%, 40.1%, and 39.5% of the buildings belong to Cluster (class) 1, 2, and 3, respectively, and there is one unidentified building.
- SR/LOD3.0 Building Level measured annual and monthly data have been used to calibrate each building, and the error has been discussed on building levels for five different temporal resolutions.

We demonstrate if significant improvements in model accuracy can be obtained even using simple uncertainty models and less streamed data.

4.4. Bayesian Calibration Framework

Bayesian calibration has been proposed and used by other studies ([31, 38]). Equation 1 was employed to analyze the uncertainty in every introduced spatial resolution; the analysis was carried out through a formulation introduced by Kennedy and O'Hagan [61]. We calculate the posterior probability $P(Q_{SR/LODn}|E_{SR/LODny})$ with prior probability $P(Q_{SR/LODn})$ and the likelihood function $(P(E_{SR/LODny})|(Q_{SR/LODn}))$. The five unknown parameters go through the calibration process to demonstrate whether the simulation outputs are compatible with the measured data on different spatial resolutions. It has been decided to calibrate Maximum Temperature Set-point, Infiltration, Average Walls U_{value}, Roof U_{value}, and window U_{value}, which are each divided into five levels with a uniform prior probability distribution respectively, and this leads to 625 input combinations.

The Bayesian calibration provides which input combination, $Q_{SR/LODn}$, is most likely to be correct, given the simulation model and the metered data $E_{SR/LODny}$, where y is the temporal measurement resolution in the training set.

The Bayesian inference equation is as follows:

$$P(Q_{SR/LODn}|E_{SR/LODny}) = \frac{P(E_{SR/LODny}|Q_{SR/LODn})P(Q_{SR/LODn})}{P(E_{SR/LODny)}} \tag{1}$$

and

$$P(E_{SR/LODny}) = \sum_{Q_{T_{max}}} \sum_{Q_{N_{inf}}} \sum_{Q_{U_{wall}}} \sum_{Q_{U_{roof}}} \sum_{Q_{U_{window}}} P(E_{SR/LODny}|Q_{SR/LODn}) \times P(Q_{SR/LODn}) \tag{2}$$

In reality, many factors can affect the likelihood function, and the explicit form does not exist. As a consequence, we assume the likelihood function $P(E_{SR/LODny}|Q_{SR/LODn})$ can be described by a Gaussian normal distribution function as shown below:

$$P(E_{SR/LODny}|Q_{SR/LODn}) = \frac{1}{\sigma_{SR/LODn}\sqrt{2\pi}} \times \left(\exp - \frac{(E_{SR/LODn_y} - \mu_{SR/LODn_y})^2}{2\sigma^2_{SR/LODn}} \right) \tag{3}$$

where $E_{SR/LODny}$ is the measured cooling (Wh/m^2) of the individual spatial resolution of the corresponding measurement time y; $\mu_{SR/LODny}$ is the simulated cooling of each building under given the specific input combination $Q_{SR/LODn}$; the standard deviation $\sigma_{SR/LODn}$ accounts for the inherent variability of energy consumption in the spatial resolution and is estimated from the standard deviation of the measured consumption of the buildings under the same spatial resolution subgroups. The likelihood function is described for each building separately for each spatial resolution. As a result, we obtained posterior distribution for each building and each spatial resolution group. Statistics for each spatial resolution and temporal resolution are given in the table. The campus model is calibrated annually and monthly with three years of measured data.

Validation is performed by comparing the baseline simulation result as well as the calibrated simulation result to the measurement data (five different temporal resolutions), using the a CV(RMSE) defined in Equation 4 to measure how well the model fits the measured values at validation period 2017-2019 and compared with following standards by the ASHRAE Guideline 14-2014.

Table 6: Characteristic of each Spatial Resolution.

Spatial Resolution	**Aggregation**	**Buildings**	$\sigma_{SR/LOD_{annual}}$	$E_{SR/LOD_{annual}}$
SR/LOD2.0	All Buildings	70	22,421.0	401,175.2
SR/LOD2.1	Before 1980	55	30,681.6	348,577.6
	1980- 2004	8	8,564.0	358,693.2
	2004-2007	3	57,551.1	743,923.0
	2007-2013	4	29,365.9	801,754.3
SR/LOD2.2	RL	11	9,148.0	281,174.1
	H	9	24,840.9	333,159.8
	OA	4	20,498.4	409,266.5
	CA	24	10,496.7	342,759.6
	PA	22	77,119.0	846,757.6
SR/LOD2.3	Class 1	13	9,052.8	288,728.6
	Class 2	29	39,298.0	517,035.8
	Class 3	28	10,147.9	326,582.1

$$CV(RMSE)_{SR/LOD2.n} = \frac{100}{y_{mean}} \times \sqrt{\left(\frac{\sum_{i=1}^{N_{SR/LOD2.n}} (y_{i_{metered}} - y_{i_{simulation}})^2}{N_{SR/LOD2.n}}\right)} \tag{4}$$

where $N_{SR/LOD2.n}$ is the number of cooling measurements for each spatial resolution and sub-groups defined for them, $y_{i_{metered}}$ is the metered cooling data for the i_{th} spatial resolution and subgroups, while $y_{i_{simulation}}$ stands for the simulated cooling demand and y_{mean} is the mean of the $N_{SR/LOD2.n}$ metered chilled water (Wh/m^2) for each group presented in the Table 6.

Then the CV(RMSE) is calculated for each building (Equation 5) to discuss the difference between aggregation approaches and understand how well clustering is predicting cooling demand compared to building level calibration.

$$CV(RMSE)_{BuildingID} = \frac{100}{y_{mean}} \times \sqrt{\left(\frac{\sum_{i=1}^{N_{SR/LOD3}} (y_{i_{metered_{BuildingID}}} - y_{i_{simulation_{BuildingID}}})^2}{N_{SR/LOD3}}\right)} \tag{5}$$

where $N_{SR/LOD3}$ is the number of cooling measurements for each building, $y_{i_{metered_{BuildingID}}}$ is the metered cooling data for each building, while $y_{i_{simulation_{BuildingID}}}$ stands for the simulated cooling demand for each building, and y_{mean} is the mean of the $N_{SR/LOD3}$ metered chilled water (Wh/m^2) for each building.

5. Results

In this section, the results of parametric simulations are presented to analyze the impact of the level of detail used for data aggregation during the calibration process. The results provided are based on geometry model G/LOD1.2.2 and TZ/LOD0. Bayesian calibration was performed using TR/LOD0, and results are provided for different temporal and spatial resolutions. The same calibration is performed using monthly data only; however, the marginal distribution for the prior and the posterior distribution of the calibration parameters had approximately the same range and distribution, meaning that the data is non-informative about the calibration parameters.

5.1. Results for SR/LOD3 Building Level Spatial Resolution

Building scale (SR/LOD3) performance of the Bayesian calibration with the annual temporal resolution is discussed, and errors are provided for each building in Figure 8. We assess errors in cooling consumption data [Wh/m^2]. Before performing Bayesian calibration on the building level (SR/LOD3), the annual cooling consumption prior to the Bayesian process is provided. After performing Bayesian calibration, the resulting marginal probability mass distributions are obtained and presented as a heat map for the five variables of each building in Figure 7. Based on the posterior distribution of each parameter, the most likely parameter is assigned to each building in the simulation file, and CV(RMSE) is calculated for the calibrated model. The posterior distribution also shows that window property doesn't depart far from their initial uniform distributions for most buildings because it has a more negligible effect on cooling usage except for ten buildings out of 70 buildings.

