Time-dependent solar aperture estimation of a building: comparing grey-box and white-box approaches

Xiang Zhang [1,*], Christoffer Rasmussen [2], Dirk Saelens [1], Staf Roels [1]

1 KU Leuven, Department of Civil Engineering, Building Physics and Sustainable Design Section, Kasteelpark Arenberg 40 Bus 2447, 3001 Heverlee, Belgium
2 Technical University of Denmark, Department of Applied Mathematics and Computer Science, Section for Dynamical Systems, Asmussens Allé, Building 303B, 2800 Kgs. Lyngby, Denmark

Abstract. This paper proposes a B-splines integrated method combining in-situ data with grey-box modelling to estimate buildings' dynamic solar gain more efficiently than the conventional white-box model and much more precisely than the classic grey-box model. Solar gain, referring to the overall indoor energy gain supplied by solar radiation, plays a vital role in the indoor energy balance. Estimating dynamic solar gain precisely is essential to building energy optimisation, e.g. , in model predictive control. However, in almost all existing grey-box modeling works, a constant solar gain factor (solar aperture; *gA*) is assumed to estimate dynamic solar gain, which almost certainly will result in solar gain prediction errors, especially in buildings with unevenly distributed windows. To fill this gap, this study presents an advanced B-splines integrated grey-box model, using customized B-splines to advance the constant *gA* assumption toward its nature of time-dependence and precisely characterize the dynamic solar gain conclusively. On-site measured datasets of a portable site office (PSO) representing a 'simplified' building, under two scenarios with windows fully or partially uncovered, serve as test cases. To verify the physical interpretation of outcomes estimated by the proposed method, based on the said test cases, the proposed B-splines integrated grey-box model is compared with a classic white-box simulation. It is concluded that the proposed method can reveal the main trends and key dynamic features of solar gain very well, but still has some limitations of quantifying 'local' details with acceptable variations. Nevertheless, given that the proposed method merely asks for a very limited amount of low-frequency data, the proposed method is considered as a much more effective alternative to the classic white-box approach, which requires massive and often hard-to-collect input data.

Highlights.

1. Constant solar aperture (*gA*) assumption, commonly used in reduced-order statistical models, is demonstrated as less suitable to predict solar gains in buildings, especially for cases with unevenly distributed windows.
2. An advanced B-splines integrated grey-box modelling technique is proposed and verified by white-box simulation outcomes, to offer reliable dynamic solar gain estimations.
3. Only a short (e.g., one-week) and low frequency (e.g., hourly) dataset with limited number of parameters (e.g., seven) is required for this enhanced technique.
4. Limitations of four techniques for estimating dynamic solar gain are discussed, which include the advanced and original B-spline integrated grey-box, traditional grey-box, and white-box (simulation) models.

Keywords. Solar gain; dynamic solar aperture (*gA*); grey-box model; building energy simulation; direct normal irradiance; comparative study

Nomenclature

Abbreviations

ACF	auto-correlation function
AIC	Akaike information criterion
BES	building energy simulation
BIC	Bayesian information criterion

* Corresponding author, KU Leuven, Department of Civil Engineering, Building Physics and Sustainable Design.
Email: jason.zhang.kul@outlook.com; jason.zhang@kuleuven.be.

CI	confidence interval
CP	cumulated periodogram
ctsm-r	continuous time stochastic modelling for R
Dymola	dynamic modelling laboratory
IDEAS	integrated district energy assessment by simulation
LogL	loglikelihood-value
ML	maximum likelihood
MPC	model predictive control
PASSYS	passive solar component and system testing
PRBS	pseudo random binary sequence
PSO	portable site office
RC	resistance-capacitance
SDE	stochastic first-order differential equation

Mathematical notation

$Bs_{i,m}$	the ith B-spline of order m
ϕ_i	scaling factor of the ith B-spline
S	spline function
ω	Wiener process
σ_w	standard deviation of Wiener process
σ_s	standard error of the spline parameter
e	measurement error
θ	a vector of length m, containing the spline parameters
B	a n-by-m matrix, containing m basis splines with their corresponding values for each of the n observations
ρ	correlation matrix of the spline parameters
Φ	the cumulative distribution function of the normal distribution

