Energy & Buildings 344 (2025) 115971

Contents lists available at ScienceDirect

Energy & Buildings

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

Building energy simulation calibration accuracy and modelling complexity: Implications for energy performance improvement

Laurence Peinturier *, David C.H. Wallom

Oxford e-Research Centre, Department of Engineering Science, University of Oxford, Oxford, United Kingdom

A R T I C L E I N F O

Keywords:
Building Energy Simulation
Calibration Accuracy
Modelling Complexity
Level of Detail
Energy Conservation Measure
Energy Performance

A B S T R A C T

Building Energy Simulation (BES) offers valuable virtual assessment of decarbonisation strategies but faces implementation challenges due to the complex balance between accuracy, modelling detail, and data availability. There is a significant lack of understanding in how these factors interact across different building characteristics and how they impact the reliability of energy conservation measure (ECM) modelling. This study proposes a novel integrated framework that correlates input data quality, modelling complexity, and ECM reliability across diverse building types and sizes, and introduces a quantitative scoring system to objectively balance modelling effort against prediction accuracy. The methodology employs three levels of modelling detail (LOD), progressively simplifying input parameters while verifying calibration accuracy through one-at-a-time modifications. A systematic evaluation of fifteen ECMs is then conducted on each LOD model. The framework is validated using three University of Oxford case study buildings of different sizes, types, and ages. Results reveal larger buildings permit greater simplification while maintaining accuracy with goodness of fit values averaging 7.1 % for simplified models in the largest case study building compared to 8.6 % in the smallest. Process-driven buildings show greater resilience to schedule-related simplifications with sensitivity values averaging 0.08 compared to 0.17 for occupancy-driven buildings. Deep retrofit and renewable measures maintain high reliability with simplified models, scoring highest on the developed scoring system, while control-based ECMs require higher modelling fidelity. This comprehensive framework provides a novel methodology for optimising BES implementation based on building characteristics and analysis objectives, facilitating broader adoption of BES tools for net-zero strategy development and policy support.

Nomenclature

Nomenclature and abbreviations	
AC	air conditioning
AHU	air handling unit
AMY	actual meteorological year
ASHP	air source heat pump
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
AWB	Andrew Wiles building
BES	building energy simulation
BIM	building information modelling
BMS	building management system
CHP	combined heat and power
CIBSE	Chartered Institution of Building Services Engineers
CL	confidence level
COP	coefficient of performance

(*continued on next column*)

Nomenclature (*continued*)

Nomenclature and abbreviations	
CS	case study
CV(RMSE)	coefficient of variation of root mean square error
DEC	display energy certificate
DHW	domestic hot water
ECM	energy conservation measure
EPC	energy performance certificate
GOF	goodness of fit
GSHP	ground source heat pump
HH	Holywell House
HVAC	heating, ventilation, and air conditioning
IAQ	indoor air quality
IC	influence coefficient
IEA	International Energy Agency
IEA EBC	IEA Energy in Buildings and Communities Programme
IP	input parameter

(*continued on next page*)

* Corresponding author.
E-mail address: laurence.peinturier@eng.ox.ac.uk (L. Peinturier).

https://doi.org/10.1016/j.enbuild.2025.115971
Received 1 April 2025; Received in revised form 18 May 2025; Accepted 1 June 2025
Available online 4 June 2025

Nomenclature (*continued*)

Nomenclature and abbreviations	
IR	infrared
LED	light emitting diode
LOD	level of detail
NMBE	normalised mean bias error
OAT	one at a time
PHARM	Department of Pharmacology building
POEI	post-occupancy evaluation and intervention
PV	photovoltaic
SA	sensitivity analysis
TMY	typical meteorological year
UBEM	urban building energy modelling
UK	United Kingdom
UKGBC	UK Green Building Council
UL	unregulated loads
VAV	variable air volume
Variables	
Abs.Sav.	absolute saving
Abs.U.	absolute uncertainty
DeviationScore	deviation score
IP	input parameter
m_i	measured data point
n	number of simulated data points
N_p	total number of data points
OP	output
p	number of parameters
ReliabilityScore	reliability score
Rel.Sav.	relative saving
Rel.U.	relative uncertainty
S	sensitivity
Sav.	saving
Score	score
s_i	simulated data point
TimeScore	time score
$t-stat$	t-statistic
UncertaintyScore	uncertainty score
$U(R)$	random uncertainty
U.Range	uncertainty range
$U(S)$	systematic uncertainty
w_R	reliability weight
w_T	time weight
Functions and operators	
Δ	difference (delta)
$log(\cdot)$	log function
$max(\cdot)$	maximum
$min(\cdot)$	minimum
$t^{-1}(\cdot)$	inverse t-distribution
Subscripts and superscripts	
B	base
ECM_{LOD_j}	ECM-LOD model
LOD_j	LOD model
Energy	energy type

1. Introduction

Nations have pledged to achieve carbon neutrality by 2050 through the Paris Agreement [1]. The energy sector is responsible for more than three quarters of global emissions, with building operations alone accounting for a third of global final energy demand [2,3]. The International Energy Agency (IEA) states that at least 20 % of existing buildings must be zero-carbon-ready by 2030 to reach the Paris Agreement goals [3]. This target is particularly crucial since an estimated 90 % of current buildings will still be in use by 2050 [4]. Non-domestic buildings, though only accounting for 21 % of the total global building floor area, contribute to 50 % of building-related emissions due to their high energy intensity, largely driven by their heterogeneous nature and the increased demand for better indoor environmental quality [5–7].

Building Energy Simulation (BES) is defined as 'the use of computational mathematical models to represent the physical characteristics, expected or actual operation, and control strategies of a building (or buildings) and its (their) energy systems' [8]. BES can be an innovative method to support and derisk building decarbonisation strategies. It can assist building stakeholders and policymakers in virtually evaluating different decarbonisation scenarios and cost options before implementing any major physical changes, providing assurance that energy conservation measures (ECMs) and retrofits will meet their expectations [9]. However, while BES shows great potential, it is inherently complex and time-consuming, hindering its wider adoption. The balance between modelling accuracy and complexity needs to be better optimised to maximise BES's effectiveness while minimising human and computational efforts [10]. Moreover, the quality of a BES model and therefore its reliability for ECM modelling depends on the quality and accessibility of the input data used to create the model [11]. While modern buildings often come with Building Management Systems (BMS) and comprehensive documentation, older buildings typically lack mechanisms to generate high-resolution data.

Three main BES modelling approaches exist: white-box, black-box, and grey-box modelling [11]. Black-box modelling is data-driven and relies on machine learning algorithms to relate key input parameters to simulated outputs [12]. It is advantageous for its shorter development times and minimal need for detailed building-specific knowledge, however does not explicitly link inputs to physical building parameters, making it unsuitable for ECM modelling [11,12]. Grey-box modelling is a hybrid approach, offering greater interpretability than black-box modelling, but can only account for changes in aggregated input parameters, restricting its scope of application [11,12]. White-box modelling is a physics-based approach, describing building characteristics based on conservation of mass, energy, and momentum, and provides high accuracy and detailed insights into energy use and performance predictions [11,12]. However, this approach tends to be over-parameterised, requiring more data than are typically available, making it time-consuming and expertise-intensive to develop [11]. Further research is needed to simplify white-box approaches in terms of data collection and modelling complexity while maintaining accuracy, ECM modelling capabilities, and recognition of individual building characteristics. The remainder of this study focuses exclusively on white-box modelling approaches due to their suitability for ECM evaluation and detailed energy performance analysis.

Creating accurate yet simplified white-box BES models requires balancing accuracy, sensitivity, and modelling complexity. Model accuracy verification occurs through calibration against actual building energy use and conditions [11]. Previous studies have investigated spatial modelling simplifications at whole-building, floor, and zone levels [10,13–15]. While Cheng et al. [13] explored data collection efforts versus model reliability in single zones, limiting whole-building applicability, Kong et al. [10] analysed whole-building approaches but overlooked model bias. Zone-typing, which merges thermal zones of similar activity [15], can accelerate modelling processes but requires careful implementation to avoid accuracy loss through over-simplification [16]. Studies have also explore temporal modelling simplifications [17,18], though rarely addressing data sourcing or calibration purposes. The concept of using simplified models with different levels of detail (LOD) has gained popularity in Urban Building Energy Modelling (UBEM) in recent years. This approach helps accelerate the modelling process at urban scale [19–22]. However, studies typically analyse either urban-scale LOD or zone-typing exclusively, making it more challenging to assess individual parameter simplification impacts. Further research is needed on single-building scale modelling detail variations and their effects on accuracy and complexity. Sensitivity analysis (SA) can help evaluate isolated impacts of modelling simplifications by quantifying how individual input parameter variations affect simulated outputs [23]. While numerous studies employ SA to examine model inputs [17,18,24–27], they either focus exclusively on accuracy or neglect the trade-offs of modelling simplifications [10,13], suggesting opportunities to leverage SA for reducing complexity and computational costs. In terms of ECMs and retrofits, identifying optimal building refurbishment approaches which minimise costs and maximise energy savings is crucial. Validated BES models can test virtually various measures and assess savings, costs, and efforts, as demonstrated in multiple studies [17,28–32]. However, modelling uncertainty can have a significant impact on the confidence of ECM saving predictions [33]. Uncertainty analysis in BES has been explored in several studies to address this challenge [34–37]. However, these studies typically address either calibration accuracy or uncertainty quantification in isolation. Limited work has been conducted on systematically connecting calibration accuracy, modelling complexity, and ECM reliability across different building types and sizes [13]. This integration is crucial for developing practical guidance on which level of modelling detail is appropriate for different building characteristics and intervention strategies. Furthermore, while studies like Sun et al. [37] have applied multi-criteria optimisation models for retrofit decision-making, they do not address the practical challenges of modelling resource constraints and do not explore how framework guidelines might need to vary across different building typologies. There remains a critical need for a holistic methodology that balances modelling accuracy with practical implementation constraints while accounting for building-specific characteristics [38].

While a number of previous studies have explored aspects of calibration accuracy, modelling complexity, or ECM reliability individually, no comprehensive framework exists that connects all three dimensions across different building characteristics. This study presents two key novel contributions to BES methodology. First, it proposes an integrated framework that systematically quantifies the relationships between calibration accuracy, modelling complexity, and ECM reliability across diverse building characteristics. Second, it introduces a quantitative scoring system that objectively balances modelling effort against ECM saving prediction reliability. The methodology is verified using three case study (CS) buildings of different sizes and types from the University of Oxford, United Kingdom (UK). The paper first describes the methodology, then presents and critically discusses results to identify patterns across findings and CS buildings, before concluding with key learnings and future work directions.

2. Methods

This section defines the methods adopted for this study, which methodological framework is illustrated in Fig. 1. The framework is structured in two parts. The first investigates how modelling simplifications affect BES accuracy through a three-level modelling approach (LOD 3, LOD 2, and LOD 1). This staged approach is adapted from UBEM simplification studies by Johari et al. [19] and Bishop et al. [21]. LOD 3 represents the most comprehensive model in terms of input data quality and resolution. LOD 2 and LOD 1 represent scenarios with progressively limited data access or reduced modelling resources.

Fig. 2 summarises the data collection methods and modelling resolutions for each LOD level, adapted from reviews by Mantha et al. [39] and Coakley et al. [11]. Simplifications for LOD 2 and LOD 1 are implemented by modifying one parameter at a time (OAT) while verifying calibration accuracy, and a sensitivity analysis is performed to identify the most influential input parameters on simulated energy outputs. The validated OAT simplifications are then combined to create the general LOD 2 and LOD 1 models, with modelling times and computational efforts quantified for each LOD model.

The second part of the framework examines how modelling simplifications affect ECM reliability. Various ECMs are modelled on each LOD model to assess their impact on energy use, with results subjected to uncertainty analysis to examine relationships between modelling accuracy, complexity, and ECM reliability. All results are compared across the three different CS buildings under study.

The following sections detail the calibration methods and input parameter categorisation (2.1), creation of LOD 3 models and OAT modifications (2.2), OAT sensitivity analysis (2.3), development of final LOD 2 and LOD 1 models (2.4), ECM modelling (2.5), ECM uncertainty analysis (2.6), and case study context (2.7).

