Applied Energy 298 (2021) 117174

Contents lists available at ScienceDirect

Applied Energy

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

Feature assessment frameworks to evaluate reduced-order grey-box building energy models

Mohammad Haris Shamsi [a,*], Usman Ali [a,c], Eleni Mangina [b], James O'Donnell [a]

[a] School of Mechanical and Materials Engineering, UCD Energy Institute, University College Dublin, Belfield, Dublin 4, Ireland
[b] School of Computer Science and UCD Energy Institute, UCD, Dublin, Ireland
[c] Sustainable Authority of Ireland, Dublin 2, Ireland

ARTICLE INFO

Keywords:
Energy modelling
Building performance simulation
Grey-box models
Scalability
BEPS
Flexibility
Interoperability

ABSTRACT

With a drive towards achieving an integrated energy system, there is a need for holistic and scalable building modelling approaches for the commercial building stock. Existing grey-box modelling approaches often fail to produce a generalised network structure, which limits the suitability of models for different applications. Furthermore, existing feature assessment frameworks provide limited opportunities to quantify the potential of model characteristics in terms of flexibility, scalability and interoperability. Considering the diversity of the possible characterisation approaches, this study aims to define and assess a set of basic and derived features for reduced-order grey-box models through a generalisable framework that would act as a decision support tool for the identification of appropriate model characteristics. This research proposes an integrated methodology to test and evaluate model features, namely, scalability, flexibility, and interoperability for reduced-order grey-box models and formulates test-cases with the available commercial reference buildings published by the Department of Energy of the United States. The model scalability errors lie between 3.42% and 4.35% that indicates the suitability of implementing a zone level model for model predictions at the whole building level. The model flexibility error decreased from 5.73% to 4.78% when considering a trade-off between accuracy and complexity. These frameworks produce scalable and flexible models that facilitate urban energy modelling of building stocks and subsequent evaluation of retrofit strategies. Furthermore, the devised models aid the implementation of heat demand reduction scenarios in a building cluster to achieve an integrated energy system.

1. Introduction

The built environment accounts for over one-third of global final energy consumption and nearly 40% of total direct and indirect CO_2 emissions [1]. The associated energy demand continues to rise, mainly driven by improved access to energy in developing countries, greater ownership and use of energy-consuming devices, and rapid growth of building floor area. Energy-related CO_2 emissions from buildings have experienced a gradual increase in recent years after flattening between 2013 and 2016 [2]. Direct and indirect emissions from electricity and commercial heat used in buildings rose to an all time high of 10 $GtCO_2$ in 2019 [1]. Several factors, including growing energy demand for heating and cooling with rising air-conditioner ownership and extreme weather events, have been instrumental in this rise [3].

Buildings often involve complex interconnected systems that depend on several dynamic factors, for instance, weather and occupancy. Furthermore, building system optimisation can require balancing of contradictory objectives based on energy efficiency and overall performance [4]. Building Energy Performance Simulation (BEPS) tools offer a suitable platform to conduct performance-based analysis and optimisation that takes into account the complex model inter-dependencies, internal and external inputs as well as the associated performance objectives. BEPS tools have been extensively used over different stages of the building life cycle, ranging from design to post-construction operation stages [5]. While these tools take into account the multitude of complex inter dependencies, the generated models are often non-scalable and non-generalisable [6]. Moreover, model development using BEPS tools is resource intensive; models require geometric and non-geometric building data. This would be impractical when dealing with individual buildings on a large scale [7]. Moreover, the developed models would be computationally expensive as each individual building differs in terms of structural parameters and the nature of operation [8].

* Corresponding author.
E-mail address: mohammad.shamsi@ucdconnect.ie (M.H. Shamsi).

https://doi.org/10.1016/j.apenergy.2021.117174
Received 2 March 2021; Received in revised form 22 April 2021; Accepted 23 May 2021
Available online 3 June 2021

List of Abbreviations

ANOVA	ANalayis Of VAriance
BEPS	Building Energy Performance Simulation
CTSM	Continuous Time Stochastic modelling
CVRMSE	Coefficient of Variation Root Mean Square Error
DOE	Department of Energy
GOF	Goodness Of Fit
HVAC	Heating, Ventilation and Air Conditioning
IWEC	International Weather Energy Calculation
KPI	Key Performance Indicator
MAPE	Mean Absolute Percentage Error
MLE	Maximum Likelihood Estimation
NMBE	Normalised Mean Bias Error
PHDR	Peak Heat Demand Reduction
RC	Resistance Capacitance

List of Symbols

A_e	Effective window area in m^2
C_e	External heat capacity in kWh/°C
C_i	Internal heat capacity in kWh/°C
P_h	Radiator flux in kW
P_s	Solar radiation flux in W/m^2
R^2	Coefficient of Determination
R_{ea}	Thermal resistance between building envelope and ambience in °C/kW
R_{ie}	Thermal resistance between internal environment and building envelope in °C/kW
T_a	Ambient temperature in °C
T_e	External temperature state in °C
T_i	Internal temperature state in °C

It is of paramount importance to identify and develop modelling approaches that are scalable and offer flexibility in model design. Such approaches would facilitate the evaluation of energy efficiency measures at individual or urban building level. Furthermore, the modelling approach should offer opportunities for evaluating energy optimisation scenarios when considering control or smart grid applications. Grey-box modelling is one such approach that represents the actual behaviour of the system and delivers the advantages of data-driven and physical modelling approaches. The grey-box approach inherits the model description of physical modelling approaches and the computational efficiency of data-driven modelling approaches [9].

While grey-box building energy models have been widely implemented, the applicability of these models has often been specific to particular applications and stakeholders [10]. Furthermore, scalability of these models is limited by the network order that defines the level of complexity incorporated in the devised model [6]. Reduced-order grey-box models can counter these drawbacks through achieving a trade-off between the network order and the desired accuracy [11]. There is a need for a generalisable framework that systematically defines and evaluates the potential of the grey-box model features for various applications. Such framework would produce scalable and flexible models that address the challenges posed by individual, dynamic simulation models.

One of the most significant features when modelling a building stock is the scalability of the devised model, which relates to a model's ability to represent extended system variables, for instance, when modelling urban building stocks [12]. As the size of the problem can significantly grow depending on the analysed system, it becomes crucial to handle the growing amount of data while retaining the required characteristics of different subsystems. A limited number of studies exist in literature that focus on defining the scalability potential of a particular energy modelling approach (Table E.1). Heo, Augenbroe, Graziano, Muehleisen and Guzowski devised a scalable methodology for large scale building improvement using normative energy models at the individual and aggregate building levels [13]. Although the devised methodology is scalable, the research does not account for the scalability potential of normative energy models. Another recent study by Manfren and Nastasi devised an integrated methodology to validate and monitor the building energy performance of a residential building. The authors linked parametric performance analysis to model calibration using inverse modelling. However, the study fails to account for the scalability of the analysis technique and the possibility to scale models from an individual building to building clusters for large scale performance assessment [14]. Heidarinejad, Mattise, Dahlhausen, Sharma, Benne, Macumber et al. formulated urban-scale reduced-order building energy models using highly influential thermal variables in a white-box modelling process [15]. Although the results provide an essential overview of scalability, none of these studies exclusively devise an assessment framework to measure scalability of the employed modelling approach, which is crucial to devise urban building stock models.

Another significant model feature is the flexibility associated with the modelling approach, which relates to a model's ability to incorporate modified design variables with as less effort as possible, for instance, when integrating grey-box networks with district heating networks [16]. This feature allows the energy modeller to introduce perturbations in the original model to determine how the energy and cost performance are affected by various energy efficiency measures. Previous studies devised several tool sets to perform optimisation, sensitivity analysis, and uncertainty analysis for evaluation of the optimal energy efficiency solution (Table E.1). These tool sets often require numerous simulations and huge computational resources [17]. A comparison usually overlooked in these studies involves the flexibility potential of the deployed modelling approach, which aims to define a generic model that is flexible to design variations [18]. For instance, reduced-order grey-box energy models provide more flexibility (compared to black-box models) in optimising the building operation when implementing different design scenarios. Bourdeau, Zhai, Nefzaoui, Guo and Chatellier evaluated the flexibility of data-driven techniques for modelling and forecasting building energy consumption [19]. Another recent publication dealt with the flexibility of building energy modelling approaches for energy performance prediction [20]. Both these studies concluded that hybrid modelling approaches (grey-box models) provide significantly more flexibility than data-driven modelling yet a comprehensive framework to assess the model flexibility still remains absent.

The building stock remains noticeably disconnected in terms of energy transactions due to the existing differences in individual building characteristics and their nature of operation [21]. Advanced integrated approaches are required to model the transaction of energy services (e.g., prosumers) between different buildings in the stock [22]. This could facilitate intelligent trade-offs between comfort/quality of service and consumption [23]. The interoperability feature defines the integration process (Table E.1) and is pertinent to buildings rather than a particular modelling approach. Reduced-order grey-box models facilitate the evaluation of the interoperability potential through the use of grey-box parameters (thermal resistance and thermal capacitance). As opposed to white-box and black-box modelling approaches, grey-box model parameters directly define the heat storage characteristics of the building, which determine the thermal energy transfer potential of the building cluster. Razmara, Bharati, Shahbakhti, Paudyal and Robinett proposed a bi-level optimisation framework to transact energy between commercial buildings and distribution grids [24]. The optimisation framework involved a dynamic model for buildings and operational models for distribution grids. However, the study did not account for the interoperability potential of each individual building. Beil, Hiskens and Backhaus discussed the role of buildings for serving the grid through demand response and ancillary services [25]. The study dealt with the interaction between the individual building and the grid, which could be extended to account for any building to building interactions. Previous research mostly accounts for interactions between the building cluster and the power grid; building to building interactions are often not taken into account. Building to building interactions are essential to achieve an integrated energy system and thereby, enhance the overall efficiency of the system [26]. To address these challenges, there is a need for an interoperability framework that provides guidelines to decide whether or not a particular building will be suited for system integration scenarios.

Owing to the above arguments, this study aims to address following significant gaps in knowledge associated with building energy modelling through the implementation of reduced-order grey-box energy models:

1. The first gap relates to model scalability issues as buildings differ in their individual characteristics and the implemented modelling approach becomes irreproducible for other buildings. White-box and black-box modelling approaches provide limited scalability when modelling individual buildings or urban building stocks. On the other hand, grey-box models eventually retain their original structure and facilitate the development of scalable models.
2. The second gap concerns model flexibility issues that address the building model operation or execution after the introduction of any perturbations. Furthermore, previous flexibility assessment frameworks usually focus on either white-box or black-box approaches. While white-box models are complex and computationally inefficient, black-box models require extensive historical data and perform poorly when the input data are outside the bounds of the training set. Grey-box models facilitate an integrated implementation to represent a building using physical relationships and statistical data treatment.
3. The third and final gap focuses on the lack of an integrated modelling approach for the optimised operation of a building cluster (interoperability) while addressing the underlying modelling approach. Although previous studies have addressed these gaps individually in brief, a consolidated framework to construct an overall understanding is still absent. Grey-box models facilitate swift integration of thermal network models with other energy systems, for instance, district heating networks.

This research proposes an integrated methodology to test and evaluate model features, namely, scalability, flexibility and interoperability for reduced-order grey-box models. These features are instrumental in the development of an appropriate model as per a defined problem. The approach introduces novelty in defining, testing and evaluating the model features to produce scalable and flexible models. When considering the scalability feature, this study proposes a zonal approach to propagate the building level dynamics at various modelling scales. The approach assesses zone model scalability relative to existing multiple zones. For flexibility, the proposed approach assesses the effect of design perturbations on the grey-box modelling approach. For interoperability, the study proposes a framework to evaluate how significantly individual buildings can participate and enhance the interoperability between buildings.

This integrated approach formulates assessment frameworks to examine the respective set of features. Furthermore, the methodology addresses the generalisation issues of grey-box networks. The paper establishes the significance of grey-box model features through data analytics and experimental simulations. As the devised approach is generalisable, the application could be extended to any test case scenario irrespective of country, building type or available data. The assessment of model features enhances the effectiveness of a particular energy model that represents the optimal fit for a certain purpose or situation.

The paper consists of the following sections: Section 2 describes the devised overarching methodology to test and evaluate model features. Section 3 details sub-methodologies that outline individual assessment frameworks for different features and introduces different case studies to implement the devised methodology. Section 4 lists the interpretation of the results and the associated limitations. Section 5 describes the conclusion and future work.

2. Overarching methodology for feature assessment

The devised methodology proposes a generalisable assessment framework to evaluate reduced-order grey-box model features and thereby, identify a balance between complexity and accuracy for various applications (Fig. 1). The framework deals with an integrated analysis of different domains in defining and evaluating the model features such as scalability, flexibility and interoperability. One of the significant features of the framework is the use of heat demand patterns as a pre-assessment criterion, which forms a part of the “Data analysis” procedure in the “Grey-box model development” process. It has been well established that individual building dynamics closely relate to the heat demand patterns for any building stock [27]. Furthermore, these patterns depend on the local climate, heat transfer through the building envelope, daily operations and the occupancy schedules.

The assessment framework proposes a five step process to define the scalability, flexibility and interoperability potentials of reduced-order grey-box models as outlined below.

1. Model feature definition identifies all the plausible definitions of respective features using previous literature, technical reports, expert views and model requirements.
2. Data collection identifies and collects the required data for grey-box model development and model feature testing processes.
3. Grey-box model formulation develops the grey-box network to assess the respective feature.
4. Model feature testing identifies the required set of experiments to assess individual features.
5. The Key Performance Indicator (KPI) identification associates the performance measures to each test feature, which indirectly measures the potential of the respective feature.

2.1. Model feature definition

This procedure defines the derived features to be tested under the devised assessment framework. Model feature definitions are extracted from previous studies, technical reports, expert views and model requirements [28].

2.2. Data collection

Data collection refers to a standardised procedure for collecting data pertinent to building stock modelling. Although this study identifies a set of variables to carry out the assessment task, it is worthwhile to note that the variable set might change depending on the user and the type of the devised grey-box model. Furthermore, any additional variables will only enrich the grey-box model with further information. The devised approach categorises the variables into two sets, namely, mandatory and optional variables. The mandatory variables comprise weather variables, building site information, building physical parameters and building operation variables. The optional variables include building renovation history and Heating, Ventilation and Air Conditioning (HVAC) system information. While mandatory variables decide the initial order (complexity) of the grey-box network, the optional variables enrich the dynamics of the formulated grey-box structure. Further details about the variable sets are outlined in Table 1.

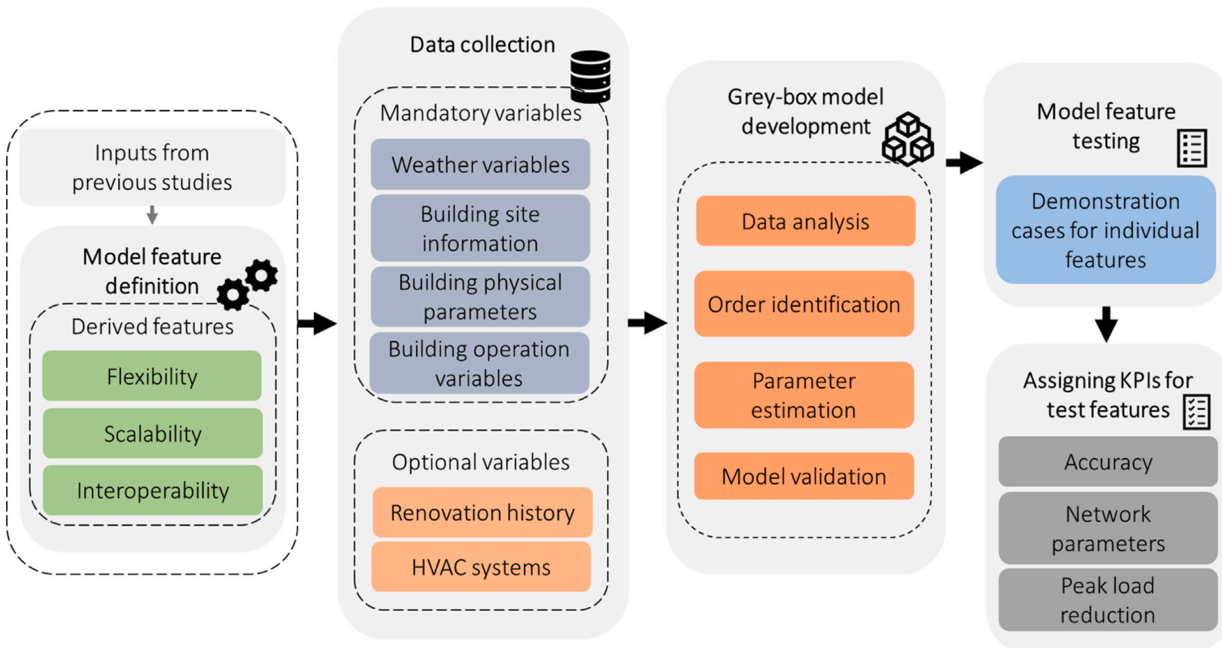

Fig. 1. Overarching feature assessment methodology to test model features of reduced-order grey-box models.