Physical parameters

A	glazed area of window [m^2]
C	heat capacity [J/K]
DNI	direct normal irradiance [W/m^2]
DHI	diffuse horizontal irradiance [W/m^2]
g	solar transmittance [-]
gA	solar aperture [m^2]
gA_{dni}	solar aperture tailored to direct normal radiation [m^2]
gA_{dhi}	solar aperture tailored to diffuse horizontal radiation [m^2]
GHI	global horizontal irradiance [W/m^2]
prm	number of parameters [-]
R	thermal resistance [K/W]
Sg	total solar gain [W]
Sg_{dir}	direct solar gain [W]
Sg_{dir}	diffuse solar gain [W]
t	time [s]
T	temperature [°C]
T_k^*	measured temperature [°C]
ws	wind speed [m/s]
ϕ_s	heat transfer of solar energy through opaque envelope [W]
ϕ_h	heating input [W]
θ_z	zenith angle of the sun [°]

Subscripts

95	For 95-percentage confidence interval
a	ambient air
dir	direct
dif	diffuse
e	building envelope
ea	interaction between e and a
k	discrete time step
i	internal air
ia	interaction between i and a
ie	interaction between i and e
m	internal thermal mass

s1	PSO scenario 1
s2	PSO scenario 2
sim	BES-model simulation
t	time

1. Introduction

1.1 Grey-box modelling of building thermal dynamics

Energy is an irreplaceable resource for people's daily life with expanding demands, and the current building industry accounts for around one-third of total final energy consumption, and for almost 30% of direct and indirect CO_2 emissions [1,2]. Thus, to reduce energy consumption and associated carbon emissions, understanding and enhancing building energy performance have become increasingly significant. Building energy simulation models, such as TRNSYS, Energy+, Modelica (Dymola) amongst others, are nowadays common tools to simulate and optimize the thermal and energetic behavior of a building. They are typical examples of so-called white box models, which make use of detailed prior information to explicitly model a building's energy use and thermal response. For existing buildings, this detailed input data is often not available and simplified models identified based on data collected on ite, are to be preferred [3]. We can distinguish between black box and grey box simplified models. Black box models disregard any prior knowledge and solely aim at describing input/output relations. Artificial neural networks [4] and linear regression models [5] are examples. Grey box models embed limited prior knowledge in a stochastic framework. Overall, they are found to be more robust and provide better extrapolation performance with limited training data. Recently, grey-box modeling has been widely applied in energy performance assessment of buildings [6,7], model predictive control [8–10], fault detection and diagnosis [11], the optimization of smart grids [12], etc. The key advantages of grey-box modeling are: 1) relative low needs on data size compared to white-box models, and 2) good physical interpretability of parameter estimations compared to black-box models. The said merits of the grey-box model are contributed by the fact that grey-box model combine the prior physical knowledge with statistical information embedded in the data [8,13]. Thus, the performance of the grey-box model leverages the merits and minimizes the weaknesses of both white- and black-box models [14]. The qualitative comparison of the three-modeling approach is tabulated in Table 1. It is not hard to find that – for existing buildings – grey-box models show the best comprehensive performance, in terms of low data requirement, low computational cost, physical interpretation of estimated outcomes, etc.

Table 1 Qualitative comparison of three modeling approaches.