2.1. Calibration methods and input parameter categorisation

All simulations run in this study employ a white-box modelling approach using DesignBuilder, the graphical user interface for EnergyPlus [40,41]. The calibration of the models' simulated electricity and gas outputs is validated using the coefficient of variation of root mean square error (CV(RMSE)) and normalised mean bias error (NMBE) functions on a monthly basis over the year of 2023. The CV(RMSE) and NMBE functions can be found in Eqs. (1) and (2) [42]. The CV(RMSE) quantifies how well a model fits real data while the NMBE identifies overall model bias [11].

$$CV(RMSE) = \frac{\sqrt{\sum_{i=1}^{N_p} (m_i - s_i)^2 / N_p}}{\bar{m}} \tag{1}$$

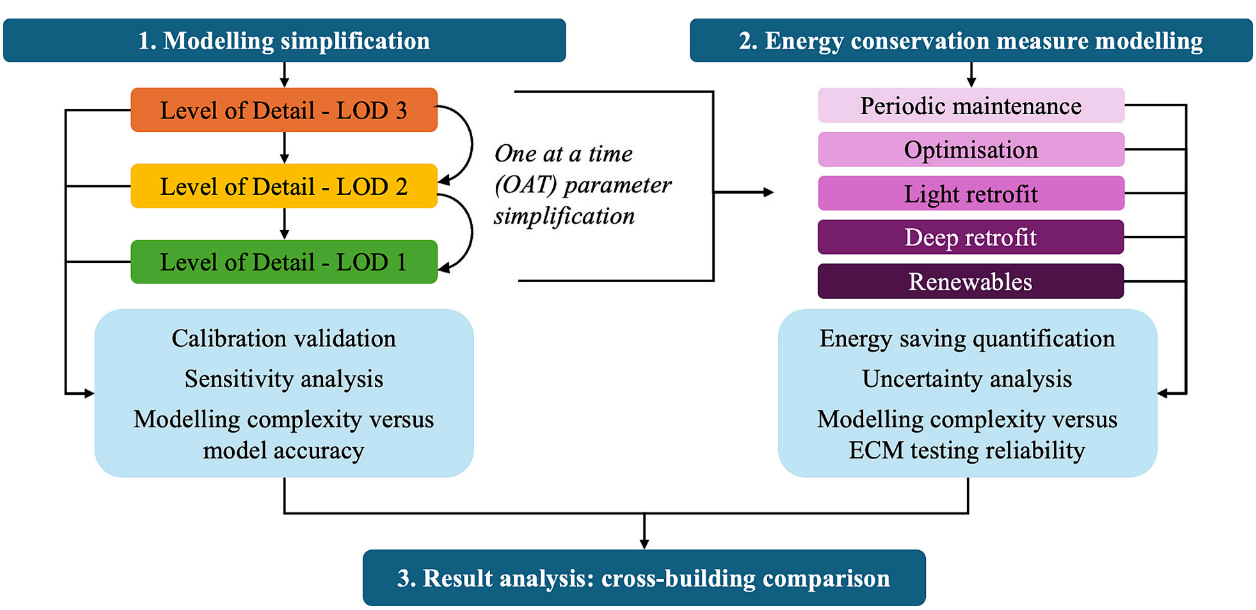

Fig. 1. Overview of methodological framework.

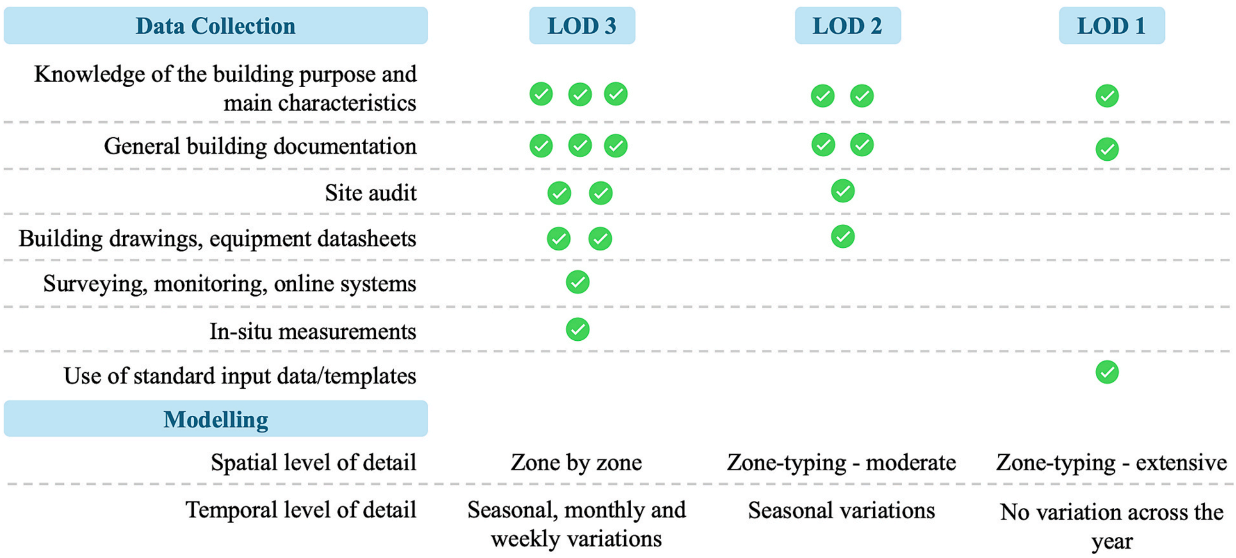

Fig. 2. Summary of data collection and levels of modelling detail for each LOD model. The number of green ticks indicates the degree of availability for each data collection method presented. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$NMBE = \frac{\sum_{i=1}^{N_p}(m_i - s_i)}{\sum_{i=1}^{N_p}(m_i)} \tag{2}$$

N_p is the number of data points ($N_p = 12$ for a monthly resolution), m_i and s_i are the measured and simulated data points respectively, and $\overline{m}$ is the average of the measured data points. All BES models in this study have a calibration considered validated if the CV(RMSE) for both electricity and gas is below 15 % and NMBE between −5% and 5 %, as recommended by ASHRAE Guideline 14-2014 [42]. The CV(RMSE) and NMBE functions only involve comparison of simulated and measured energy outputs. Another critical aspect of calibration validation is ensuring the simulated building model's indoor environmental conditions match with real-life conditions [25]. The models' indoor temperatures are therefore also verified. Finally, for easier comparison analysis, the Goodness of Fit (GOF) function is also computed for the final LOD 3, 2, and 1 models. The GOF function combines both the variance and bias of a model, as defined in Eq. (3) [23].

$$GOF = \frac{\sqrt{2}}{2} * \sqrt{CV(RMSE)^2 + NMBE^2} \tag{3}$$

The IEA Energy in Buildings and Communities (EBC) annex 53 identifies six key factors for building energy performance analysis: (1) climate, (2) building envelope, (3) building services and energy systems, (4) building operation and maintenance, (5) occupant activities and behaviour, and (6) indoor environmental quality [43]. Building on this framework and previous studies on BES input parameter resolutions [10,44], 12 main categories of input parameters (IP) are established for the OAT model simplification process, summarised in Fig. 3. Weather data is handled separately from the IP categories as no modelling simplifications are applied to this parameter. All BES models use the same Actual Meteorological Year (AMY) weather file obtained from the climate data interface Oikolab [45], which provides post-processed historical weather data from national weather agencies. While using a Typical Meteorological Year (TMY) weather file from a nearby prominent location could be a potential simplification due to easier access, TMY files are typically updated infrequently. This raises the risk of underestimating current-year weather impacts, which can significantly affect BES modelling accuracy beyond calibration thresholds [46,47].

2.2. LOD 3 model development and OAT modifications

The input data sources and modelling resolution details for creating the LOD 3 models are summarised in Table 1. The models undergo manual calibration through an iterative adjustment process in which energy service contributions are systematically varied based on sub-meter data. While more time-intensive than automated methods, manual calibration reduces uncertainty and provides better understanding of input–output correlations by maintaining direct control over

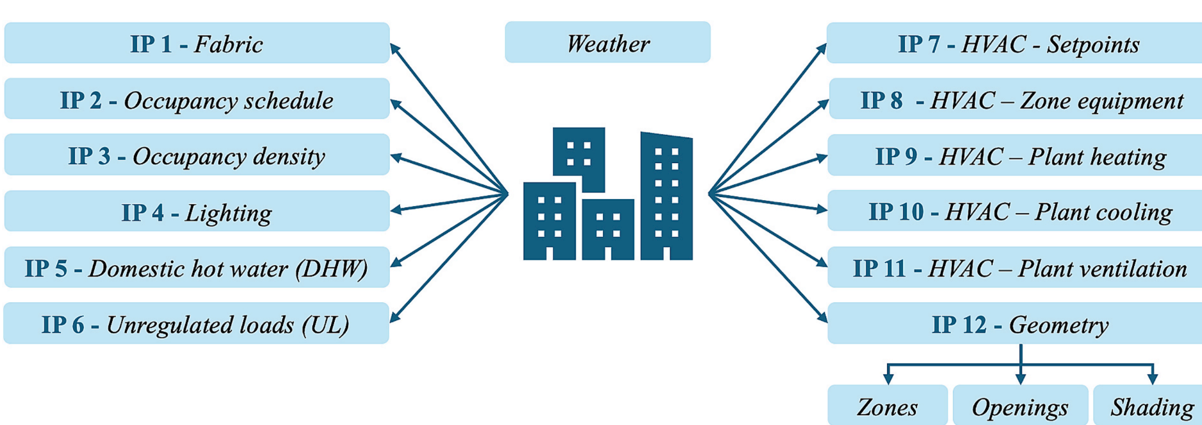

Fig. 3. Input parameter categorisation used in this study for BES modelling.

Table 1
Summary of input parameter sources and modelling information for LOD 3 models, and OAT simplification details for LOD 2 and LOD 1.

Category	Input Parameter Source	OAT simplification	OAT simplification
	LOD 3	***LOD 2***	***LOD 1***
IP 1: Fabric	U-value measurements, architectural drawings, building documentation	Uniform U-values across the building	U-values based on construction regulations [49]
IP 2 and 3: Occupancy schedule	Entrance card data, Wi-Fi data, online room booking system, site audit – zone-level hourly schedules with seasonal variation	9 am – 5 pm Monday – Friday for all zones except rooms which follow business-specific schedules	9 am – 5 pm Monday – Friday for all zones
IP 3: Occupancy density	Entrance card data, Wi-Fi data, CO_2 sensors, site audit	Simplified based on zone types and floor areas	Based on CIBSE standard occupancy densities [50]
IP 4: Lighting	Lighting drawings	Simplified based on zone types and floor areas	Based on CIBSE standard lighting power densities [50]
IP 5: Domestic Hot Water (DHW)	Monthly water bills, site audit, datasheets	Simplified based on zone types and floor areas	Based on ASHRAE standard hot water usage [51]
IP 6: Unregulated loads (UL)	Building documentation, site audit, online surveys	Simplified based on zone types and floor areas	Based on CIBSE standard UL power densities [50]
IP 7: HVAC* – Setpoints	Online Building Management System (BMS), indoor air quality monitoring, discussion with building users	Uniform setpoints except special zones	Uniform setpoints for all zones
IP 8: HVAC – Zone equipment	Datasheets, site audit, building documentation, discussion with building managers	Simplified based on zone types and floor areas	Uniform HVAC equipment and flow rates for all zones
IP 9, 10, 11: HVAC – Plant heating, cooling, ventilation	Online BMS, datasheets, site audit, building documentation, discussion with building managers	As LOD 3	Modelled as one asset of equivalent size
IP 12: Geometry	Detailed 3D architectural Building Information Model (BIM), architectural drawings, site audit	Zone-typing: merging neighbouring zones of similar activity by groups of four maximum	Merging neighbouring windows. Simplified external geometry. Further zone-typing.

**HVAC – Heating, ventilation, and air conditioning*

data changes [11]. Based on calibration results, model parameters are fine-tuned to improve alignment with metered data while preserving physical realism, ensuring the LOD 3 models accurately represent building energy performance across all major systems.

The LOD 3 models are progressively simplified by modifying one parameter at a time in two stages – LOD 2 and LOD 1. The OAT simplifications details for each IP category can be found in Table 1. Some heating, ventilation, and air conditioning (HVAC) IP categories for LOD 2 remain similar to LOD 3. This limitation stems from previous research indicating that HVAC plant equipment simplifications should be minimised due to the complex thermal dynamics involved in BES [14,47]. Similarly, geometry simplifications are restricted to zone-typing without further floor- or building-level zone merging. Raftery et al. [16] demonstrated that simplifying geometry beyond zone-typing can cause significant accuracy issues. Fig. 4 illustrates the zone-typing progression from LOD 3 to LOD 2 to LOD 1 geometries. This process follows ASHRAE Standard 90.1 guidelines [48], supplemented by Raftery et al. [16] recommended criteria: thermal zones may be combined if (a) they share similar space functions, (b) they have significant inter-zone surface area compared to the total area of the combined zone, (c) the resulting combined zone is not excessively large in any dimension, and (d) the resulting combined zone does not have excessive floor area.

2.3. OAT sensitivity analysis

To better understand the influence of input parameter changes on simulated energy outputs, an OAT local sensitivity analysis (SA) is performed for each IP category, using the influence coefficient (IC) function, defined in Eq. (4) [26].

$$IC = \left| \frac{\Delta OP \div OP_B}{\Delta IP \div IP_B} \right| \quad (4)$$

ΔOP and ΔIP are the changes in output (energy) and input respectively, and OP_B and IP_B are the initial model outputs and inputs respectively. Table A1 in Appendix A summarises the changes in input parameters for each IP category, LOD, and CS building.

The absolute sensitivity (S) for each IP and type of energy (electricity or gas) is taken as the average of the ICs when modifying IPs from LOD 3 to LOD 2 and from LOD 2 to LOD 1. ICs are computed as [*LOD 3 → LOD 2 + LOD 2 → LOD 1*] rather than [*LOD 3 → LOD 2 + LOD 3 → LOD 1*] to capture the stepwise progression of modelling simplifications. The resulting absolute sensitivity for each IP is defined in Eq. (5).

$$S = \frac{IC_{LOD3 \rightarrow LOD2} + IC_{LOD2 \rightarrow LOD1}}{2} \quad (5)$$

2.4. Final LOD 2 and LOD 1 models and calibration verification

Based on each OAT model's calibration results, the final LOD 2 and LOD 1 models are created. These combine all validated OAT modifications for their respective LOD level. Fig. 5 illustrates this combination process. If an OAT model's calibration is not validated for a specific LOD, the input parameter from the previous LOD is retained in the final combined model. The calibration accuracy of the resulting LOD 2 and LOD 1 models is verified, along with simulated indoor zone temperatures to ensure proper HVAC system simulated operation. Additionally, modelling time, data collection time, simulation execution time, and memory usage are quantified for each LOD model.