Table 1
Data variable classification as mandatory and optional variables required for grey-box development and the feature assessment procedure. These variable sets could be acquired through measurements using databases such as TABULA that consist of the building stock information at the urban level [29]. Another source of data collection could be reference buildings such as U.S. Department of Energy (DOE) commercial reference buildings [30].

Mandatory Variables			
Variable Type	Variable	Units	Source
Weather	Ambient Temperature Global Solar Radiation Wind Speed	°C kWh/m^2 m/s	IWEC File
Building Site Information	Location No. of Buildings Building Type	- - -	
Building Physical Parameters	No. of Floors Interior Floor Area Window Area No. of Zones	- m^2 m^2 -	Design Plans
Building Operation Variables	Heat Demand Profiles Internal Temperature Profiles Building Space Use	kW °C %	Measurement Sensors Design Plans
Optional Variables			
Renovation History	Fabric Renovation	-	Retrofit Reports
HVAC System	Embedded HVAC System Ventilation System	- -	Design Plans

2.2.1. Mandatory variables

The mandatory variable set comprise the following variables that are instrumental in the identification of the basic structure of any grey-box model.

Weather Variables: These variables mainly constitute ambient temperature, global solar radiation and wind speed. Weather variables mainly act as inputs to the grey-box model and do not play any role in grey-box model identification. Weather information is easily available for any major location in the world in the form of International Weather Energy Calculation (IWEC) [31].

Building Site Information: The site information mainly includes the location, number of buildings in the neighbourhood and building type. The number of buildings is only relevant when implementing energy management scenarios in an interconnected building cluster. Although these variables do not affect the model formulation process, these variables aid the identification of the required complexity in representing individual buildings or a cluster of buildings. For instance, when buildings are connected via a district heating network, similar types of buildings could be represented through a grey-box network with similar complexities.

Building Physical Parameters: These variables consist of gross interior floor area, number of floors, window to wall ratio (WWR), and number of zones.

1. The gross interior floor area (m^2) is defined as the total interior floor area of a building's space, measured from the inside surface of the exterior walls or from the interior surface of walls in common with adjoining buildings. The floor area, number of floors and heat demand profiles are used to validate the network order.

2. Number of floors is used together with the floor area to model the heat flow dynamics inside any building. For instance, a high rise office building might consume more energy although each floor represents a similar demand usage profile with limited complexity [27].
3. The WWR (%) parameter influences the network order from two aspects. When windows are a part of the building envelope, the network complexity reduces as a result of reduction in the number of capacitance in the network. Furthermore, a large number of windows inside any building reduces the heterogeneity in temperature profiles, thus, reducing the complexity in the grey box network. This parameter represents an approximate value of the window to wall ratio existing in the building envelope and hence, the methodology does not use the exact values in any calculations.
4. Number of zones provide an approximation of existing zones to facilitate the process of network order identification. The spaces inside a commercial building are often divided into zones for optimised utilisation of HVAC systems. A single zone represents an area possessing similar thermal characteristics; the knowledge of this parameter simplifies the development of grey-box models.

Building Operation Variables: These variables form the most crucial variable set and constitute building heat demand patterns, internal temperature profiles and building space use.

1. Building heat demand (kW) and indoor temperature profiles (°C) usually reflect the ongoing activities inside any building. These profiles, defined on an hourly timestamp, capture the link between demand patterns and network order.
2. Building space use (%) specifies the individual proportions of spaces used for various functions (for instance, offices, storage and toilets).

2.2.2. Optional variables

The optional variable set consists of the following variables and enriches the grey-box network with added information when available.

Renovation History: This variable lists all the past retrofits applied to the building and relates to the increased heterogeneity in the building envelope. As any fabric related renovation alters the building dynamics, this information could identify any required modifications in the structure of the formulated grey-box model. Simply changing the network parameter values might lead to unrealistic description of the grey box network.

HVAC System Information: This variable mainly constitutes of information regarding the existing ventilation strategies and the presence of embedded heating or cooling systems in the building fabric. These factors alter the building dynamics, thus, increasing the complexity associated with the grey-box model. While ventilation causes heat transfer due to the movement of conditioned air across different spaces, embedded HVAC systems alter the dynamics of the building by increasing the heat capacity associated with the area where the systems are installed.

2.3. Grey-box model development

This process identifies the structure of grey-box models (RC networks). A typical resistance capacitance (RC) network represents the dynamics of the building assuming a steady-state heat transfer across the building envelope. Such networks have seen wide implementation for a number of optimisation and control studies [32]. The RC networks use resistance parameters to represent the thermal resistance and capacitance parameters to represent the heat storing capacity of the building envelope. Grey-box model development comprises four sequential steps, namely, data analysis, order identification, parameter estimation and model validation (Fig. 2). These processes are outlined below and are further illustrated under each sub-methodology.

2.3.1. Data analysis

The data analysis procedure involves an initial statistical analysis of building heat demand patterns to describe any existing patterns in the respective profiles. This step is crucial in the generation of generalised reduced-order grey-box models and represents a novel process in the proposed methodology. As the demand profiles only give an indication of the network order, the profile variations need to be established using tests of statistical significance such as ANalayis Of VAriance (ANOVA). This technique could be used to infer whether means of two or more groups are significantly different from each other [33]. This technique compares the means of different samples to identify the impact of one or more factors. Furthermore, the ANOVA procedure could be enhanced through additional available information that reflects possible correlations with the factors under investigation. For this particular research, a cluster of buildings represents the different groups in ANOVA. Each sample constitutes one single day. ANOVA is formulated using the following set of equations.

$$Null\ Hypothesis\ H_o : \mu_1 = \mu_2 = \cdots = \mu_k \tag{1}$$

$$Alternate\ Hypothesis\ H_A : \mu_k \neq \mu_m \tag{2}$$

where $(\mu_1, \mu_2, \ldots \mu_k)$ represent individual group means, k represents the number of a group and μ_m represents the group mean when $(\mu_1 = \mu_2 .... = \mu_{k-1})$. When comparing different demand profiles, ANOVA would establish the existence of any variations. The alternate hypothesis suggests that there are at least two group means that are statistically significantly different from each other. ANOVA usually involves the calculation of F-test statistic that represents the ratio of two quantities expected to be roughly equal under the null hypothesis. Large F- test values signify the variability in group means is large compared to the within group variability, which is essential to reject the null hypothesis. F-test values are used in conjunction with p-values to establish the significance. For instance, A large F-test value with a small p-value (less than the defined significance levels) represents a rejected null hypothesis.

2.3.2. Order identification

The order identification procedure outlines the underlying structure of the grey-box network. This process uses a generalised methodology devised by the authors in a previous publication [27]. The methodology considers thermal zoning of spaces inside the building where each thermal zone possesses similar characteristics and can be represented using a similar RC network. The order identification procedure uses the results of ANOVA analysis from the data analysis procedure using the following set of rules.

1. Gather the ANOVA results from the data analysis procedure. Check for individual building heat demand patterns to identify whether a lumped parameter model is sufficient to represent the building dynamics.
2. Normalise the heat demand patterns using floor area. Identify buildings with relatively high values of normalised heat demand and assign separate RC network branches for exterior walls and interior mass.
3. Identify any recurring heat demand fluctuation in hourly profiles. If yes, assign a separate thermal zone, else end the identification procedure.

This process could be further enriched with available data regarding the past retrofits and existing HVAC systems. For instance, radiant heating or cooling systems alter the heat capacity of the floor, which might require another RC branch for appropriate representation of the building dynamics. Similarly, fabric renovation of only one side of the building might require a separate RC branch to represent the associated wall dynamics.

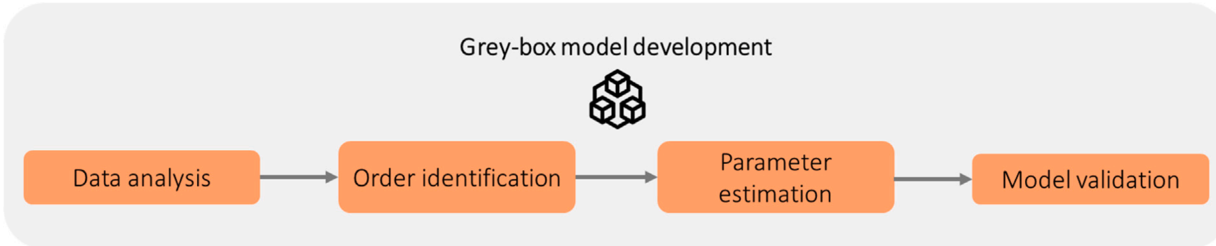

Fig. 2. Grey-box model development to identify the model network order for individual buildings.

2.3.3. Parameter estimation

Parameter estimation (or model calibration) procedure involves the estimation of network parameters, namely, thermal resistance and thermal capacitance and the associated time constants (Fig. 3). Although there exist several parameter estimation procedures, this study implements the parameter estimation procedure using Continuous Time Stochastic modelling (CTSM) in R programming language that uses the Maximum Likelihood Estimation (MLE) and automates the estimation procedure [34]. By using a continuous time formulation of the dynamics and discrete time measurements, the tool bridges the gap between physical and statistical modelling. It is possible to generate both pure simulation and k-step prediction estimates of the states and the outputs, filtered estimates of the states and, for nonlinear models, smoothed estimates of the states.

2.3.4. Model validation

This procedure acts as a measure of the goodness of fit for the model. This procedure further checks if a model satisfies the assumptions and provides reasonable estimates from a physical point of view. Error KPIs give an indication of the suitability of the model and aid in understanding the effects not properly described by the model. This study validates the model using KPIs identified in Section 2.5 to find errors in the prediction of internal temperature profiles and heat demand patterns as these form a fundamental part of the state equation.

2.4. Model feature testing

The model feature testing procedure identifies the required experimentation to test each individual model feature (scalability, flexibility and interoperability) on the basis of two crucial processes, namely, Analyse and Specify (Fig. 4). These processes identify sub-methodologies for feature testing which then combine with the overarching methodology to lay out a framework for feature assessment. This section first describes a general implementation of the feature testing procedure, which is then modified and structured for each individual feature. The Analyse and Specify processes follow a general set of guidelines as outlined in [35] for test case generation. These guidelines have been modified to account for various scenarios in the building simulation domain.

The Analyse process comprises a set of four sub-processes that aim to analyse the test feature under consideration (Fig. 4 and Table 2). The first sub-process includes the analysis of test feature that defines the model feature using previous literature (Section 2.1). The second sub-process involves the analysis of test subject, i.e. the building cluster to identify and associate the building parameters and activities. The third sub-process analyses the underlying modelling approach (grey-box networks) that represents the building dynamics. The last sub-process analyses the testing procedures that could be used to specify and quantify the model feature.

The Specify process comprises a set of four sub-processes that aim to specify the testing scenarios for buildings under consideration (Fig. 4 and Table 3). The first sub-process identifies a high level scenario and various sub-scenarios to be tested for the considered modelling approach. The second sub-process defines the input data required to carry out the testing scenarios (Section 2.2). The third sub-process lays out the sequence of testing activities for variations in the testing model. The last sub-process involves the validation of testing results against set standards, for instance, ASHRAE Guideline 14.

The feature testing process stands out in itself and is crucial in the identification of sub-methodologies for testing various features as highlighted in Section 3. Within the BEPS domain and on account of testing grey-box model features, the test guidelines provide insights into the implementation of experimental testing procedure for each feature. The individual feature assessment sub-methodologies indicate how the testing process could be modified to account for different types of features.

2.5. KPI identification

The KPI identification process is the last procedure of the overarching methodology and identifies a set of performance measures for each individual feature. These metrics assess the performance of the reduced-order grey-box modelling approach and further evaluate each feature in terms of the model suitability. KPIs are valuable not only to describe a specific characteristic but also facilitate a comparison with other model traits designed for similar scopes [36]. The majority of BEPS studies define KPIs for assessing building energy performance [37]. These KPIs often include energy consumed, CO_2 emissions, efficiency, energy cost and building comfort. Few studies define KPIs associated with the modelling approach that mainly include various definitions of accuracy for the energy calibration procedure [38].

This study implements the following list of KPIs that include a combination of the aforementioned categories (building energy performance KPIs and model calibration KPIs). The list of significant KPIs depends on the feature to be tested, which is highlighted under each feature assessment sub-methodology. It is worthwhile to mention that the following KPI list is selective and different combinations of KPIs could be used to measure the model feature under consideration.

1. Normalised Mean Bias Error (NMBE) metric represents the normalised average of errors of a sample space.
2. Coefficient of Variation Root Mean Square Error (CVRMSE) establishes the variability of the errors between measured and simulated values. As per the ASHRAE Guideline 14, CVRMSE values of less than 25% represents an acceptable model fit. ASHRAE Guideline 14 defines well-accepted criteria to measure the accuracy of energy simulations [39] (Table D.1).
3. Mean Absolute Percentage Error (MAPE) measures the absolute size of the error in terms of percentage. It represents the average of the absolute percentage errors of forecasts.
4. Coefficient of Determination or R^2 metric determines the closeness of the simulated values to the regression line of the measured values.
5. Goodness Of Fit (GOF) metric defines the overall fit of the devised model using a combination of other metrics, namely, CVRMSE and NMBE.

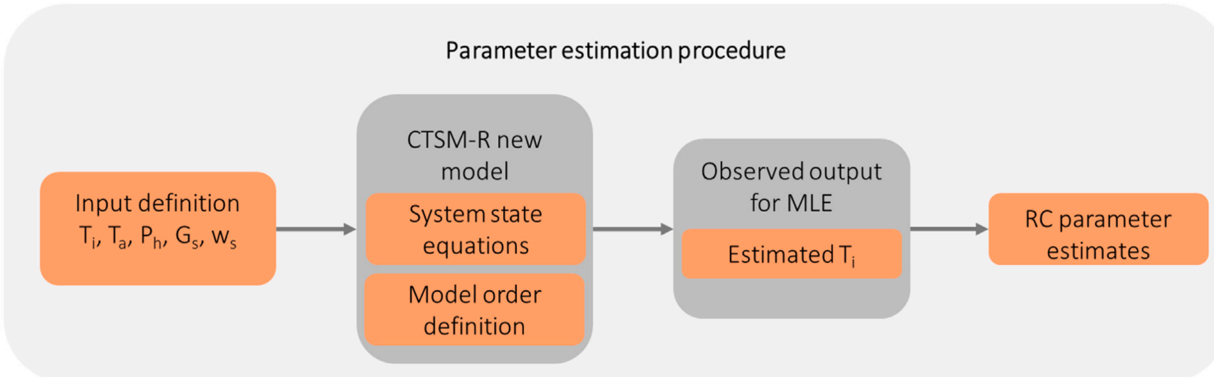

Fig. 3. Parameter estimation procedure using CTSM-R to identify the thermal resistance and thermal capacitance values.