This observation indicates that less influential parameters can be treated as a constant value instead of being included among the probabilistic parameters. Although there is no standard for urban energy modeling, ASHRAE sets the allowable maximum percentage error of the annual calibrated model for electricity use to be 5% for building energy modeling. With the calibration process of each building, the CV(RMSE) for all buildings decreased at least below 30% while only for five buildings, the ASHRAE standard of CV(RMSE) is achieved for annual cooling consumption (Figure 9. Figure 8 shows that 50% of the buildings have annual CV(RMSE) over the average error of 42% error when we define unknown inputs based on literature (deterministic), not probabilistic approach as given in the Table 3. The highest and the lowest CV(RMSE) of annual cooling consumption with a deterministic approach are obtained as 102% and 10% for Building 21 and Building 51, respectively. The simulation model over-predicts the cooling load for most of the buildings. The reason is that the simulation model uses only one thermal zone for the whole building. The results have been discussed on higher temporal resolutions, and it is observed that the calibration by using annual data introduced more bias to hourly data and resulted in higher errors compared to the deterministic approach.

5.2. Results for SR/LOD2-Aggregated Spatial Resolution

We aggregated data for different spatial resolutions and calculated posterior distribution using Bayesian calibration methodology. We try to understand which data aggregation methodology gives the closest predictions to the building level calibration by performing this aggregation. Recent studies have shown the quality of the Bayesian approach. Still, because of the lack of data granularity on the building level, they could not discuss the performance of the approach on the building level. The Bayesian approach used in these studies has also been only discussed annually, again because of privacy concerns and data availability.

Figure 10 shows us the simulation results error compared to the measured data for each spatial resolution group before the Bayesian calibration was applied. CV(RMSE) for each spatial resolution and subgroup are calculated by comparing the sum of the measured data and the sum of the simulation results of each building (before and after calibration). The results are provided for five temporal resolutions to understand how the error changes with the increase of the resolution before the Bayesian calibration is applied. It is seen with the deterministic parameters for each

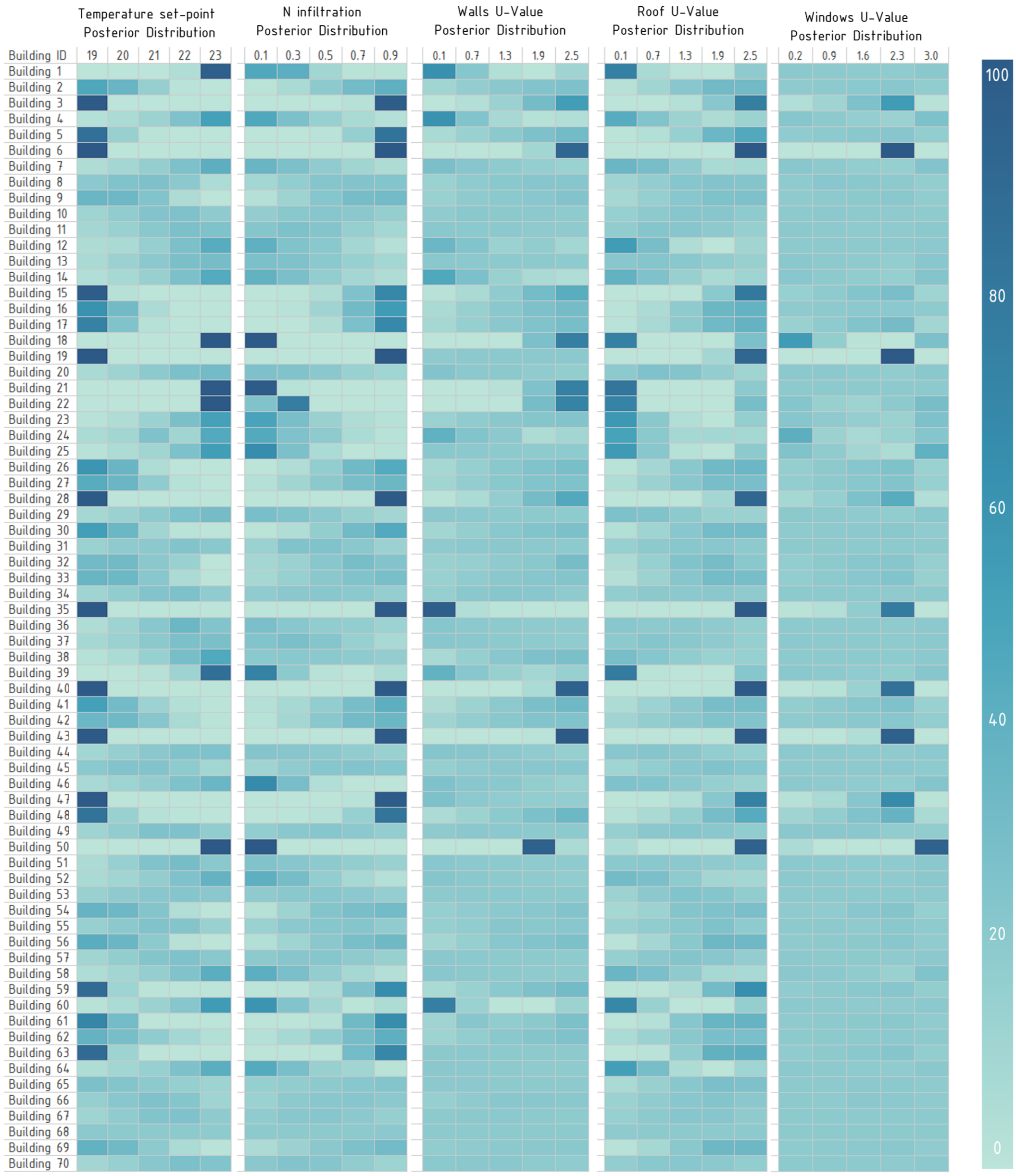

Figure 7: Marginal posterior distribution of Maximum Temperature set-point, Infiltration rate, Walls U-value, Roof U-value, and Window U-Value visualized in a heat map. Each row represents one building, and color intensity represents probability (%).The prior distribution is constant as 20% for given inputs. Calibration is performed by using each building's annual cooling consumption data available for three years.

Figure 8: CV(RMSE) of each building before Bayesian Calibration at Annual Temporal Resolution (TR/LOD0).