Model	Demands for data amount & quality	Computational cost	Physical interpretation	Accuracy	Expertise required	Application Scenario
White box	Very high	Very high	Excellent	Excellent	Very high	(Mainly in) design stage
Grey box	Medium	Medium	Good	Good	High	After construction (based on 'real' in situ data)
Black box	High	Medium-High	Poor	Good	Medium	

Specifically to estimate solar gains, the part of energy gain supplied by the sun penetrating through the glazed building envelope, BES-models (white-box) are a popular technique. However, the BES model demands massive data as input, such as building geometry, properties of the envelope, information about the surroundings and weather, and certain expertise level. Without sufficient data and expertise, it might not only lead to poor simulation accuracy [15,16], but even can result in an incorrect or no outcome. Thus, a grey-box model-based method of estimating dynamic solar gains could serve as a more efficient and reliable alternative, which will be valuable to reduce the cost of assessing the thermal dynamics of buildings, owing to its low demands of data-size. Most importantly, since a grey-box model is based on on-site measurement data, the estimated outcomes of the grey-box model stay closer to the 'reality'. Black-box models are out of interest here, because of their poor physical interpretation. In addition, grey-box modelling is generally regarded to

have an enhanced extrapolation properties, compared to the black-box models [17]. The reader is referred to [18] for elaborated information of said three model types. In brief, grey-box model is able to 'mine' the maximum information from limited data to estimate unknown parameters of interests, while keeping an acceptable physical interpretation of the estimated outcomes. Therefore, developing a grey-box model-supported technique on the dynamic solar gain is the key focus of this study.

Practically, in grey-box models, buildings are commonly represented as resistance-capacitance (RC) systems [18] by analogy with electric circuits[19], mathematically represented as a system of stochastic first-order differential equations (SDEs) [20], making the equations and parameters involved physically interpretable. Statistical tools such as maximum likelihood (ML) and Kalman filter algorithm are adopted for the parameter estimations [21]. It is noteworthy that, the estimation outcomes (of grey-box models) for each parameter such as C_i, C_e and R_{ie} has a mean value with a standard error [22]. Eq. (1-3) gives an example of two states (i.e. *Ti* and *Te*) grey-box model for modeling a single thermal zone building, which is visualized in Fig. 1.

$$dT_i = \frac{1}{R_{ie}C_i}(T_e - T_i)dt + \frac{1}{C_i}(Sg + \boldsymbol{\phi_h})dt + \sigma_{w,i}d\omega_i \quad (1)$$

$$dT_e = \frac{1}{R_{ie}C_e}(T_i - T_e)dt + \frac{1}{R_{ea}C_e}(\boldsymbol{T_a} - T_e)dt + \sigma_{w,e}d\omega_e \quad (2)$$

$$\boldsymbol{T^*_{ik}} = T_{ik} + e_k \quad (3)$$

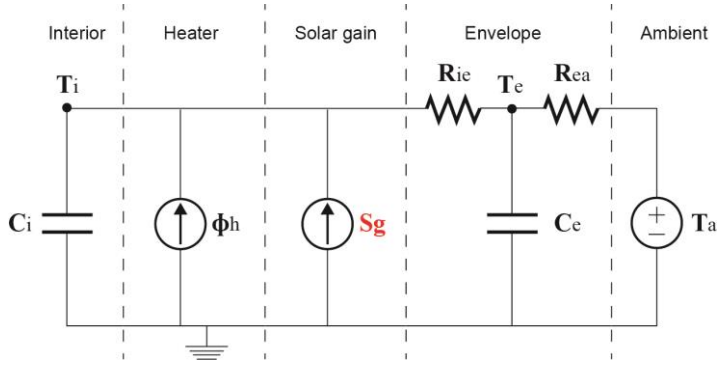


Fig. 1 A single zone two-states thermal RC-network model (*TiTe* model).