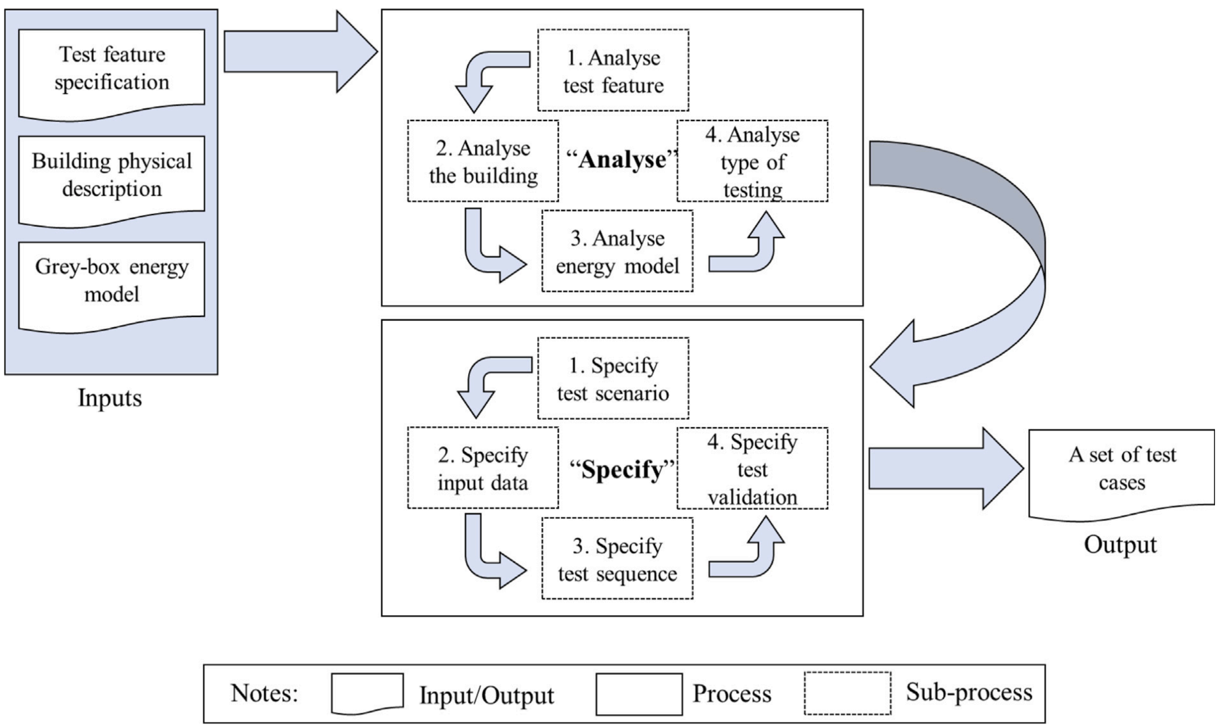

Fig. 4. Model feature testing implementation process to identify the required experimentation for individual features.

Table 2
Analyse process to identify the experiments for model feature testing.

Sub-process	Description	Input	Output
Analyse test feature	Model test feature definition to understand, verify and validate the requirements	Test feature definitions from literature	Understanding of requirements, constraints and an overview of how to test in general
Analyse the building	Building description to identify the associated parameters and building activities	Building related inputs (physical as well as operational)	Understanding of the dynamics to derive tests for the building
Analyse grey-box energy model	Grey-box model to represent the heat energy dynamics	Grey-box model	Understanding of testing strategy to assess the associated complexity
Analyse type of testing	Testing type analysis to specify and quantify the model feature	Requirement specification and grey-box energy model	Understanding the type of testing to prepare an appropriate design strategy

6. Thermal Resistance Capacitance (RC) parameters of a grey-box model define the building dynamics. While the thermal capacitance (C, J/K) value indicates the heat storing capacity of a building's mass, the thermal resistance (R, K/W) controls the response of a building when subjected to a temperature change [40]. These parameters are estimated using the state-space equations for the formulated RC networks as defined in Section 2.3.3.

A high thermal mass (C value) would result in a more stable environment, which will be resistant to changing temperatures. The multiplication of these two parameters results in time constant, which is calculated for the building as a whole. The higher the time constant, the longer it will take to alter the internal conditions irrespective of excitation source (sudden weather events or the building system).

Table 3
Specify process to identify the experiments for model feature testing.

Sub-process	Description	Input	Output
Specify test scenario	Specify a high level scenario and various sub-scenarios for testing	Requirement specification and grey-box energy model	A set of test scenarios
Specify input data	Specify different sets of input data to be used in testing (simulated and measured)	Requirement specification and building physical and operational inputs	Different sets of input data
Specify test sequence	Specify a sequence of testing activities for each scenario/variation in the energy model	Requirement specification, grey-box energy model and building physical and operational inputs	Different sets of test sequences
Specify validation	Specify baselines and boundary levels to validate results against set standards	A set of test scenarios	A complete set of test cases

7. Peak Heat Demand Reduction (PHDR) has been extensively used as a performance metric for demand alteration measures. PHDR metric represents the percentage decrease in the peak heat demand of a building or a group of buildings to provide required comfort levels. With the appropriate application of control strategies, the peak heat demand could be reduced to a high extent. This indicator is used to analyse the maximum heat demand of a system in comparison with the average heat demand [40].

3. Sub-methodologies for model feature assessment

The sub-methodologies section formulates individual assessment frameworks for the considered model features. The devised sub-methodologies follow a similar procedure to the overarching methodology while taking into account specific feature definitions, data requirements, generalised grey-box model generation, specific feature testing procedures and appropriate KPI assignments. A distinct characteristic of these sub-methodologies lies in the identification of the feature testing process using the generalised structure as illustrated in Section 2.4. Each sub-methodology further includes a case study formulated using the U.S DOE commercial reference buildings in order to enhance the comprehensibility of each framework. These reference buildings comprise 16 building types that represent approximately 70% of the commercial buildings in the United States across 16 locations, and correspond to all U.S. climate zones. These buildings are available as EnergyPlus models that are representative of realistic building characteristics and construction practices [30]. EnergyPlus is a whole building simulation program widely used to model the energy consumption and water use in buildings.

The DOE reference building EnergyPlus models provide the data required for grey-box model order identification and model parameterisation (Fig. 5). The grey-box model order identification procedure uses mandatory and optional variables, defined in Sections 2.2.1 and 2.2.2, which are extracted from the EnergyPlus input IDF files of DOE reference buildings. The IDF files contain the information regarding building physical parameters, building space usage and HVAC systems. Building operation variables are obtained through energy simulations using IDF files in EnergyPlus (v 7.2). The outputs of these simulations provide heat demand and internal temperature profiles of the reference building buildings. The network parameterisation procedure uses these variables to identify the parameters of the formulated grey-box model. In the case of the representative example of DOE reference buildings, the EnergyPlus simulation solely focuses on the generation of indoor temperature profiles and space head demand. When the real-time building monitoring data is available, the grey-box models could be directly calibrated and hence, these EnergyPlus models would no longer be required.

Each reference building corresponds to varied uses of the interior space (Fig. 6, Figs. A.1 and A.2). For instance, standalone retail DOE reference building comprises core retail space, front retail space and back storage where core retail space represents 69.8% of the total space use inside the buildings (Fig. 6(a)). These spaces act as separate thermal zones; each zone represents a similar profile of the underlying building dynamics. While some reference buildings such as standalone retail, strip mall, quick service restaurant, small office etc. have a simplified homogeneous building space use, other reference buildings such as primary school, hospital, large hotel etc. are more complex and have a heterogeneous building space use. An initial ANOVA analysis classifies these reference buildings on the basis of existing variations in the daily heat demand profile of each reference building (Table 4). The ANOVA test establishes whether the daily variations in heat demand are statistically significant or not. The significance levels are used in each individual representative case studies (Sections 3.1.1, 3.2.1 and 3.3.1) to aid the model identification procedure of the reduced-order model. Fig. 6 illustrates the zone floor use of six reference buildings. The zone floor use of the other twelve reference buildings is included in Appendix A.

3.1. Sub-methodology for scalability assessment

The sub-methodology for the scalability feature outlines a scalability assessment framework and follows a five step process similar to the proposed overarching methodology (Fig. 7). The first step defines scalability using previous literature [15]. The second step formulates the list of variables required to carry out the scalability assessment task. The third step involves the development of the grey-box model using the procedure as outlined in Section 2.3. The fourth step focuses on the scalability testing procedure, which uses the Analyse and Specify processes (Fig. 4).

The type of experimentation required to assess scalability involves multi-level modelling that assesses the relevance of individual zone models in representing various zones inside a building. After an initial analysis of the scalability definitions, one significant requirement would be to associate scalability with grey-box modelling at different levels, initiating with the simplest zone and then identifying the model suitability for representation of multiple zones, the whole building and building clusters. Depending on the complexity of individual zones, the Analyse process determines the initial order of the grey-box model, which would further be updated depending on the complexity of individual zones and buildings under consideration. Based on the above outlined processes, this process identifies a bottom-up energy modelling analysis as the testing strategy.

The Specify process defines the modelling levels to be tested. Grey-box models differ in their respective structures based on the underlying dynamics. Moreover, the grey-box network order is dependent on the number and type of existing zones inside the building. Different modelling levels correspond to different boundaries of dynamics existing in the building and are selected to identify thermal zones with similar dynamics. The Specify process identifies the following test scenarios to implement the multi-level modelling.

1. Zonal modelling: This modelling level forms the baseline model, which is used to compare the model prediction results with other modelling levels. Zonal modelling level represents a single zone volume inside the building. This zone volume corresponds to the floor area with the highest proportion of space usage for a particular building function.

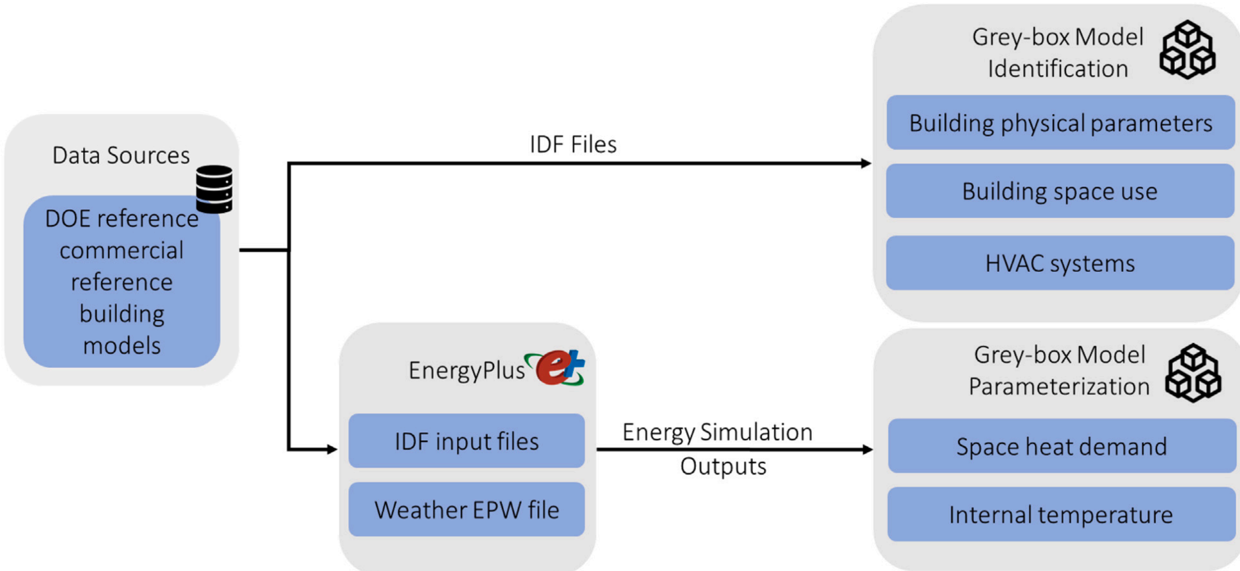

Fig. 5. Process workflow illustrating the use of DOE reference building EnergyPlus models for grey-box model identification and model parameterisation.

Table 4
DOE commercial building reference buildings demonstrating the proposed feature assessment frameworks. The significance level represents the level of variations in the heat demand profile of the reference building.

Building Type	Floor Areat (m^2)	Number of Floors	ANOVA Test p-values (Daily Variations, $\alpha = 0.10$)	Significance level
Large Office	46,320	12	0.091	Significant
Medium Office	4,982	3	0.291	Not Significant
Small Office	511	1	0.323	Not Significant
Warehouse	4835	1	0.545	Not Significant
Stand-alone Retail	2,319	1	0.433	Not Significant
Strip Mall	2,090	1	0.159	Not Significant
Primary School	6,871	1	0.025	Significant
Secondary School	19,592	2	0.085	Significant
Supermarket	4,181	1	0.027	Significant
Quick Service Restaurant	232	1	0.191	Not Significant
Full Service Restaurant	5,500	1	0.103	Not Significant
Hospital	511	5	0.055	Significant
Outpatient Health Care	3,804	3	0.069	Significant
Small Hotel	4,013	4	0.156	Not Significant
Large Hotel	11,345	6	0.071	Significant
Midrise Apartment	3,135	4	0.661	Not Significant

2. Multi-zone modelling: This modelling level forms the next phase of the test sequence where the identified zonal grey-box model represents multiple zones inside the same building. Different zonal grey-box networks have the same structure but different network parameters. This sequence updates the model parameters based on the heat dynamics of the respective zones.
3. Whole building modelling: This test sequence further updates the zone model to represent the heat dynamics at the whole building level. The whole building modelling level aggregates the zone level building operational parameters to update the network parameters.
4. Building cluster modelling: This is the last test sequence that assesses scalability based on two aspects. The first aspect deals with the identification of the cluster of buildings that represent a similar variation in heat demand profiles. The ANOVA technique analyses the significance of variations as described in Section 2.3.1. The second aspect assesses the suitability of the identified whole building grey-box model in representing similar building clusters.

The last step of the scalability assessment framework identifies a suitable list of KPIs from the list described in Section 2.5. The KPIs for assessing scalability include CVRMSE, NMBE, MAPE, R^2 and GOF. The multi-level modelling results are compared against these metrics to establish the scalability potential.

3.1.1. Scalability assessment of DOE commercial reference buildings

To demonstrate the application of the devised scalability framework, this study analyses four DOE reference buildings, namely, small office, medium office, full service restaurant and quick service restaurant as these buildings belong to the non-significant category of variations in the daily heat demand profile (Table 4). The DOE reference buildings provide the data for grey-box model order identification and network parameterisation (Fig. 5).

Small office reference building comprises mainly two types of office spaces, a core space and an unconditioned attic (Fig. 6(e)). Medium office reference building consists of two types of office spaces, a core space and a conditioned plenum (Fig. 6(f)). Both these reference buildings consists of two separate thermal zones. Both full service (Fig. 6(d)) and quick service (Fig. 6(c)) restaurant reference buildings comprises a dining area, a kitchen area and an unconditioned attic. While the dining area for the full service restaurant is approximately three times of the kitchen area, the dining and kitchen areas represent equal proportions of the building space use for the quick service restaurant.