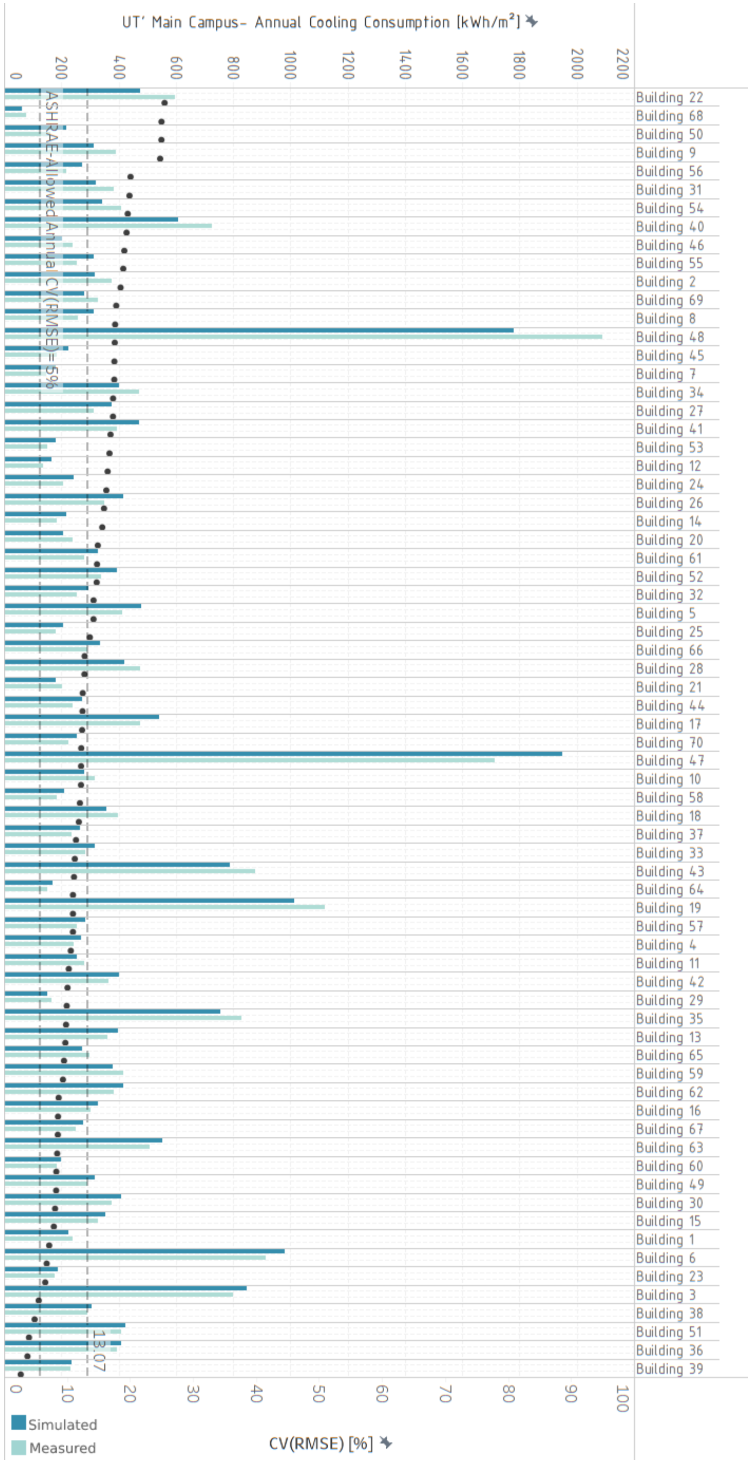

Figure 9: CV(RMSE) of each building after Bayesian Calibration at Annual Temporal Resolution (TR/LOD0).

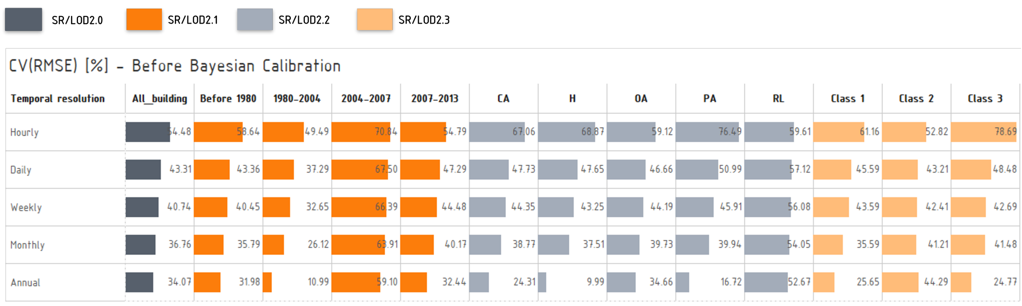

Figure 10: CV(RMSE) of each spatial resolution before Bayesian Calibration at all Temporal Resolutions.

building that CV(RMSE) of our predictions changes between 10% to 60% depending on the subgroup that we defined for each spatial resolution. When the discussion is performed for all buildings, the CV(RMSE) is 34%, meaning that without calibration applied, the total annual cooling prediction for 70 buildings differs 34% from actual cooling consumption.

As seen in the provided results before the calibration process,CV(RMSE) of daily, weekly, and monthly temporal resolutions are closer to each other. The highest difference between hourly and annual temporal resolution is obtained for housing, classroom-Academics, and Public Assembly and Multi-Purpose. The occupancy and schedules can explain why these primary use types are hard to predict. Nevertheless, this provided insight into data aggregation and how the primary use type plays an essential role in different temporal resolutions. The same behavior also has been seen for different subgroups under different spatial resolutions, such as SR/LOD2.1 the construction year 1980-2004, and SRLOD2.3 unsupervised clustering, class 3. These are the subgroups with the high number of buildings from Public Assembly and Multi-Purpose and Classroom and Academics.

Aggregated prediction CV(RMSE) of the 70 buildings are shown in Figure 11 against the aggregated measured data for all subgroups with the probabilistic parameters. It is found that when we look into overall annual consumption for 70 buildings before and after calibration, the error is improved for each group; for some aggregated groups, the improvement in the CV(RMSE) is high after calibration. Comparing the aggregated data from these groups to improve the cooling demand predictions obtained through UBEM and predict annual consumption, we have a max 15% error for all campus buildings after applying the probabilistic approach. However, it is also observed that the results did not improve for subgroups that had less than 20% CV(RMSE) with the deterministic approach (before Bayesian calibration).

Although this probabilistic Bayesian calibration improves the results on an aggregated level, it would result in at least a +-40% error change on the building-level prediction. The annual CV(RMSE) is either improved or worsened for these 70 buildings on the campus. Figure 12 shows building distribution with the error change for each group. Negative change means that the calibration process worsened the annual CV(RMSE) of predicted cooling consumption, and positive shows the opposite, an improvement for each building.

All these approaches can be used for forecasting annual cooling consumption under different scenarios. However, discussing the probabilistic approach for different primary use types should be considered. Based on the sensitivity analyses performed for form and climate, specific parameters were decided to be calibrated. However, our results show that for different primary use types, there is a need to include other input parameters to better the performance of the simula-

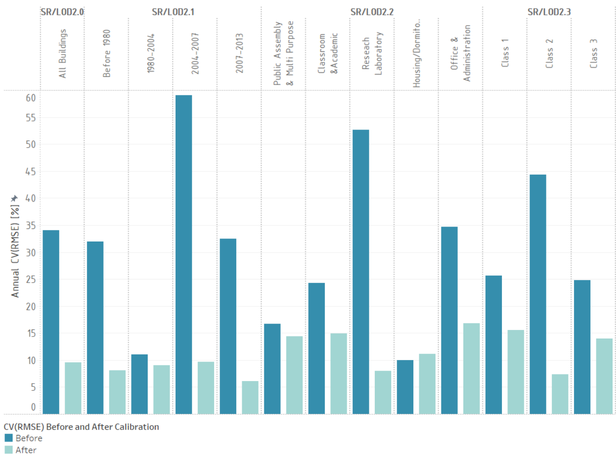

Figure 11: Annual CV(RMSE) of each spatial resolution after Bayesian Calibration performed against cooling data at annual temporal resolution (TR/LOD0).

tion results, such as occupancy density. Because of the number of people and the schedules of the housing/dormitory, Public Assembly and Multi-Purpose and classrooms, and Academics are hard to predict. Therefore the application of the Bayesian calibration did not improve our results. On the other hand, the user behavior in office and research lab buildings is better predictable and structured, leading to better predictions on the cooling demand than different primary use types within the campus.

