Energy & Buildings 292 (2023) 113127

Contents lists available at ScienceDirect

Energy & Buildings

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

Thermal modeling of existing buildings in high-fidelity simulators: A novel, practical methodology

J.A. Borja-Conde [a,*], K. Witheephanich [b], J.F. Coronel [c], D. Limon [a]

[a] Department of Automatic Control and System Engineering, University of Seville, Seville, Spain
[b] Department of Electrical and Electronic Engineering, Munster Technological University, Cork, Ireland
[c] Department of Energy Engineering, University of Seville, Seville, Spain

ARTICLE INFO

Keywords:
Simplified building modeling
Data-based grey-box modeling
Building thermal model
Automated model identification
Building energy efficiency
Existing buildings
TRNSYS
GenOpt

ABSTRACT

Optimizing efficiency in the operation of the HVAC system of existing buildings requires the construction of a thermal dynamic model of the building, which may be challenging because architectural metadata may be missing or obsolete. Based on a suitable set of measured data, this paper presents a novel practical methodology to create and automatically derive thermal models of existing buildings in high-fidelity simulators for energy management. To this end, the philosophy of grey-box strategies is followed to simplify the modeling and avoid the requirement of architectural metadata, facilitating and expediting the process. First, a building model with a highly reduced number of parameters is constructed by exploiting the existing similarities in the materials of the buildings and simplifying their elements to a simple one-layer parameterization. Then, the parameters of the derived model are iteratively updated while minimizing the error between the real temperature evolution and that generated by the model being identified. For this purpose, data of the room air temperature, estimated occupancy, weather conditions, and variables of the HVAC system are assumed to be available in suitable zones of the building to apply the creation and identification processes of the model, allowing that a whole digital twin of the building is constructed. The methodology is presented by its application to a real case study: the Nimbus Research Centre building at Munster Technological University, located in Cork (Ireland). The high-fidelity simulator software TRNSYS is used for the modeling task, together with the GenOpt optimization program. The results demonstrate that the proposed methodology yields a highly accurate model of the building, capable of representing reality with RMSE values consistently below 0.6 °C during open-loop validation periods of up to four days. The findings suggest that this methodology may outperform other modeling techniques reported in the literature. Importantly, the proposed technique is less complex and time-consuming to implement than many of the alternatives.

1. Introduction

The energy consumed in buildings represents up to 40% of the total energy consumption in developed countries, of which HVAC (Heating, Ventilation and Air Conditioning) systems comprise about 50% [1,2]. Furthermore, buildings contribute to 36% of energy-related greenhouse gas emissions [3]. This has motivated international policies to improve and specifically in heating and cooling systems.[1]

To improve the energy efficiency of buildings, different tasks can be performed, such as improving the thermal insulation of buildings [6] or installing solar panels to produce clean energy for self-consumption, among others. However, these tasks require a significant investment and time. On the other hand, a feasible solution would be to apply advanced methodologies to operate HVAC systems, which may not require new equipment or a large investment and thus it is appropriate for existing buildings: In [7], the great potential of implementing advanced

* Corresponding author.
E-mail addresses: jaborja@us.es (J.A. Borja-Conde), kritchai.witheephanich@mtu.ie (K. Witheephanich), jfc@us.es (J.F. Coronel), dlm@us.es (D. Limon).

[1] For instance, the European Union has included the need to increase energy efficiency in buildings to reduce energy use in recent plans for the EU's climate transition [4], where heating and cooling are presented as one of the key factors [5].

https://doi.org/10.1016/j.enbuild.2023.113127
Received 20 February 2023; Received in revised form 21 April 2023; Accepted 27 April 2023
Available online 8 May 2023

management methodologies in existing buildings is highlighted, as existing buildings represent the vast majority of the entire building stock. In particular, it has been reported that profiting from new technological advances in energy management can reduce energy use in buildings from 13% to 28% on average [8]. Therefore, applying these techniques only to old buildings would lead to a great deal of energy savings that could be achieved in the short term.

However, the implementation of these advanced strategies is closely linked to the availability of an energy model of the building to take predictions into account, to check the viability of the management policies, etc.

In addition, the model could be used as a Digital Twin of the building if it has the capability not only to represent the real system, continuously updated with real-time system data, but also to allow faithful prediction of how the system will evolve [9], as long as high accuracy of the model is provided. Furthermore, it could also be used for other tasks, such as anomaly and fault detection, testing and training control systems before application in the real system, even dimensioning of a new HVAC system prior to installation if it had to be updated, etc. [8].

1.1. Literature review

For the aforementioned purpose, white-box (WB) models are one of the suitable popular options—as long as a good parametrization of the building is developed [10]. There are mature white-box modeling and simulation tools, highlighted by their high fidelity, such as EnergyPlus [11], Modelica [12], IDA ICE [13], eQUEST [14], or TRNSYS [15]. They have been widely used to model the thermal and energetic evolution of buildings [16–19].

These tools require a detailed description of the building and its construction techniques, needing the specification of a large set of architectural metadata parameters, such as layers, thickness, conductivity, capacity, density or convective coefficient of materials in walls, windows, etc. Therefore, they will provide high fidelity as long as the parameter values are close to reality. According to [20], there is evidence that a more detailed and complex model does not necessarily translate into a more accurate model of the real building as a result of increasing the quantity and accuracy of the required set of architectural design parameters. This is because design parameters are not real parameters due to the enormous range of particularities that exist in the manufacturing of materials, the construction of the building, its surrounding conditions, unexpected energy losses, etc. [10,21]. Therefore, it has been shown that, in order to enhance the model performance, it is not necessary to add more specific and particular architectural details, but to calibrate a suitable subset of parameters [22].

Because of this, there are several works in the literature that couple white-box modeling in high-fidelity (HF) simulators with a calibration process.

Some studies implement a calibration of energy exchanges due to air infiltration. In [23], a domestic building is modeled using TRNSYS, and the air infiltration change rate is calibrated, resulting in an improved simulation-based energy assessment. In [24], the authors have used TRNSYS to model a school center based on complete architectural metadata. They had the added difficulty that the buildings had defective window insulation, increasing energy losses due to infiltration. Then, they applied a deterministic calibration approach of the parameters corresponding to the infiltration, yielding a significant reduction of the thermal model error. In [25], an extensive work is carried out to model a public library in TRNSYS. The construction properties are set by the architectural metadata, while only infiltration values are used to calibrate the thermal model of the building. This study is notably extended in [26], where the calibration parameters of the building thermal model are not only the infiltration, but also the capacitance of each thermal zone.

Other studies focus on the calibration of the construction material properties, but constrain the possible results to an interval around the values provided by the manufacturers. In [27], a two-story test building is modeled in EnergyPlus in compliance with detailed architectural documentation. They set several calibration parameters related to internal gains, infiltration, construction materials, such as thickness or conductivity, etc., which are constrained to a maximum first design error of around 25%. A similar case study is developed in [28], where the building is modeled in IDA ICE. Similarly, in [22], an automated procedure for calibration is proposed starting from an initial detailed model, which is focused on power consumption. In [29], a historical building modeled in EnergyPlus is calibrated by comparing two different methods. The calibration parameters chosen in the study are conductivity, thermal and solar absorption, and specific heat of the walls. One method calibrates from air temperature measurements and the other method estimates using air as well as surface temperature measurements. Calibration leads to a significant reduction in model error for both methods, although the development of good initial models is mentioned as one of the key factors.

The works mentioned above require detailed and complete technical information extracted from architectural plans and metadata to set a starting point in the modeling and calibration process. Although this information is very useful, obtaining it is a tedious task, which limits the application of the methodology to other buildings [30]. Furthermore, problems become particularly challenging in relatively old buildings, where architectural metadata may not be available or not reliable.

To avoid this issue, another popular option is the use of black-box (BB) models instead. These are generated using input–output data in pure data-driven methods, disregarding physical relations or architectural metadata, and require a limited number of parameters and complexity [10,31]. To this end, there are a wide range of model structures suitable, such as linear regression (LR), neural networks (NN), support vector machine (SVM), etc., as reviewed in [10,32–34]. However, black-box techniques have clear disadvantages. For instance, the parameters do not usually have physical meaning—so they are not interpretable for building operators. Furthermore, they require long training and validation periods and are limited to building operation conditions covered during the training period [10], so that good accuracy will be obtained whenever a wide range of different operating scenarios are forced on the real system over long periods of time, which is not usually desirable or even possible.