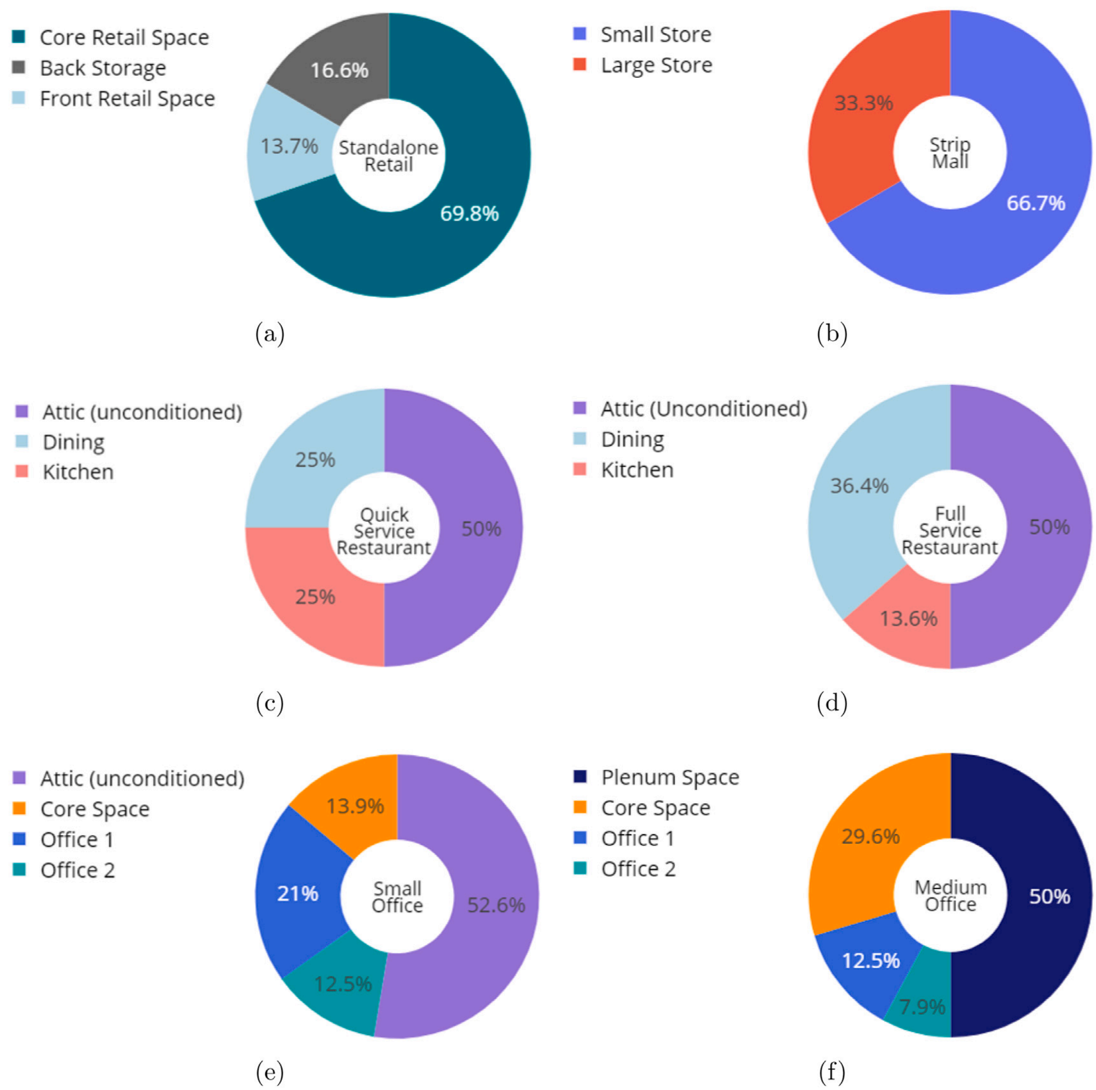

Fig. 6. Proportion of individual space use (% of total area) of the investigated DOE reference buildings: (a) Standalone retail space use with core retail function as the largest proportion amongst all. (b) Strip mall space use with major proportion occupied by small stores. (c) Quick service restaurant space use with equal proportions dedicated to dining and kitchen area. (d) Full service restaurant space use with larger proportion dedicated to the dining area. (e) Small office space use with major proportion dedicated to offices. (f) Medium office space use with major proportion representing the plenum space to facilitate air circulation.

Scalability is first assessed for the small office DOE reference building. To demonstrate the applicability for building stock modelling, the framework extends the modelling process to include the medium office DOE reference building. After the identification of the scalability definition, the data collection process extracts the required inputs from the EnergyPlus model of the small office. The crucial data inputs include the following variables [27].

1. T_a Outside Dry Bulb Temperature in °C.
2. G_s Global Solar Irradiation in kW/m^2.
3. T_i Internal Temperature Profiles in °C.
4. P_h DOE reference building Space Heat Demand in kW.
5. Zone Floor Area in %.

These variables are extracted from the EnergyPlus output file after carrying out an initial simulation over a four week period for the city of San Francisco. The grey-box model development process uses these variables to identify the grey-box network for the small office DOE reference building (Fig. 2). When implementing the grey-box model development process, the data analysis step outlines the ANOVA test results. Corresponding to a significance level of 0.10, the variations are found to be non-existent in the heat demand profile of small office (p-value = 0.323, Table 4). The non-significant ANOVA results indicate that the demand profile is stable across the entire day without any existing fluctuations. The order identification step suggests a lumped parameter model would be sufficient to represent the small-office building dynamics. Heat demand fluctuations can be directly linked to a space inside the building that needs to be assigned a separate special thermal zone, thereby requiring a new RC branch altogether. Corresponding to a floor area of 511 m^2 (more than half of which is unconditioned), normalising the demand with respect to the floor area yields high demand levels per m^2 of space use. Henceforth, separate RC network branches are assigned to exterior walls and interior mass. As the demand fluctuations are deemed insignificant by the ANOVA test, a second order grey-box model would be sufficient to represent the building dynamics (Fig. 8). Furthermore, the building does not have any embedded heating system and has a homogeneous fabric installed all throughout the building envelope. The parameter estimation step uses CTSM-R to parameterise the identified second order grey-box model and predict the internal temperature profiles (Fig. 3). The model validation step involves the use of identified KPIs (discussed below) to indicate the suitability of the model. The testing period starts from 15/01/2020 and ends on 14/02/2020 as this period comprise extreme cold temperatures.

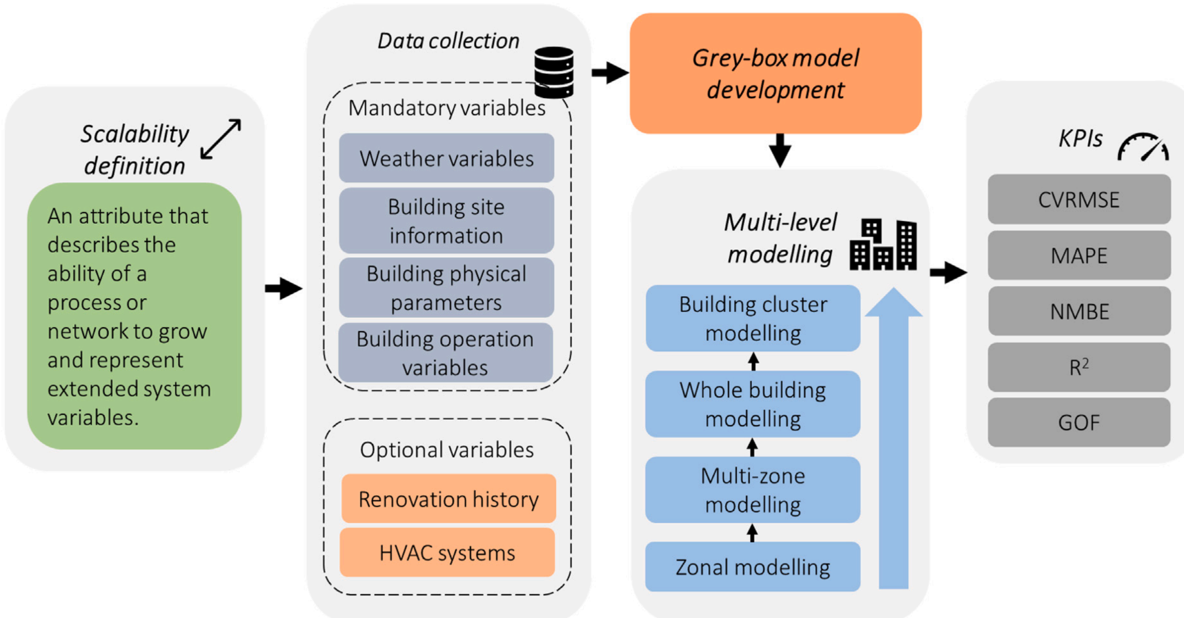

Fig. 7. Scalability assessment framework (sub-methodology) to identify zone, building and stock level scalability of reduced-order grey-box models.

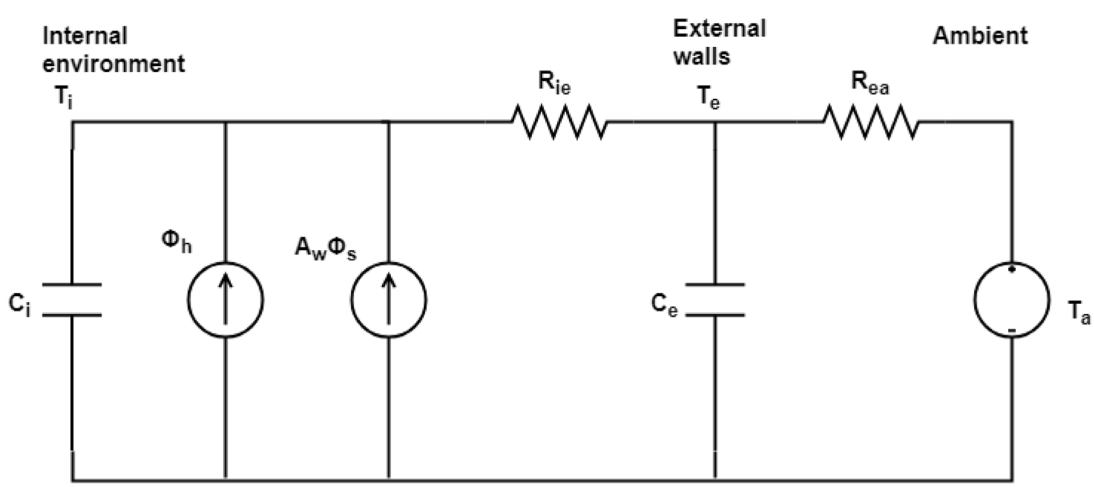

Fig. 8. The implemented second order network for the small office DOE reference building. T_i, T_e and T_a represent the states of the internal, special zone, heater, exterior and ambient elements in the network. R_{ie} and R_{ea} represent the thermal resistances of the network. C_i and C_e represent the thermal capacities of the network. ϕ_h and ϕ_s are the heater and solar radiation flux elements. A_w represents the effective window area.

The small office DOE reference building comprises five zones (zone 1 to zone 5); each of which is simulated individually to obtain the internal temperature profiles. The simulation process considers the zone 1 office grey-box network as the baseline model to compare against other zone (zone 2 to zone 5) grey-box model predictions (Fig. 9). The set-point temperatures vary among different zones in the building. It is worthwhile to mention that individual zone models involve a similar second order grey-box network (with updated parameters). Zone 1 represents the first office category of the small office space use. As evident from the DOE archetype and grey-box temperature profiles, a second order model is able to trace the dynamics of the zone with a CVRMSE of 3.65%. The corresponding values of NMBE, MAPE and R^2 are 1.92%, 2.65% and 0.95 respectively. These test case implementation uses these metric values as the baseline to compare other zone predictions. Zone 2 belongs to the similar office category and the model predictions correspond to slightly higher KPI values when compared to zone 1 KPIs (Table 5). The R^2 values are similar for both zones. When a similar zone 1 office grey-box model with updated parameters predicts the internal temperature profiles of zone 3 and zone 4, the KPIs again experience an increase in individual values indicating more complex zone dynamics. Zone 5, which represents the core space, observes KPI values between those of zones 1–2 and zones 3–4. The multi-level modelling process further implements the zone 1 office grey-box model at the whole building level. To obtain the internal temperature profile at the building level, the simulation process averages the temperatures of individual zones. When considering the second-order model for the whole building, the KPI values average out in between zone 1 and zone 4. This clearly indicates the suitability of implementing a zone level model for model predictions at the whole building level.

To demonstrate the applicability of the framework for building stock modelling, the modelling process extends the identified grey-box model to represent a medium office, which has a similar zone volume. A combined ANOVA test of small office and medium office reference buildings establishes that these reference buildings belong to the same group. The test statistics (F-value of 1.77 and p-value of 0.243) corresponding to a significance level of 0.10 nullifies the test hypotheses and hence, a similar grey-box model is used to predict the internal temperature profile for the medium office DOE archetype

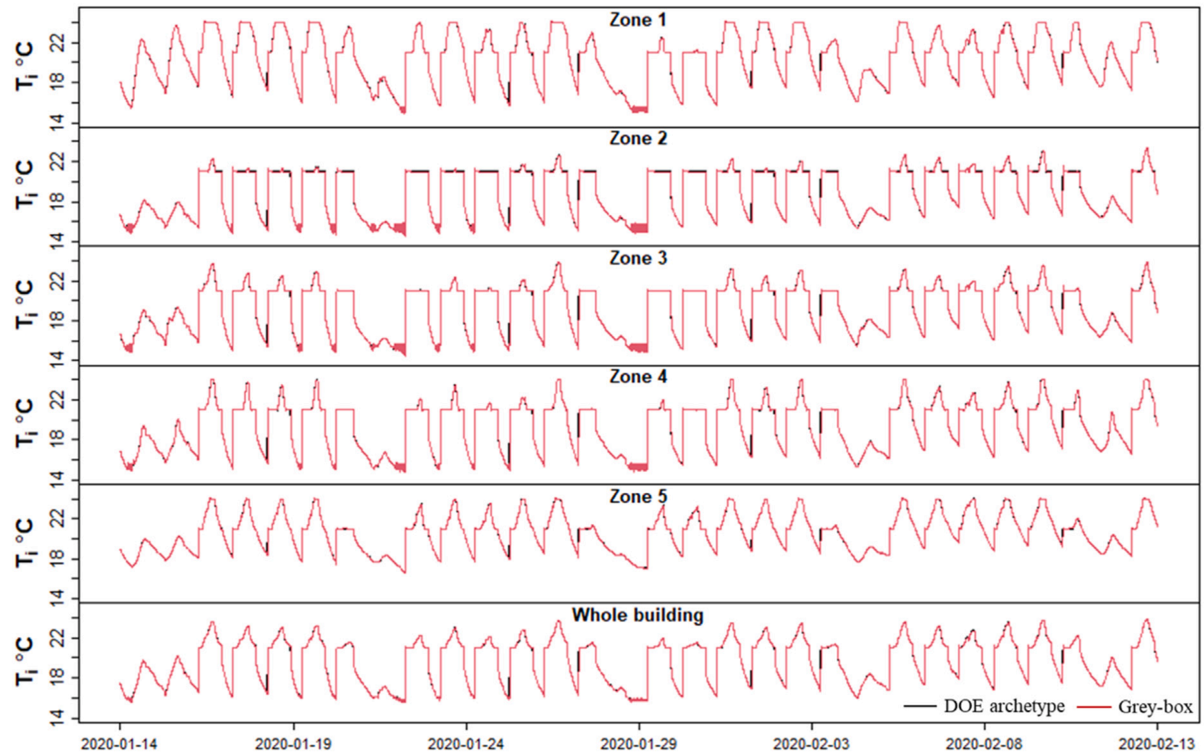

Fig. 9. DOE archetype and grey-box internal temperature profiles at the whole building level and respective zone level for the small office DOE archetype over a four week period.

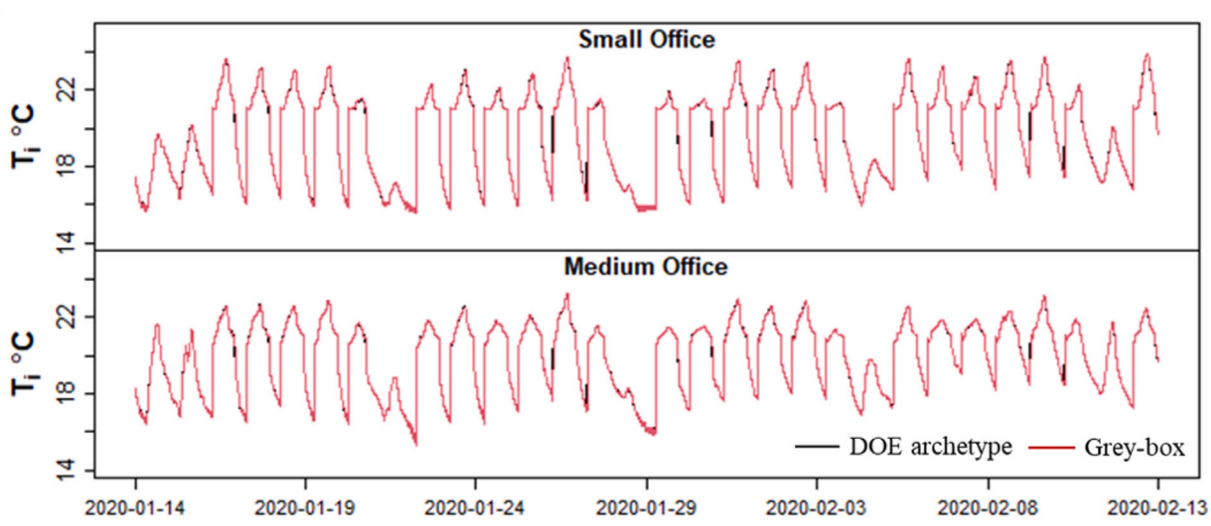

Fig. 10. DOE archetype and grey-box internal temperature profiles for the small office and medium office DOE archetypes at the whole building level over a four week period.