5.3. Comparison between SR/LOD2 and SR/LOD3

This section is created to discuss how close our cooling demand predictions are when we perform Bayesian calibration with annual aggregated cooling consumption against building-level calibration. The previous section discussed how using probabilistic Bayesian calibration of the unknown parameters improved the simulation error on annual cooling consumption for both calibrating five parameters of 70 buildings by using measured annual data at building level (SR/LOD3) and aggregated level (SR/LOD2). CV(RMSE) is calculated for each building by using annual measured data of each building and simulation results when the probabilistic parameters are decided using aggregated data. Figure 13 provides annual CV(RMSE) for 70 buildings after probabilistic Bayesian calibration is applied using different spatial resolutions compared to calibrating each building with annual building level data. In Figure 13, we see that building level calibration results in CV(RMSE) between 17.23% (upper Quartile) and 9.4% (lower Quartile) with a median of 12% for 70 buildings.

The closest CV(RMSE) distribution for buildings is calculated when we used aggregated data based on primary use type. Among the three box plots, except spatial resolution SR/LOD2.2, two have compact interquartile ranges and long tails, indicating CV(RMSE) at the building scale deviates but generally performs well. There are buildings with much better estimates within each spatial resolution group. We then conducted a t-Test of means between the distributions of CV(RMSE) to assess if reported CV(RMSE) differs when different spatial resolutions are used to perform annual calibration compared to building level calibration (Table 7). We observe the statistically insignificant relationship between CV(RMSE) of SR/LOD3 and other SR/LOD2.n;

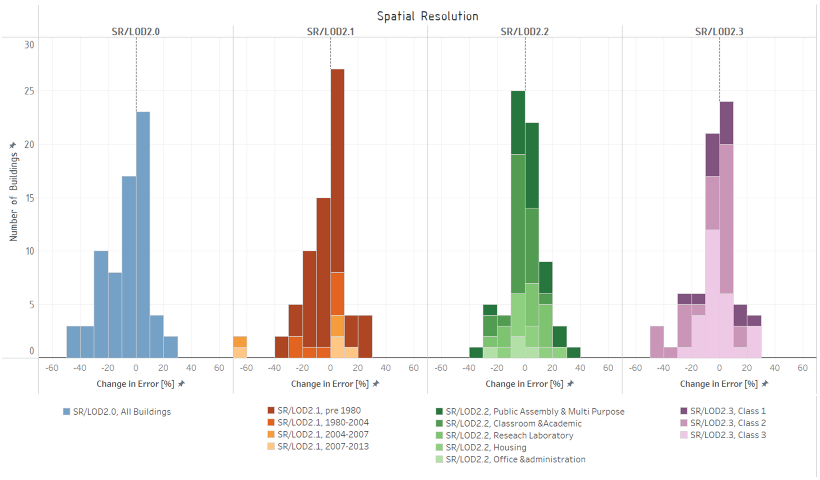

Figure 12: How the annual simulation error improved and worsened for the buildings based on aggregation approach.

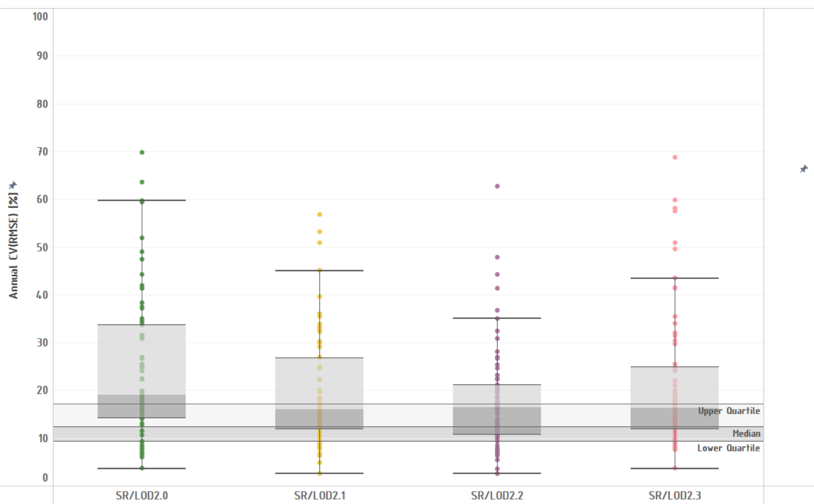

Figure 13: Annual CV(RMSE) distribution after Bayesian Calibration performed against annual cooling data compared to the building level error for 70 Buildings.

So any measured data used to predict the annual cooling demand is giving significantly different than what annual building level CV(RMSE) is. It is also concluded that the difference between obtained CV(RMSE) from aggregated level data based on construction year and based on the clustering method is not significant (p-value:0.66). So this shows that if the UBEM tool is used to forecast the annual consumption of the buildings, these two methodologies can be swapped. Based on the t-test results, the difference between SR/LOD2.0 and SR/LOD2.1 is also not significant. This can be explained by 55 buildings belonging to the group built before 1980. The only difference between SR/LOD2.0 and SR/LOD2.1 is 15 buildings. The least favorite aggregated level data is a spatial resolution based solely on location, SR/LOD2.0.

6. Discussion

Validation of a UBEM tool's result is a challenging task because of the lack of measurement data availability. We performed Bayesian calibration using metered data for different spatial and temporal resolutions in this study. We discussed how the reported error changes when we validate our results for the given method. It is hard to conclude from previous studies if the discussed methodology will perform well on energy predictions by looking into reported errors, first because of inconsistent error reporting and second the privacy issue of reporting the error. Previous studies performed result validation at different temporal and spatial resolutions because of the data availability. The typical approach is to validate results on the zip code level or based on the archetype. Here are the answers to our research questions listed as key findings:

What are the relevant and frequently used levels of detail in UBEM? The geometry of the buildings in the UBEM has shown a different level of detail; however, the discussed case studies have leaned toward the common use of G/LOD1.1, which is described as detailed 2.5D extrusion with 2D geometry of the building's detailed footprint, including 19 case studies out of 54. The geometry details show various approaches, from a very fine modeling approach to a very coarse approach, based on the purpose of the study. Within the group of G/LOD1.1, there are subgroups with more details about the model geometry, which can be found in Appendix A. Our discussion showed that the details of the number of floors of the given geometry are the most commonly used property to report the model's accuracy. G/LOD2 is the least widely used because of the roof's complexity. The sub-categories in Appendix A can be extended to include more studies in building energy modeling to compare them; however, it is essential to remind our audience that our literature only has UBEM.

The thermal zone is the basis for the heat transfer calculations in the energy models. We create thermal zones for the geometry to reflect the different spaces in the building. Our literature review for UBEM showed that the most commonly used thermal zoning is Z/LOD1.0 which has a Zone per Floor/Space in 3D models. Moreover, 9 out of 23 studies performed their model accuracy discussion using TZ/LOD1.0 in their model. The least used approach is the TZ/LOD3, which has detailed Internal Zoning and is not feasible considering the computational cost and time required to run UBEM simulations. However, it is challenging to discuss which level of detail is necessary to get more accurate predictions compared to the computing effort in UBEM.