In order to unify the advantages of both white-box and black-box models, grey-box (GB) modeling techniques are used, where the model structure is established from physical laws, while model parameters are identified from input-output data [31]. Traditionally, these methods are based on simple resistance-capacitance (RC) model structures, as reviewed in [32]. Often, the simplification of the modeling is one of their focus [35]. In comparison with black-box models, RC models have the advantages of being physically more interpretable and not requiring such a wide range of different operating scenarios. However, non-linear dynamics are not well modeled, and there is no consensus on the optimal model complexity, since lower-order models may not be able to catch the thermal dynamics, but in the same time, higher-order models may lead to be over-fitted to training data [36]. On the other hand, in comparison with white-box models, RC models are less arduous to develop and have fewer parameters, at the expense of lower accuracy and less representativeness of nonlinear dynamics.

More recently, in order to enhance the traditional grey-box model, a further and deeper rapprochement between grey-box and black-box modeling methods has been proposed. To this end, researchers introduce prior physical knowledge in more sophisticated model structures—typically used in black-box techniques—than RC ones. For example, in [37], a physics-informed linear regression model (PILR) is compared to machine learning methods, and the physics-informed model is concluded to be superior to the others. In [38], physically consistent neural networks (PCNN) were proposed, concluding that the proposed model clearly outperformed RC models. However, they presented it as a limitation that are only physically consistent with respect to control inputs

Table 1
Compilation of advantages and disadvantages of the different types of modeling and their corresponding structure model.

	White-Box	Black-Box	**Grey-Box**		
Feature	HF Sim.	NN, LR, SVM, etc.	RC	PCNN/PILR	**HF Sim.**
• Easy and quick modeling	✗	✓	✓	✓	✓
• Physically consistent	✓	✗	✓	✓	✓
• Reduced number of parameters	✗	~	~	~	✓
• Catch non-linear dynamics	✓	✓	✗	✓	✓
• Small number of scenarios for training	✓	✗	✓	✗	✓
• Good trade-off between complexity and accuracy	✗	✗	✗	~	✓

✓ It presents advantages in this regard. ✗ It presents disadvantages in this regard. ~ It depends on other aspects.

and exogenous temperatures; otherwise, they cannot guarantee the robustness of the model anymore [39]. Also, some exogenous factors and non-linear dynamics, such as catching the disturbances of solar gains through the windows or of the occupancy, are also presented as a difficulty. Lastly, similarly to [40], the quality of the solution can vary significantly should an initialization with unrealistic values be developed; that is, PCNNs do not always recover physically consistent parameters from data. Thus, some level of engineering insight is required to properly use the presented methodologies.

Other recently popular field of research to identify building thermal models automatically is symbolic regression, with promising results [41–43]. At its core, symbolic regression could be considered a black-box approach. However, it differs from other black-box methods because it is focused on discovering mathematical models and equations from data, which can provide interpretability by offering insights into the relationship between input and output variables. Nonetheless, symbolic regression is a relatively new technique in machine learning, and while it has shown promise in various fields, it is still under study and there are still some challenges to overcome, such as scalability to large datasets, overcoming overfitting, modeling non-linear dynamics, etc. [41].

Recapping the conclusions presented throughout this section and based particularly on [10,32,33,36,38,39,41], a compilation of the advantages and disadvantages of each type of modeling is shown in Table 1.

1.2. Contribution

In this paper a methodology to derive structured Grey-Box models in High-Fidelity Simulator (denoted as GB-HF Sim. in the rightmost column in Table 1) is presented. The model structure is constructed in high-fidelity simulators, such as TRNSYS or EnergyPlus (typically used for white-box modeling) but—motivated by the philosophy of grey-box strategies—simplifying the model and avoiding the requirement of architectural metadata. Notice that PCNN and PILR models, presented above, are aimed to get grey-box models from black-box model tools, while the proposed method in this work is also aimed to get a grey-box model, but from the White-Box model tools (i.e. high-fidelity simulators).

Grey-box RC models are also aimed at obtaining grey-box models but using simple resistance-capacitance structures of the building, which cannot describe the inherent non-linear dynamics. In contrast, the grey-box models using high-fidelity simulator, proposed in this paper, cope with this issue, thanks to the use of high-fidelity simulation tools. This, together with the proposed simplified modeling, allows one to achieve a very good trade-off between complexity of the model and accuracy, as it is demonstrated in the real case study.

Furthermore, thanks to the very nature of such simulators, a small number of operation scenarios for the training process is sufficient, contrary to the PCNN/PILR methods. This, together with the model simplifications considered in this paper, helps to avoid over-fitting problems.

Fig. 1. Nimbus Research Centre building.

In the proposed method, the model of the building is simplified by minimizing the number of parameters, thanks to a strategy that exploits the topology of the building, e.g., taking advantage of the similarities between rooms and seeking equivalent one-layer walls that represent the multi-layer ones. This strategy is comparable to that of the lumped parameter models [35,44] and the technique of clustering zones [45,46]. In addition, on the basis of this simplified building model, an automated parameter identification process[2] is performed using historical operation data, as does grey-box modeling, thus, without resorting to architectural metadata, unlike white-box methods. The only real parameters that are supposed to be known are primary information such as the main dimensions of the rooms and windows (height, width, and depth), location, orientation, etc.[3]

The proposed method is applied to a real case study: the Nimbus Research Centre building at Munster Technological University, located in Cork (Ireland), shown in Fig. 1. The building is modeled in the TRNSYS simulation program [15]. The validation results obtained show high fidelity of the model.

To the best knowledge of the author, this is the first combined use of these simulators with grey-box modeling methods, in addition to being applied to a real building.

The rest of the paper is organized as follows; First, the case study is presented in Section 2, which includes a description of its zones and layout, available data collected based on IoT technologies [47], and a proposed classification of zones for modeling. In Section 3, the setup of the model and the main objectives are presented. Then, in Section 4, the modeling and identification strategy is explained. A zone-level occupancy estimator is proposed in Section 5. Finally, the paper ends with

[2] Note that the term "calibration of parameters"—used in the mentioned works that couple detailed white-box models with a calibration process—is henceforth avoided, using "identification of parameters" instead, since, unlike the articles cited above, the initial values of the model are completely unknown.

[3] Note that this assumption would not be a drawback in comparison with typical grey-box modeling, as most of them also need this primary information, which is not counted as using architectural metadata [33].

Fig. 2. Second floor building layout regarding zones types. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Fig. 3. First floor building layout regarding zones types.

the presentation of the application in a real case study and the discussion of the corresponding final results in Section 6, followed by some conclusions in Section 7.

2. Case study description

The method proposed in this paper will be introduced by means of a real case study: the Nimbus Research Centre building at Munster Technological University, located in Cork (Ireland), shown in Fig. 1

2.1. Description of building zones and layout

The case study is a two-story building where there are several rooms for different uses, such as offices, meetings, seminars, etc. The building has been used as a test bed case study in several projects [48,49]. In it, a set of wireless sensor networks have been deployed to measure energy and thermal variables, which are suitable for modeling and control.

In Figs. 2 and 3 the layouts of the first and second floors, respectively, are shown. These are divided into zones, which are classified according to the set of measurements available for each one and with associated radiator control valve (RCV, allows for automated control) and thermostatic radiator valves (TRV, allows for manual control). The ones with more sensors are shown in color in the layout and are described in Section 2.3.

- Zone 1 (in red) is a kitchen.
- Zone 2 (in green) is a room where there are approximately thirty computer workstations.
- Zone 3 (in yellow) is an office for three or four people.
- Zone 4 (in blue) is an office for six or seven people.

The zones shown in white and gray are described in Section 2.3.

2.2. Installed sensors and available data

Some of the building zones have been equipped with sensors to obtain suitable measurements, based on IoT technologies [47], for energy monitoring and management. The types of sensors used are listed below.

- Zone air temperature sensor: Measures the dry bulb temperature in the zone where it is installed.
- Estimated occupancy: The number of people in a room is estimated by presence sensors together with CO_2 balances. Occupancy may be used to set other internal gains, such as computers and lighting.
- Heating system exchanges: Heat transferred from the HVAC system to the room is estimated by combining measurements of water flow and inlet and outlet temperatures in the fan coil.

Additionally, the following variables that affect the whole building are also measured:

- Outside air temperature sensor: Measures the temperature of the dry bulb in the outside air.
- Beam solar radiation: Direct radiation from the Sun on a horizontal surface is measured using a pyranometer.
- Diffuse solar radiation: Diffuse radiation from the Sun on a horizontal surface is measured using a pyranometer.

Radiation measurements must be combined with solar azimuth and zenith angles, using timestamp, location, and building orientation.