Table 5
KPI values for the small office DOE reference building to assess the scalability of formulated model.

KPI	Zones					Small office building	Medium office building
	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5		
NMBE	1.92	2.33	2.81	2.95	2.44	2.59	3.47
CVRMSE	3.65	3.94	4.12	4.23	3.42	3.68	4.35
MAPE	2.65	2.79	3.44	3.52	2.61	2.83	3.22
R^2	0.95	0.95	0.92	0.92	0.94	0.94	0.91

(Fig. 10) over the similar four week period. The KPI values are only slightly higher than small office archetype prediction KPIs indicating that the second order network is able to trace the dynamics of the medium office reference building.

The scalability assessment is further extended to include full service and quick service restaurant reference buildings (Appendix B). It is worthwhile to mention that although the third order network reduces the prediction errors for both full service and quick service, the reduced CVRMSE values could be considered insignificant when compared to the CVRMSE values for the second order model (Table B.1).

3.2. Sub-methodology for flexibility assessment

The sub-methodology for the flexibility feature outlines a flexibility assessment framework and follows a five step process similar to the proposed overarching methodology (Fig. 11). The first step defines flexibility using previous literature [20]. The second step formulates the list of variables required to carry out the scalability assessment task. The third step involves the development of the grey-box model using the procedure as outlined in Section 2.3. The fourth step focuses on the flexibility testing procedure, which uses the Analyse and Specify processes (Fig. 4). Feature testing is a crucial step in the flexibility assessment framework as this process develops the required experimentation to test and evaluate flexibility.

The type of experimentation required to assess flexibility involves design variation modelling that assesses the relevance of the formulated grey-box model to accommodate variations introduced in the building design. After an initial analysis of the flexibility definitions, one significant requirement for the grey-box network would be to inherit the operational characteristics of the modified building. While considering the inherent regularities of the observed data, it is crucial for the testing process to ensure a balance between the model complexity and the required scope. The process requires weather variables (outside dry bulb temperature and global solar irradiation), building physical parameters (zone volume and floor area) and building operational parameters (internal temperature profiles and space heat demand). Based on the building dynamics and the intended scope of application, the Analyse process determines the initial order of the grey-box model, which would further be updated depending on the perturbations introduced in the building design. Possible design variations might include fabric enhancements, improved accuracy of HVAC systems or other perturbations that directly affect the building heat dynamics.

The Specify process defines the test sequences for design variation modelling. As any perturbations in the system will directly affect the grey-box network parameters, the test sequence design considers a trade-off analysis to avoid the over-fitting of models. The Specify process identifies the following test scenarios to implement the design variation modelling. These test scenarios are selected based on the standardised modelling procedure of identifying a basecase of a building and implementing a parametric analysis to test several design variations. Furthermore, these scenarios represent corresponding changes in the grey-box network due to the introduced design variations.

1. Basecase building design: This test sequence focuses on an individual building to identify a basecase design. The baseline model is crucial to discern the impacts of the variations in building design. This sequence defines the initial complexity of the model based on the building function.
2. Variations in building design: This test sequence sets out different scenarios for variations in building design. The scenarios could act as plausible retrofits to improve the building energy efficiency. Any design variation will have a direct effect on the grey-box network order and network parameters.
3. Grey-box model variations: This test sequence translates the variations in building design to the building grey-box model. The sequence broadly focuses on two network aspects, namely, network order and network parameters. The network order evaluates the flexibility in terms of complexity and identifies whether the current model is flexible enough to represent the perturbations. Network parameters evaluate flexibility in terms of updating the current model parameters with as little effort as possible.
4. Trade-off analysis: This is the last test sequence that establishes the significance of model flexibility using a trade-off analysis between model complexity and required scope. A good fit of model predictions to empirical data does not necessarily provide an indication of model validity. If the model is flexible enough to fit a large proportion of potential empirical outcomes, finding a good fit becomes less meaningful. Furthermore, increased flexibility introduces higher uncertainty in the parameter estimates as higher model order increases the number of model parameters. Hence, this test sequence identifies an optimal network order to balance the uncertainty and systematic errors in model predictions.

The last step of the flexibility assessment framework identifies a suitable list of KPIs from the list described in Section 2.5. The KPIs for assessing flexibility include CVRMSE, NMBE, R^2 and GOF. The design variation modelling results are compared against these metrics to assess the flexibility potential.

3.2.1. Flexibility assessment of DOE commercial reference buildings

To demonstrate the applicability of the devised flexibility framework, this study analyses two DOE reference buildings, namely, strip mall and standalone retail reference buildings as these belong to the non-significant category of variations in the daily heat demand profile (Table 4). The DOE reference buildings provide the data for grey-box model order identification and network parameterisation (Fig. 5). The strip mall mainly comprises small and large stores, where small stores occupy two-thirds of the total mall space. The standalone retail building comprises core retail space, front retail space and back storage, where core retail space represent approximately 70% of the total space use.

After the identification of the flexibility definition, the data collection process extracts the required inputs from the EnergyPlus model of the strip mall and standalone retail building. The crucial data inputs include the following variables [27].

1. T_a Outside Dry Bulb Temperature in °C.
2. G_s Global Solar Irradiation in $\mathrm{kW/m^2}$.
3. T_i Internal Temperature Profiles in °C.
4. P_h DOE reference building space heat demand in kW.
5. Roof and wall insulation thickness in m.
6. Window U-values in $\mathrm{W/m^2K}$.

Grey-box model development process follows a similar procedure as outlined in the scalability assessment framework. Corresponding to p-test values of 0.159 and 0.4433, the ANOVA test establishes that heat demand variations are not significant for the strip mall and standalone retail reference buildings. Hence, a lumped parameter model would effectively represent the dynamics of these buildings. However, owing to the presence of slightly varying zone volume uses, it would be ideal to assign separate networks for zones representing the majority of building space use. Therefore, the selected reference test cases are assigned an initial second order network to predict the internal temperature profiles. The parameter estimation step uses a similar CTSM-R procedure to parameterise the model and predict the internal temperature profiles. The model validation step uses the identified KPIs to indicate suitability of each model. The testing period starts from 14/01/2020 and ends on 13/02/2020 for the city of San Francisco as this period comprises extreme cold temperatures.

Flexibility assessment involves design variation modelling that introduces perturbations in the base case model. The first test reference building is the strip mall, which is simulated in EnergyPlus initially to obtain the base case results. The construction elements in the base case model follow ASHRAE 90.1-1989 standards that specify values of roof insulation thickness, wall insulation thickness and window U-factor as 0.0857 m, 0.0507 m and 6.88 $\mathrm{W/m^2K}$. To introduce variations in building design, these parameters are updated to ASHRAE 90.1-2004 construction standards. The new values of roof insulation thickness,

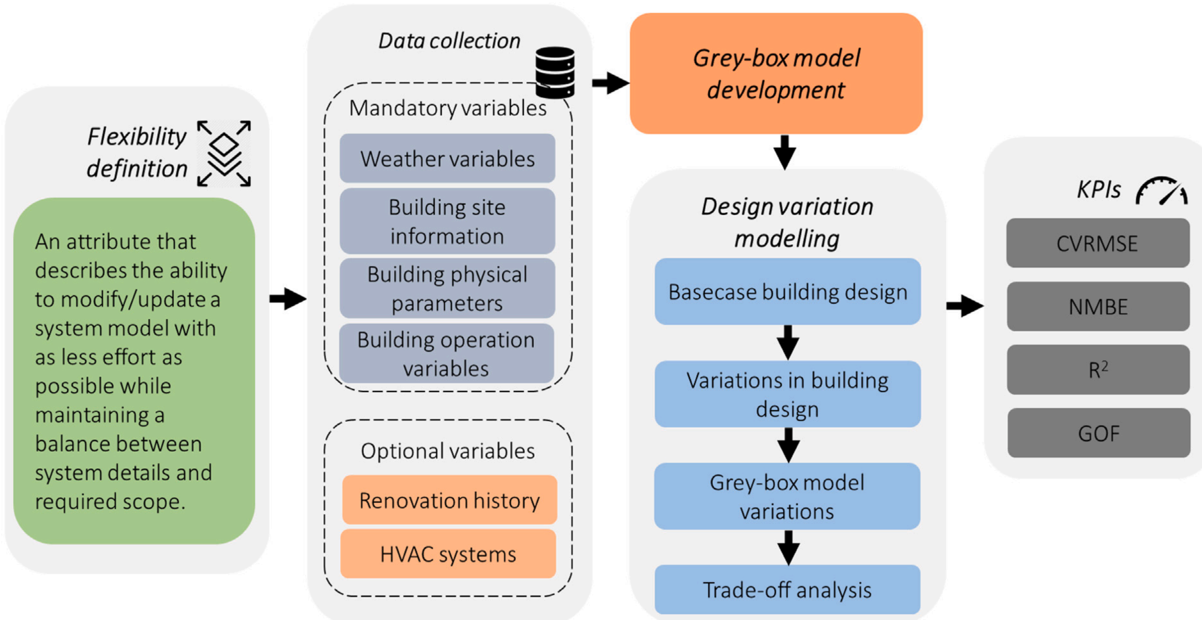

Fig. 11. Flexibility assessment framework (sub-methodology) to identify the trade-off between complexity and accuracy for reduced-order grey-box models.

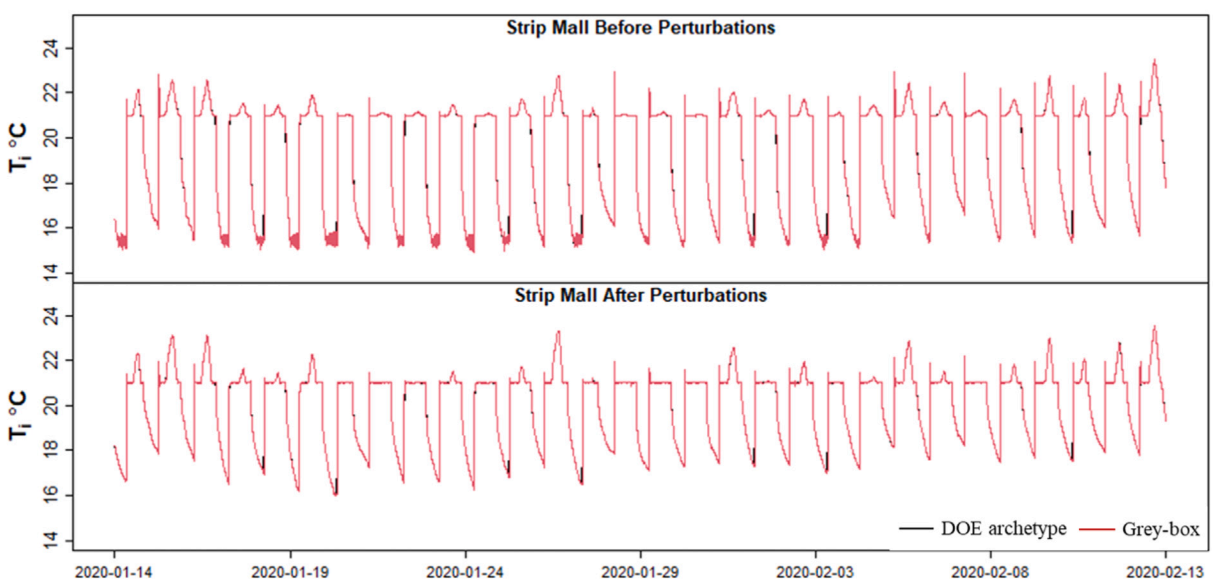

Fig. 12. DOE archetype and grey-box internal temperature profiles for the strip mall DOE archetype at the whole building level before and after the introduction of design perturbations over a four week period.

wall insulation thickness and window U-factor are 0.1246 m, 0.0532 m and 5.835 W/m^{2}K. These new values could act as plausible retrofits for old energy inefficient buildings.

Design variation modelling process uses a similar grey-box network to trace the modified building dynamics. As evident from the DOE archetype and grey-box internal temperature profiles before and after the model perturbations, a second order model effectively traces the original and modified building dynamics of the strip mall DOE archetype (Fig. 12). The values of NMBE, CVRMSE and GOF experience a slight decrease after the design variations are introduced in the base case model of the strip mall (Table 6). This can be attributed to the fact that enhancement in fabric design (improved wall and roof insulation) stabilises the building dynamics (peak heat demand levels are significantly reduced). It is worthwhile to mention that the second order model is unable to trace the timestamps when the strip mall is initially excited before any perturbations are introduced (Fig. 12). This suggests that a third order model might be a better fit for this reference building. However, when considering a trade off between model complexity and accuracy, the improvement in the grey-box model accuracy (decreased CVRMSE of 4.78%) is not significant enough to introduce another state variable in the original model. Also, these errors at the excitation timestamps are almost non-existent in the DOE archetype and grey-box temperature profiles of the reference building after the design perturbations are introduced (Fig. 12).

A similar process is repeated to assess the grey-box model flexibility of the standalone retail DOE reference building (Appendix C). The results indicate that base case model of standalone retail reference building might benefit from an enhanced network complexity. However, an increased model order would not have a significant impact on

Table 6
KPI values for the strip mall and standalone retail DOE reference buildings to assess the flexibility of the formulated model.

KPI	Strip Mall Building		Standalone Retail Building	
	Before perturbation	After perturbation	Before perturbation	After perturbation
NMBE	4.55	3.63	3.86	2.21
CVRMSE	5.73	4.61	4.47	3.04
GOF	0.05	0.04	0.04	0.03
R^2	0.77	0.86	0.87	0.82

the accuracy of the formulated model for the after perturbation scenario (Fig. C.1).

3.3. Sub-methodology for interoperability assessment

The sub-methodology for the interoperability feature outlines an interoperability assessment framework and follows a five step process similar to the proposed overarching methodology (Fig. 13). The first step defines interoperability using previous literature [22]. The second step formulates the list of variables required to carry out the interoperability assessment task. The third step involves the development of the grey-box model using the procedure as outlined in Section 2.3. The fourth step focuses on the interoperability testing procedure, which uses the Analyse and Specify processes (Fig. 4). Feature testing is a crucial step in the interoperability assessment framework as this process develops the required experimentation to test and evaluate interoperability.

The type of experimentation required to assess interoperability involves scenario optimisation modelling that assesses the interoperability potential of a community of buildings using grey-box models. The interoperability feature is more associated with building clusters rather than the employed modelling approach. However, this feature demonstrates how grey-box models would facilitate the interoperability characterisation process, and thereby, illustrates a crucial modelling characteristic of these models. An initial analysis of the interoperability definition indicates that there is a need to characterise individual buildings in a cluster using model parameters and building function. The process requires weather variables (outside dry bulb temperature and global solar irradiation), building physical parameters (floor area), model parameters and building operational parameters (internal temperature profiles and space heat demand). Based on the building dynamics, the Analyse process determines the optimal order of the grey-box model for individual buildings in the cluster. As interoperability deals with the optimisation of building systems, a possible testing strategy would involve scenario optimisation modelling that aims to reduce the heat demand variations of the building cluster on an aggregate basis.

The Specify process defines the test sequences for scenario optimisation modelling, which facilitates individual building level integration. Variations in heat demand among a community of buildings will allow for demand balancing that can lead to a drop in the overall heat demand. The Specify process identifies the following test scenarios to implement the scenario optimisation modelling. These scenarios are selected so as to reduce the overall heat demand of the system. Henceforth, the test scenarios need to identify and characterise the buildings based on pre-defined building metrics. The characterisation facilitates the optimisation of the overall heat demand in the building cluster.