2.5. ECM models and energy saving quantification

A set of 15 different ECMs are modelled on each final LOD model. These ECMs are organised into five categories: periodic maintenance, optimisation, light retrofit, deep retrofit, and renewables. The latter four categories are derived from commercial and office building retrofit recommendation guidelines provided by the UK Green Building Council (UKGBC) [49,50]. Periodic maintenance involves regularly inspecting building operating systems to identify or prevent potential equipment or controls failures that could lead to increased energy usage. Optimisation is an iterative process of adjusting controls, setpoints, or operating schedules to maximise energy efficiency. Light retrofit measures involve basic modification or replacement of existing building elements. Deep retrofit measures represent significant works that fundamentally change elements of a building's structure and/or services. Finally, renewable measures involve the installation of renewable energy systems. Table 2 provides the complete list of modelled ECMs. The total annual absolute and relative energy savings of each ECM model are quantified using the formulae in Eqs. (6) and (7) respectively.

$$Abs.Sav._{ECM_{LOD_j}} = Energy_{LOD_j} - Energy_{ECM_{LOD_j}} \quad (6)$$

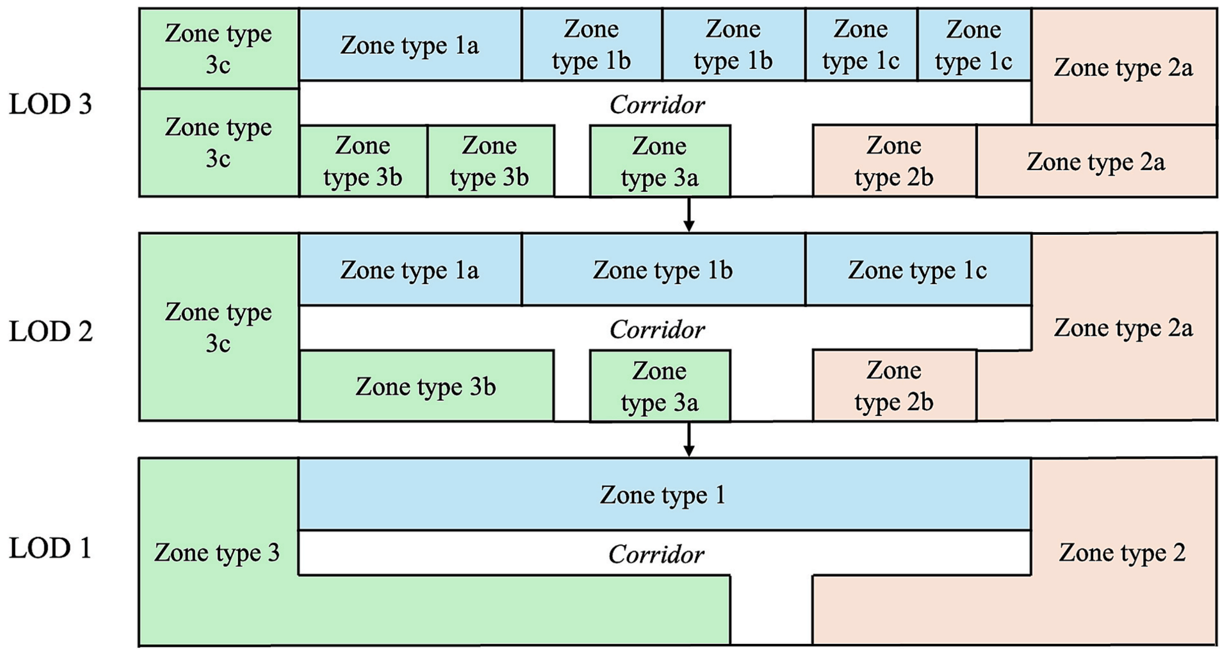

Fig. 4. Geometry simplification process through zone-typing.

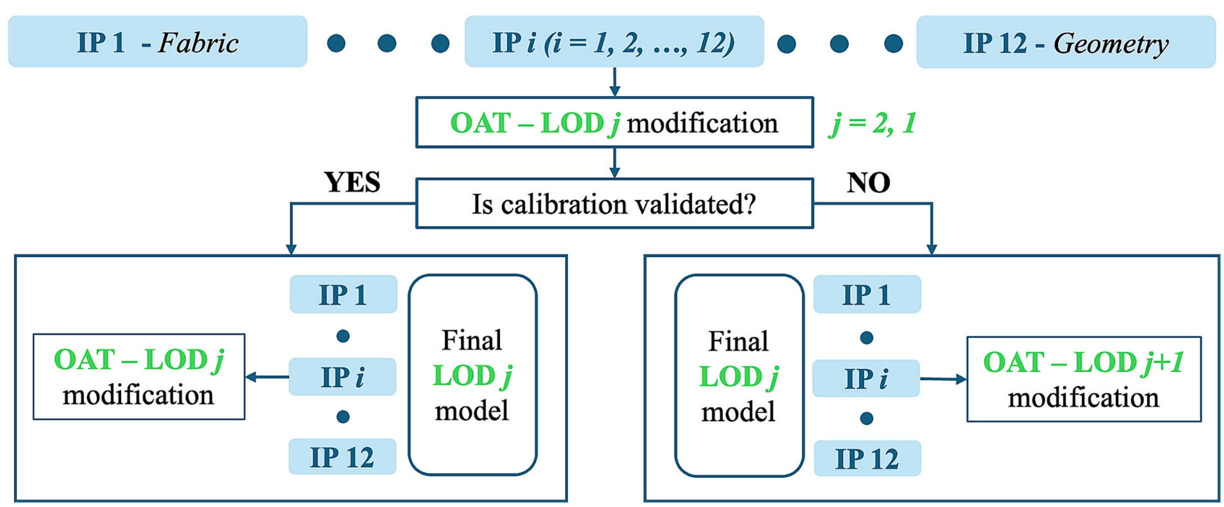

Fig. 5. OAT combination process to create final LOD 2 and LOD 1 models.

$$Rel.Sav._{ECM_{LOD_j}} = \frac{Abs.Sav._{ECM_{LOD_j}}}{Energy_{LOD_j}} \tag{7}$$

Energy can refer to electricity, gas or total energy, and $j = 3, 2, 1$.

2.6. ECM results uncertainty and reliability

An uncertainty analysis is conducted on the resulting energy savings for each ECM model, derived from the inherent uncertainty in the LOD models as reflected by the CV(RMSE) and NMBE functions. This analysis is adapted from ASHRAE Guideline 14 uncertainty assessment methodology [42] and helps better understand the reliability of ECM modelled saving predictions based on modelling detail level, building size, and building type. The first component of model uncertainty is random uncertainty, which reflects the difference between model predictions and actual measurements and is directly associated with the CV (RMSE) function. Random uncertainty is computed as shown in Eq. (8). It is scaled by a factor of $\sqrt{2}$ due to the combination of both the inherent uncertainty in the initial LOD_j model and the uncertainty of the ECM_{LOD_j} model. The second component of model uncertainty is the systematic uncertainty, which reflects the bias in the model and is directly computed as the absolute value of the NMBE error, as shown in Eq. (9). The total uncertainty is the combination of the random and systematic uncertainties, as shown in Eq. (10).

$$U_{LOD_j}(R) = \sqrt{2}{*}CV(RMSE)_{LOD_j} \tag{8}$$

$$U_{LOD_j}(S) = |NMBE|_{LOD_j} \tag{9}$$

$$U_{LOD_j} = \sqrt{U_{LOD_j}(R)^2 + U_{LOD_j}(S)^2} \tag{10}$$

The confidence in predicted ECM results is accounted for through the t-statistic, which is a measure of the probability that an ECM saving result is statistically valid, assuming a confidence level of 95 % [42]. The t-statistic is computed as shown in Eq. (11).

Table 2
Summary of modelled ECMs.

ECM	Description
Periodic Maintenance	
Systems and/or control maintenance	Equipment or control adjustments – identified via site audits and discussion with facilities managers
Optimisation	
Off-site cloud computing	Remove local servers and outsource to cloud computing for more energy efficient operations [49]
Overnight power use	Automatic equipment shutdown when not in use (e.g. wake-on-LAN systems [51]. Assumed to reduce overnight UL equipment use by 30 % [52]
HVAC setpoints and setbacks	Decrease heating setpoint to 19 °C
HVAC schedules	Heating running from 20 Oct to 01 May from 7 am to 7 pm
Light retrofit	
HVAC pump motor replacement	Increase heating pumps motors efficiency by 10 %
Low energy lighting	Replace all lights with light-emitting diode (LED) (2.5 W/m^2)
Deep retrofit	
Building airtightness	Reduce air infiltration rate by 30 %
Window replacement	Triple glazing: 4/12/4/12/4 mm with argon
Roof insulation	Reduce U-value by an absolute 0.15 W/m^{2}K
Wall insulation	Reduce U-value by an absolute 0.1 W/m^{2}K
Floor insulation	Reduce U-value by an absolute 0.08 W/m^{2}K
ASHP* for DHW	Replace DHW electric or gas boiler with ASHP of equivalent size (COP** of 3.5, hot water flow and return of 60 °C/40 °C)
Heat decarbonisation	Replace gas boilers with ASHPs of equivalent size (COP of 3.5, hot water flow and return of 60 °C/40 °C)
Renewables	
Rooftop Photovoltaic (PV)	Install 150 W/m^2-available roof area. The total PV area installed is proportional to building's total floor area for consistent comparison

**ASHP – Air source heat pump*
***COP – Coefficient of performance*

$$t-stat = t^{-1}\left(\frac{1+CL}{2},[n-p]\right) = t^{-1}\left(\frac{1+0.95}{2},11\right) = 2.201 \tag{11}$$

n is the number of simulated data points ($n = 12$ since simulations are run monthly) and p is the number of parameters, taken as 1 for calibrated simulations [42].

Combining Eqs. (6), (10) and (11), the relative and absolute uncertainties associated to a modelled ECM saving (electricity or gas) for a specific LOD can be derived as defined in Eqs. (12) and (13).

$$Rel.U_{LOD_j} = t-stat^{*}U_{LOD_j} \tag{12}$$

$$Abs.U_{ECM_{LOD_j}} = Rel.U_{LOD_j}{}^{*}Abs.Sav._{ECM_{LOD_j}} \tag{13}$$

Finally, a scoring system is developed to assess the balance between ECM saving results reliability and modelling complexity. The scoring relies on two main components: total modelling time and reliability of the modelled ECM results. This system allows for different scoring weights depending on whether accuracy or time is prioritised. The final score associated with a specific modelled ECM is derived using Eq. (14), with scores ranging from 0 to 100.

$$Score_{ECM_{LOD_j}} = \left[TimeScore_{LOD_j}{}^{*}w_T + ReliabilityScore_{ECM_{LOD_j}}{}^{*}w_R\right] \tag{14}$$

w_T and w_R are the weights to assign priorities on modelling time or model reliability ($w_T + w_R = 1$). These are chosen to be 30 % and 70 % respectively for the purpose of this exercise. They can however be modified according to the BES user's priorities.

The time score reflects the total amount of time spent on collecting input data, creating the model, and running the simulation. It is computed as shown in Eq. (15).

$$TimeScore_{LOD_j} = \frac{\max\left(\log(time)_{LOD_{1,2,3}}\right)^{*}1.3 - \log(time)_{LOD_j}}{\max\left(\log(time)_{LOD_{1,2,3}}\right)^{*}1.3 - \min\left(\log(time)_{LOD_{1,2,3}}\right)^{*}0.7}{}^{*}100 \tag{15}$$

where:

time is the total time for data collection, modelling, and simulation execution. The time is normalised based on the maximum and minimum modelling times across LODs, with the maximum and minimum interval window extended by +/- 30 % so that a model with the lowest modelling time does not necessarily get 100 % as modelling times can be subjective and vary across BES users and computer used.

The reliability score is computed as shown in Eq. (16). It is an average of the score associated to the model's uncertainty and the deviation of ECM modelling results from the most 'accurate' model, LOD 3.

$$ReliabilityScore_{ECM_{LOD_j}} = \frac{UncertaintyScore_{LOD_j} + DeviationScore_{ECM_{LOD_j}}}{2}{}^{*}100 \tag{16}$$

The uncertainty score is taken as the average of the uncertainty scores for electricity and gas outputs. It reflects the overall uncertainty associated to a selected LOD and is computed as shown in Eq. (17), using Eq. (12). The higher the relative uncertainty for a specific model, the lower the uncertainty score.

$$UncertaintyScore_{LOD_j} = \frac{\left(1 - Rel.U_{LOD_j-elec}\right) + \left(1 - Rel.U_{LOD_j-gas}\right)}{2} \tag{17}$$

The deviation score reflects the deviation of a specific ECM model electricity and gas saving result compared to its LOD 3 equivalent which is considered to be the most reliable ECM result. The deviation score for any ECM tested on a LOD 3 model is therefore 1. The deviation score for LOD 2 or LOD 1 is then computed as shown in Eq. (18), using Eqs. (19), (20) and (21).

$$DeviationScore_{ECM_{LOD_j}} = \frac{DeviationScore_{ECM_{LOD_j-elec}} + DeviationScore_{ECM_{LOD_j-gas}}}{2} \tag{18}$$

Fig. 6. Google map satellite images of the three CS buildings: AWB, PHARM, and HH (left to right).