2.3. Zones classification in modeling

According to the availability of the variables measured in each zone, these are classified into four different groups of zones in terms of modeling:

- Full-Sensorized Zones (FS Zones): These are the zones where all energy variables are measured: Zone air temperature, estimated occupancy and heating system exchanges. Therefore, these are the zones that can be fully modeled. Zone 1 and Zone 2 comprise this group.
- Almost-Full-Sensorized Zones (AFS Zones): These are the zones where all energy variables but occupancy are measured: Zone air temperature and heating system exchanges. Since occupancy is not known, this could be estimated. Zone 3 and Zone 4 comprise this group.
- Low-Sensorized Zones (LS Zones): These are the zones where only zone air temperature is measured. These zones are represented by a white background color in Fig. 2.
- Non-Sensorized Zones (NS Zones): These are the zones where it is not possible to measure any variable. These zones are represented by a gray background color in Fig. 2.

3. Model setup and objectives

3.1. Building simplification

The development of a dynamic model of a building is typically a complex task, since it usually requires a precise description of every material and construction technique of the building, such as the layers of which the walls are composed and its materials, the type of windows and the glass properties, etc. In addition, their properties must be obtained from architectural metadata, which are typically difficult or even impossible to obtain. This is particularly hard in existing buildings, where technical information is missing or obsolete. Therefore, this methodology is not appropriate to obtain models for existing buildings, which is the main objective of this paper.

To this end, it is proposed to use a simplified building model based on some assumptions that are commonly addressed in practice. Under this simplification, the model obtained may provide less detailed results, but this allows one to estimate the evolution of the main variables of the building necessary for energy analysis and building management.

These assumptions are developed according to three main principles. The first is to avoid detailed building geometry, by assuming that all building elements can be modeled as uniform, without any irregularity (e.g., a wall that has protruding components is supposed to be smooth). The second is to seek simpler equivalent elements, which means that instead of modeling the elements layer-by-layer, they can be modeled by an equivalent single layer that encompasses all. And the third principle is to take advantage of similarities between the elements that form the building (e.g., windows, internal walls, external elements, as in walls or roofs, and ceilings between floors). These principles are closely related to the philosophy of grey-box models, as motivated in the Introduction.

For the proposed modeling process, the following assumptions are made, which are validated in Section 6:

Assumption 1. Building elements with multiple layers of construction are modeled as one-layer elements.

Assumption 2. Rooms are modeled with rectangles and smooth surfaces.

Assumption 3. Building elements are categorized into a reduced number of common groups. Within each group, all elements are supposed to have the same materials and layers. The only difference will be their dimensions (height and width).

Assumption 4. The natural air changes per hour due to air infiltration from outside the room are approximately the same in all rooms.

Assumption 5. The ratio between zone capacitance and zone volume is approximately the same in all rooms.

These five assumptions are generally applicable to most buildings, greatly simplifying the modeling process. As long as these assumptions are met, the methodology can be scaled. The objective is to obtain a simplified but analogous model to the real building to be used for energy analysis and building management.

For example, in the case study, according to Assumption 3 there are four groups: “Windows”, “Internal walls”, “External elements”, and “Ceilings”.

3.2. Identification of parameters of the building model

The set of parameters to identify are the ones of the one-layer elements, the natural air changes per hour, and the ratio between zone capacitance and volume. Since they are unknown, they are initially set as standard values. Then, the available data collected using the installed sensors from different experimental scenarios are used to identify them. From the zones to be identified, it is necessary to know the zone air temperature, the heat transferred from the HVAC system and the occupancy. In addition, it is necessary to know the zone air temperature of the adjacent zones to the identified one.

Since the zones that meet these requirements are the FS Zones, their parameters are identified and then analyzed to validate the fulfillment of Assumptions 1 to 5, as explained in detail in Section 4.3. This can be done provided that all the common group types defined in Assumption 3 are present in the FS Zones; otherwise, the missing groups cannot be identified.

Once the above assumptions are validated, the results of the identified parameters can be extrapolated to the entire building, as explained in Section 4.4.

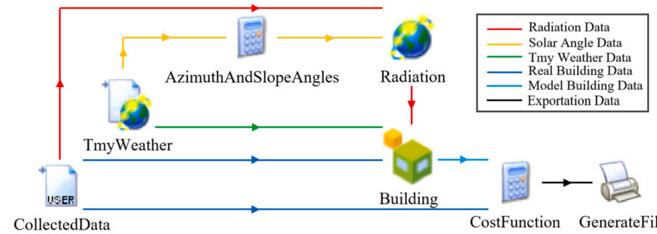

Fig. 4. TRNSYS project in TRNSYS Simulation Studio.

3.3. Estimation of occupancy

Once the building model has been identified and validated, this can be used to estimate those signals that are not measured. For instance, in the AFS zones, all significant variables are measured except occupancy. Based on the model, the occupancy of these zones can be estimated, as shown in Section 5.

4. Modeling and identification (based on TRNSYS)

The high-fidelity simulator software TRNSYS (TRaNsient SYstem Simulation program) [15] is used to model the building.[4] This software is compatible with SketchUp 3D modeling software [50], which facilitates the introduction of architectural information and dimensions of the building, as will be seen below.

Note that the features of the construction materials are unknown at this point. This means that they will first be set using generic materials.

4.1. TRNSYS project

In this project, the plugin for SketchUp TRNSYS3D [51] has been used to model the building in TRNSYS.[5]

Based on the layout and main dimensions of the rooms and windows (height, width and depth), the building can be modeled zone by zone. This then generates a building file to be directly imported into TRNBuild.

Once the building model is available in TRNBuild, a TRNSYS project is developed in the TRNSYS Simulation Studio. This project, shown in Fig. 4, has the following blocks:

- *CollectedData*: This is a Type9 block that reads the text file in which the collected data are located. In this text file, the data history of the values of the variables measured in the zones and outside the building—presented in Section 2.2—is found. The corresponding values for each iteration are the output of the block.
- *TmyWeather*: This is a Type15 block that reads a typical meteorological year file. It is used to complete weather data that have not been measured in the building.
- *Radiation*: This is a Type16 block that calculates the radiation on each wall orientation using the radiation measurements in the building and the solar azimuth and zenith angles provided in the typical meteorological year file.
- *Building*: This is a Type56 block that imports the TRNBuild file.
- *CostFunction*: This is an Equation Block where a cost function is calculated. It matches the cost function defined in Equation (1), which measures the discrepancy between the real temperature evolution collected in the zones and the simulated one.

[4] The use of TRNSYS is only a proposal without loss of generality. Other simulation programs, such as EnergyPlus or IDA ICE, could be used.

[5] Please note that the use of TRNSYS3D and SketchUp to model the building is optional. There are alternative—although more tedious—ways of developing the model directly using TRNBuild. The result would be equally valid.

- *GenerateFile*: This is a Type25 block that generates a text file in which the simulation results are saved. Specifically, the value of the cost function is saved.

4.2. Identification variables

The features of the referred layers and the capacitance-volume ratio in the TRNSYS model are set as decision variables in an optimization procedure to identify the model. However, it is necessary to determine exactly which identification variables to set during the process since a one-layer element may have several physical values that can be modified, e.g. thickness, thermal resistance, capacitance, etc.

Note that the chosen parameters must be compatible so that they do not result in redundancy that complicates the identification process. Since two different physical values can influence the same system variable, e.g. increasing thickness increases thermal resistance, only one of them should be set as a identification variable.

Under these considerations, the proposed identification parameters are as follows.

- Ratio between zone volume and capacitance.
- Thickness of the external wall and roof.
- Thickness of the internal wall and the ceiling between levels.[6]
- Density of the external wall and roof.
- Density of the internal wall and the ceiling between levels.[6]
- Natural air changes per hour due to infiltration of outside air.
- Convective coefficient of the external wall and roof.
- Convective coefficient of the internal wall and the ceiling between levels.[6]

Thickness and density of the material are chosen since the former will allow the thermal resistance to be adjusted and the latter will allow the capacity. Alternatively, although less recommended, the identification parameters corresponding to the material density could be replaced by the material convective coefficient, even though the resulting thickness will be different. In this case, the convective coefficient would adjust the thermal resistance, and the thickness would adjust the capacity.

It must be taken into account that the identification parameters must be adapted according to the common groups of building elements as specified in Assumption 3.

The parameters of the model are assumed to be time invariant. Although it would be more desirable to define some of them with variable values during simulation, for instance, the natural air changes per hour and the convective coefficient, it has been assumed that this is an acceptable error, since the complexity of the optimization problem would increase exponentially. This assumption, traditionally accepted in grey-box modeling [21,52], will be validated by checking that the results are good enough and that the resulting values of these variables are not very sensitive depending on the results of the other optimization variables.