What is the level of detail for spatial and temporal resolution necessary to validate the urban energy model with a deterministic approach? How the reported error differs for each approach? One of the earliest studies in the literature by Davila et al. (2016) [31] validated the MIT UMI tool on the hourly and annual scale at the aggregate level by zip code. The obtained error was between 5 and 20% for aggregated data based on primary use type; however, they found a higher averaged absolute error of 40% for 23 zip codes with individual errors ranging

Table 7: Results for t-Test of means between the distributions of annual CV(RMSE) for spatial resolutions- Dataset:70 Buildings

p-value	SR/LOD2.0	SR/LOD2.1	SR/LOD2.2	SR/LOD2.3	SR/LOD3
SR/LOD2.0	**1.00**	0.25	<0.005	0.07	<0.005
SR/LOD2.1		**1.00**	0.08	0.67	<0.005
SR/LOD2.2			**1.00**	0.09	<0.005
SR/LOD2.3				**1.00**	<0.005
SR/LOD3					**1.00**

between 5% to 94%. A study performed by Sokol et al. (2017) [32] using the EnergyPlus engine presented their calibrated and non-calibrated results. Calibration was performed with two temporal resolutions, annually and monthly, and errors were reported as annual errors. They have shown that monthly calibration improved the error by 10% on the annual level error. According to the ASHRAE standard, the RMSE of monthly energy consumption of a computer model must be about 15%. However, as seen from the literature, this is hardly possible and needs an adjustment depending on the purpose of the model used. Some studies provided errors only to validate their results. Nageler et al. (2018) [43] discussed the monthly CV(RMSE) both on the building level and archetype level, but the discussion on the building level was performed only for two buildings; thus, it makes it hard to compare the archetype level error versus building level error. While building level CV(RMSE) is 24.9% and 40.2% for two buildings, the aggregated level error is reported as 21.4% for 34 buildings. The other validated results by Nutkiewicz et al. (2018) [42] present the quality of their approach for 22 buildings and prove that the CV(RMSE) is nearly doubled from aggregated to building level. They also discussed the error change for different temporal resolutions and provided how the error doubles from monthly resolution to hourly resolution.

According to our findings, reporting the accuracy of the prediction with aggregated data on an annual level shows that the error changes between 10% up to 60% depending on the subgroup in each aggregated data, and the CV(RMSE) of the highest temporal resolution, hourly, changes between 50% to 79%. This shows that even on an hourly level, our predictions are not bad for aggregated data. But on building level CV(RMSE), our hourly and annual CV(RMSE) with deterministic approach can go up to 400% and 102% for a building within 70 buildings. Therefore, when we report the accuracy of our methodology by using aggregated data, it is seen that any aggregation methodology will report lower errors compared to the building-level error. Additionally, our results present that the change between annual, monthly, and hourly CV(RMSE) can be high depending on the aggregation approach. However, it is seen that daily, weekly, and monthly CV(RMSE) show closer results. Therefore, if the measured daily/weekly data is unavailable, monthly data should be enough to discuss the error of the UBEM before any calibration is applied.

How does performing Bayesian calibration with aggregated data improve the predictions of cooling demand? Which approach gives the closest results to the building level?

It is expected that while some of the building-level predictions are close to the measurement data, some will have much higher errors because of the nature of the simulation tools. But, the aggregated or archetypes level predictions averages these errors and provide us with lower errors. Unfortunately, the lack of study on this topic makes it hard to compare UBEM tool and their performance. Because of data availability, a limited number of studies discuss how well the UBEM predicts the building level consumption with and without calibration processes.

Recently, Risch et al.(2021) [24] discussed the performance of the Bayesian calibration on building level error, yet it was only three buildings. Rashidfarokhi (2021) also performed Bayesian calibration on two buildings and reported the absolute percentage error improved from 33% to 13% and from 35% to 15% by re-calibrating posterior distributions. Tardioli et al. (2020) also performed Bayesian calibration for 326 individual buildings and provided validation error within +/-20% for 70% of the buildings.

In this research, we discussed the annual aggregated level of measured data to perform Bayesian calibration, reported errors of these approaches for both aggregated simulation results, and compared it with building-level results for 70 buildings. Our results show that aggregating the annual measured data based on primary use type gives the closest median error to the building-level calibration result. Furthermore, using both aggregated level data and building level data improved the model predictions compared to the deterministic approach. But when we perform the t-test, the results proved that all aggregations error significantly differs from building level error, proving that without solving the privacy issue and calibrating input parameter with building data, we should avoid projections on the building level with the help of UBEM tools.

Providing building-level comparisons without performing building-level calibration can be misleading. More importantly, the results suggest that any analysis based on an UBEM should be careful when interpreting model predictions at a higher resolution. Some studies suggest that aggregating multiple buildings' energy consumption provides acceptable accuracy. However, this should be discussed since ASHRAE currently does not have a standard to evaluate prediction accuracy at higher spatial resolutions. Many studies discussed their accuracy based on these criteria [42, 62, 23]. Therefore, it is important to have criteria for higher spatial resolutions. Future studies should consider improving the current methodology by discussing different likelihood functions and input distributions.

How should we report our results so everyone can benefit from the accuracy of the methodology discussed?

Our findings and literature review showed that the reporting of the model accuracy is essential. There is no single key attribute that is in control of the accuracy of the model. When the error is reported on annual temporal resolution and for aggregated data, the tool can be reported as a practical tool, and methodology can be seen as a valid approach to predict future consumption; however, the chances of errors at the individual building resolution can get extreme, increased up to 102% on a building level. While Bayesian calibration has been reported to show improvement on UBEMs, it is rarely discussed on the building level. When it is discussed, the given results are for a smaller group of buildings. Therefore, this study provides valuable information on how reporting can be misinterpreted. Consequently, we recommend reporting the errors systematically and providing details on each level of detail presented in this study. Thus, future studies can decide the most suitable UBEM and calibration method for their application. Also, created data management tools for these processes should be openly shared within UBEM study groups to improve efficiency.

7. Conclusions

Adapting proper spatial and temporal resolutions to use UBEM for energy policy assessment and scenario analysis to mitigate the impact of the climate crisis is essential. This study used a validated Bayesian calibration approach on a higher temporal resolution to examine the optimal temporal resolution of the metered data that keeps the model's accuracy and requires less computational effort. Bayesian Calibration was performed by using annual and monthly metered data, but the results were not informative for the monthly data, therefore they were excluded. Although this probabilistic Bayesian calibration by using annual data improves the results on an aggregated level, our results have shown that it results in at least a +-40% error change on the building-level prediction, meaning that the annual CV(RMSE) is either improved or worsened for these 70 buildings on the campus. Building level calibration results in CV(RMSE) between 17.23% (upper Quartile) and 9.4% (lower Quartile) with a median of 12% for 70 buildings. The impact of data availability and granularity on simulation results is noticeable; therefore, the discussion of different aggregate-level resolutions should be studied to allow data collection for higher data quality and release available data under secured conditions. If the calibration is performed by using building-level data, then the strategies could be developed for each building's savings. However, it should be avoided to provide single building-level savings by using aggregated data.