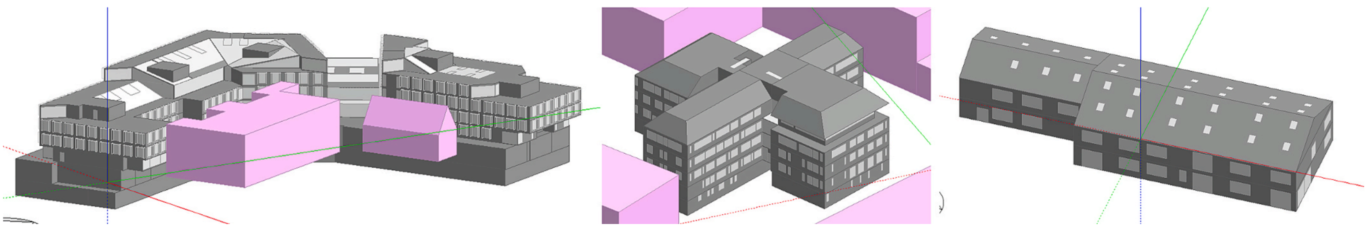

Fig. 7. DesignBuilder LOD 3 models of the three CS buildings: AWB, PHARM, and HH (left to right).

$$DeviationScore_{ECM_{LOD_{j-energy}}} = 1 - \frac{\Delta Sav._{LOD_{j-energy}} - U.Range_{LOD_{3-energy}}}{Abs.Sav._{ECM_{LOD_{3-energy}}}} \tag{19}$$

$$\Delta Sav._{LOD_{j-energy}} = \left| Abs.Sav._{ECM_{LOD_{j-energy}}} - Abs.Sav._{ECM_{LOD_{3-energy}}} \right| \tag{20}$$

$$U.Range_{LOD_{3-energy}} = Abs.Sav._{ECM_{LOD_{3-energy}}} {}^{*}U_{LOD_{3-energy}}(R) \tag{21}$$

The deviation score is higher if the difference in modelled savings ($\Delta Sav._{LOD_{j-energy}}$) falls within the uncertainty range of LOD 3 ($U.Range_{LOD_{3-energy}}$).

2.7. Case study context

The methodology for this study is tested on three CS buildings within the University of Oxford's estate. These buildings significantly differ in size, age, and type, enabling verification of the methodological framework across varying building characteristics and identification of patterns in the findings. Real-life buildings are essential for this validation process, as they provide actual energy data against which calibration accuracy can be verified. The CS buildings are shown in Fig. 6, their DesignBuilder LOD 3 models in Fig. 7, with general information summarised in Table 3.

The first CS building is the Andrew Wiles Building (AWB), the University's Mathematical Institute. Detailed information on this building appears in a previous study examining BES benefits for energy performance gap identification [53]. Despite its large size, AWB features repeating geometry patterns in zone construction. Over 300 zones are offices with similar lighting, HVAC, and unregulated load (UL) equipment. Occupancy modelling is more complex, with varying schedules between office and teaching spaces. The second CS building is the Department of Pharmacology (PHARM). More details on this building are available in a previous study on modelling simplifications and BES calibration accuracy [47]. Though not the largest in size, PHARM requires the most time for data collection and modelling. The building exhibits significant variations in lighting, UL, and HVAC equipment types across zones, along with diverse operating schedules, necessitating careful zone-by-zone modelling to represent actual building conditions. The third CS building is Holywell House (HH), a small office building. Additional information on the building appears in a previous study on post-occupancy evaluation and intervention (POEI) protocols for energy performance assessment [54]. HH is the least complex CS building to model, with fewer than 40 zones. All metered energy data for all three CS buildings are obtained via the University's estates services online energy dashboard.

3. Results

This section presents findings from testing the proposed methodological framework on the three CS buildings. First, modelling simplification results are presented, including calibration accuracy of the LOD 3 models (3.1), sensitivity analysis and calibration verification of OAT models (3.2–3.3), and validation of the final LOD 2 and 1 models (3.4). Section 3.5 then explores the key relationships between modelling complexity and accuracy. Finally, Sections 3.6 and 3.7 analyse ECM modelling outcomes, quantifying energy saving potentials and evaluating ECM reliability across the different LOD models through uncertainty analysis.

3.1. LOD 3 models calibration

The 2023 metered and simulated energy consumption across all CS buildings, along with their corresponding CV(RMSE) and NMBE values are summarised in Fig. 8. All models achieve CV(RMSE) values below 15 % and NMBE values within ± 5 %, meeting ASHRAE Guideline 14 calibration criteria. CV(RMSE) and NMBE values are generally higher for gas than electricity, suggesting thermal dynamics may be more difficult to model accurately. The highest CV(RMSE) appears in HH's simulated gas demand. As the smallest CS building, HH highlights potential challenges in modelling smaller facilities where individual zones and systems may have proportionally larger impacts on overall performance, leaving less room for error.

3.2. OAT modifications calibration verification

The resulting CV(RMSE)s and NMBEs are computed for each OAT model, with 72 models run in total. Most OAT models pass the ASHRAE calibration thresholds, with a few exceptions.

For AWB, models failing calibration are IP 1 (Fabric), IP 3 (Occupancy Density), and IP 6 (UL) at the simplest level, LOD 1. As a modern building with complex architecture and large window-to-wall ratio, AWB's accuracy may be affected by generic U-values. IP 3 – LOD 1 and IP 6 – LOD 1 use generic CIBSE figures for occupancy and UL peak power densities. Since AWB is primarily occupancy-driven, accurately representing these parameters may be crucial for validated calibration. For these three OAT models, the LOD 2 equivalents are retained in the final

Table 3
Overview of the three CS buildings: characteristics, energy performance, and input data information.

	AWB	PHARM	HH
Characteristics			
Construction date	2013	1989	1960s – refurbished in 2022
Main use	Teaching and office	Laboratory and office	Office
Useful floor area (m^2)	21,090	6,436	1,124
Total number of modelled zones	443	329	37
Space heating	Three ground source heat pumps (GSHP) backed up by four gas-condensing boilers	Three gas-condensing boilers and one combined heat and power (CHP) generator	Two gas-condensing boilers
Space cooling	None except air conditioning (AC) units in served rooms	AC units at different setpoints (in laboratories and offices)	AC units in some of the offices
Space ventilation	Three air handling units (AHU) serving variable air volume (VAV) boxes	One AHU serving 11 close-control spaces with monitored temperature and relative humidity	One AHU serving two meeting rooms
UL overnight use	Computers are equipped with wake-on-LAN systems	Estimated 30 % of overnight UL use which can be mitigated, excluding 24/7 UL equipment	
Energy performance			
Energy usage driver	Occupancy-driven	Process-driven	Occupancy-driven
Display Energy Certificate (DEC) or Energy Performance Certificate (EPC) score	C	F	B
2023 electricity consumption	52.9 kWh/m^2	170.5 kWh/m^2	47.5 kWh/m^2
2023 gas consumption	8.4 kWh/m^2	252.7 kWh/m^2	59.5 kWh/m^2
BMS functionalities	Elaborate	Basic	No BMS
Identified control or equipment failures	Photocell sensors, VAV boxes, and shading panels	Hot water loops operating at higher temperatures – recorded average temperatures of 20 °C during winter	Oversizing of boilers –recorded average temperatures of 21 °C during winter
Building input data			
Building data resolution and availability	High	Moderate	Low
Electricity data resolution	Hourly	Hourly	Hourly
Gas data resolution	Monthly	Monthly	Hourly
Energy data disaggregation level	High	Moderate	Low
Metered energy services	Lighting, UL, IT equipment, ventilation, cooling, electric heating, total gas, total electricity	CHP production, heating controls, ventilation, general power, total gas, total electricity	Standby load, total gas, total electricity

LOD 1 model. All of PHARM's OAT models pass calibration. Modelling simplifications for this CS building were explored as part of a previous study, which highlighted that process-driven buildings like PHARM can be more robust to schedule-related simplifications compared to occupancy-driven buildings [47]. However, these types of buildings tend to feature complex HVAC systems with limited simplification potential. Based on previous findings, no further plant HVAC simplifications are made to ensure successful calibration. For HH, models failing validation are IP 1 (Fabric) at both LOD 2 and LOD 1. As the smallest building, HH has minimal margin for error when simplifying parameters. Its relatively small envelope indicates the importance of accurate U-value knowledge, which cannot be averaged (LOD 2) or based on regulatory standards (LOD 1). The over-simplified occupancy schedule (IP 2 – LOD 1) also fails calibration. Like AWB, HH is occupancy-driven and may require more accurate occupancy modelling. Finally, the IP 11 – LOD 1 model assumes no AHU modelling since it represents a minor part of the HVAC system serving only two zones. However, this causes a non-negligible reduction in simulated electricity consumption, invalidating the IP 11 – LOD 1 model.

3.3. OAT sensitivity analysis

The absolute sensitivity results for simulated electricity and gas outputs of the OAT models are shown in Fig. 9. An IP category may show high sensitivity yet still pass calibration, as sensitivity merely indicates how parameter changes affect simulated energy outputs.

AWB's electricity consumption is highly sensitive to IP 2, IP 4, and IP 6, reflecting the significant impact of occupancy, lighting, and UL in this occupancy-driven building. While AWB also shows high gas consumption sensitivities, its gas consumption per unit area is minimal since it primarily uses electric heating. Consequently, changes to building fabric, occupancy schedules, UL, or HVAC equipment may have amplified effects on relative gas usage. PHARM shows high electricity sensitivity for IP 6, likely due to diverse equipment operating schedules, but generally displays the lowest overall sensitivities among the three CS buildings. This suggests process-driven buildings may be more robust to changes, particularly those involving less predictable factors like occupancy. HH exhibits high sensitivities across numerous IPs for both electricity and gas, likely due to its compact size making outputs more responsive to minor changes. Its electricity consumption is particularly sensitive to IP 2, IP 4, and IP 8, suggesting strong dependence on occupant activity and individual HVAC zone equipment. For gas consumption, HH shows notably high sensitivity to IP 1, highlighting potential limited tolerance for fabric modifications in smaller buildings. The sensitivity analysis reveals that IP sensitivity to simulated energy outputs appears largely influenced by building size and usage patterns, emphasising the importance of considering individual building characteristics when developing modelling simplification strategies.

3.4. Final LOD 2 and LOD 1 models calibration verification

The CV(RMSE) and NMBE values across the three LODs are presented in Fig. 10. CV(RMSE) values increase as LOD decreases, indicating reduced accuracy with simpler models. The CV(RMSE) increase tends to be more pronounced between LOD 3 and LOD 2 than between LOD 2 and LOD 1, suggesting a non-linear relationship in modelling complexity across the three LOD stages. HH consistently shows the highest CV (RMSE) values, further highlighting the challenges in calibrating models for smaller facilities. Interestingly, NMBE values demonstrate varying trends, suggesting that model bias is not directly related to simplification level. Despite these variations, all models across all LODs successfully meet the ASHRAE Guideline 14 calibration thresholds.

Fig. 11 shows the average daily indoor temperatures for various space types across all three CS buildings and LODs providing validation of HVAC system simulation accuracy. When compared against a suggested thermal comfort temperature range, all LOD models maintain

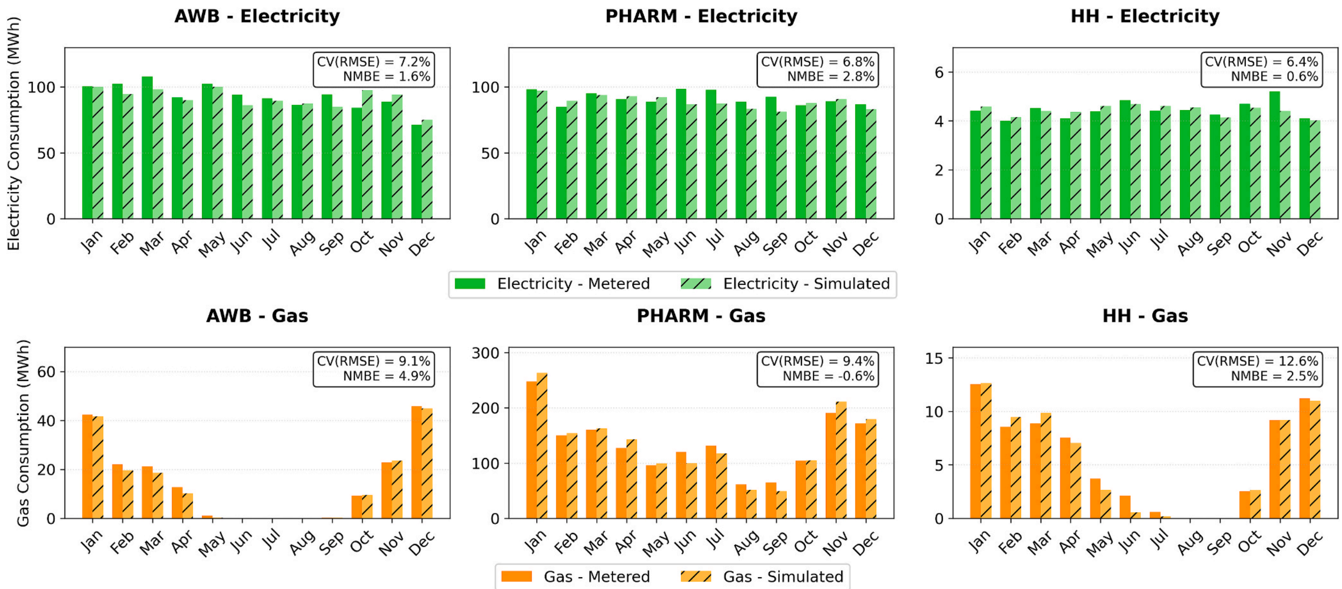

Fig. 8. Monthly metered and simulated electricity and gas consumption and resulting CV(RMSE) and NMBE for each LOD 3 model.