4.3. Parameter identification based on GenOpt

Once the TRNSYS model with generic parameters has been deployed according to Assumptions 1 to 5, and the identification variables have been defined, the appropriate values of these variables are identified using the available data collected with the installed sensors.

4.3.1. Identification process

In the identification process the parameters of the derived model are iteratively updated while minimizing the error between the real temperature evolution and that generated by the model being identified. Should the error be small enough, the resulting layer may be considered to have the same impact on the zone air temperature as the real element that it replaces.

Since there are several full-sensorized zones, this redundancy can be exploited in the identification procedure considering the following scenarios.

Scenario 1: An identification process is applied independently to each FS Zone in order to reinforce the validation of the assumptions made, specifically Assumptions 3, 4 and 5. Then, different identified values for the same materials will be obtained in each of them. This allows one to check the congruence between these results: If the identified parameters for the same materials are similar, then the assumptions are probably correct; otherwise, some reconsiderations should be made, like reviewing the categorization of building elements (for example, to analyze if adding another kind of building element is necessary), or increasing the data set used, etc. See Remark 1.

Scenario 2: When Scenario 2 is performed, a identification process is applied simultaneously to all FS Zones, while sharing the variables between zones of the corresponding common building elements in order to obtain the intermediate values that most resemble reality. In the same way as before, it will be necessary to prove the congruence between the results of the application jointly and independently. The analysis of the resultant error between the evolution of the air temperature of the model zone and the real building is another validation method that will be developed.

It is important to note that while one or more zones are identified, all external variables, except the air temperature of the identified zones, are set to the corresponding ones in the real system using the collected data described in Section 2.2.

Remark 1. The Identification process applied independently to each FS Zone (Scenario 1) is an optional step used to reinforce the validation of the assumptions made by proving congruence. It might be possible to avoid performing this step and just develop the Identification process applied simultaneously to all FS zones (Scenario 2). The assumptions are valid provided that the error between actual and simulated temperature evolution is small enough.

In order to evaluate the performance of the resulting model, a cost function is defined to measure the estimation error. The function corresponds to the root mean square error (RMSE)—a performance index commonly found in the literature, as used by [53]—between the evolution of real building measurements and the temperatures based on the simulation of the model. The cost function is expressed in Equation (1).

$$RMSE_p(^\circ\mathrm{C}) = \sqrt{\frac{\sum_{i=1}^{N}(T_{real,i} - T_{estimated,i})^2}{N}}, \tag{1}$$

where $T_{real,i}$ is the value of the real temperature of the zone at sample i expressed in degrees Celsius (°C), and $T_{estimated,i}$ is the corresponding estimate. N is the total number of iterations corresponding to a determined operation period p.

4.3.2. Identification process in TRNSYS with GenOpt

Once the identification strategy, its variables, and its performance evaluation cost function are defined, performing the identification procedure based on the resulting TRNSYS project is not immediate, as TRNSYS does not allow the physical values of the elements to be changed as user input in an automated way. This hinders the iterative process required in the identification.

However, this issue is solved thanks to the GenOpt optimization program [54]. GenOpt can automatically change the TRNSYS configuration files externally to set suitable values of the parameters, execute a simulation test and read the file with the simulation results from where the cost function is obtained (see Fig. 5).

GenOpt program receives the identification parameters and a template equal to the configuration files of the TRNSYS project. Note that

[6] For simplicity, features of internal walls and ceilings are lumped.

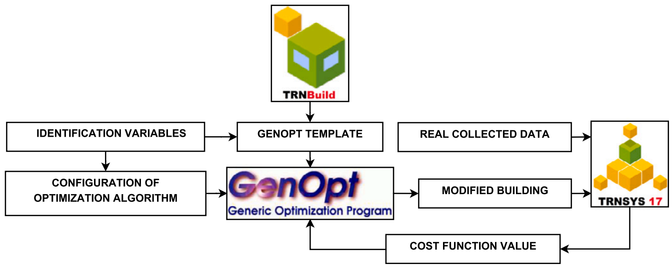

Fig. 5. Connection diagram between TRNSYS and GenOpt.

the values of the identification parameters are set with a label instead of its number, which has to be developed previously. Thus, GenOpt can insert the corresponding values in each iteration and send the file to TRNSYS for execution.

When the execution in each iteration is complete, TRNSYS will generate a file in which the value corresponding to the cost function is located. Thus, GenOpt can adjust the parameters, i.e., the decision variables, and obtain the value of the associated cost function. Based on this, GenOpt can iteratively take appropriate values of the parameters in order to minimize the cost function. For this, GenOpt allows the user to choose an optimization algorithm from a wide collection available.

4.3.3. Training and validation processes

The identification process is divided into two phases:

- Training Phase: From the available data of the real building, a subset is taken to carry out the identification of the parameters. These are the so-called *training data*. Note that the identified parameters are those that make the model fit better to the training data.
- Validation Phase: In this phase, the identified model is validated using the so-called *validation data*, which corresponds to a time period different from the training data. To do this, an open-loop simulation of the model with the identified parameters needs to be performed, where the only values to be set are the initial temperatures of the zones being simulated. The identified model is considered to be valid if its simulated evolution fits sufficiently well with the validation data.

This procedure can be applied to both identification scenarios: when the identification is performed independently to every FS Zone and when this is done to all the FS Zones together.

If the validation process was not satisfactory, i.e. the identified model did not fit the real one, then, there would be a need to reconsider some aspects, such as the categorization of building elements (for example, to consider adding other kinds of building elements) or increasing the training data set according to a suitable analysis of the residuals.

4.4. Modeling the whole building

Once Assumptions 1 to 5 are validated for FS Zones, the entire building can be modeled using the resulting identified parameters in those zones. In virtue of Assumption 3, all groups of building elements have the same materials and layers, so the materials identified in FS Zones will match with those of the other zones (AFS, LS and NS Zones). And analogously, because Assumptions 4 and 5 specify that natural air changes per hour due to infiltration and that the ratio between zone capacitance and volume is the same in all zones, so infiltration and zone capacitance can be calculated. The result is that the entire building is modeled and identified.

5. Data-based occupancy estimation

Based on the identified model of the building, it is possible to design an estimator to determine the values of missing variables of interest that have not been measured.

For instance, in the AFS Zones the occupancy is the only variable of interest that has not been measured, and this could be estimated. In this case, the temperature measurements of some surrounding zones are missing (about 30%, due to NS zones), but they can be approximated by the measured temperature of its neighbors for estimation purposes, since the contribution of this error to the evolution of the temperature of the AFS-zone is negligible compared to the effect of the occupancy.[7]

5.1. Occupancy estimator

In order to demonstrate this, a simple estimation policy has been used. In this case, the parameter to be estimated, the occupancy, is calculated as the linear combination of the estimation error and the accumulated estimation error. The estimation error is the mismatch between the real measured temperature and the estimated one. This is described as follows.

$$error_i = T_{real,i} - T_{model,i}, \tag{2}$$

$$Occ_i = K_1 \cdot error_i + K_2 \cdot \sum_{k=0}^{i} error_k, \tag{3}$$

where $T_{real,i}$ is the value of the real temperature of the zone and $T_{model,i}$ is the value of the simulated temperature of the zone with the estimated occupancy Occ_{i-1} at sample i. K_1 and K_2 are the suitable gains.

5.2. Validation of the estimated occupancy

A byproduct of the occupancy estimator is the possibility of implementing an additional step to reinforce the validation of Assumptions 3, 4 and 5, that is, the extension of the identified parameters in FS Zones to the entire building. This can be done, for example: (i) by analyzing its consistency with the use of the building, e.g., checking if the resulting estimated occupancy is coherent with the occupancy schedule of offices and the number of people who usually occupy the room), (ii) by comparing it with the occupancy measurements in other zones, etc.

6. Application to the case study and results

The proposed methodology is applied to a real case study: the Nimbus Research Centre building at Munster Technological University, located in Cork (Ireland), shown in Fig. 1.

In this case study, by exploiting the similarities between building elements, according to Assumption 3, four common groups can be considered: Windows, internal walls, external elements (walls and roofs) and ceilings between floors. Then the whole model is characterized by only the six parameters described in Table 2.

The available data for modeling have been collected for 13 days in November, since this is a month when the HVAC system is needed to be operating and the weather is less regular, so the operating range is greater.