Where *Ti* , *Ta* and *Te* refer to the indoor and ambient air and envelope temperature, and *Rie* and *Rea* are the thermal resistances against heat transfer from the said different temperature (states), with ventilation heat losses included [23]. For further specification, see the Nomenclature. The heat capacities of the interior mass and envelope are denoted as *Ci* and *Ce*. {ωi,t}, and {ωe,t} are independent standard Wiener processes, and $\sigma^2_{w,i}$ and $\sigma^2_{w,e}$ are the incremental variances of the Wiener processes [13]. $\boldsymbol{T^*_{ik}}$ is the observed internal temperature and e_k is the residuals between predicted and measured indoor temperature at k^{th} observation. ϕ_h and *Sg* are the energy flux (in W) from the heating system and the solar radiation respectively. A two states (also called second-order) model is generally regarded as a sufficient model for modeling building thermal dynamics, in most of studies [24–29]. One state (first-order) grey-box model might not be sufficient to 'catch' the key dynamics [13], and few cases requires three states model or above. For more details on statistical grey-box modelling in general the reader is referred to [22] and for applications on buildings specifically to [3,13,18,23]. As stated, the solar gain (*Sg*) is the research object of this work, aiming at proposing a grey-box model-based technique of gauging dynamic solar gain effectively and precisely.

1.2 Buildings' dynamic solar gain estimation

For most buildings, the solar gain (*Sg*) supplies considerable energy to the buildings' indoor heat balance [30], which significantly impacts on the heating or cooling loads of buildings [31]. Solar gain through windows refer to the total indoor energy gain contributed by the solar irradiance transmitted through the glazed building surfaces [32], and can be written as Eq. (4).

$$Sg_t = \sum(g_{i,t} * A_i) * GHI_t = Sg_{dir_t} + Sg_{dif_t} \quad (4)$$

with GHI_t the global horizontal irradiance (W/m²) at time *t*, and $g_{i,t}$ and A_i respectively the time-dependent solar transmittance (-) and the (effective) glazed area (m²) of the window *i*. In BES (white-box) modelling, the details of $g_{i,t}$ and A_i for each glazing are essential, as input data, to simulate accurately the dynamic solar gain, while in grey-box modelling, the model is asked to fit the unknown parameters (e.g. $g_{i,t}$ and A_i values) based on limited on-site measured data by maximum likelihood estimation. In practice, it is almost infeasible to estimate $g_{i,t}$ and A_i values for each window, and a single (aggregated) solar aperture (*gA*) is used to simplify (and replace) $\sum(g_{i,t} * A_i)$ in Eq. (4). This simplification leads to Eq. (5), which is widely used in literature.

$$Sg_t = gA * GHI_t = Sg_{dir_t} + Sg_{dif_t} \quad (5)$$

The solar aperture (gA-value) is also referred to as solar gain factor, solar transmittance, or effective window area. It is generally regarded as 'the equivalent area of a perfectly transparent surface that transmits the same amount of solar energy as the actual windows of the space' for a particular building [33]. *gA* accounts both for direct and indirect solar gains [34], as shown in Eq. (5), since total solar gain (Sg_t) consists of both direct (Sg_{dir_t}) and diffuse parts (Sg_{dir_t}). The decomposition of Sg_t will be elaborated in section 3.1.3. Omitting the subscript '*t*' for gA in Eq. (5), refers to the common assumption of a time-invariant value in almost all grey-box modeling works in literature. For example, Eq. (5) embedded grey-box models are used to determine building energy performance [35–37], serve as control model in model predictive control (MPC) [9,38,39] for optimizing building energy performance, and etc. However, a constant value *gA*-value is not in line with the dynamic nature of solar aperture (*gA*), determined by both the angle dependent solar transmittance (Fig. 2) and effective glazed area of window. Specifically, the solar transmittance (g) generally decreases when incidence angle enlarges [40,41]. In addition, the effective glazed area also varies over time [42], which is jointly determined by windows' distribution and shading effects from the building itself or nearby obstacles.