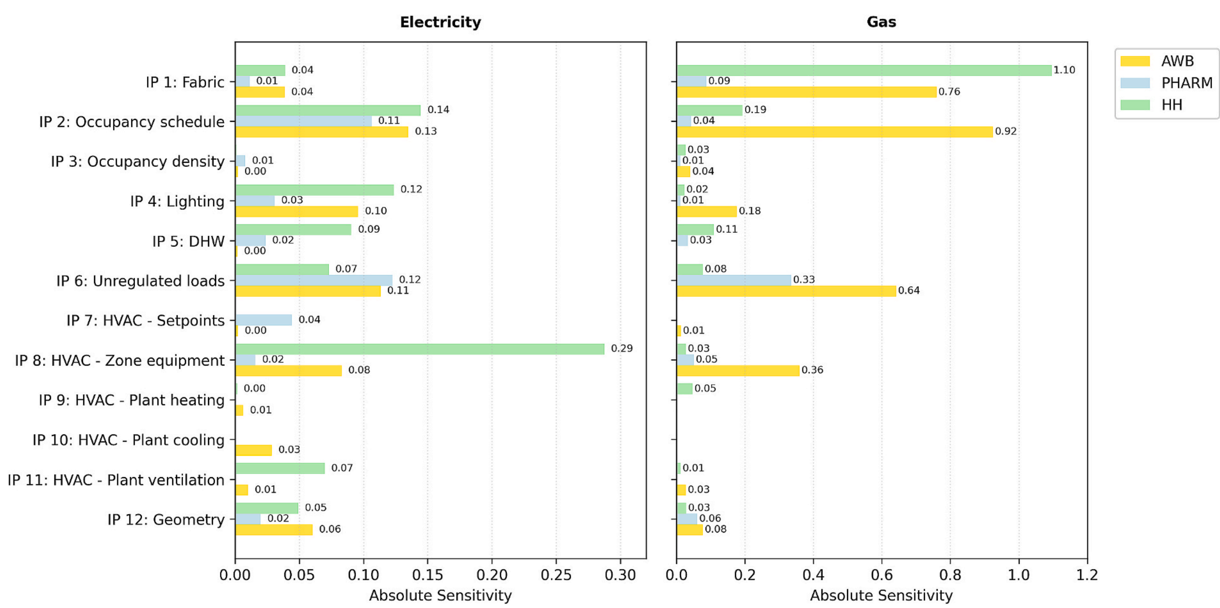

Fig. 9. Absolute sensitivities of each OAT model on simulated electricity and gas consumption.

temperatures within acceptable limits, with one exception: AWB's LOD 2 common areas during summer, due to the absence of cooling systems in the building.

3.5. Balancing accuracy and modelling complexity

The modelling, data collection, and simulation execution times alongside memory usage for each LOD model are summarised in Fig. 12. While increasing levels of simplification tends to lead to decreased modelling and simulation times and reduced memory usage, these reductions are not uniform across CS buildings and LODs. The efficiency gains from simplification vary based on individual building characteristics.

Fig. 13 plots the GOF for each LOD against the total modelling and simulation times on a logarithmic scale. On average, increasing time invested in modelling and simulation generally leads to decreased GOFs and thus improved accuracy. However, this trend shows diminishing returns, particularly for electricity consumption. Furthermore, while electricity data points closely align with their predicted trend, gas consumption shows more scattered distribution, highlighting the greater uncertainty in simulating gas outputs compared to electricity.

Fig. 14 plots the average GOF and total simulation and modelling time versus floor area for each LOD and CS building. Larger buildings tend to achieve lower GOFs, suggesting better predictive accuracy due to aggregation effects. However, this relationship is not linear, as demonstrated by PHARM, which despite being smaller than AWB, achieves lower average GOFs for LOD 2 and 3. Similar to the relationship between accuracy and modelling time, the correlation between accuracy and building size shows diminishing returns, particularly for LOD 3 and 2. This suggests that accuracy may not necessarily increase with floor area, likely due to variations in building usage patterns. It is also interesting to observe that the difference between HH LOD 2 and LOD 1 GOFs is relatively small compared to the much larger gap between LOD 3 and LOD 2. This may suggest that smaller buildings experience a significant

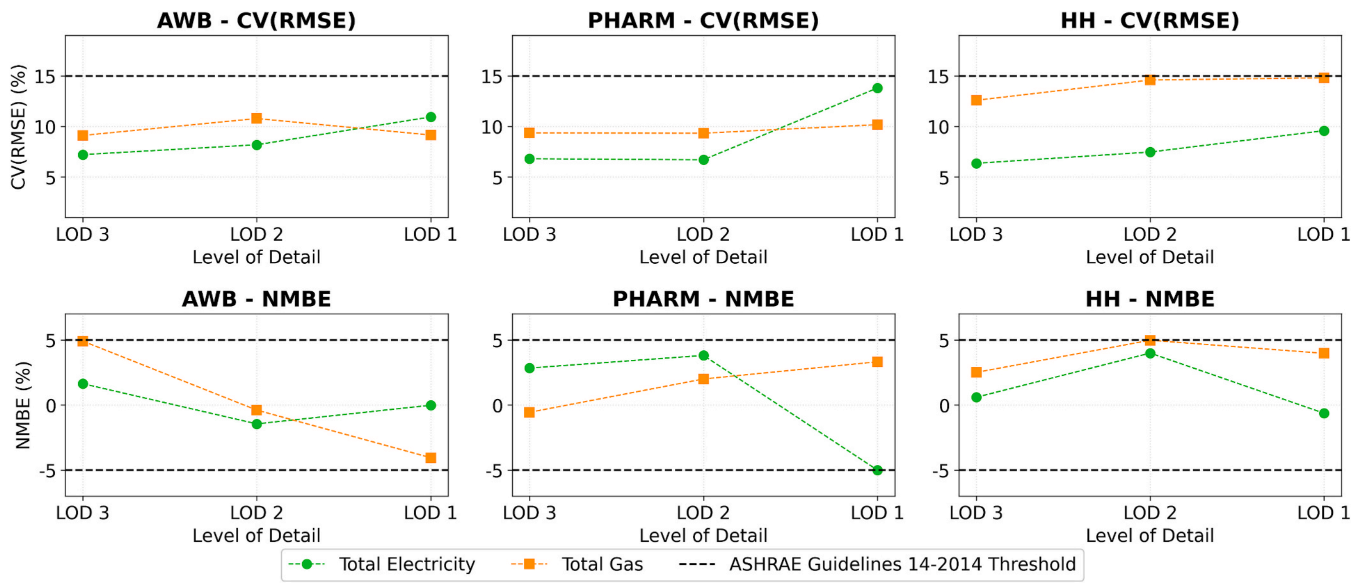

Fig. 10. CV(RMSE) and NMBE for each LOD model.

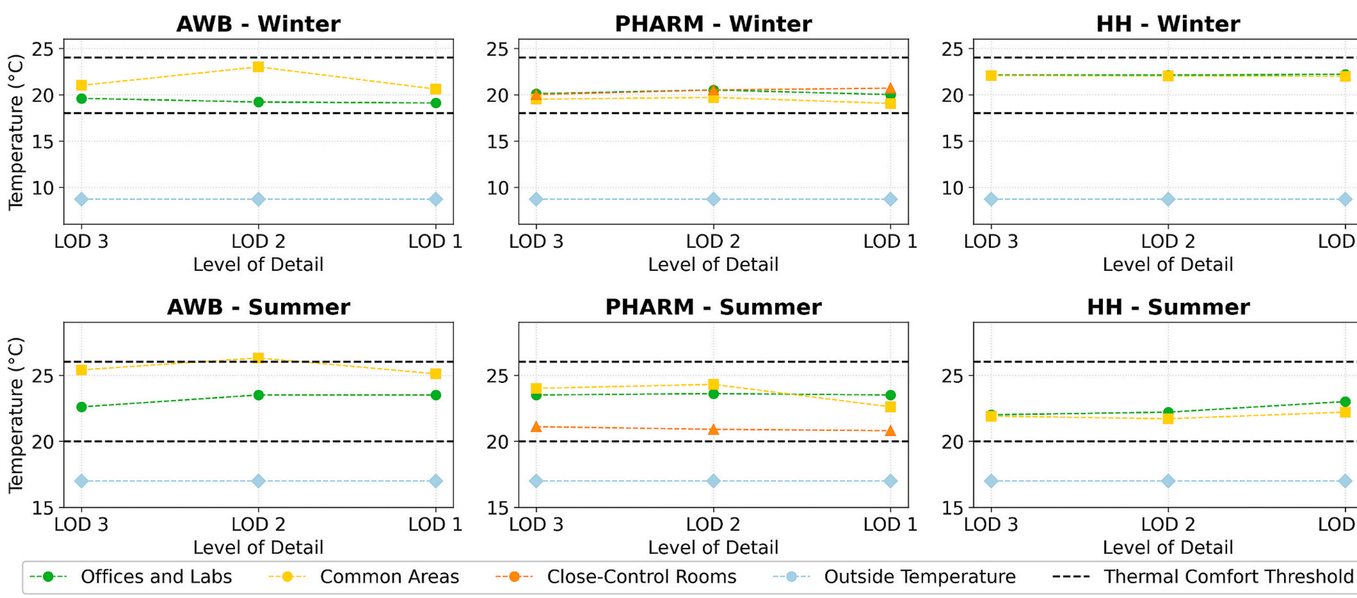

Fig. 11. Winter (top) and summer (bottom) average indoor temperatures for each LOD model.

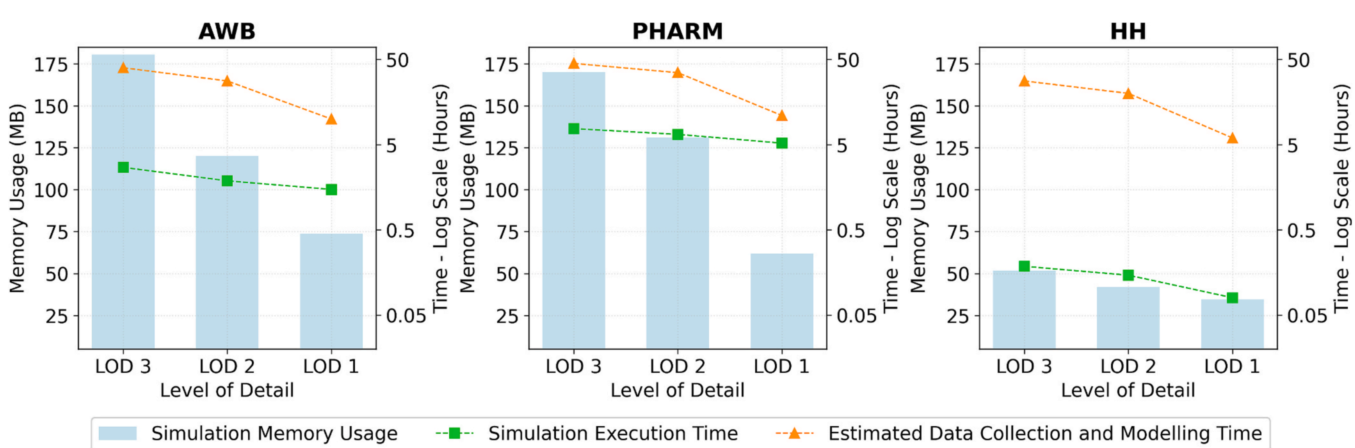

Fig. 12. Simulation memory usage, data collection, modelling and execution time for each LOD model.

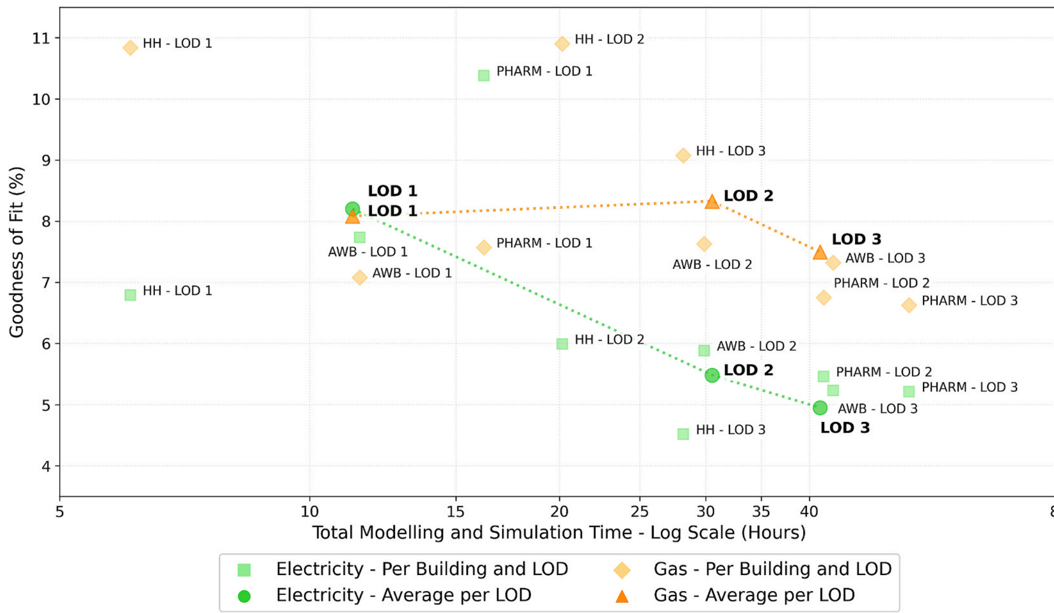

Fig. 13. GOF versus total modelling and simulation time for each LOD model.