The proposed approach based on the calibrated urban building energy model would be mainly wanted by municipalities, urban planners, utilities, and engineering consultancy firms who might show intense interest in implementing energy policy assessment and scenario analysis. Likewise, it would also be possible to implement large-scale building performance mapping and labeling to building retrofit targets and for building stock renovation and energy conservation. In this research, reported accuracy on different levels with aggregated data on an annual level show that the error changes between 10% up to 60% depending on the subgroup in each aggregated data, and the CV(RMSE) of the highest temporal resolution, hourly, changes between 50% to 79%. Performing calibration by using four different approaches on aggregating data has shown that the difference between obtained CV(RMSE) from aggregated level data based on construction year and based on the clustering method is not significant (p-value:0.66). So this shows that if the UBEM tool is used to forecast the annual consumption of the buildings, these two methodologies can be swapped when data is not available on the building level to calibrate the model. Overall, we suggest that the urban energy prediction accuracy on annual temporal resolution can be increased significantly when the Bayesian calibration is performed by using building-level data; however, whenever privacy is a concern, then the data should be provided by aggregating them based on primary use type, showing CV(RMSE) of 21.50% (upper Quartile) with a median of 19% for 70 buildings.

The next step in developing this methodology should include discussing the input numbers and having more discrete inputs defined as the prior distribution. Although the proposed methodology is not suitable for using measured monthly data because of including only 5 discrete inputs for the prior distribution and providing posterior distribution for the unknown parameter, it should be considered to calibrate input parameters seasonally, which will lead to more computational effort and cost by giving more accurate posterior distributions. Also, Our model had precise information on the heated area by using UT' campus online data platforms and subtracting mechanical rooms, circulation, and alteration areas, which is challenging in practice is challenging due to a lack of information. Therefore, occupancy as a calibration parameter should be considered to be calibrated to decrease the error in simulation results against measured consumption when the data is not available.

Limitations of the study should be addressed in future work about the proposed modeling technique and the analysis. The Gaussian normal distribution is initialized to depict the likelihood function for unknown parameters in the proposed Bayesian methodology and calibration framework. Our results show that this assumption is reasonable when we perform the method with annual temporal resolution; however, it led to the inaccurate description of the posterior distribution of the variables. An iterative calibration process may be an alternative to improve the effectiveness of the assumption on likelihood functions; for example, the posterior distribution function from the annual calibration can be used as a new prior for the calibrated input in the calibration framework for monthly calibration. Also, the efficient granularity of the input data set is worth to be discussed; changing the prior distributions to include a broader range instead of five steps would be beneficial. Additionally, the other limitation of the study is to not explore the zoning part in CitySim. We would like to explore the future of building scale calibration to see how these errors could be lessened for cooling demand predictions by discussing different zoning details, applying higher temporal resolution data to calibrate the model, and considering more computationally efficient studies that can be available to everyone.

In addition, while efforts are being made to improve UBEM models to reduce computational requirements and resources for predicting the energy demand of cities, future research should also prioritize refining these models to more accurately predict the performance of solar energy systems in urban environments. This includes taking into account factors such as building orientation, shading, and local weather patterns, especially under the effects of climate change. By enhancing our understanding of how solar energy systems can perform in varying urban contexts, we can promote the wider adoption of renewable energy sources and contribute to a more sustainable future.

Appendix A. Introduced LoDs

Appendix A.1. Building Geometry Layer

Building geometry used in urban energy modeling is categorized into three subcategories as and explained in the provided figure. The literature review reveals that in urban energy modeling case studies, either LOD1 or LOD2 is used to create the energy models. The other observation is, defined LOD3 has openings in the city model; however, in energy modeling, this attribute can be also given LOD1 and LOD2 geometries. Therefore, the sub-levels are created to reflect the energy modeling 3D approach and LOD1 and LOD3 are excluded from the geometry layer of the UBEM, contrary to defined geometries by Biljecki et al. (2016) [28].

- G/LOD1.0.0: Simplified 2.5D extrusion- 2D geometry of the building simplified footprint extruded to their respective heights
- G/LOD1.0.1: G/LOD1.0.0 and including the number of floors
- G/LOD1.0.2: G/LOD1.0.0 and including window openings
- G/LOD1.0.3: G/LOD1.0.0 and G/LOD1.0.1
- G/LOD1.1.0: Detailed 2.5D extrusion- 2D geometry of the building detailed footprint extruded to their respective heights
- G/LOD1.1.1: G/LOD1.1.0 and including the number of floors
- G/LOD1.1.2: G/LOD1.1.0 and including window openings
- G/LOD1.1.3: G/LOD1.1.0 and G/LOD1.1.1
- G/LOD1.2.0: Detailed 2.5D extrusion- 2D geometry of the building's detailed footprint extruded to their detailed heights
- G/LOD1.2.1: G/LOD1.2.0 and including the number of floors
- G/LOD1.2.2: G/LOD1.2.0 and including window openings

- G/LOD1.2.3: G/LOD1.2.0 and G/LOD1.2.0
- G/LOD2.0.0: 3D Geometry-The simplified 3D geometry of the buildings accounting for different shapes of a roof as opposed to the prismatic flat roof.
- G/LOD2.0.1: G/LOD2.0.0 and including the number of floors
- G/LOD2.0.2: G/LOD2.0.0 and including window openings
- G/LOD2.0.3: G/LOD2.0.0 and G/LOD2.0.1
- G/LOD2.1.0: 3D Geometry-The actual detailed 3D geometry of the buildings accounting for different shapes of a roof as opposed to the prismatic flat roof.
- G/LOD2.1.1: G/LOD2.1.0 and including the number of floors
- G/LOD2.1.2: G/LOD2.1.0 and including window openings
- G/LOD2.1.3: G/LOD2.1.0 and G/LOD2.1.1
- G/LOD2.2.0: 3D Geometry-The actual detailed 3D geometry of the buildings with detailed heights and corresponding shapes of the roofs as opposed to the prismatic flat roof.
- G/LOD2.2.1: G/LOD2.2.0 and including the number of floors
- G/LOD2.2.2: G/LOD2.2.0 and including window openings
- G/LOD2.2.3: G/LOD2.2.0 and G/LOD2.2.1

Appendix A.1.1. Building-Thermal Zoning Layer

Thermal zoning is divided into four sub-categories as:

- TZ/LOD0: Single Zone per Building- Each building volume is a single thermal zone
- TZ/LOD1: Zone per Floor/Space–Separate thermal zone to have different adjacency and exposure.
- TZ/LOD2.0: Core-Perimeter Zoning for the whole volume– Accounts for the impact of different orientations.
- TZ/LOD2.0: Core-Perimeter Zoning per Floor/Space – Accounts for the impact of different orientations, different adjacency, and exposure
- TZ/LOD3: Detailed Internal Zoning – Further divides interiors spaces following the building's interior layout e.g.

Figure 1 provides visualization. The subset for the level of detail is based on previous literature in energy modeling.