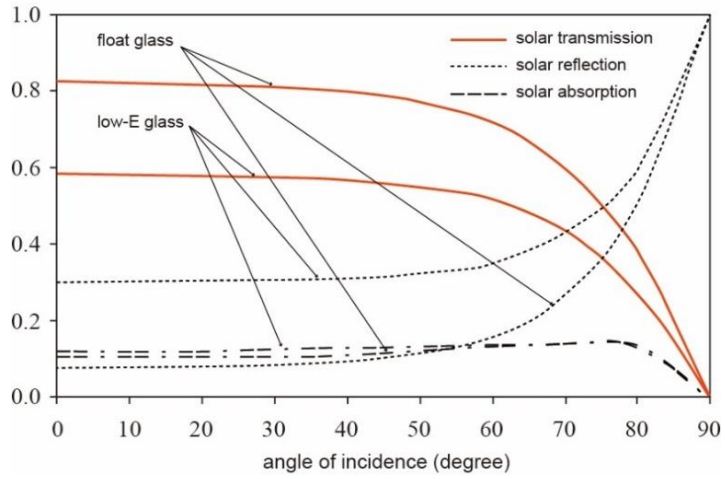


Fig. 2 Example of incidence angle dependent solar properties of glazing layers (modified from [40]).

Hence, the extensively used constant *gA* assumption in grey-box modeling is significantly inconsistent with the 'reality', especially in buildings with irregularly distributed windows. For instance, for a particular building that merely has windows on the east and west walls, its daily time-dependent *gA* curve will show small values for most of the day but can have large values in the early morning and late afternoon. So, the real *gA* curve of this particular building will obviously not be constant. Therefore, some researchers have reported that assuming *gA* constantly might yield 'unqualified' grey-box models [43], while other suggest that a qualified grey-box model for model predictive control (MPC) requires on-site measured [38] or pre-simulated [44] dynamic solar gain data, as extra input, rather than estimating solar gains by Eq. (5).

1.3 Problem statement

To avoid estimation (or prediction) uncertainties caused by assuming *gA* as a constant value, Rasmussen et al. [33] have proposed a technique to estimation daily gA_t curve (the blue curve in Fig. 3) for a specific building. Similar to the example discussed in the previous section 1.2, The case studied [33] has no windows on the southern and northern walls. As expected, Fig. 3 shows that the constant *gA* assumption (the red line in Fig. 3) considerably mismatch with the reality (the blue curve in Fig. 3).

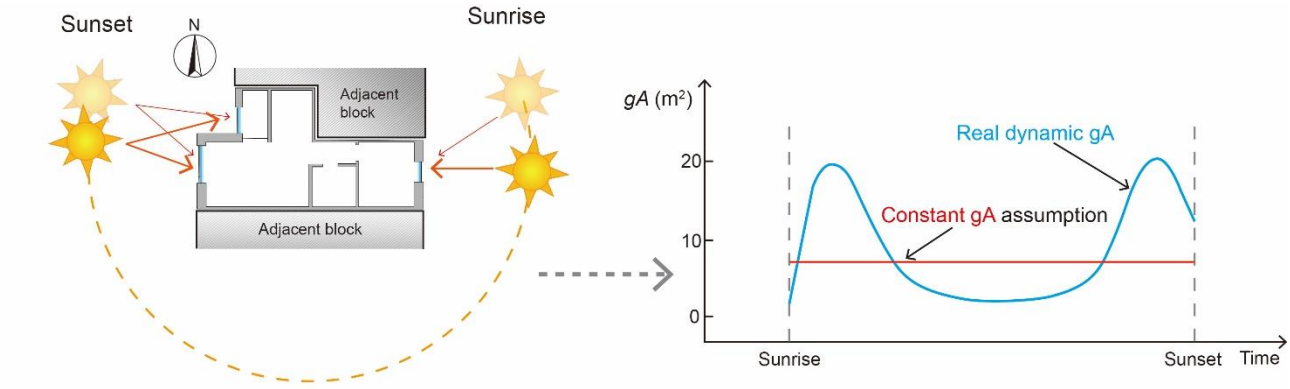


Fig. 3 Dynamic solar aperture curve (*gA*) analysis of the study case in [33] (modified from [33]).