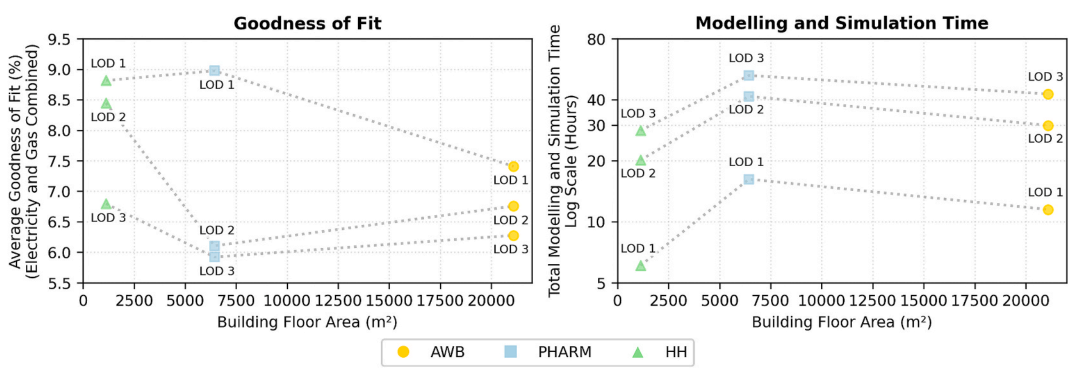

Fig. 14. Average GOF (left) and total modelling and simulation time (right) versus building floor area for each LOD model.

accuracy loss even with moderate modelling simplifications. On the other hand, PHARM shows the opposite pattern, with minimal difference between LOD 3 and LOD 2 GOFs but a larger gap between LOD 2 and 1, demonstrating its potential resilience to moderate modelling simplifications. AWB exhibits a more uniform distribution of GOF differences across all three LODs. Finally, while larger buildings tend to require more modelling and simulation time, this relationship is not linear either. For example, PHARM, which despite being smaller than AWB, required more modelling and simulation time, due to its complex HVAC systems and laboratory operations, in contrast to AWB's more repetitive zone structure and equipment configurations.

3.6. ECM model saving results and associated uncertainty

The total annual relative energy savings for each modelled ECM are presented in Table B1 in Appendix B. Fig. 15 plots absolute energy savings for ECMs yielding annual savings above 1 %, with associated absolute uncertainties. Gas savings from the heat decarbonisation ECM are plotted separately due to their significantly larger magnitude.

AWB's energy services are primarily control-based, with results suggesting benefits from more regular periodic maintenance. With electricity use largely driven by IT equipment, significant energy savings come from shifting to off-site cloud computing. Window replacement significantly reduces gas usage, likely due to the building's large window-to-wall ratio. Overall, AWB shows limited savings, reflecting its more modern, control-based operations compared to PHARM and HH. PHARM benefits mainly from HVAC- and fabric-related ECMs, reflecting its older construction. Due to malfunctioning BMS controls, heat decarbonisation would not only eliminate gas consumption but also significantly reduce electricity use. HH benefits from diverse ECMs including lighting, off-site computing, HVAC, and fabric improvements. Building age appears to significantly impact ECM effectiveness, with older buildings like HH and PHARM showing greater saving potentials across diverse categories.

Uncertainty ranges for ECM savings increase as LOD decreases, indicating lower result fidelity in simpler models. Despite increased uncertainty, the relative ranking of ECM effectiveness remains fairly consistent across LODs, particularly for deep retrofit and renewable measures, likely because these wholesale system changes are less sensitive to precise modelling of existing systems. However, some outliers remain. For HVAC- and fabric-related measures, LOD 2 and LOD 1 saving results tend to exceed LOD 3 results. As models simplify, thermal zone aggregation through zone-typing may mask localised variations in zone-level HVAC and fabric definition, suggesting higher predicted savings due to reduced spatial resolution. These discrepancies require careful consideration when selecting appropriate LOD models for ECM analysis. Additionally, larger absolute savings correlate with larger uncertainties; this balance between model accuracy and ECM reliability is

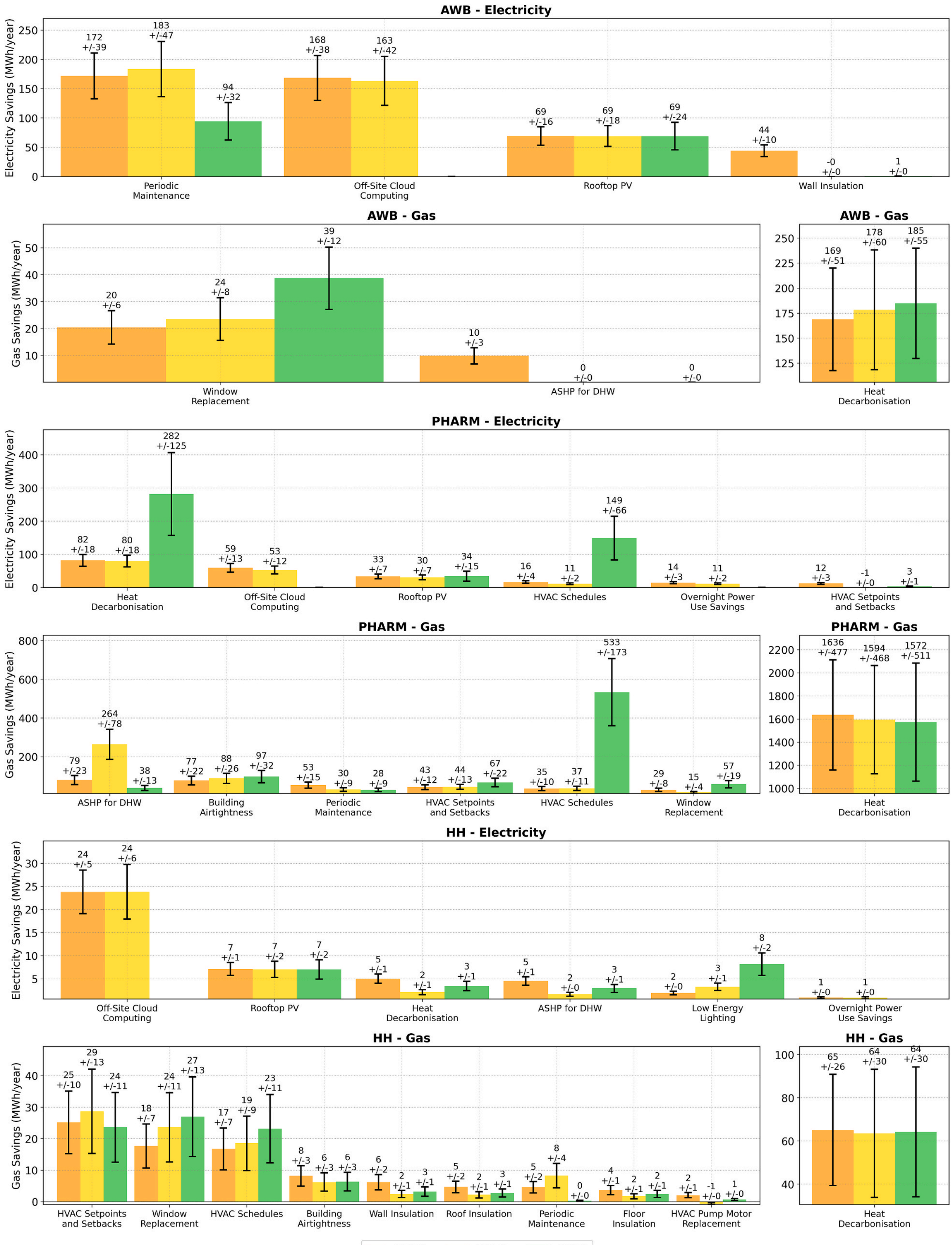

Fig. 15. Total electricity and gas savings of ECMs leading to more than 1% savings for each LOD model.

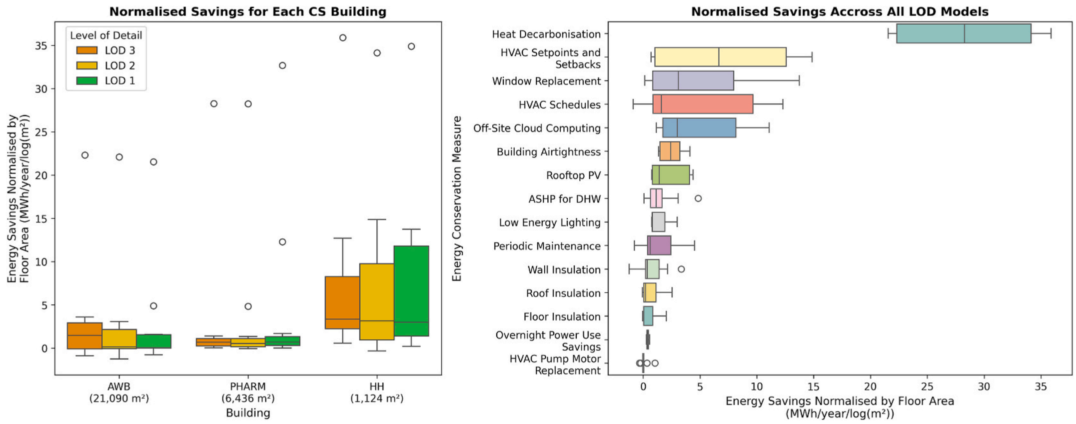

Fig. 16. Box plot of total normalised savings per floor area for each CS building (left) and average ECM normalised savings across all LOD models (right).

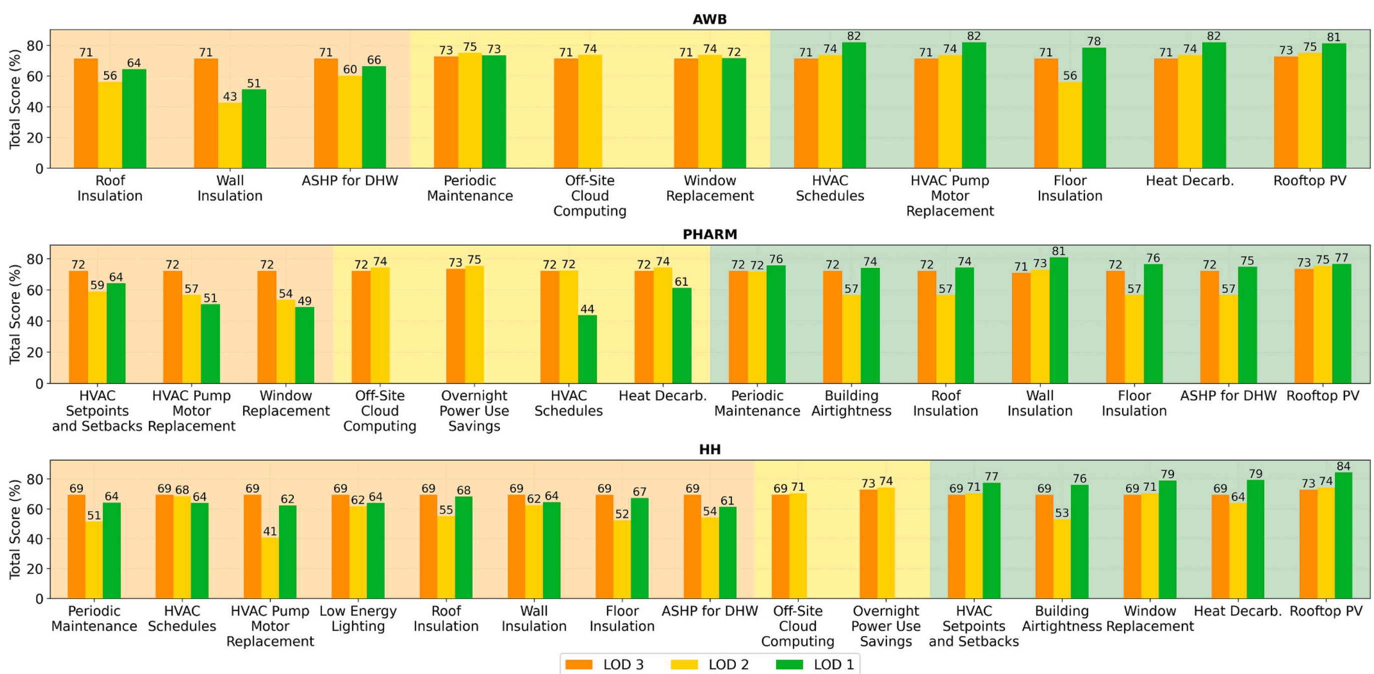

Fig. 17. Total score per ECM and LOD for each CS building ($\boldsymbol{w_T} = 30\%, \boldsymbol{w_R} = 70\%$).

further explored in Section 3.7.