[7] In addition, note that an air-conditioned building has all the zone temperatures within a small range, so temperature jump between boundary zones will always be less than 5 °C approximately, whereas the jump between inside and outside temperatures is going to be clearly higher. In this case, their disturbances can be ignored.

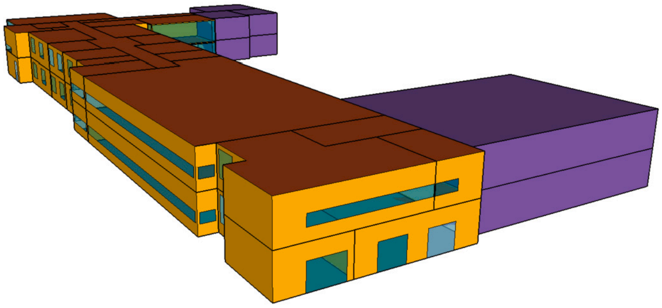

Fig. 6. 3D model in *SketchUp* using *TRNSYS3D* plugin.

6.1. TRNSYS project development

As explained in Section 4.1, TRNSYS3D is used to model the building. Then, the construction elements are drawn with SketchUp, and, next, TRNSYS3D automatically sets the types of construction by analyzing their positions and shapes (roofs, floors, internal and external walls, ceilings, windows, etc.). It is important to note that these types must be classified according to Assumption 3 (common groups of building elements). According to these types, elements are assigned generic materials of a provided template (in the case study, a generic Irish template is used). The features of these materials are not manually changed to match those of the actual building since, initially, the materials are assumed to be unknown and are set as random generic materials. Then, a feature of TRNSYS3D is used for surface matching, automatically identifying the adjacent elements between zones, their external elements, the boundaries, etc. This is essential information when analyzing the thermal evolution of the building. The resulting building is shown in Fig. 6.

Once the building is modeled in SketchUp and imported into TRNBuild, the resulting file is set in the building block (Type56) of the TRNSYS project in TRNSYS Simulation Studio, Fig. 4.

Then, the files to be set in the TRNSYS Simulation Studio would be the collected data file (Type 9)—which contains the sensors measurements at the Nimbus Center—and the typical meteorological year file (Type 15)—which contains weather data collected at Cork Airport [55]. The sample time for the data and simulation is 10 minutes.

To simulate internal gains of the building, the following considerations are taken:

- *Persons*: The number of people is set using the estimated occupancy data (done by presence sensors together with CO_2 balances). The energy gain of each person in the room is set according to ISO 7730:2005 [56] with the type of activity *seated, light work, typing.*
- *Computers*: The number of computers running in the offices is equal to the number of people in the room. The computer power is set to 230 W according to the TRNSYS documentation.
- *Artificial lighting:* the lights are on wherever there is someone in the room. Consumption is 13 W/m^2 (EVG direct).
- *Heating power*: Heat transferred from the HVAC system to the room (estimated by combining measurements of water flow and inlet and outlet temperatures in the fan coil) is set as a convective gain.

6.2. Results of parameter identification using full-sensorized zones

The identification problem is performed according to a hybrid generalized pattern search algorithm with a particle swarm optimization algorithm, available in GenOpt [54]. The data collected available are divided into a nine-day period for the Training Phase (from 6 to 15 of November), and a four-day period for the Validation Phase (from 15 to 19 of November).

6.2.1. Resulting parameter values

As mentioned in Section 3.2, to identify the features of the equivalent one-layer elements and the ratio between zone capacitance and volume, the FS zones are used. And according to Section 4.3, the parameter identification process is applied both independently to each FS Zone and simultaneously to all FS Zones to analyze the congruence between all the results, as follows.

Scenario 1: In Table 2 the resulting values of the identification process for each FS Zone are shown in the first two columns (*Zone 1* and *Zone 2*). Taking into account that it is possible to prove congruence if results of the zones have similar values for the same and equivalent parameters, it can be checked that the differences between the results are small enough to consider that they are congruent. Therefore, Assumptions 1 to 5 would be validated in this case. Specifically Assumption 3, where all elements of each defined group are supposed to have the same materials and layers, and Assumptions 4 and 5, where all zones have the same natural air changes per hour due to infiltration and the same ratio between zone capacitance and volume.

Scenario 2: Once the FS Zones have been identified and verified for congruence, the next step is to identify these zones with the same parameters simultaneously to obtain intermediate values that minimize the sum of cost functions of each zone. The results are shown in Table 2 in fourth column (*Zones 1 and 2 together*), in comparison with the mean values between the Zone 1 and Zone 2 independently, shown in third column ($(Zone1 + Zone2)/2$). Again, it can be checked that the differences between the results are small enough to consider them congruent.

6.2.2. Zone temperature evolution fitting

In the modeling and identification process, there are two types of results to show: (i) model response after the Training Phase is applied and (ii) model response when the Validation Phase is performed.

To compare the results between simulations, the cost according to Equation (1) is taken into account.

The results shown in this section are those corresponding to Scenario 2 (modeling and identification of both FS Zones simultaneously).

Zone 1: Results corresponding to Training Phase are shown in Fig. 7. In the upper plot, the real zone temperature evolution is shown in red, and the simulated one is shown in blue. In the lower plot, the main heat gains of the real building are shown (heating system exchange is shown in red, and occupancy is shown in blue). The maximum error between the temperature of the real and simulated zone is less than 1 °C, the mean absolute error (MAE) is 0.3 °C, and the median absolute deviation is 0.26 °C, with a $RMSE_{9days}$ value of 0.381 °C. Similarly, in Fig. 8 the results are shown when the Validation Phase is applied. The maximum error between the temperature of the real and simulated zone is less than 1 °C, the mean absolute error is 0.27 °C, and the median absolute deviation is 0.25 °C, with a $RMSE_{4days}$ value of 0.349 °C.

Zone 2: Obtained results are similar to those of Zone 1, as shown in Fig. 9, for Training Phase, and Fig. 10, for Validation Phase. In the first, the maximum error between temperatures is less than 1.2 °C, the mean absolute error is 0.36 °C, and the median absolute deviation is 0.33 °C, with a $RMSE_{9days}$ value of 0.449 °C. In validation, the maximum error is less than 1.5 °C, the mean absolute error is 0.46 °C, and the median absolute deviation is 0.38 °C, with a $RMSE_{4days}$ value of 0.587 °C.

Therefore, it can be concluded that the resulting error between the temperature evolution of the real data and those generated with the identified model is small enough ($RMSE_{4days} < 0.6$ °C). In white-box models, for example, in [53], although with a seven-day validation period, using a highly more complex model and starting from architectural drawings and fabrication details available, the RMSE value (0.27 °C$\leq RMSE_{7days} \leq$1.5 °C) is comparable to that of this paper. In [57], for a two-day period, with a totally detailed model in terms of architectural metadata, the RMSE value ($RMSE_{2days} \leq$1.59 °C) is also of the same order. On the other hand, for PCNN and RC models, for example in [38], the mean absolute error is 0.88 °C and 1.48 °C, respectively,

Table 2
Identification results. Values of parameters.

	Initial	Scenario 1			Scenario 2 - Final
Parameter	Random Values	Zone 1	Zone 2	$(Zone1 + Zone2)/2$	Zones 1&2 together
Ratio Capacitance/Volume ($kJ/m^3 \cdot K$)	5.00	28.52	23.56	26.04	26.09
Thickness External Elem. (cm)	0.50	0.110	0.125	0.118	0.093
Convect. Coef. Ext. Elem. ($W/m^2 \cdot K$)	50.0	14.34	19.10	16.73	15.33
Thickness Internal Elem. (cm)	0.50	0.123	0.130	0.127	0.128
Convect. Coef. Int. Elements ($W/m^2 \cdot K$)	50.0	23.31	21.90	22.64	21.63
Natural air changes per hour (l/h)	0.50	0.175	0.198	0.188	0.180
Open-loop $RMSE_{9days}$ in training (°C)		0.377	0.433	0.405	0.415
Open-loop $RMSE_{4days}$ in validation (°C)		0.360	0.562	0.461	0.468

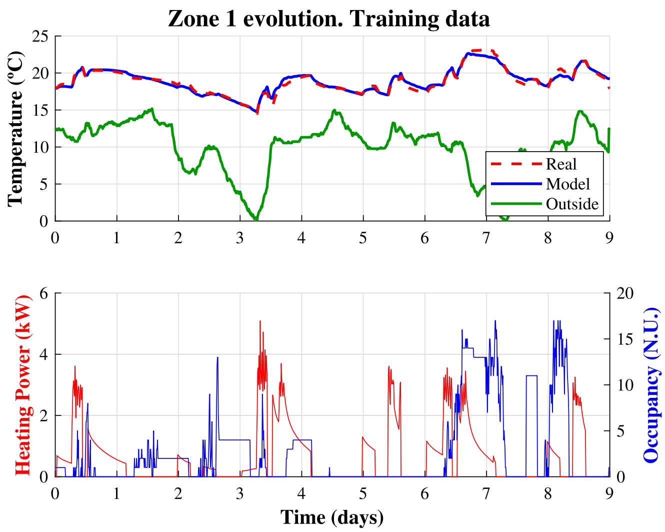

Fig. 7. Identified vs. Real model according Training Phase: Open-loop simulation from 6 to 15 of November. Zone 1.