1. Building cluster: This test sequence lists individual building in a cluster, estimates grey-box parameters and performs an initial ANOVA analysis using the heat demand profiles.
2. Heat demand characterisation: This sequence characterises the buildings on the basis of heat demand profiles. The demand profiles are a direct indication of the kind of ongoing activities inside a building. The characterisation leads to two categories, namely, buildings with a variable demand pattern and buildings with a stable demand pattern. Buildings with similar demand profiles provide limited opportunities to implement any kind of demand scenario optimisation. Furthermore, buildings with cyclic demand patterns would facilitate demand balancing alterations.
3. C-value characterisation: This test sequence derives a relative characterisation of individual buildings on the basis of grey-box model parameters. The C-value affects the response time of the building's thermal mass when subjected to a change in operational parameters (for instance, temperature). A relatively high value of C increases the response time of building mass to excitation from the mechanical systems. The time constant indicates the response time of the building's interior to a temperature difference between indoor and ambient environment. The devised categories include buildings with a high C-value and buildings with a low C-value (high and low values represent a relative comparison).
4. Heat demand alteration scenarios: This is the last test sequence that identifies demand alteration scenarios based on the aforementioned characterisations. Heat demand alterations are easier to implement in a building with a high value of C coupled with low variations in heat demand. A high C building experiences a delayed response and thus will maintain the required comfort conditions even when the temperature is not within the acceptable limits. Moreover, the low demand fluctuations allow the building to jump back to the desired temperature range without increasing the overall demand. This sequence identifies the time instances when peaks occur in heat demand profiles. Building with high variations in heat demand provide such instances and therefore, any control measures in other buildings should be introduced at these timestamps.

The last step of the interoperability assessment framework identifies a suitable list of KPIs from the list described in Section 2.5. The KPIs for assessing interoperability include CVRMSE, RC parameters and PHDR. The scenario optimisation modelling results are compared against these metrics to assess the interoperability potential.

3.3.1. Interoperability assessment of DOE commercial reference buildings

To demonstrate the applicability of the devised interoperability framework, this study analyses four DOE reference buildings, namely, primary school, midrise apartment, small hotel and supermarket. The DOE reference buildings provide the data for grey-box model order identification and network parameterisation (Fig. 5). Two of these reference buildings (primary school and supermarket) belong to the significant category of variations in the daily heat demand profile (Table 4). The other two reference buildings (midrise apartment and small hotel) belong to the non-significant category of variations in the daily heat demand profile (Table 4). While midrise apartment and small hotel reference buildings represent simplified building use (Figs. A.1(d) and A.1(c)), primary school and supermarket reference buildings represent dynamic use of the building space (Figs. A.1(a) and A.1(e)). It is important to note that the interoperability feature is specifically assessed for an individual building or a group of buildings using grey-box model parameters rather than for an individual model.

After the identification of the interoperability definition, the data collection extracts the required inputs from the EnergyPlus model of

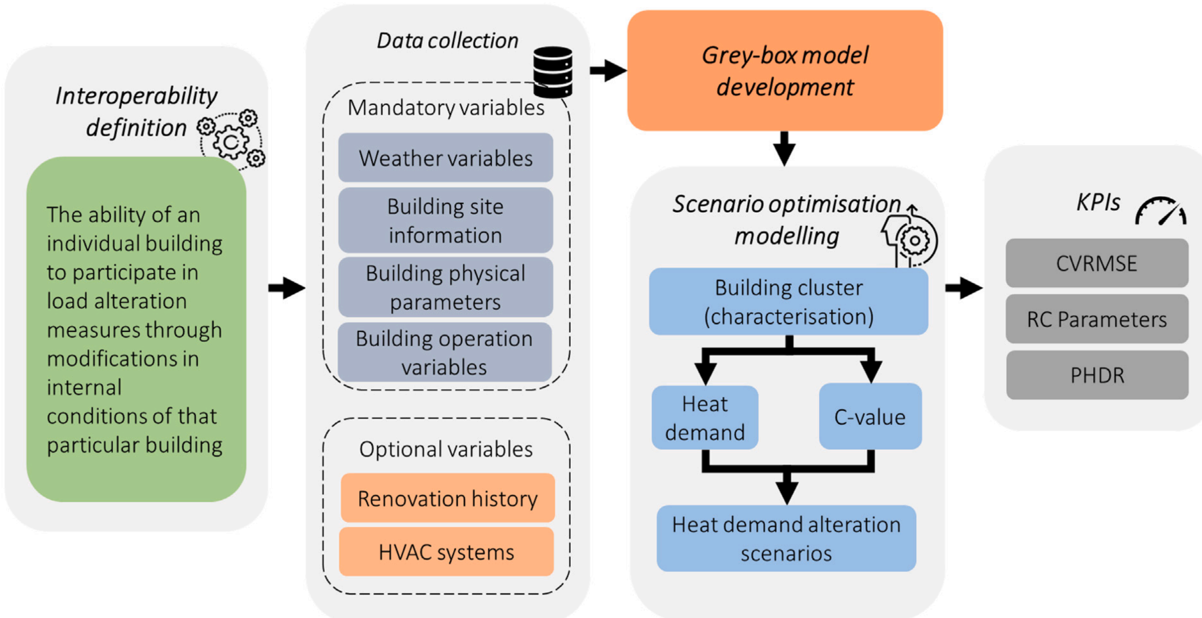

Fig. 13. Interoperability assessment framework (sub-methodology) to formulate load reduction scenarios using reduced-order grey-box models.

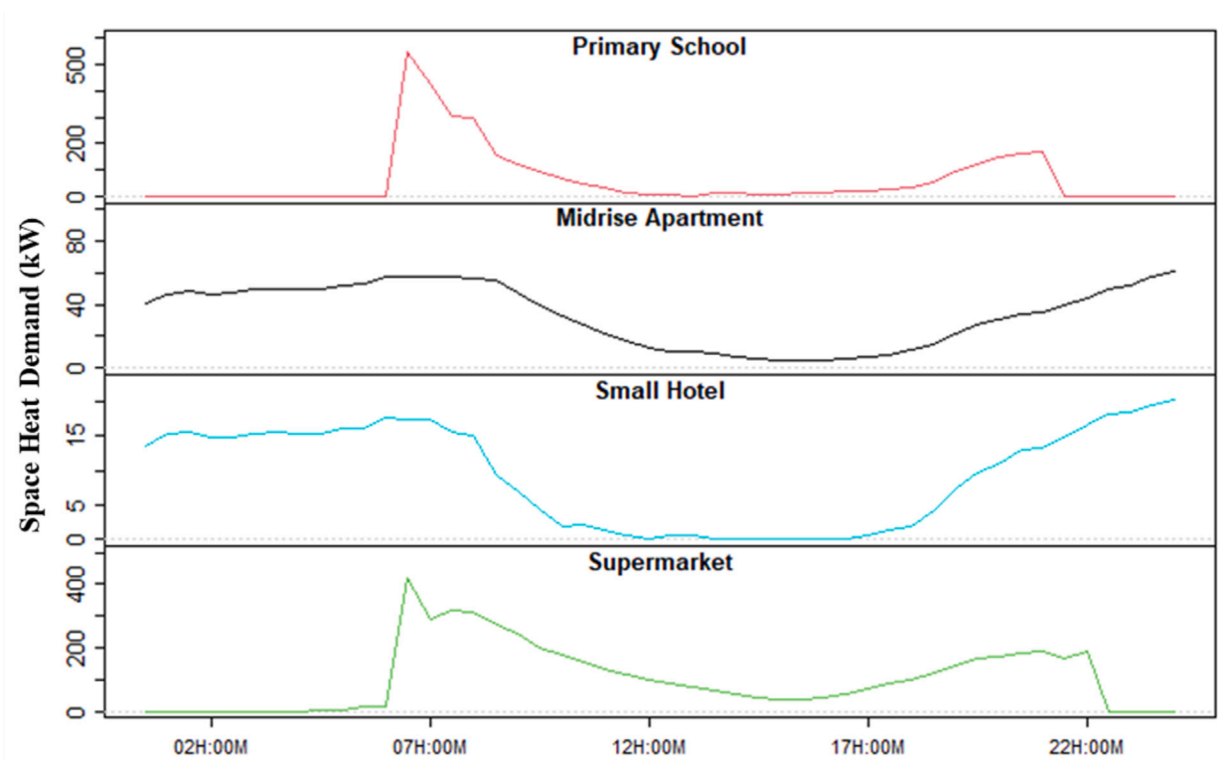

Fig. 14. Daily heat demand profile for primary school, mid-rise apartment, small hotel and supermarket DOE reference buildings for possible interoperability characterisation.

the primary school, midrise apartment, small hotel and supermarket reference buildings. The crucial data inputs include the following variables [27]:

1. T_a Outside Dry Bulb Temperature in °C.
2. G_s Global Solar Irradiation in $\mathrm{kW/m^2}$.
3. T_i Internal Temperature Profiles in °C.
4. P_h DOE reference building space heat demand in kW.

Grey-box model development follows a similar procedure as outlined in the scalability and flexibility assessment frameworks. Corresponding to p-test values of 0.025 and 0.027, the ANOVA test establishes that heat demand variations are significant for the primary school and supermarket reference buildings. The heat demand variations are found to be non-significant for midrise apartment and small hotel reference buildings (p-values of 0.661 and 0.156). While a third order

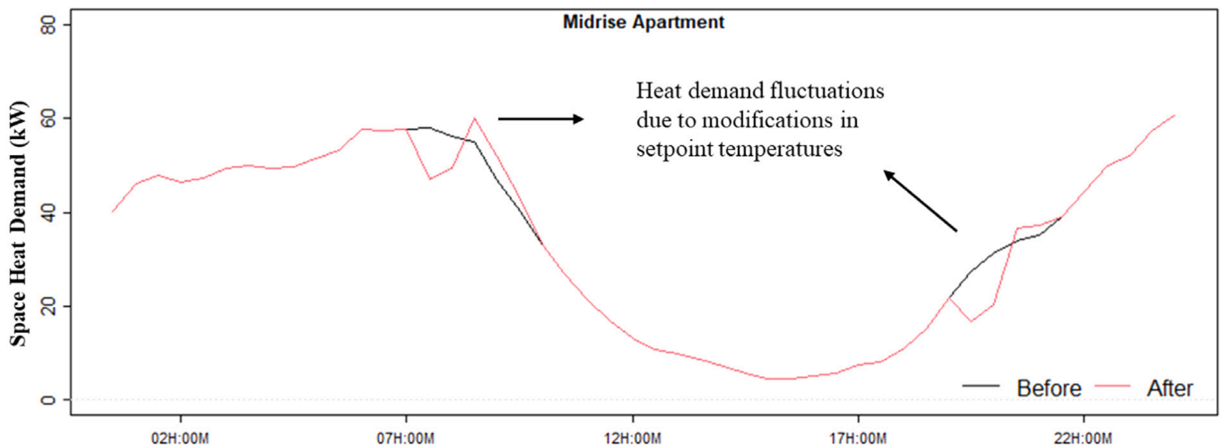

Fig. 15. Daily heat demand profile for the midrise apartment DOE reference building before and after the introduction of demand alteration scenarios. The heat demand fluctuations represent time instants when the set-point temperature is varied.

model is assigned to the primary school and supermarket reference buildings, a second order model is used to represent the midrise apartment and small hotel reference buildings. It should be noted that grey-box model identification is not crucial to assessing interoperability as only the model parameters are required. These parameters yield similar overall aggregated values when considering various orders of the grey-box network. The parameter estimation step uses CTSM-R to identify the network parameters. The testing period involves a single day (14/01/2020) during which the heat demand profiles are recorded on an hourly basis for San Francisco city. This day is representative of a midweek day with cold temperatures.

Interoperability assessment involves scenario optimisation modelling that introduces demand alteration scenarios in the considered building cluster (primary school, midrise apartment, small hotel and supermarket). The building cluster undergoes a characterisation process on the basis of heat demand and network C-values. When comparing the variation in heat demand over 24 h between different reference buildings, primary school and supermarket are found to possess large variations in heat demand compared to midrise apartment and small hotel (Fig. 14). Furthermore, the demand variations between midrise apartment and small hotel are found to be insignificant. This indicates that the means of both test groups are the same and similar variations exist in the profiles. As such, demand alteration scenarios could not be implemented between midrise apartment and small hotel and the interoperability potential between these buildings is not significant. However, a high demand balancing potential exists between primary school or supermarket and midrise apartment. Similarly, demand alterations could also be achieved with the small hotel building.

Further building characterisation involves network C-values as these directly affect the response time of the building's thermal mass. A relatively high value of C increases the response time of building's mass. Owing to this explanation, midrise apartment has the highest response time closely followed by the supermarket (Table 7). Small hotel and primary school reference buildings have a low response time. Buildings with relatively larger C value respond slower to changes in temperature and thus maintain the required comfort conditions even when the setpoint temperature is lower than required. Additionally, the low demand fluctuation allows the building to achieve the setpoint temperature without increasing the overall heat demand. Therefore, midrise apartment provides more opportunities for the implementation of heat demand alterations.

As the overall aim is to reduce the peak systems demand, demand alteration measures are usually applied during the time instance of peaks occurring in the heat demand. The primary school and supermarket reference buildings provide these instances as these buildings have high variations in heat demand profiles. Based on the above observations, a simulation experiment is conducted using the grey-box models to simulate the hourly heat demand and determine the building response when subjected to a step change in temperature. Simulations are performed using a second order network for the midrise apartment with a corresponding CVRMSE of 4.33% for internal temperature predictions. The setpoint temperature of the midrise apartment is assumed to be 21.1 °C. The step changes in setpoint temperature are introduced at 07:00 (until 08:00) and 19:00 (until 20:00) for the midrise apartment. The peak heat demand occurs at these timestamps in the heat demand profiles of primary school and supermarket. At 07:00, the modification of the setpoint temperature to 20.1 °C reduces the corresponding heat demand from 56 kW to 44 kW (Fig. 15). The high C-value of the midrise apartment delays the heat transfer from the walls, thereby, maintaining the desired comfort conditions. When the setpoint is restored back to the original value, the heat demand in the next hour increases from 54 kW to 60 kW. Although these modifications result in an increased demand at the next timestamp, the overall peak demand of the system reduces by 10% during that hour (PHDR = 10%). A similar phenomenon is experienced at 19H:00 when the temperature setpoint is again reduced by 1 °C, which yields a PHDR of 6% during that hour (Fig. 15). It is worthwhile to mention that these setpoint temperature modifications introduce fluctuations in the heat demand profile of the midrise apartment building. However, the system peak heat demand reduces at the same time when the building cluster is considered to function together.

The DOE reference building case studies highlight the importance of feature assessment in building energy performance modelling. The devised feature assessment frameworks lay out a foundation to define, assess and evaluate three crucial features of reduced-order grey-box energy models. The feature assessment will eventually facilitate the possibilities of scaling up networks, the evaluation of numerous design scenarios and the integration of individual building level components with other energy systems.

4. Discussions

The BEPS domain has experienced a significant increase in the number of underlying modelling techniques and approaches over the past few years. These devised techniques use different variants of the established white-box, grey-box and black-box models. Considering the overall spectra of the modelling techniques, it becomes crucial to evaluate the characteristics that identify model applicability and suitability for a specific application. This research defines an integrated framework to assess reduced-order grey-box model features. The highlights of the framework involve generalised grey-box model development and feature testing processes that identify the experimentation

Table 7
Thermal capacitance and time constant values of the selected DOE reference buildings to assess the interoperability of the formulated cluster.

DOE Archetype	Thermal Capacitance (kWh/° C)	Time Constant (h)	Hourly Variations in Heat Demand
Midrise Apartment	75.24	101.25	Not Significant
Small Hotel	32.55	39.04	Not Significant
Primary School	27.19	44.05	Significant
Supermarket	72.08	103.79	Significant

requirements to assess the scalability, flexibility and interoperability of grey-box networks. The devised DOE reference building case studies provide an effective and a holistic overview of the respective feature potential of each building using pre-defined KPIs. Although the proposed approach lays out a strong foundation for model feature assessment, there are certain challenges associated with the feature testing procedures of scalability, flexibility and interoperability.

When implementing multi-level modelling to assess the scalability feature, the test building loses a few dynamics upon the aggregation of the zone level dynamics to obtain the building level profiles. Although these dynamics could be revived through an additional state variable, this eventually increases the complexity of the existing grey-box network. As such, the process requires a trade-off analysis to identify the significance of increased model complexity over accuracy. For buildings with a simplified zone volume use (office building), a zone model would yield accurate results at the whole building level. However, a similar scenario might not be observed for a building with varied zone volume use (hospital), which suggests the non-scalable nature of these buildings.