The corresponding calculated solar gains (Sg_t) under dynamic and constant *gA* assumptions are visualized in Fig. 4. Although, both assumptions might yield the same total amount of daily solar gain (in *J*) (the red and blue zones in Fig. 4 are almost equal in size), the estimated solar gain dynamics (in *W*; red and blue curve in Fig. 4) differ significantly. This discrepancy (between blue and red curves in Fig. 4) could force grey-box model to 'believe' there is a certain amount of Sg_t around noon, when almost no solar gain exists in reality, leading to an over-estimation of the predicted indoor temperature. For most of buildings, especially with evenly distributed windows, the common used constant *gA* assumption might enlarges certain prediction uncertainties.

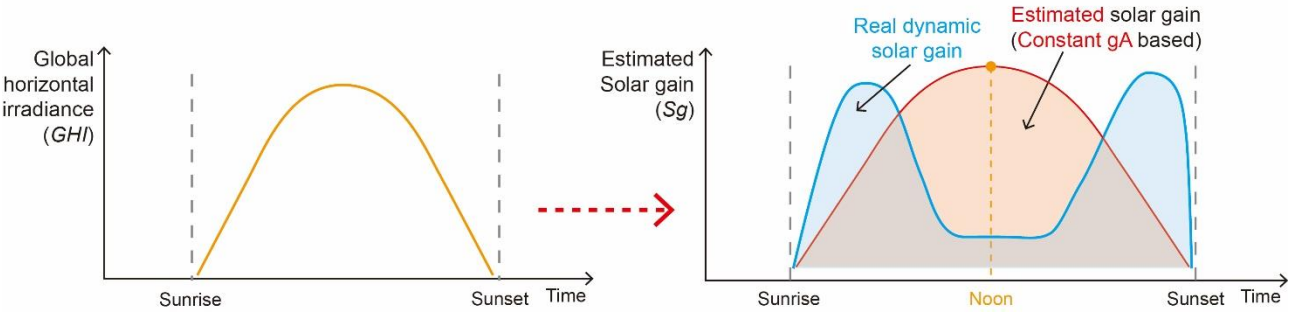


Fig. 4 Dynamic solar gain estimation gap analysis based on constant *gA* assumption.

1.4 Research aims and methods

As stated above, efficient and precise estimation of *Sg* dynamics (e.g., the blue curve in Fig. 4) is vital for model predictive control (MPC), fault detection and diagnosis, etc. Thus, this study aims to propose a grey-box model-based method to gauge the daily dynamic *gA* curve (e.g., the blue curve in Fig. 3) efficiently and reliably. As visualized in Fig. 3, Rasmussen et al. [33] has successfully integrated B-splines into grey-box models to estimate the daily gA_t dynamic curve for a specific building (Fig. 3). However, the method proposed in [33] hasn't been compared to the typical white-box simulation method, leading to the unknown reliability of this method. To ensure the comparability between BES-simulation and estimated outcome of grey-box, the grey-box model proposed in [33] should also be enhanced to ensure the clear physical interpretation of estimated gA_t. This is mainly attributed to the grey-box model suggested in [33] has a certain risk to embed the part of solar energy penetrating opaque envelope (e.g. walls and roof) into estimated gA_t, which is not

fully in line with the definition of the solar aperture gA, which merely accounts for the part of solar energy penetrating window glazing [33]. Details of three main types of solar gains will be elaborated on in section 3.1.3. At this stage, in short, the work aims to examine the reliability of the modeling technique proposed in [33] by a physical interpretation of the estimation offered by the technique in [33]. However, since the version of the model used in [33] has a certain risk of estimating gA_t with 'ambiguous' physical interpretation, the current study firstly enhances the grey-box model of [33] to systematically fix the physical interpretation problem and refines gA_t to gA_{dni_t} curve. Secondly, the estimated outcome of the enhanced grey-box model proposed in this study will be compared with a traditional BES-model simulation, verifying the reliability of the proposed technique in this study. The readers refer to section 3.1.3 more details and the definition of gA_{dni_t}. Fig. 5 visualize the flow chart of this study, and the remaining content of the paper is organized as follows: first, two datasets measured under two scenarios of a portable site office (PSO) are described in Section 2. Section 3 clarifies the B-splines integrated grey-box model, the model selection process, and the two grey-box models selected for the said two datasets. In Section 4, two BES-models of two mentioned PSO scenarios are constructed in Dymola (Modelica). Section 5 compares the estimation outcomes from grey-box and white-box approaches. Finally, the key lessons learnt in this study are summarized in Section 6.