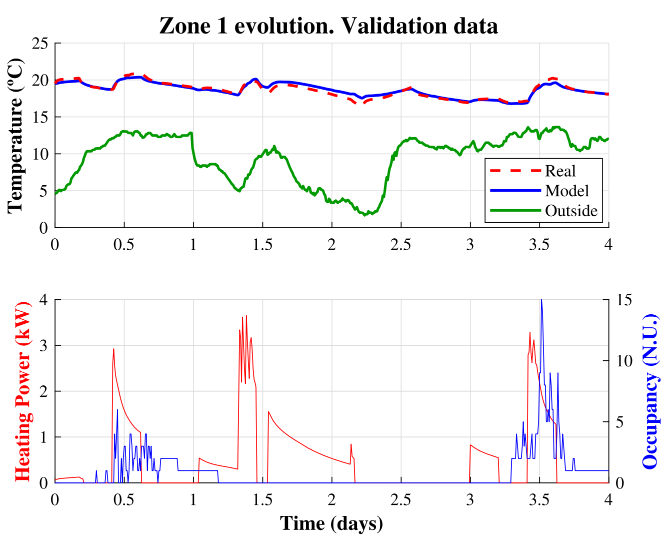

Fig. 8. Identified vs. Real model according Validation Phase: Open-loop simulation from 15 to 19 of November. Zone 1.

for a three-day validation period, similar to that obtained in this work (0.36 °C for a four-day open-loop validation period).

6.3. Results of estimation of occupancy of AFS zones

In this case, the evolution of the resulting estimated occupancy across the entire time period must be analyzed. To do that, the known

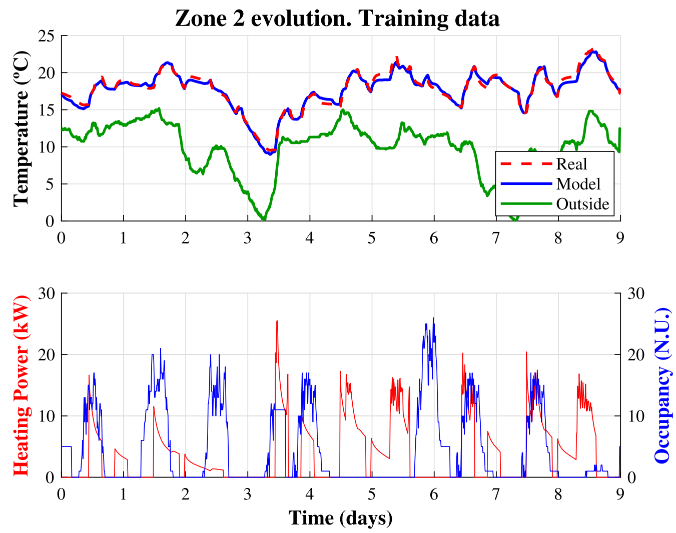

Fig. 9. Identified vs. Real model according Training Phase: Open-loop simulation from 6 to 15 of November. Zone 2.

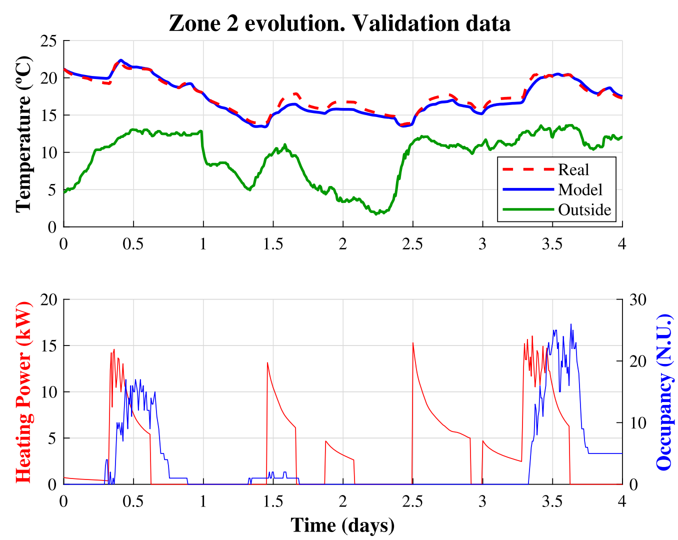

Fig. 10. Identified vs. Real model according Validation Phase: Open-loop simulation from 15 to 19 of November. Zone 2.

real occupancy of other zones is used to compare them with the identified ones, along with the number of people who usually occupy the zones.

Zone 3: The results are shown in Fig. 11. In the upper plot, the real zone temperature evolution is shown in red, and the simulated one is shown in blue. As can be seen, the estimator updates the occupancy in order to fit the simulated temperature with the real one. In the lower

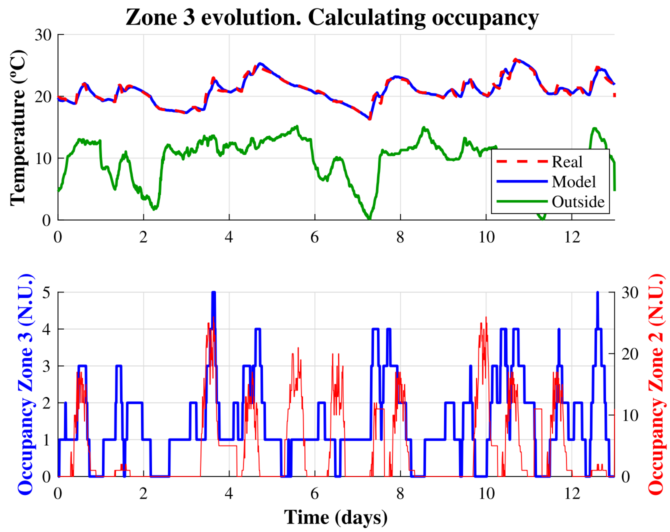

Fig. 11. Extrapolated model vs. Real model when occupancy is calculated. Zone 3.

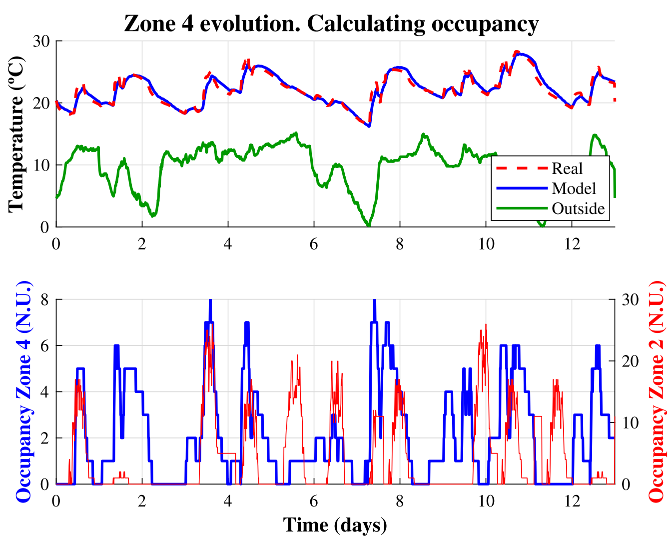

Fig. 12. Extrapolated model vs. Real model when occupancy is calculated. Zone 4.

plot, the identified occupancy of the target zone is shown in blue (left axis), and the real occupancy of Zone 2 is shown in red (right axis). It can be observed that the number of people usually occupying the zone is three or four, and the peaks of the identified occupancy fit the corresponding peaks of Zone 2, as expected. Also, people in Zone 3 and Zone 2 arrive and leave the building at similar times, and the controller never sets a *negative occupancy*, which is physically impossible.

Zone 4: Obtained results are similar to those of Zone 3, as shown in Fig. 12, but with 6 people usually occupying the room instead of 3.