When implementing design scenario modelling to assess flexibility, one significant challenge with grey-box models is to address the issue of over fitting the model parameters. A good fit of model predictions to the measured data is often considered for validating the model. However, it should be noted that if a model is flexible enough to fit several design scenarios, identification of a good fit becomes less meaningful. With enhanced building dynamics, the associated grey-box model need to be compensated with additional complexity at the same level. Furthermore, the accessibility of these design scenarios could be enhanced through a grey-box model parameter database that corresponds to the introduced design variations.

When implementing scenario optimisation modelling to assess interoperability, the type of buildings in the considered building cluster define the demand alteration scenarios. It is crucial to identify a set of buildings that have varied building operation for interoperability to actually work. Furthermore, time constant is as crucial as the C-value but is often overlooked when assessing the building's thermal mass. For instance, a low thermal mass building with low conductance would respond more quickly to internal heat gains than a high thermal mass building with higher conductance although both buildings represent a similar time constant.

Another limitation concerns the demonstration cases using the DOE reference buildings. Although the DOE reference buildings represent reasonably realistic building characteristics and nature of operation, the use of simulated heat demand profiles is closer to an almost ideal building function. As commercial buildings often differ in their day to day operations, these profiles do not represent the entire system dynamics. Furthermore, although this study focuses mainly on heating loads, the developed grey-box networks could also be used to estimate the cooling loads. These networks combine a dynamic simulation of heating as well as cooling loads (both expressed in terms of thermal power). Forecasting of cooling loads would use a similar set of energy balance equations as for the heating loads [41]. Another limitation relates to the version updates of the EnergyPlus software specific to the considered case study. Any major shifts could potentially cause issues with reproducibility.

5. Conclusions and future work

Buildings could play an instrumental role towards achieving an integrated energy system. Building to grid and building to building integration possesses enormous potential to enable consumer and energy-related benefits. To facilitate this integration, there is a need for holistic and scalable building modelling approaches to model the commercial building stock. This study proposes a framework to evaluate the scalability, flexibility and interoperability associated with grey-box models through the assessment of the building energy model accuracy and complexity. Grey-box model scalability could directly be linked to the building space use (floor area dedicated to a specific building function). The grey-box model flexibility results provide clear distinction between model accuracy and the desired complexity when introducing perturbations in the base model. Interoperability results indicate that it is possible to reduce the system peak demand in a building cluster using grey-box model parameters.

The developed framework allows rapid and accurate creation of grey-box building energy models at various modelling levels. The results of this study could support the current need for the assessment of consumption patterns of the commercial building stock. The devised grey-box model development approach entails a generic model structure and provides additional flexibility in terms of modelling the design variations. When analysing buildings with varied nature of operation, it becomes crucial to associate key performance indicators that determine the effects of variations in model design. The devised flexibility assessment framework could be implemented to study the post-retrofit heat consumption patterns at the individual building as well as the district scale. The generalised model structures facilitate component modelling for larger and complex energy systems. Energy modellers and practitioners could use these grey-box models to enhance their understanding of system parameters that lead to significant errors in energy predictions. The devised approach further aids demand management in buildings through the use of HVAC and water heating systems.

The control systems for energy-related components often fail to deliver an optimal energy operation within residential and commercial buildings. This can be attributed to the disintegrated nature of building operation and building control systems both within and outside the building envelope. Furthermore, even with sophisticated energy modelling approaches, deployed control strategies and dispatch of loads are still rudimentary and are neither cost effective nor scalable. The interoperability framework using building grey-box model parameters enables building to building integration and interoperability. The interoperability framework facilitates the evaluation of individual building's ability to participate in the energy integration process. Building managers and grid operators could achieve significant reductions in peak heat demand as buildings and industrial energy use/consumption drive system peak demand. Ultimately, interoperable buildings should enhance the deployment of distributed generation.

Future work could integrate the quantification of uncertainties in building energy systems due to uncertain design parameters (network parameters) or inherent uncertain parameters (weather). When considering grey-box building energy models, uncertainty propagation usually involves the implementation of inverse quantification techniques to associate confidence intervals with network parameters. Integration of an uncertainty framework would enhance the value of this research. Future work could further involve the use of real-time buildings to test the devised feature assessment frameworks for scalability, flexibility and interoperability. Moreover, model validation could be improved using ANSI/RESNET Standard 1201-2016 for calibration methods.

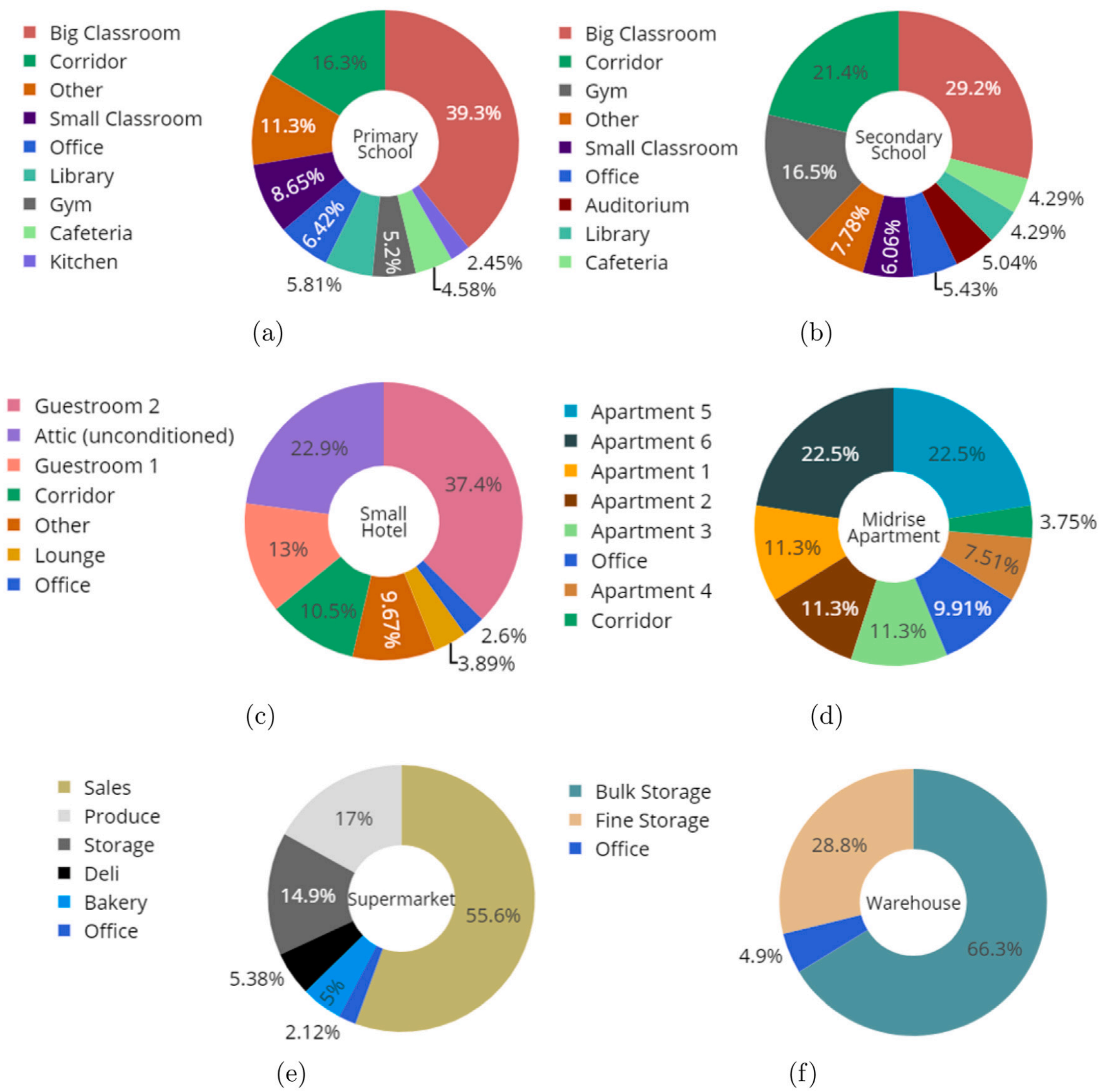

Fig. A.1. Proportion of individual space use (% of total area) of the investigated DOE reference buildings: (a) Primary school space use with majority of the space dedicated to classrooms, categorised as big and small. (b) Secondary school with majority of the space dedicated to classrooms and gym. (c) Small hotel space use with majority of the space dedicated to guestrooms divided into two categories. (d) Midrise apartment space use with six categories of apartment categorised on the basis of orientation. (e) Supermarket space use with the majority of the space dedicated to sales area and, (f) Warehouse space use with majority of the space use dedicated to bulk storage.

CRediT authorship contribution statement

Mohammad Haris Shamsi: Conceptualization, Methodology, Results, Validation, Writing - original draft. **Usman Ali:** Data curation. **Eleni Mangina:** Supervision, Reviewing. **James O'Donnell:** Supervision, Writing - review & editing.

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.

Acknowledgements

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI), Ireland under the SFI Strategic Partnership Programme Grant number SFI/15/SPP/E3125. The opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Science Foundation Ireland.

The author would also like to acknowledge, IBPSA Project 1, an international project conducted under the umbrella of the International Building Performance Simulation Association (IBPSA). Project 1 will develop and demonstrate a BIM/GIS and Modelica Framework for building and community energy system design and operation.

Appendix A. Description of the building cluster

The building cluster consists of 16 commercial prototype building models published by the U.S. Department of Energy. These models are instrumental in the development of commercial building codes and standards. The 16 reference building types represent a majority of the commercial buildings (70%) in the U.S. across different climate zones. Each reference building represents a different zone volume use in terms of the total floor area. The corresponding EnergyPlus models provide a consistent baseline for comparison to evaluate the existing building energy performance.

The zone floor use of six reference buildings is illustrated in Section 3. Figs. A.1 and A.2 illustrate the zone floor use of the remaining twelve reference buildings. Except the warehouse, other reference

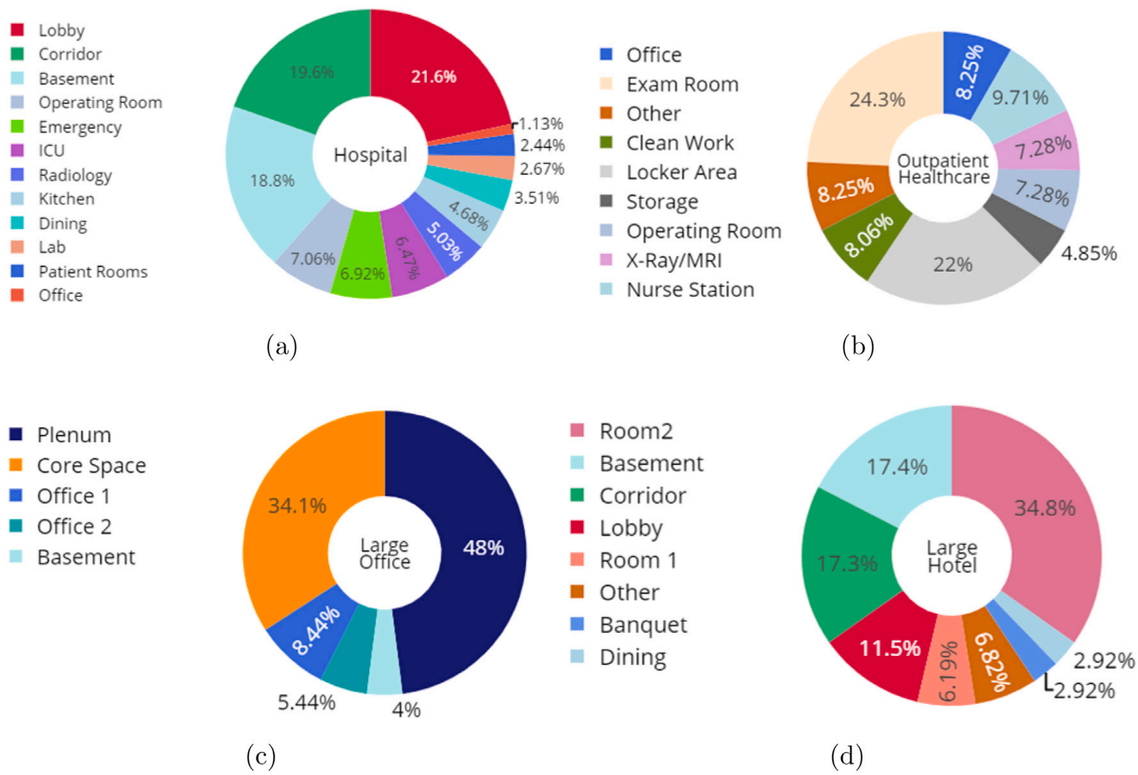

Fig. A.2. Proportion of individual space use (% of total area) of the investigated DOE reference buildings: (a) Hospital space use representative of varied building functions. (b) Outpatient healthcare space use with major proportion dedicated to offices and exam rooms. (c) Large office space use with majority portion dedicated to plenum and core areas and. (d) Large hotel space use with rooms representing the majority.

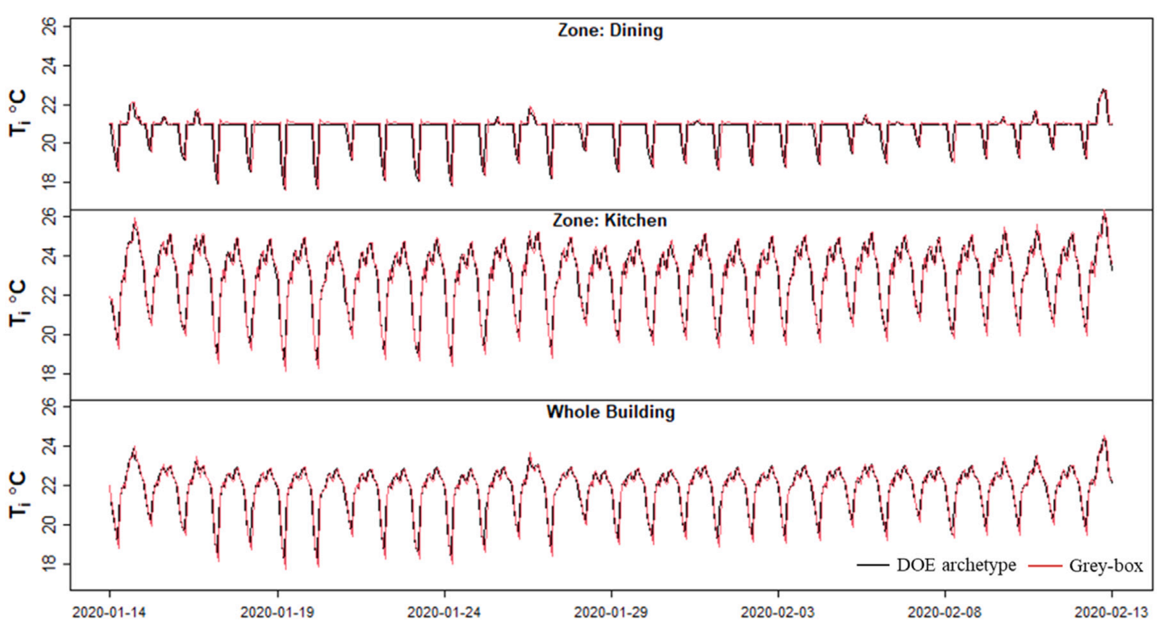

Fig. B.1. DOE archetype and grey-box internal temperature profiles at the whole building level and respective zone level for the full service restaurant DOE archetype over a four week period.

buildings represent multi-zone floor use with at least four different zones.

Appendix B. Scalability assessment of DOE commercial reference buildings

The scalability assessment is further extended to include full service and quick service restaurant reference buildings. These two reference buildings mainly comprise dining and kitchen areas as two separate zones with temperature profile variations only slightly comparable to both office reference buildings (Fig. B.2). A second-order model represents different zones and predicts the individual zone internal temperature profiles inside the full service restaurant reference building. As evident from the individual zone temperature profiles, the identified model is able to trace the internal dynamics of the respective zones although with reduced accuracy levels (Fig. B.1).