Fig. 16 shows the box plots of total saving potentials for each CS building and LOD, and energy saving potentials per ECM averaged across all models. Savings are normalised by floor area. HH demonstrates the highest energy saving potential but also the greatest variability per floor area, suggesting higher sensitivity to interventions, likely due to its small size and occupancy-driven nature. AWB shows similar patterns. PHARM's tight interquartile range indicates more predictable and robust ECM saving results across LODs, again highlighting the resilience of process-driven buildings to modelling simplifications. However, PHARM shows the lowest normalised saving potentials per floor area, suggesting its energy use depends heavily on hard-to-reduce equipment demands given the building's lab-based nature. In terms of average energy savings per ECM, heat decarbonisation demonstrates the highest saving potential per unit area. HVAC controls, window retrofitting, and off-site cloud computing also yield significant normalised savings, though with greater variability in predicted results. The remaining measures show more modest saving potential but with smaller variability ranges, indicating more certain predicted outcomes.

3.7. Balancing accuracy, modelling complexity, and ECM modelling reliability

The total scores per ECM, LOD, and CS building, derived from the scoring methodology presented in Section 2.6, are presented in Fig. 17. The ECMs with an orange background are the ones for which LOD 3 has the highest score compared to LOD 2 and 1. Similarly, the ECMs with a yellow and green background have LOD 2 and LOD 1 ranked the highest, respectively. PHARM and AWB benefit from the most modelling simplifications all the while maintaining ECM results fidelity for most ECMs except fabric measures for AWB and HVAC measures for PHARM. HH seems to be more limited in the extent to which its model can be simplified while maintaining ECM modelling reliability. Most ECMs have the highest score for LOD 3, in particular HVAC- and fabric-related measures, similar to AWB and PHARM. Across all buildings, deep retrofit and renewables measures consistently score the highest for LOD 1, demonstrating great opportunity to leverage the use simplified BES models to virtually test major large-scale interventions.

4. Discussion

This study addresses the key relationships between building characteristics, modelling complexity, and ECM reliability through a detailed investigation of three diverse CS buildings. While the limited sample size necessitates caution when generalising findings, the analysis reveals valuable patterns suggesting BES processes may be simplified while maintaining accuracy and confidence for ECM modelling. However, it also indicates that simplifications and modelling approaches need to be carefully tailored based on building size, type and intended analysis purpose. The following sections propose preliminary guidelines based on observations from the three CS buildings, recognising that further validation across a broader building sample would strengthen these insights.

4.1. Building size and modelling complexity

The results highlight a non-linear relationship between building size and the accuracy of difference levels of modelling detail. Smaller buildings ($< 2{,}000$ m^2) tend to require higher modelling details to maintain accuracy, as highlighted by their higher resulting GOFs across all LODs (e.g. HH). This increased sensitivity to simplifications may be attributed to the proportionally larger impact of individual zones and building systems on overall energy use. In comparison, larger buildings benefit from averaging effects allowing for greater simplification without compromising accuracy. This is particularly evident for AWB (21,090 m^2), where zone-typing and greater systems aggregation maintained acceptable calibration accuracy. However, this relationship shows diminishing returns, suggesting an optimal range for simplification strategies.

4.2. Building type and modelling complexity

Simplification strategies are not only guided by building size but also by building type. Occupancy-driven buildings such as AWB and HH, although very different in size, both show larger sensitivities to schedule-related simplifications, suggesting the need for higher model detailing of occupancy patterns, internal gains, and occupancy-dependent systems. This heightened sensitivity affects the variability of ECM effectiveness to a larger extent. In contrast, process-driven buildings such as PHARM maintain better accuracy with schedule-related simplifications and show an overall more consistent ECM performance across all LODs. However, these buildings may require a higher level of modelling detail of specialised systems. These findings suggest that building type may be as important a factor as building size when determining appropriate simplification strategies.

4.3. ECM modelling reliability

The analysis of ECM modelling reliability reveals a relationship between simplification potential and the type of intervention under study. Deep retrofit and renewables measures are reliably tested on LOD 1 models and show consistent results across all LODs, suggesting a more robust savings predictions even with increased modelling uncertainty. On the other hand, control- and schedule-based ECMs tend to require higher modelling fidelity, especially for smaller buildings. Finally, fabric- and HVAC-related improvements demand detailed modelling regardless of building size or type, with a significant variation in ECM savings results across LODs, highlighting the significance of thermal dynamics for BES accuracy and utility. These findings provide valuable guidance for BES users in selecting appropriate modelling strategies based on the purpose of the BES.

4.4. Practical implications for BES implementation

Overall, this study shows that modelling simplifications should be based both on building characteristics and analysis objectives. The confidence of predicted ECM results can be assessed with uncertainty management, which is directly derived from the BES model predictive accuracy. Uncertainty results can inform decision-making processes when developing net-zero building strategies. By proposing a novel scoring system balancing modelling time and ECM reliability, this study also highlights the importance of maintaining detailed modelling for certain specification applications while enabling greater modelling simplifications for initial assessments and major or large-scale interventions.

5. Conclusion

This study aims to advance understanding of Building Energy Simulation (BES) processes for wider adoption and improved energy conservation measure (ECM) effectiveness for policy development. It investigates the balances between modelling complexity and accuracy, and the impact of modelling simplifications on ECM reliability. A modelling simplification methodological framework is developed and is aimed to be applicable across various building types, ages, and sizes. Three different levels of modelling detail (LODs) are developed and derived by simplifying one building input parameter at a time in two stages and evaluating the impact on calibration accuracy using the CV (RMSE) and NMBE metrics. A sensitivity analysis is performed to understand the influence of simplifications on simulated energy outputs. Time and modelling effort are also investigated to identify the trade-offs between modelling time and accuracy. Various ECMs are modelled on each LOD model and simulated savings and uncertainty associated to ECM results are quantified, to assess the reliability of predictive ECM savings across the different LODs. Finally, a scoring system is developed to guide BES users in identifying the adequate balance between modelling time and effort, and ECM reliability. The methodological framework is verified against three case study (CS) buildings from the University of Oxford.

This study highlights that simplification strategies should be tailored to both building characteristics and analysis objectives. Smaller buildings may require higher fidelity models while larger buildings allow for more extensive simplifications. Additionally, contrary to occupancy-driven buildings, process-driven buildings tend to allow for greater occupancy-based simplifications. The methodology also demonstrates that deep-retrofit and renewable measures can be reliably tested with simplified models, while control, HVAC, and fabric ECMs may require more detailed modelling due to the intricate thermal dynamics involved in buildings energy use. By providing a systematic framework and quantitative scoring metrics that balance modelling effort against reliability, this study offers practical guidelines for implementing simplified BES approaches that maintain confidence in results while reducing resource requirements.

There are key areas for future research which emerge from this study. First, expanding the proposed methodology to a larger and more diverse building sample would strengthen the generalisability of the presented findings. Automated simplification protocols for different building archetypes could then be developed using machine learning to optimise simplification processes. Identified strategies could also be compared across different climate zones. Additionally, ECM effectiveness and reliability could be investigated under different future climate scenarios to ensure the resilience of simplified models against changing conditions. Finally, key findings from this study can be put into the further development of practical tools and guidelines, including its integration with building performance standards to ensure current methodologies involved in the Energy Performance Certificate (EPC) assessment process for example account for the particularities of different building characteristics.

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

During the preparation of this work, the lead author used Claude AI in order to improve the readability and language of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Data availability

The data that support the findings of this study are openly available online at https://data.mendeley.com/datasets/ysk4trkgpt/1.

CRediT authorship contribution statement

Laurence Peinturier: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **David C.H. Wallom:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This work was supported by the John McCall MacBain Graduate Scholarship.

The authors would like to thank the University of Oxford Estate Services for their collaboration. We thank Professor Frances Platt, Head of the Department of Pharmacology, and Lukas Hofer, the department's Building and Safety Manager, for their valuable support. Special thanks to Nick Bees from CPW for providing U-value measurement equipment and conducting on-site testing at the Department of Pharmacology. We are grateful to Dr. Keith A. Gillow from the Andrew Wiles Building Professional Services Staff and Hugh Darnell, Manager of the Andrew Wiles Building Facilities and Services, for their assistance. We also thank Laing O'Rourke for providing the initial AWB architectural 3D model. Our appreciation goes to Professor Jesus Lizana and Dr. Scot Wheeler from the Department of Engineering Science for their help gathering input data for Holywell House. Finally, we wish to thank Dr. Phil Grunewald for his continuous support throughout this research.

Appendix A. . Input parameter changes for sensitivity analysis

Table A1
Input parameters and values for each IP category, CS building and LOD – data used to compute ICs.

		AWB			PHARM			HH		
	LOD	***3***	***2***	***1***	***3***	***2***	***1***	***3***	***2***	***1***
IP category	***Description***									
IP 1: Fabric	Area-weighted envelope U-value (W/m^{2}K)	2.96	2.90	0.80	0.98	0.89	0.80	0.99	1.10	2.18
IP 2: Occupancy schedule	Average occupied hours (hours/week)	25.4	26.7	29.5	31.7	38.0	45.0	23.5	36.9	45.0
IP 3: Occupancy density	Average occupancy density (people/m^2)	0.10	0.07	0.20	0.12	0.11	0.13	0.13	0.07	0.10
IP 4: Lighting	Average lighting power density (W/m^2)	8.1	8.9	7.5	14.0	9.5	10.0	6.2	4.3	7.0
IP 5: DHW	Number of DHW loops	5	1	1	2	2	2	1	1	1
	Average daily litres per area (L/day/m^2)	0.02	0.02	0.20	4.00	5.00	2.10	0.26	0.26	0.20
IP 6: UL	Average peak power density (W/m^2)	18.0	16.0	25.0	56.0	51.4	55.0	13.5	11.3	20.0
IP 7: HVAC – Setpoints	Total zone area with special indoor thermal conditioning (m^2)	6,310	6,310	0	168.4	86.0	67.2	0	0	0
IP 8: HVAC – Zone equipment	Average radiator capacity (W/m^2-heated area)	82.7	51.0	47.4	95.3	113.2	99.7	170.0	105.7	85.6
	Average AC capacity (W/m^2-cooled area)	N/A	N/A	N/A	205.9	136.1	76.1	65.0	63.4	88.5
	Average ventilation air flow rate (m^3/s/m^2-ventilated area)	0.002	0.003	0.005	N/A	N/A	N/A	N/A	N/A	N/A
IP 9: HVAC – Plant heating	Number of modelled plant heating assets (boiler, CHP, GSHP)	7	7	2	4	4	4	2	2	1
IP 10: HVAC – Plant cooling	Number of modelled plant cooling assets (chillers, GSHP)	3	3	1	1	1	1	0	0	0
IP 11: HVAC – Plant ventilation	Number of modelled plant ventilation assets (AHUs)	3	3	1	1	1	1	1	1	0
	Total plant ventilation air flow rate (m^3/s/m^2-ventilated area)	N/A	N/A	N/A	N/A	N/A	N/A	0.005	0.005	0.000
IP 12: Geometry	Total number of modelled zones	443	197	78	329	151	87	37	21	11

Appendix B. . Relative energy savings for each ECM model

Table B1
Total relative energy savings for each ECM modelled on each LOD.

	AWB			PHARM			HH		
LOD	***3***	***2***	***1***	***3***	***2***	***1***	***3***	***2***	***1***
Energy Conservation Measure									
Periodic maintenance	13.6 %	12.3 %	5.6 %	1.6 %	1.0 %	1.0 %	4.0 %	7.6 %	0.3 %
Off-site cloud computing	13.0 %	12.2 %	N/A	1.9 %	1.6 %	N/A	14.0 %	11.7 %	N/A
Overnight power use	N/A	N/A	N/A	0.5 %	0.4 %	N/A	0.8 %	0.5 %	N/A
HVAC setpoints and setbacks	N/A	N/A	N/A	2.0 %	1.6 %	2.6 %	21.4 %	25.1 %	20.2 %
HVAC schedules	0.0 %	0.0 %	0.8 %	1.9 %	1.8 %	25.0 %	14.4 %	16.3 %	19.6 %
HVAC pump motor replacement	0.0 %	0.0 %	0.1 %	0.1 %	0.0 %	0.0 %	1.8 %	−0.5 %	0.6 %
Low energy lighting	N/A	N/A	N/A	N/A	N/A	N/A	0.9 %	0.7 %	3.6 %
Building airtightness	N/A	N/A	N/A	3.1 %	3.2 %	3.7 %	6.9 %	5.3 %	5.4 %
Window replacement	1.5 %	2.0 %	3.2 %	1.1 %	0.4 %	1.9 %	13.7 %	19.4 %	22.8 %
Roof insulation	0.0 %	0.0 %	0.1 %	0.4 %	0.4 %	0.6 %	4.2 %	1.9 %	2.3 %
Wall insulation	3.5 %	−0.7 %	0.0 %	0.5 %	0.5 %	0.7 %	5.6 %	2.3 %	3.4 %
Floor insulation	−0.3 %	0.1 %	0.0 %	0.1 %	0.0 %	0.1 %	3.4 %	1.4 %	2.1 %
ASHP for DHW	1.3 %	0.4 %	0.3 %	2.7 %	10.7 %	1.4 %	4.3 %	1.5 %	2.3 %
Heat decarbonisation	10.3 %	9.8 %	8.4 %	63.5 %	63.2 %	68.0 %	59.4 %	57.2 %	57.4 %
Rooftop PV	5.5 %	5.3 %	5.3 %	1.2 %	1.1 %	1.2 %	6.0 %	5.9 %	5.6 %

References

[1] IPCC, 'Climate Change 2022: Mitigation of Climate Change: Summary for Policymakers'. 2022.