7. Conclusions

Consistent with the need to develop easily extensible methodologies to model existing buildings, this paper presents a novel practical methodology to create and automatically derive thermal models of existing buildings developed in high-fidelity simulators for energy management and without resorting to architectural metadata.

Following the philosophy of grey-box modeling, the objective is not to model the real building, but to model a simplified but analogous one, which uses only a highly reduced number of parameters, facilitating and expediting the process. Thus, the building has been modeled avoiding the use of detailed building geometry, seeking simpler equivalent elements, and taking advantage of similarities between zones. For modeling and simulating the building, a high-fidelity program, traditionally used for white-box modeling, TRNSYS, is used.

The identification of the model parameter from the available data has been done by means of the optimization program GenOpt combined with TRNSYS. For this purpose, measurements of air temperature, estimated occupancy, weather conditions and variables of the HVAC system are assumed to be available in some rooms to apply the creation and identification processes of the building model, resulting in that the energy features of the real building are identified.

The proposed methodology is applied to a real case study—the Nimbus Research Centre building at Munster Technological University—obtaining an accurate model which is able to represent reality with an $RMSE_{4days}$ value always less than 0.6 °C. These results—perfectly acceptable according to the literature—validate this methodology.

Furthermore, an estimator is proposed to determine the values of missing occupancy measured of some rooms, resulting in an estimate very similar to what is expected, thus increasing the level of confidence in the resulting model.

The overall results demonstrate that the identified model proposed in this work can replicate the real building and predict its thermal and energy evolution. This allows a digital twin of the building with sufficient predictive power to be constructed and used by a higher-level controller such as MPC to optimize the efficiency in the operation of the building. Implementing those controllers in the building could be a future work to carry out.

Future research should explore the potential of combining the proposed methodology with machine learning techniques, such as Symbolic Regression and Neural Networks, to further enhance the accuracy and robustness of building thermal behavior modeling.

CRediT authorship contribution statement

Jose Antonio Borja: Conceptualization, Methodology, Software, Investigation, Validation, Data curation, Writing and Editing Original and Revised draft preparation. **Kritchai Witheephanich:** Resources, Supervision, Reviewing. **Juan Francisco Coronel:** Reviewing. **Daniel Limón:** Conceptualization, Resources, Supervision, Reviewing.

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.

Data availability

The real data used in this study was obtained through the TOPAs project [47]. For information on the resources used in this work, please contact the corresponding author.

Acknowledgement

This work was supported in part by the Junta de Andalucía and the European Regional Development Fund (ERDF) under Grant P20_00546, by the Agencia Estatal de Investigación (AEI) under Grant PID2019-106212RB-C41/AEI/10.13039/501100011033 and by MU-Spain under Grant FPU20/02023. The authors wish to acknowledge the support of the European Unions Horizon 2020 research and innovation program under grant agreement No. 676760 - TOPAs in providing access to the dataset used in this work. European Union's programme Erasmus+ supported the collaboration that led to this work [Erasmus+ 2021-22].

References

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

[2] M. González-Torres, L. Pérez-Lombard, J.F. Coronel, I.R. Maestre, D. Yan, A review on buildings energy information: trends, end-uses, fuels and drivers, Energy Rep. 8 (2022) 626–637, https://doi.org/10.1016/j.egyr.2021.11.280.

[3] European Commission and Directorate-General for Communication, Making our homes and buildings fit for a greener future, Publications Office 2021.

[4] European Council, "Fit for 55": Council agrees on higher targets for renewables and energy efficiency, Consilium 2022, https://europa.eu/!f6vqPd, 2022.

[5] European Commission, Heating and cooling. Topics, https://energy.ec.europa.eu/topics/energy-efficiency/heating-and-cooling_en, 2022.

[6] L. Aditya, T. Mahlia, B. Rismanchi, H. Ng, M. Hasan, H. Metselaar, O. Muraza, H. Aditiya, A review on insulation materials for energy conservation in buildings, Renew. Sustain. Energy Rev. 73 (2017) 1352–1365, https://doi.org/10.1016/j.rser.2017.02.034.

[7] N. Hashempour, R. Taherkhani, M. Mahdikhani, Energy performance optimization of existing buildings: a literature review, Sustain. Cities Soc. 54 (2020) 101967, https://doi.org/10.1016/j.scs.2019.101967.

[8] J. Drgoňa, J. Arroyo, I. Cupeiro Figueroa, D. Blum, K. Arendt, D. Kim, E.P. Ollé, J. Oravec, M. Wetter, D.L. Vrabie, L. Helsen, All you need to know about model predictive control for buildings, Annu. Rev. Control 50 (2020) 190–232, https://doi.org/10.1016/j.arcontrol.2020.09.001.

[9] F. Bordeleau, B. Combemale, R. Eramo, M. van den Brand, M. Wimmer, Towards model-driven digital twin engineering: current opportunities and future challenges, in: Ö. Babur, J. Denil, B. Vogel-Heuser (Eds.), Systems Modelling and Management, Springer International Publishing, Cham, 2020, pp. 43–54.

[10] M. Killian, M. Kozek, Ten questions concerning model predictive control for energy efficient buildings, Build. Environ. 105 (2016) 403–412, https://doi.org/10.1016/j.buildenv.2016.05.034.

[11] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W. Buhl, Y. Huang, C.O. Pedersen, R.K. Strand, R.J. Liesen, D.E. Fisher, M.J. Witte, J. Glazer, EnergyPlus: creating a new-generation building energy simulation program, Energy Build. 33 (4) (2001) 319–331, https://doi.org/10.1016/S0378-7788(00)00114-6.

[12] M. Wetter, Modelica-based modelling and simulation to support research and development in building energy and control systems, J. Build. Perform. Simul. 2 (2) (2009) 143–161, https://doi.org/10.1080/19401490902818259.

[13] N. Björsell, A. Bring, L. Eriksson, P. Grozman, M. Lindgren, P. Sahlin, A. Shapovalov, M. Vuolle, IDA indoor climate and energy, in: Proc. of the 6-th IBPSA Conference, 1999, pp. 1035–1042.

[14] J.J. Hirsch, et al., eQUEST introductory tutorial, version 3.64, Camarillo, CA, 2010.

[15] S. Klein, et al., TRNSYS 17 Manual: A TRaNsient System Simulation Program, Solar Energy Laboratory, Univ. of Wisconsin-Madison, Madison, USA, 2014, http://sel.me.wisc.edu/trnsys.

[16] X. Li, J. Wen, Review of building energy modeling for control and operation, Renew. Sustain. Energy Rev. 37 (2014) 517–537, https://doi.org/10.1016/j.rser.2014.05.056.

[17] J. Santos-Herrero, J. Lopez-Guede, I. Flores-Abascal, Modeling, simulation and control tools for nZEB: a state-of-the-art review, Renew. Sustain. Energy Rev. 142 (2021) 110851, https://doi.org/10.1016/j.rser.2021.110851.

[18] A. Alwisy, S. BuHamdan, M. Gül, Evidence-based ranking of green building design factors according to leading energy modelling tools, Sustain. Cities Soc. 47 (2019) 101491, https://doi.org/10.1016/j.scs.2019.101491.

[19] F. Bano, V. Sehgal, Finding the gaps and methodology of passive features of building envelope optimization and its requirement for office buildings in India, Therm. Sci. Eng. Prog. 9 (2019) 66–93, https://doi.org/10.1016/j.tsep.2018.11.004.

[20] E. Fabrizio, V. Monetti, Methodologies and advancements in the calibration of building energy models, Energies 8 (4) (2015) 2548–2574, https://doi.org/10.3390/en8042548.

[21] G. Reynders, J. Diriken, D. Saelens, Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals, Energy Build. 82 (2014) 263–274, https://doi.org/10.1016/j.enbuild.2014.07.025.

[22] T. Yang, Y. Pan, J. Mao, Y. Wang, Z. Huang, An automated optimization method for calibrating building energy simulation models with measured data: orientation and a case study, Appl. Energy 179 (2016) 1220–1231, https://doi.org/10.1016/j.apenergy.2016.07.084.

[23] C. Aparicio-Fernández, J.L. Vivancos, P. Cosar-Jorda, R.A. Buswell, Energy modelling and calibration of building simulations: a case study of a domestic building with natural ventilation, Energies 12 (17) (2019) 3360, https://doi.org/10.3390/en12173360.

[24] A. Ogando, N. Cid, M. Fernández, Energy modelling and automated calibrations of ancient building simulations: a case study of a school in the northwest of Spain, Energies 10 (6) (2017) 807, https://doi.org/10.3390/en10060807.