Appendix A.1.2. Model Calibration/Validation-Temporal Resolution Layer

Temporal resolution refers to the discreet resolution of the measured data with respect to the time used for validation. The temporal resolution layer is divided into three categories based on the availability of collected measured data and discussed in the literature:

- TR/LOD0: Annual Measured Data
- TR/LOD1: Monthly/bimonthly Data
- TR/LOD2: Weekly/Daily Data
- TR/LOD3: Hourly Data

Appendix A.1.3. Model Calibration/Validation-Spatial Resolution Layer

The spatial resolution layer is divided into four main categories [29] Mather et al. suggested and we extended it by adding sub-categories to the aggregated data since they have been presented in the literature as:

- SR/LOD0: No calibration
- SR/LOD1: Archetype Level
- SR/LOD2: Aggregate Level
- SR/LOD2.0: Aggregated Level based on location (ZIP, District)
- SR/LOD2.1: Aggregated Level based on construction year
- SR/LOD2.2: Aggregated Level based on primary use type
- SR/LOD2.3: Aggregated Level based on supervised/unsupervised clusters
- SR/LOD3: Building Level – Metered energy data for each building being simulated in UBEM used for calibration building by building

While the definition of temporal resolution of measured data is quite clear and consistent in the literature, spatial resolution is discussed differently. So this area also needs more explicit study.

Appendix A.1.4. Model Accuracy Layer

The metrics used for model accuracy were observed and presented as:

- (%): Percentile Error or Percentage Error,
- (R^2):The Coefficient of Determination
- (CVRMSE): The coefficient of Variation of the Root Mean Squared Error
- (NMBE): Normalised mean bias error
- (MAPE): Mean Absolute Percentage Error
- (MPE): Mean Percentage Error
- (NRMSE): Normalized Root Mean Square
- (TEDI): Total Thermal Energy Demand Intensity Error

References

[1] United nations population division (2019).
URL https://population.un.org/wpp/DataQuery/

[2] 2019 global status report for buildings and construction, accessed: 2020-10-14 (2019).
URL https://www.worldgbc.org/sites/default/files/

[3] C. Zhu, W. Tian, P. D. Wilde, B. Yin, Approximate bayesian computation in parameter estimation of building energy models, Environmental Science and Engineering Proceedings of the 11th International Symposium on Heating, Ventilation and Air Conditioning (ISHVAC 2019) (2020) 391–399doi:10.1007/978-981-13-9528-4_40.

[4] C. Zhu, W. Tian, B. Yin, Z. Li, J. Shi, Uncertainty calibration of building energy models by combining approximate bayesian computation and machine learning algorithms, Applied Energy 268 (2020) 115025. doi:10.1016/j.apenergy.2020.115025.

[5] Z. Zhang, A. Chong, Y. Pan, C. Zhang, K. P. Lam, Whole building energy model for hvac optimal control: A practical framework based on deep reinforcement learning, Energy and Buildings 199 (2019) 472–490. doi:10.1016/j.enbuild.2019.07.029.

[6] Z. Zhang, A. Chong, Y. Pan, C. Zhang, S. Lu, K. P. Lam, A deep reinforcement learning approach to using whole building energy model for hvac optimal control, 2018 Building Performance Modeling Conference and SimBuild co-organized by ASHRAE and IBPSA-USA (2018).

[7] M. H. Kristensen, R. Choudhary, S. Petersen, Bayesian calibration of building energy models: Comparison of predictive accuracy using metered utility data of different temporal resolution, Energy Procedia 122 (2017) 277–282. doi:10.1016/j.egypro.2017.07.322.

[8] R. Gagnon, L. Gosselin, S. Decker, Sensitivity analysis of energy performance and thermal comfort throughout building design process, Energy and Buildings 164 (2018) 278–294. doi:10.1016/j.enbuild.2017.12.066.

[9] J. S. Hygh, J. F. Decarolis, D. B. Hill, S. R. Ranjithan, Multivariate regression as an energy assessment tool in early building design, Building and Environment 57 (2012) 165–175. doi:10.1016/j.buildenv.2012.04.021.

[10] D. G. Sanchez, B. Lacarrière, M. Musy, B. Bourges, Application of sensitivity analysis in building energy simulations: Combining first- and second-order elementary effects methods, Energy and Buildings 68 (2014) 741–750. doi:10.1016/j.enbuild.2012.08.048.

[11] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data, Renewable and Sustainable Energy Reviews 37 (2014) 123–141. doi:10.1016/j.rser.2014.05.007.

[12] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: A review, Renewable and Sustainable Energy Reviews 23 (2013) 272–288. doi:10.1016/j.rser.2013.03.004.

[13] N. Abbasabadi, M. Ashayeri, Urban energy use modeling methods and tools: A review and an outlook, Building and Environment 161 (2019) 106270. doi:10.1016/j.buildenv.2019.106270.

[14] F. Campolongo, J. Cariboni, A. Saltelli, An effective screening design for sensitivity analysis of large models, Environmental Modelling & Software 22 (10) (2007) 1509–1518. doi:10.1016/j.envsoft.2006.10.004.

[15] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global sensitivity analysis: the primer, Wiley, 2008.

[16] J. Vazquez-Canteli, A. D. Demir, J. Brown, Z. Nagy, Deep neural networks as surrogate models for urban energy simulations, Journal of Physics: Conference Series 1343 (11 2019). doi:10.1088/1742-6596/1343/1/012002.

[17] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data, Renewable and Sustainable Energy Reviews 37 (2014) 123–141. doi:10.1016/j.rser.2014.05.007.

[18] D. Hou, I. Hassan, L. Wang, Review on building energy model calibration by bayesian inference, Renewable and Sustainable Energy Reviews 143 (2021) 110930. doi:10.1016/j.rser.2021.110930.

[19] J. Keirstead, M. Jennings, A. Sivakumar, A review of urban energy system models: Approaches, challenges and opportunities, Renewable and Sustainable Energy Reviews 16 (6) (2012) 3847–3866. doi:10.1016/j.rser.2012.02.047.

[20] T. Hong, Y. Chen, X. Luo, N. Luo, S. H. Lee, Ten questions on urban building energy modeling, Building and Environment 168 (1 2020). doi:10.1016/j.buildenv.2019.106508.

[21] H. Li, S. Wang, H. Cheung, Sensitivity analysis of design parameters and optimal design for zero/low energy buildings in subtropical regions, Applied Energy 228 (2018) 1280–1291. doi:10.1016/j.apenergy.2018.07.023.

[22] X. Faure, T. Johansson, O. Pasichnyi, The impact of detail, shadowing and thermal zoning levels on urban building energy modelling (ubem) on a district scale†, Energies 15 (2 2022). doi:10.3390/en15041525.

[23] 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 (Switzerland) 13 (2021) 1–22. doi:10.3390/su13041595.

[24] S. Risch, P. Remmen, D. Müller, Influence of data acquisition on the bayesian calibration of urban building energy models, Energy and Buildings 230 (1 2021). doi:10.1016/j.enbuild.2020.110512.

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

[26] R. Nouvel, A. Mastrucci, U. Leopold, O. Baume, V. Coors, U. Eicker, Combining gis-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support, Energy and Buildings 107 (2015) 204–212. doi:10.1016/j.enbuild.2015.08.021.

[27] A. Oraiopoulos, B. Howard, On the accuracy of urban building energy modelling, Renewable and Sustainable Energy Reviews 158 (4 2022). doi:10.1016/j.rser.2021.111976.

[28] F. Biljecki, H. Ledoux, J. Stoter, An improved lod specification for 3d building models, Computers, Environment and Urban Systems 59 (2016) 25–37. doi:10.1016/j.compenvurbsys.2016.04.005.

[29] A. Mathur, P. Fennell, R. Rawal, I. Korolija, Assessing a fit-for-purpose urban building energy modelling framework with reference to ahmedabad, Science and Technology for the Built Environment 27 (2021) 1075–1103. doi:10.1080/23744731.2021.1941248.