[2] H. Ritchie, P. Rosado, and M. Roser, 'CO_2 and Greenhouse Gas Emissions', *Our World Data*, Dec. 2023, [Online]. Available: https://ourworldindata.org/co2-and-greenhouse-gas-emissions.

[3] IEA, 'Buildings'. 2023. [Online]. Available: https://www.iea.org/topics/buildings.

[4] C. Maduta, G. Melica, D. D'Agostino, P. Bertoldi, Towards a decarbonised building stock by 2050: the meaning and the role of zero emission buildings (ZEBs) in Europe, *Energy Strategy Rev.*, 44 (Nov. 2022) 101009, https://doi.org/10.1016/j.esr.2022.101009.

[5] 'Total floor area by use in the Net Zero Scenario, 2010-2030 – Charts – Data & Statistics', IEA. [Online]. Available: https://www.iea.org/data-and-statistics/charts/total-floor-area-by-use-in-the-net-zero-scenario-2010-2030.

[6] IPCC, 'IPCC Sixth Assessment Report'. [Online]. Available: https://www.ipcc.ch/report/ar6/wg3/chapter/chapter-9/.

[7] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, *Energy Build.*, 40 (3) (Jan. 2008) 394–398, https://doi.org/10.1016/j.enbuild.2007.03.007.

[8] T. Hong, J. Langevin, and K. Sun, 'Building simulation: Ten challenges', *Building Simulation*, vol. 11, no. 5. Tsinghua University Press, pp. 871–898, Oct. 2018. doi: 10.1007/s12273-018-0444-x.

[9] Y. Lu, P. Li, Y.P. Lee, X. Song, An integrated decision-making framework for existing building retrofits based on energy simulation and cost-benefit analysis, *J. Build. Eng.*, 43 (Nov. 2021), https://doi.org/10.1016/j.jobe.2021.103200.

[10] D. Kong, et al., Evaluation of the impact of input-data resolution on building-energy simulation Accuracy and computational load—A case study of a low-rise office building, *Buildings* 13 (4) (Mar. 2023) 861, https://doi.org/10.3390/buildings13040861.

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

[12] Y. Li, Z. O'Neill, L. Zhang, J. Chen, P. Im, and J. DeGraw, 'Grey-box modeling and application for building energy simulations - A critical review', *Renewable and Sustainable Energy Reviews*, vol. 146. Elsevier Ltd, Aug. 2021. doi: 10.1016/j.rser.2021.111174.

[13] S. Cheng, Z.D. Tekler, H. Jia, W. Li, A. Chong, Evaluating different levels of information on the calibration of building energy simulation models, *Build. Simul.*, (Feb. 2024), https://doi.org/10.1007/s12273-024-1115-8.

[14] K.-U. Ahn, Y.-J. Kim, C.-S. Park, I. Kim, K. Lee, BIM interface for full vs. semi-automated building energy simulation, *Energy Build.*, 68 (Jan. 2014) 671–678, https://doi.org/10.1016/j.enbuild.2013.08.063.

[15] M. Shin, J.S. Haberl, Thermal zoning for building HVAC design and energy simulation: a literature review, *Energy Build.*, 203 (Nov. 2019) 109429, https://doi.org/10.1016/j.enbuild.2019.109429.

[16] P. Raftery, M. Keane, J. O'Donnell, Calibrating whole building energy models: an evidence-based methodology, *Energy Build.*, 43 (9) (Sep. 2011) 2356–2364, https://doi.org/10.1016/j.enbuild.2011.05.020.

[17] 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 (Sep. 2017) 277–282, https://doi.org/10.1016/j.egypro.2017.07.322.

[18] F. Khayatian, A. Bollinger, and P. Heer, 'Temporal Resolution of Measurements and the Effects on Calibrating Building Energy Models', Mar. 08, 2022, *arXiv*: arXiv: 2011.08974. doi: 10.48550/arXiv.2011.08974.

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

[20] G. Agugiaro, S. Hauer, and F. Nadler, 'Coupling of CityGML-based Semantic City Models with Energy Simulation Tools: some Experiences', 2015. Accessed: Feb. 23, 2025. [Online]. Available: https://www.semanticscholar.org/paper/Coupling-of-CityGML-based-Semantic-City-Models-with-Agugiaro-Hauer/85293ec56faa9c5d007fc66722cd13510747d2bd.

[21] D. Bishop, M. Mohkam, B. L. M. Williams, W. Wu, and L. Bellamy, 'The Impact of Building Level of Detail Modelling Strategies: Insights into Building and Urban Energy Modelling', *Eng*, vol. 5, no. 3, Art. no. 3, Sep. 2024, doi: 10.3390/eng5030118.

[22] S. Elhadad, C. H. Radha, I. Kistelegdi, B. Baranyai, and J. Gyergyák, 'Model Simplification on Energy and Comfort Simulation Analysis for Residential Building Design in Hot and Arid Climate', *Energies*, vol. 13, no. 8, Art. no. 8, Jan. 2020, doi: 10.3390/en13081876.

[23] A. Chong, Y. Gu, H. Jia, Calibrating building energy simulation models: a review of the basics to guide future work, *Energy Build.*, 253 (Dec. 2021) 111533, https://doi.org/10.1016/j.enbuild.2021.111533.

[24] F. Randaxhe, V. Lemort, S. Bertagnolio, F. Randaxhe, V. Lemort, Evidence-based calibration of a building energy simulation model: application to an office building in Belgium. In *International Conference for Enhanced Building Operations*, 2012.

[25] M. Derakhti, W. O'Brien, S. Bucking, Impact of measured data frequency on commercial building energy model calibration for retrofit analysis, *Sci. Technol. Built Environ.*, 28 (5) (May 2022) 628–644, https://doi.org/10.1080/23744731.2021.1991177.

[26] F.S. Westphal, R. Lamberts, 'building Simulation Calibration Using Sensitivity Analysis' (2005).

[27] K. Menberg, Y. Heo, R. Choudhary, Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information, *Energy Build.*, 133 (Dec. 2016) 433–445, https://doi.org/10.1016/j.enbuild.2016.10.005.

[28] F. Ascione, N. Bianco, R.F.D. Masi, G.M. Mauro, G.P. Vanoli, Energy retrofit of educational buildings: transient energy simulations, model calibration and multi-objective optimization towards nearly zero-energy performance, *Energy Build.*, 144 (Jun. 2017) 303–319, https://doi.org/10.1016/j.enbuild.2017.03.056.

[29] I. El-Darwish, M. Gomaa, Retrofitting strategy for building envelopes to achieve energy efficiency, *Alex. Eng. J.*, 56 (4) (Dec. 2017) 579–589, https://doi.org/10.1016/j.aej.2017.05.011.

[30] S. Fathi, A. Mahravan, Mitigating energy consumption of educational buildings using a novel simulation method (case study: Faculty of oil and Petrochemical Engineering, Razi University, Iran), *J. Renew. Energy Environ.*, 10 (3) (Aug. 2023) 13–28, https://doi.org/10.30501/jree.2022.349282.1394.

[31] M. Pazouki, K. Rezaie, A. Bozorgi-Amiri, A fuzzy robust multi-objective optimization model for building energy retrofit considering utility function: A university building case study, *Energy Build.*, 241 (Jun. 2021), https://doi.org/10.1016/j.enbuild.2021.110933.

[32] S.J. Guo, T. Wei, Cost-effective energy saving measures based on BIM technology: case study at National Taiwan University, *Energy Build.*, 127 (Sep. 2016) 433–441, https://doi.org/10.1016/j.enbuild.2016.06.015.

[33] T. Ekström, S. Burke, M. Wiktorsson, S. Hassanie, L.-E. Harderup, J. Arfvidsson, Evaluating the impact of data quality on the accuracy of the predicted energy performance for a fixed building design using probabilistic energy performance simulations and uncertainty analysis, *Energy Build.*, 249 (Oct. 2021) 111205, https://doi.org/10.1016/j.enbuild.2021.111205.

[34] Y. Heo, R. Choudhary, G.A. Augenbroe, Calibration of building energy models for retrofit analysis under uncertainty, *Energy Build.*, 47 (Apr. 2012) 550–560, https://doi.org/10.1016/j.enbuild.2011.12.029.

[35] C.J. Hopfe, J.L.M. Hensen, Uncertainty analysis in building performance simulation for design support, *Energy Build.*, 43 (10) (2011) 2798–2805, https://doi.org/10.1016/j.enbuild.2011.06.034.
[36] E. Prataviera, J. Vivian, G. Lombardo, A. Zarrella, Evaluation of the impact of input uncertainty on urban building energy simulations using uncertainty and sensitivity analysis, *Appl. Energy* 311 (Apr. 2022) 118691, https://doi.org/10.1016/j.apenergy.2022.118691.
[37] Y. Sun, P. Huang, G. Huang, A multi-criteria system design optimization for net zero energy buildings under uncertainties, *Energy Build.*, 97 (Jun. 2015) 196–204, https://doi.org/10.1016/j.enbuild.2015.04.008.
[38] N. Hashempour, R. Taherkhani, M. Mahdikhani, Energy performance optimization of existing buildings: a literature review, *Sustain. Cities Soc.*, 54 (Mar. 2020), https://doi.org/10.1016/j.scs.2019.101967.
[39] B.R.K. Mantha, C.C. Menassa, V.R. Kamat, A taxonomy of data types and data collection methods for building energy monitoring and performance simulation, *Adv. Build. Energy Res.*, 10 (2) (Jul. 2016) 263–293, https://doi.org/10.1080/17512549.2015.1103665.
[40] EnergyPlus, 'EnergyPlus'. 2023. [Online]. Available: https://energyplus.net.
[41] 'DesignBuilder Software Ltd'. [Online]. Available: https://designbuilder.co.uk/.
[42] ASHRAE, 'Measurement of Energy, Demand, and Water Savings - ASHRAE Guidelines 14-2014'. American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), 2014. [Online]. Available: www.ashrae.org.
[43] H. Yoshino, T. Hong, N. Nord, IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods, *Energy Build.*, 152 (Oct. 2017) 124–136, https://doi.org/10.1016/j.enbuild.2017.07.038.
[44] S. Dabirian, K. Panchabikesan, U. Eicker, Occupant-centric urban building energy modeling: approaches, inputs, and data sources - a review, *Energy Build.*, 257 (Feb. 2022) 111809, https://doi.org/10.1016/j.enbuild.2021.111809.
[45] Oikolab, 'Weather and Climate Data API for Analysts'. 2023. [Online]. Available: https://oikolab.com.
[46] C.Y. Siu, Z. Liao, Is building energy simulation based on TMY representative: a comparative simulation study on doe reference buildings in Toronto with typical year and historical year type weather files, *Energy Build.*, 211 (Mar. 2020), https://doi.org/10.1016/j.enbuild.2020.109760.
[47] L. Peinturier, D. Wallom, 'From complexity to scalability: simplifying building energy models for wider adoption', presented at the uSIM2024 -shaping net zero policies with building simulation, International Building Performance Simulation Association (2024).
[48] ASHRAE, Standard 90.1 - Energy Standard for Sites and Buildings Except Low-Rise Residential Buildings. ASHRAE, 2022.
[49] UKGBC, 'Building the case for net zero: Retrofitting office buildings', UK Green Building Council, 2024. [Online]. Available: https://ukgbc.org/wp-content/uploads/2024/01/Retrofitting-Office-Buildings-Building-the-Case-for-Net-Zero.pdf.
[50] UKGBC, 'Delivering Net Zero: Key Considerations for Commercial Retrofits', UK Green Building Council, 2022. [Online]. Available: https://ukgbc.org/wp-content/uploads/2022/05/Commercial-Retrofit-Report.pdf.
[51] O. e-Research Centre, 'Low carbon ICT'. 2013. [Online]. Available: http://www.escience2009.org/projects/low-carbon-ict/.
[52] O.T. Masoso, L.J. Grobler, The dark side of occupants' behaviour on building energy use, *Energy Build.*, 42 (2) (Feb. 2010) 173–177, https://doi.org/10.1016/j.enbuild.2009.08.009.
[53] L. Peinturier, D. Wallom, 'Energy consumption in higher education buildings: bridging the performance gap using digital twin technology',presented at the ECEEE 2024 Summer study on energy efficiency. European Council for an Energy Efficient Economy, 2024.
[54] J. Lizana, et al., Integrated post-occupancy evaluation and intervention that achieve real-world zero-carbon buildings, *Energy Build.*, 303 (Jan. 2024) 113766, https://doi.org/10.1016/j.enbuild.2023.113766.
[55] NCM, 'U-Value for Old Construction (1978-2002)'. 2024.
[56] CIBSE, CIBSE Guide F - Energy efficiency in buildings, Third edition. London: CIBSE, 2012.
[57] ASHRAE, *2007 ASHRAE Handbook - Heating, Ventilating, and Air-Conditioning Applications (I-P Edition)*. American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), 2007.