[25] A. Cacabelos, P. Eguía, J.L. Míguez, E. Granada, M.E. Arce, Calibrated simulation of a public library HVAC system with a ground-source heat pump and a radiant floor using TRNSYS and GenOpt, Energy Build. 108 (2015) 114–126, https://doi.org/10.1016/j.enbuild.2015.09.006.

[26] A. Cacabelos, P. Eguía, L. Febrero, E. Granada, Development of a new multi-stage building energy model calibration methodology and validation in a public library, Energy Build. 146 (2017) 182–199, https://doi.org/10.1016/j.enbuild.2017.04.071.

[27] V. Monetti, E. Davin, E. Fabrizio, P. André, M. Filippi, Calibration of building energy simulation models based on optimization: a case study, in: 6th International Building Physics Conference, IBPC 2015, Energy Proc. 78 (2015) 2971–2976, https://doi.org/10.1016/j.egypro.2015.11.693.

[28] C. Cornaro, F. Bosco, M. Lauria, V.A. Puggioni, L. De Santoli, Effectiveness of automatic and manual calibration of an office building energy model, Appl. Sci. 9 (10) (2019) 1985, https://doi.org/10.3390/app9101985.

[29] M. Giuliani, G.P. Henze, A.R. Florita, Modelling and calibration of a high-mass historic building for reducing the prebound effect in energy assessment, Energy Build. 116 (2016) 434–448, https://doi.org/10.1016/j.enbuild.2016.01.034.

[30] H. Gao, C. Koch, Y. Wu, Building information modelling based building energy modelling: a review, Appl. Energy 238 (2019) 320–343, https://doi.org/10.1016/j.apenergy.2019.01.032.

[31] T. Hauge Broholt, M. Dahl Knudsen, S. Petersen, The robustness of black and grey-box models of thermal building behaviour against weather changes, Energy Build. 275 (2022) 112460, https://doi.org/10.1016/j.enbuild.2022.112460.

[32] Y. Li, Z. O'Neill, L. Zhang, J. Chen, P. Im, J. DeGraw, Grey-box modeling and application for building energy simulations - a critical review, Renew. Sustain. Energy Rev. 146 (2021) 111174, https://doi.org/10.1016/j.rser.2021.111174.

[33] S. Zhan, A. Chong, Data requirements and performance evaluation of model predictive control in buildings: a modeling perspective, Renew. Sustain. Energy Rev. 142 (2021) 110835, https://doi.org/10.1016/j.rser.2021.110835.

[34] Z. Wang, Y. Chen, Data-driven modeling of building thermal dynamics: methodology and state of the art, Energy Build. 203 (2019) 109405, https://doi.org/10.1016/j.enbuild.2019.109405.

[35] A.P. Ramallo-González, M.E. Eames, D.A. Coley, Lumped parameter models for building thermal modelling: an analytic approach to simplifying complex multi-layered constructions, Energy Build. 60 (2013) 174–184, https://doi.org/10.1016/j.enbuild.2013.01.014.

[36] D. Blum, K. Arendt, L. Rivalin, M. Piette, M. Wetter, C. Veje, Practical factors of envelope model setup and their effects on the performance of model predictive control for building heating, ventilating, and air conditioning systems, Appl. Energy 236 (2019) 410–425, https://doi.org/10.1016/j.apenergy.2018.11.093.

[37] F. Bünning, B. Huber, A. Schalbetter, A. Aboudonia, M. Hudoba de Badyn, P. Heer, R.S. Smith, J. Lygeros, Physics-informed linear regression is competitive with two Machine Learning methods in residential building MPC, Appl. Energy 310 (2022) 118491, https://doi.org/10.1016/j.apenergy.2021.118491.

[38] L. Di Natale, B. Svetozarevic, P. Heer, C. Jones, Physically consistent neural networks for building thermal modeling: theory and analysis, Appl. Energy 325 (2022) 119806, https://doi.org/10.1016/j.apenergy.2022.119806.

[39] E. Atam, L. Helsen, Control-oriented thermal modeling of multizone buildings: methods and issues: intelligent control of a building system, IEEE Control Syst. Mag. 36 (3) (2016) 86–111, https://doi.org/10.1109/MCS.2016.2535913.

[40] J. Drgoňa, A.R. Tuor, V. Chandan, D.L. Vrabie, Physics-constrained deep learning of multi-zone building thermal dynamics, Energy Build. 243 (2021) 110992, https://doi.org/10.1016/j.enbuild.2021.110992.

[41] R. Rueda, L. Ruiz, M. Cuéllar, M. Pegalajar, An ant colony optimization approach for symbolic regression using straight line programs. Application to energy consumption modelling, Int. J. Approx. Reason. 121 (2020) 23–38, https://doi.org/10.1016/j.ijar.2020.03.005.

[42] J. Leprince, C. Miller, M. Frei, H. Madsen, W. Zeiler, Fifty shades of black: uncovering physical models from symbolic regressions for scalable building heat dynamics identification, in: Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Association for Computing Machinery, New York, NY, USA, 2021, pp. 345–348.

[43] Y. Ozawa, D. Zhao, D. Watari, I. Taniguchi, T. Suzuki, Y. Shimoda, T. Onoye, Data-driven HVAC control using symbolic regression: design and implementation, arXiv:2304.03078, 2023.

[44] V. Harish, A. Kumar, A review on modeling and simulation of building energy systems, Renew. Sustain. Energy Rev. 56 (2016) 1272–1292, https://doi.org/10.1016/j.rser.2015.12.040.

[45] M. Georgescu, I. Mezić, Building energy modeling: a systematic approach to zoning and model reduction using Koopman Mode analysis, Energy Build. 86 (2015) 794–802, https://doi.org/10.1016/j.enbuild.2014.10.046.

[46] F. Banihashemi, M. Weber, W. Lang, Model order reduction of building energy simulation models using a convolutional neural network autoencoder, Build. Environ. 207 (2022) 108498, https://doi.org/10.1016/j.buildenv.2021.108498.

[47] European Commission, Tools for cOntinuous building Performance Auditing (TOPAs). Horizon 2020. CORDIS EU research results - Projects & Results, 2015–2018, https://cordis.europa.eu/project/id/676760.

[48] E. O'Dwyer, L. De Tommasi, K. Kouramas, M. Cychowski, G. Lightbody, Modelling and disturbance estimation for model predictive control in building heating systems, Energy Build. 130 (2016) 532–545, https://doi.org/10.1016/j.enbuild.2016.08.077.

[49] S. Roshany-Yamchi, K. Witheepanich, J.M. Escaño, A. McGibney, S. Rea, Selective distributed model predictive control for comfort satisfaction in multi-zone buildings,

in: 2017 21st International Conference on System Theory, Control and Computing (ICSTCC), 2017, pp. 648–653.

[50] A. Chopra, Introduction to Google SketchUp, John Wiley & Sons, 2012.

[51] Thermal Energy System Specialists, LLC, TRNSYS features: suite of tools, http://www.trnsys.com/features/suite-of-tools.php.html, 2019.

[52] B. Bueno, L. Norford, G. Pigeon, R. Britter, A resistance-capacitance network model for the analysis of the interactions between the energy performance of buildings and the urban climate, Build. Environ. 54 (2012) 116–125, https://doi.org/10.1016/j.buildenv.2012.01.023.

[53] A. O' Donovan, P.D. O' Sullivan , M.D. Murphy, Predicting air temperatures in a naturally ventilated nearly zero energy building: calibration, validation, analysis and approaches, Appl. Energy 250 (2019) 991–1010, https://doi.org/10.1016/j.apenergy.2019.04.082.

[54] M. Wetter, GenOpt®Generic Optimization Program: User Manual, Lawrence Berkeley National Laboratory, Berkeley, CA, 2016, https://simulationresearch.lbl.gov/.

[55] Climate.OneBuilding, World meteorological organization. irl_st_cork.ap.039550_tmyx.2004-2018.zip, https://climate.onebuilding.org/WMO_Region_6_Europe/IRL_Ireland/, 2018.

[56] ISO, ISO 7730:2005. Ergonomics of the thermal environment – analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria, Management 3 (605) (2005) e615.

[57] V. López-Cabeza, F. Carmona-Molero, S. Rubino, C. Rivera-Gómez, E. Fernández-Nieto, C. Galán-Marín, T. Chacón-Rebollo, Modelling of surface and inner wall temperatures in the analysis of courtyard thermal performances in Mediterranean climates, J. Build. Perform. Simul. 14 (2) (2021) 181–202, https://doi.org/10.1080/19401493.2020.1870561.