Fig. B.2. DOE archetype and grey-box internal temperature profiles for the full service and quick service restaurant DOE archetypes at the whole building level over a four week period.

Fig. B.3. Higher order model DOE archetype and grey-box internal temperature profiles for the full service and quick service restaurant DOE archetypes at the whole building level over a four week period.

Table B.1
KPI values for the full service and quick service restaurant DOE archetypes to assess the scalability of formulated model.

KPI	Zones		Full Service Restaurant Building	Quick Service Restaurant Building
	Zone: Dining	Zone: Kitchen		
NMBE	−3.31	2.56	2.59	−2.83
CVRMSE	5.13	4.72	4.95	5.22
MAPE	3.78	2.88	2.83	3.57
R^2	0.85	0.94	0.87	0.82

The KPI values (except R^2) are higher for the dining zone when compared to the kitchen zone. Furthermore, the model under predicts (negative NMBE value of 3.31%) the zone temperatures for the dining zone (Table B.1). A higher R^2 for kitchen zone indicates that the second order network traces the dynamics of kitchen in a slightly better manner when compared to the dining zone. At the whole building level, the observed KPI values lie between the KPIs for the dining and kitchen zones. When a similar order network represents the quick service restaurant, the accuracy levels further experience a decline (high errors).

To enhance the prediction accuracy, we added an additional state variable to the second order grey-box network of full service and quick service restaurant. As evident from the DOE archetype and grey-box temperature profiles, a third order network improves the prediction accuracy of the model (Fig. B.3). The corresponding CVRMSE values of internal temperature predictions are found to be 3.88% and 4.12% for the full service and quick service restaurant DOE reference buildings

Fig. C.1. DOE archetype and grey-box internal temperature profiles for the standalone retail DOE archetype at the whole building level before and after the introduction of design perturbations over a four week period.

Table D.1
Acceptable accuracy limits as per the ASHRAE Guideline 14 for monthly and hourly resolutions.

Granularity of Data	Coefficient of Variation Root Mean Square Error (CVRMSE)	Normalised Mean Bias Error (NMBE)
Monthly	15%	5%
Hourly	30%	10%

Table E.1
Literature review of model feature assessment frameworks to identify definitions of scalability, flexibility and interoperability. The distinct characteristics are defined in line with [42].

Previous Literature	Model feature			Modelling scale		Distinct characteristic
	Scalab.	Flexib.	Interop.	Individual Building	Building Clusters	
[28]	✓				✓	Broad scalability definition for large energy systems
[43]	✓			✓		Integration of BIM and building simulation
[44]	✓	✓		✓		Scalability and flexibility definitions for grey-box models
[12,15]	✓			✓	✓	Reduced-order model implementation at urban level
[20]		✓		✓	✓	Comparison of modelling approaches; Advantages of hybrid approaches
[19]		✓			✓	Flexibility assessment of black-box models
[22]			✓		✓	Broad interoperability definition
[21,24]			✓		✓	Integration with the distribution grid
[45,46]			✓		✓	Demand shifting using district heating networks
[25]			✓		✓	Optimisation using HVAC systems

respectively. Although the third order network reduces the prediction errors, the reduced CVRMSE values could be considered insignificant when compared to the CVRMSE values for the second order model. Furthermore, a GOF test value of 0.237 indicates that there is no significant improvement in the dynamics of the third order model.

Appendix C. Flexibility assessment of DOE commercial reference buildings

The flexibility assessment procedure is repeated to assess the grey-box model flexibility of the standalone retail DOE reference building. As evident from the DOE archetype and grey-box internal temperature profile before and after perturbations, a second-order model effectively traces the dynamics of the standalone retail building (Fig. C.1). The KPIs for the temperature profile predictions are also well within limits for before and after scenarios (Table 6). To further evaluate the model at the excitation timestamps, we calculated the CVRMSE KPI values for hourly heat demand predictions considering both scenarios. The CVRMSE values (hourly heat demand) are found to be 18.4% and 7.1% for before perturbation and after perturbation scenarios. Although these values are within the ASHRAE 14 guidelines limits, we further investigated using a third order network to represent the base case model (before perturbation scenario) of the standalone retail reference building. While the base case model CVRMSE value for hourly heat demand predictions steeply declined to 9.7%, the CVRMSE for the after perturbation model experienced an insignificant decrease to 6.3% in the prediction of hourly heat demand. This strongly suggests that the base case model might benefit from an enhanced network complexity. However, an increased model order would not have a significant impact on the accuracy of the formulated model for the after perturbation scenario.

Appendix D. ASHRAE guideline 14 accuracy limits

See Table D.1.

Appendix E. Previous literature that defines model features in the BEPS domain

See Table E.1.

References

[1] Abergel T, Delmastro C. Tracking buildings 2020. Tech. rep., Paris: International Energy Agency; 2020, URL https://www.iea.org/reports/tracking-buildings-2020.

[2] Glasgo B, Khan N, Azevedo IL. Simulating a residential building stock to support regional efficiency policy. Appl Energy 2020;261:114223. http://dx.doi.org/10.1016/j.apenergy.2019.114223.

[3] IEA. Sustainable development scenario – world energy model. Tech. rep., IEA; 2019, URL https://www.iea.org/reports/world-energy-model/sustainable-development-scenario.

[4] Salata F, Ciancio V, Dell'Olmo J, Golasi I, Palusci O, Coppi M. Effects of local conditions on the multi-variable and multi-objective energy optimization of residential buildings using genetic algorithms. Appl Energy 2020;260:114289. http://dx.doi.org/10.1016/j.apenergy.2019.114289.

[5] Choi JH. Investigation of the correlation of building energy use intensity estimated by six building performance simulation tools. Energy Build 2017;147:14–26. http://dx.doi.org/10.1016/j.enbuild.2017.04.078.

[6] Amara F, Agbossou K, Cardenas A, Dubé Y, Kelouwani S. Comparison and simulation of building thermal models for effective energy management. Smart Grid Renew Energy 2015;6(April):95–112.

[7] Yang G, Li Z, Augenbroe G. Development of prototypical buildings for urban scale building energy modeling: A reduced order energy model approach. Sci Technol Built Environ 2018;24(1):33–42. http://dx.doi.org/10.1080/23744731.2017.1328943.

[8] Shamsi MH, Ali U, Mangina E, O'Donnell J. A framework for uncertainty quantification in building heat demand simulations using reduced-order grey-box energy models. Appl Energy 2020;275:115141. http://dx.doi.org/10.1016/j.apenergy.2020.115141.

[9] Coakley D, Raftery P, Keane M. A review of methods to match building energy simulation models to measured data. Renew Sustain Energy Rev 2014;37:123–41. http://dx.doi.org/10.1016/j.rser.2014.05.007, arXiv:j.rser.2014.05.007. URL https://www.sciencedirect.com/science/article/pii/S1364032114003232{#}!.

[10] Capizzi G, Sciuto GL, Cammarata G, Cammarata M. Thermal transients simulations of a building by a dynamic model based on thermal-electrical analogy: Evaluation and implementation issue. Appl Energy 2017;199:323–34. http://dx.doi.org/10.1016/j.apenergy.2017.05.052.

[11] Garrett A, New J. Scalable tuning of building models to hourly data. Energy 2015;84:493–502. http://dx.doi.org/10.1016/j.energy.2015.03.014.

[12] Kontokosta CE, Tull C. A data-driven predictive model of city-scale energy use in buildings. Appl Energy 2017;197:303–17. http://dx.doi.org/10.1016/j.apenergy.2017.04.005.

[13] Heo Y, Augenbroe G, Graziano D, Muehleisen RT, Guzowski L. Scalable methodology for large scale building energy improvement: Relevance of calibration in model-based retrofit analysis. Build Environ 2015;87:342–50. http://dx.doi.org/10.1016/j.buildenv.2014.12.016.

[14] Manfren M, Nastasi B. Parametric performance analysis and energy model calibration workflow integration - A scalable approach for buildings. Energies 2020;13(3):621. http://dx.doi.org/10.3390/en13030621, URL https://www.mdpi.com/1996-1073/13/3/621.

[15] Heidarinejad M, Mattise N, Dahlhausen M, Sharma K, Benne K, Macumber D, et al. Demonstration of reduced-order urban scale building energy models. Energy Build 2017;156:17–28. http://dx.doi.org/10.1016/j.enbuild.2017.08.086.

[16] Priesmann J, Nolting L, Praktiknjo A. Are complex energy system models more accurate? An intra-model comparison of power system optimization models. Appl Energy 2019;255:1–28. http://dx.doi.org/10.1016/j.apenergy.2019.113783.

[17] Tian Z, Zhang X, Jin X, Zhou X, Si B, Shi X. Towards adoption of building energy simulation and optimization for passive building design: A survey and a review. Energy Build 2018;158:1306–16. http://dx.doi.org/10.1016/j.enbuild.2017.11.022.

[18] Robinson C, Dilkina B, Hubbs J, Zhang W, Guhathakurta S, Brown MA, et al. Machine learning approaches for estimating commercial building energy consumption. Appl Energy 2017;208:889–904. http://dx.doi.org/10.1016/j.apenergy.2017.09.060.

[19] Bourdeau M, qiang Zhai X, Nefzaoui E, Guo X, Chatellier P. Modeling and forecasting building energy consumption: A review of data-driven techniques. Sustainable Cities Soc 2019;48. http://dx.doi.org/10.1016/j.scs.2019.101533, URL https://doi.org/10.1016/j.scs.2019.101533.

[20] Foucquier A, Robert S, Suard F, Stéphan L, Jay A. State of the art in building modelling and energy performances prediction: A review. Renew Sustain Energy Rev 2013;23:272–88. http://dx.doi.org/10.1016/j.rser.2013.03.004.

[21] Xue X, Wang S, Sun Y, Xiao F. An interactive building power demand management strategy for facilitating smart grid optimization. Appl Energy 2014;116:297–310. http://dx.doi.org/10.1016/j.apenergy.2013.11.064.

[22] Lawrence TM, Boudreau MC, Helsen L, Henze G, Mohammadpour J, Noonan D, et al. Ten questions concerning integrating smart buildings into the smart grid. Build Environ 2016;108:273–83. http://dx.doi.org/10.1016/j.buildenv.2016.08.022.

[23] Ahcin P, Sikic M. Simulating demand response and energy storage in energy distribution systems. In: 2010 International conference on power system technology. IEEE; 2010, p. 1–7. http://dx.doi.org/10.1109/POWERCON.2010.5666564, URL http://ieeexplore.ieee.org/document/5666564/.

[24] Razmara M, Bharati GR, Shahbakhti M, Paudyal S, Robinett RD. Bilevel optimization framework for smart building-to-grid systems. IEEE Trans Smart Grid 2018;9(2):582–93. http://dx.doi.org/10.1109/TSG.2016.2557334, URL http://ieeexplore.ieee.org/document/7457675/.

[25] Beil I, Hiskens I, Backhaus S. Frequency regulation from commercial building HVAC demand response. Proc IEEE 2016;104(4):745–57. http://dx.doi.org/10.1109/JPROC.2016.2520640.

[26] Bampoulas A, Saffari M, Pallonetto F, Mangina E, Finn DP. A fundamental unified framework to quantify and characterise energy flexibility of residential buildings with multiple electrical and thermal energy systems. Appl Energy 2021;282:116096. http://dx.doi.org/10.1016/j.apenergy.2020.116096.

[27] Shamsi MH, Ali U, O'Donnell J. A generalization approach for reduced order modelling of commercial buildings. J Build Perform Simul 2019;12(6):729–44. http://dx.doi.org/10.1080/19401493.2019.1641554, URL https://www.tandfonline.com/doi/abs/10.1080/19401493.2019.1641554.

[28] Arteconi A. An overview about criticalities in the modelling of multi-sector and multi-energy systems. Environments 2018;5(12):130. http://dx.doi.org/10.3390/environments5120130.

[29] Šijanec M, Andraž Z, Stegnar RG, Summers C, Hulme J, 06 -Bre P, et al. Energy performance indicator tracking schemes for the continuous optimisation of refurbishment processes in European housing stocks monitor progress towards climate targets in European housing stocks main results of the EPISCOPE project-final project repo. Tech. rep., Institut Wohnen und Umwelt GmbH; 2016, URL www.iwu.de.

[30] Deru M, Field K, Studer D, Benne K, Griffith B, Torcellini P, et al. U.S. department of energy commercial reference building models of the national building stock. Tech. rep., National Renewable Energy Laboratory; 2011, URL http://www.osti.gov/bridge.

[31] National Renewable Energy Laboratory (NREL). Weather data sources | energyplus. In: Energyplus.Net. 2020, https://energyplus.net/weather/sourceshttps://energyplus.net/weather/sources{#}IWEC.

[32] Berthou T, Stabat P, Salvazet R, Marchio D. Development and validation of a gray box model to predict thermal behavior of occupied office buildings. Energy Build 2014;74:91–100. http://dx.doi.org/10.1016/j.enbuild.2014.01.038, URL https://www.sciencedirect.com/science/article/pii/S0378778814000760.

[33] Tabachnick BG, Fidell LS. Experimental designs using ANOVA. Belmont, CA: Thomson/Brooks/Cole; 2007, URL https://www.worldcat.org/title/experimental-designs-using-anova/oclc/71664023.

[34] Juhl R, Møller JK, Madsen H. CTSM-R continuous time stochastic modeling in R. Tech. rep., DTU Compute; 2016, URL http://ctsm.infohttp://arxiv.org/abs/1606.00242, arXiv:1606.00242.

[35] Kosindrdecha N, Daengdej J. A test case generation process and technique. J Softw Eng 2010;4(4):265–87. http://dx.doi.org/10.3923/jse.2010.265.287.

[36] Pramangioulis D, Atsonios K, Nikolopoulos N, Rakopoulos D, Grammelis P, Kakaras E. A methodology for determination and definition of key performance indicators for smart grids development in island energy systems. Energies 2019;12(2):1–22. http://dx.doi.org/10.3390/en12020242.

[37] Li Y, O'Donnell J, García-Castro R, Vega-Sánchez S. Identifying stakeholders and key performance indicators for district and building energy performance analysis. Energy Build 2017;155:1–15. http://dx.doi.org/10.1016/j.enbuild.2017.09.003.

[38] Ruiz GR, Bandera CF. Validation of calibrated energy models: Common errors. Energies 2017;10(10):1587. http://dx.doi.org/10.3390/en10101587, URL http://www.mdpi.com/1996-1073/10/10/1587.

[39] Landsberg DR, Shonder JA, Barker KA, Haberl JS, Judson SA, Jump DA, et al. Measurement of energy, demand, and water savings. In: ASHRAE guideline 14-2014. Tech. rep., 2014, ASHRAE; 2014, URL www.ashrae.orgwww.ashrae.org/technology.

[40] Armstrong P, Leeb S, Norford L. Control with building mass-Part I: Thermal response model. Ashrae Trans 2006;449–61.

[41] Wang L, Lee EW, Yuen RK. Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach. Appl Energy 2018;228:1740–53. http://dx.doi.org/10.1016/j.apenergy.2018.07.085.

[42] Ang YQ, Berzolla ZM, Reinhart CF. From concept to application: A review of use cases in urban building energy modeling. Appl Energy 2020;279:115738. http://dx.doi.org/10.1016/j.apenergy.2020.115738.

[43] Wang S, Wainer G, Goldstein R, Khan A. Solutions for scalability in building information modeling and simulation-based design. In: SimAUD 2013 conference proceedings: symposium on simulation for architecture and urban design. 2013, p. 1–8.

[44] Bacher P, Madsen H. Identifying suitable models for the heat dynamics of buildings. Energy Build 2011;43(7):1511–22. http://dx.doi.org/10.1016/j.enbuild.2011.02.005, URL https://www.sciencedirect.com/science/article/pii/S0378778811000491.

[45] Sweetnam T, Spataru C, Barrett M, Carter E. Domestic demand-side response on district heating networks. Build Res Inf 2018;1–14. http://dx.doi.org/10.1080/09613218.2018.1426314, URL https://www.tandfonline.com/doi/full/10.1080/09613218.2018.1426314.

[46] Vanhoudt D, Claessens B, Salenbien R, Desmedt J. The use of distributed thermal storage in district heating grids for demand side management. 2017, ArXiv abs/1702.06005.