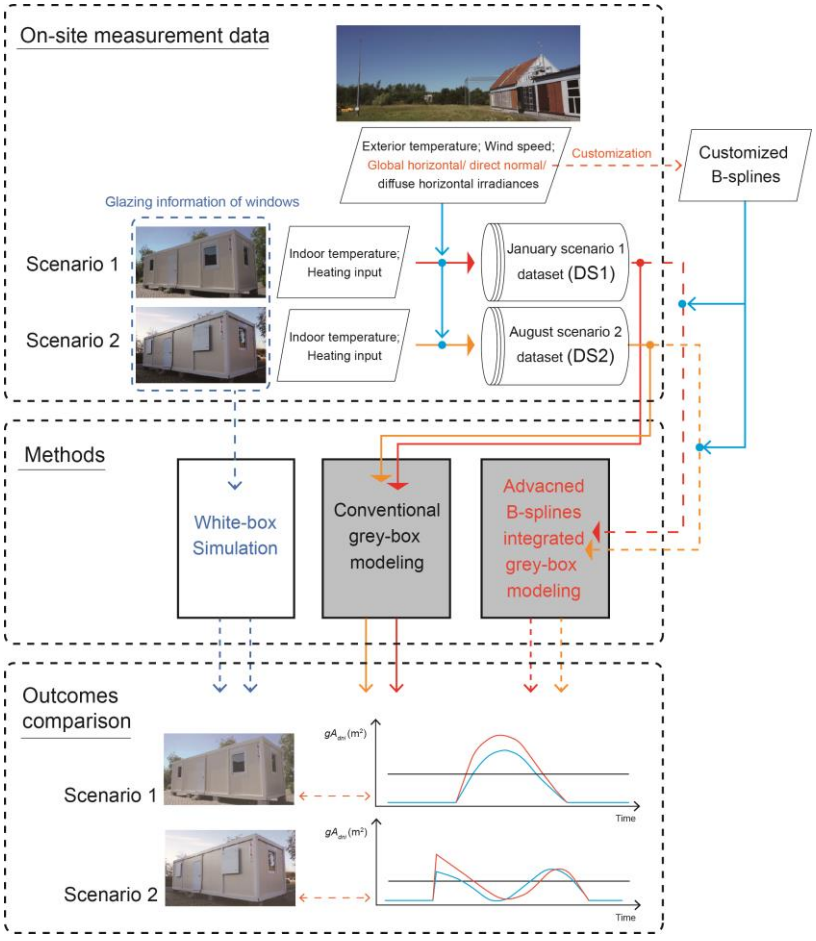


Fig. 5 The workflow of this study.

2. Description of Case Study

The energy performance of a real building is affected by plenty of factors, such as the building itself, the inhabitants-related ones, and others [45]. It is not easy to understand the cumulative impact, from multiple factors, at one time. Thus, 'simplified buildings', so-called test cells, are often used to explore the complexity of building thermal dynamics step-by-step, by controlling particular boundary conditions out of interest [46].

As an integrated part of the IEA EBC Annex-71 project: 'Building energy performance assessment based on in-situ measurements' [47]. A portable site office (PSO) is customized and constructed for this study (Fig. 6). The PSO is elevated around 30 cm above the ground to eliminate heat losses to the ground. The PSO is a standardized product serving as a temporary residence, which is treated as a full (simplified) building with overall dimensions of 9*3*3m³. It has four rooms: two main test rooms, one tiny entrance lobby, and a toilet (Fig. 7), and is located at a parking lot (50.86N, 4.68E) at the Arenberg campus, KU Leuven, Belgium. The car park is surrounded by some 1 to 4 floor(s) high buildings (Fig. 8-9), which might impose shading effects on the PSO. The main façade of the PSO is south-east (SE140°) oriented, with two windows on the main façade, and one window on each of the remaining three walls (Fig. 7), leading to in total five windows. All windows are 1.42*1.08m^2 with a glazing part of 0.98*0.72m^2, except the one in the northwest wall, which is 0.5*0.6m^2 with a glazed component of 0.33*0.4m^2 (Fig. 6-7).