[30] N. Rashidfarokhi, Calibration in urban building energy modeling (2021).
URL `http://www.teknat.uu.se/student`

[31] C. C. Davila, C. F. Reinhart, J. L. Bemis, Modeling boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets, Energy 117 (2016) 237–250. doi:10.1016/j.energy.2016.10.057.

[32] J. Sokol, C. C. Davila, C. F. Reinhart, Validation of a bayesian-based method for defining residential archetypes in urban building energy models, Energy and Buildings 134 (2017) 11–24. doi:10.1016/j.enbuild.2016.10.050.

[33] G. Tardioli, A. Narayan, R. Kerrigan, M. Oates, J. O'Donnell, D. P. Finn, A methodology for calibration of building energy models at district scale using clustering and surrogate techniques, Energy and Buildings 226 (11 2020). doi:10.1016/j.enbuild.2020.110309.

[34] 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. doi:10.1016/j.egypro.2017.07.459.

[35] A. Katal, M. Mortezazadeh, L. L. Wang, H. Yu, Urban building energy and microclimate modeling – from 3d city generation to dynamic simulations, Energy 251 (2022) 123817. doi:10.1016/j.energy.2022.123817.

[36] M. Gholami, D. Torreggiani, P. Tassinari, A. Barbaresi, Narrowing uncertainties in forecasting urban building energy demand through an optimal archetyping method, Renewable and Sustainable Energy Reviews 148 (9 2021). doi:10.1016/j.rser.2021.111312.

[37] G. Ledesma, O. Pons-Valladares, J. Nikolic, Real-reference buildings for urban energy modelling: A multistage validation and diversification approach, Building and Environment 203 (10 2021). doi:10.1016/j.buildenv.2021.108058.

[38] C. K. Wang, S. Tindemans, C. Miller, G. Agugiaro, J. Stoter, Bayesian calibration at the urban scale: a case study on a large residential heating demand application in amsterdam, Journal of Building Performance Simulation 13 (2020) 347–361. doi:10.1080/19401493.2020.1729862.

[39] T. Berthou, B. Duplessis, P. Stabat, P. Riviere, D. Marchio, Urban energy models validation in data scarcity context: Case of the electricity consumption in the french residential sector, Building Simulation Conference Proceedings 5 (2019) 3140–3147. doi:10.26868/25222708.2019.210134.

[40] A. Krayem, A. A. Bitar, A. Ahmad, G. Faour, J. P. Gastellu-Etchegorry, I. Lakkis, J. Gerard, H. Zaraket, A. Yeretzian, S. Najem, Urban energy modeling and calibration of a coastal mediterranean city: The case of beirut, Energy and Buildings 199 (2019) 223–234. doi:10.1016/j.enbuild.2019.06.050.

[41] F. H. Free, F. Battini, G. Pernigotto, A. Gasparella, A citysim urban energy simulation for the development of retrofit scenarios for a neighborhood in bolzano, italy (2019).

[42] A. Nutkiewicz, Z. Yang, R. K. Jain, Data-driven urban energy simulation (due-s): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow, Applied Energy 225 (2018) 1176–1189. doi:10.1016/j.apenergy.2018.05.023.

[43] P. Nageler, A. Koch, F. Mauthner, I. Leusbrock, T. Mach, C. Hochenauer, R. Heimrath, Comparison of dynamic urban building energy models (ubem): Sigmoid energy signature and physical modelling approach, Energy and Buildings 179 (2018) 333–343. doi:10.1016/j.enbuild.2018.09.034.

[44] R. Nouvel, M. Zirak, V. Coors, U. Eicker, The influence of data quality on urban heating demand modeling using 3d city models, Computers, Environment and Urban Systems 64 (2017) 68–80. doi:10.1016/j.compenvurbsys.2016.12.005.

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

[46] Y. Chen, T. Hong, M. A. Piette, City-scale building retrofit analysis: A case study using citybes, Building Simulation Conference Proceedings 1 (2017) 369–376. doi:10.26868/25222708.2017.071.

[47] F. Ascione, R. F. D. Masi, F. de Rossi, R. Fistola, M. Sasso, G. P. Vanoli, Analysis and diagnosis of the energy performance of buildings and districts: Methodology, validation and development of urban energy maps, Cities 35 (2013) 270–283. doi:10.1016/j.cities.2013.04.012.

[48] Ashrae standards (2020).
URL `https://www.ashrae.org/`

[49] P. Wate, V. Coors, 3d data models for urban energy simulation, Energy Procedia 78 (2015) 3372–3377. doi:10.1016/j.egypro.2015.11.753.

[50] F. Biljecki, Exploration of open data in Southeast Asia to generate 3D building models, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences VI-4/W1-2020 (2020) 37–44. doi:10.5194/isprs-annals-vi-4-w1-2020-37-2020.

[51] G. Gröger, L. Plümer, Citygml - interoperable semantic 3d city models, ISPRS Journal of Photogrammetry and Remote Sensing 71 (2012) 12–33. doi:10.1016/j.isprsjprs.2012.04.004.

[52] D. Robinson, F. Haldi, J. Kampf, P. Leroux, D. Perez, A. Rasheed, U. Wilke, Citysim: Comprehensive micro-simulation of resource flows for sustainable urban planning., Building Simulation 2009: Eleventh International IBPSA Conference 149 (2009) 1083–1090. doi:10.21105/joss.00097.

[53] Y. Chen, T. Hong, Automatic and rapid calibration of urban building energy models, Proceedings of Building Simulation 2019: 16th Conference of IBPSA (2019). doi:10.26868/25222708.2019.210190.

[54] M. H. Kristensen, S. Petersen, Choosing the appropriate sensitivity analysis method for building energy model-based investigations, Energy and Buildings 130 (2016) 166–176. doi:10.1016/j.enbuild.2016.08.038.

[55] M. H. Kristensen, R. E. Hedegaard, S. Petersen, Hierarchical calibration of archetypes for urban building energy modeling, Energy and Buildings 175 (2018) 219–234. doi:10.1016/j.enbuild.2018.07.030.

[56] S. Yang, W. Tian, E. Cubi, Q. Meng, Y. Liu, L. Wei, Comparison of sensitivity analysis methods in building energy assessment, Procedia Engineering 146 (2016) 174–181. doi:10.1016/j.proeng.2016.06.369.

[57] S. Petersen, M. H. Kristensen, M. D. Knudsen, Prerequisites for reliable sensitivity analysis of a high fidelity building energy model, Energy and Buildings 183 (2019) 1–16. doi:10.1016/j.enbuild.2018.10.035.

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

[59] Ut building information, accessed: 2021-01-14 (2021).
URL `https://utdirect.utexas.edu/apps/campus/buildings/nlogon/facilities/`

[60] K. Ahmed, A. Akhondzada, J. Kurnitski, B. Olesen, Occupancy schedules for energy simulation in new pren16798-1 and iso/fdis 17772-1 standards, Sustainable Cities and Society 35 (2017) 134–144. doi:10.1016/j.scs.2017.07.010.

[61] M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (2001).

[62] C.-K. Wang, S. Tindemans, C. Miller, G. Agugiaro, J. Stoter, Bayesian calibration at the urban scale: a case study on a large residential heating demand application in amsterdam, Journal of Building Performance Simulation 13 (3) (2020) 347–361. doi:10.1080/19401493.2020.1729862.