As summarized in Table 2, two scenarios have been studied. In the first scenario, all windows of the PSO were uncovered (Fig. 7). In the second scenario, the two windows on the main façade were shaded by, white colored, wooden shading devices (Fig. 11 and Fig. 12). A PRBS (pseudo random binary sequence) heating on-off controlling signal was imposed in the two scenarios. The PRBS signal [48] is a deterministic signal with the auto-correlation of white-noise properties and uncorrelated with other external signals (e.g. ambient weather conditions) [7,49]. Fig. 13 offers an impression of the measurement campaigns in which the heat input is imposed by switching the lamps' on/off. It is noted that, as shown in Eq. (1-3), only the bold parts, i.e., heating input ($\boldsymbol{\phi_h}$), ambient temperature ($\boldsymbol{T_a}$), and measured indoor temperature ($\boldsymbol{T^*_{ik}}$), plus additional data of solar radiation, are known to estimate certain number of physical parameters, such as C_i, C_e, R_{ie}, R_{ea}, T_e, e_k etc. Thus, although heating is not necessary in summer, heating input data is essential for the parameter estimation in grey-box modeling. Nevertheless, to avoid 'overheating' for summer scenario 2, the ten lamps used in the winter scenario 1 were reduced to six (Fig. 12).

Table 2 Two datasets measured under different scenarios with PRBS signal.

Dataset (code)	Scenarios	Measurement period	Heating setting	Frequency
Scenario 1 in January dataset (DS1)	All windows uncovered, shown in Fig. 6 and Fig. 7.	From 17 January (19:00) to 24 January 2020 (9:55); in winter	10 lamps (375W each) and 6 fans	60 mins (hourly)
Scenario 2 in August dataset (DS2)	Two windows covered shown in Fig. 11 and Fig. 12.	From 6 August (00:00) to 12 August 2020 (23:55); in summer	6 lamps (375W each) and 6 fans	30 mins

Additionally, the weather data is recorded at a weather station at the VLIET test building (Fig. 14) [50,51], which is around 300 meters away from the PSO. The short distance between the weather station and PSO makes the monitored weather data applicable to the PSO. The weather station documents exterior air dry-bulb temperature (°C), wind speed (m/s), direct normal (W/m^2), global horizontal (W/m^2), and diffuse horizontal irradiances (W/m^2), etc, in one-minute frequency. It is expected to observe more dynamics in gA_{dni_t} curve of PSO scenario 2 (see section 3.1.3), since the windows in scenario 1 (DS1 dataset) are more evenly distributed than in scenario 2 (DS2 dataset). Therefore, a common sampling time for buildings' on-site measurement data [52,53] – an hourly frequency – is selected for DS1, while DS2 is sampled on a shorter (30 mins) time interval basis. Appendix A elaborates the measurement accuracies and the sampling practice in detail. Finally, the processed DS1-2 dataset consists of seven variables namely "*Ti*", "*Te*", "ϕ_h", "*GHI*", "*DNI*", "*DHI*", and "*ws*", referring to indoor temperature (°C), outdoor temperature (°C), heating input (W), global horizontal irradiance (W/m^2), direct normal irradiance(W/m^2), and diffuse horizontal irradiance (W/m^2), and wind speed (m/s) respectively. The constructed DS1-2 datasets are visualized in Fig. 15 and Fig. 16 respectively.

Fig. 6 The image of the PSO in scenario 1.

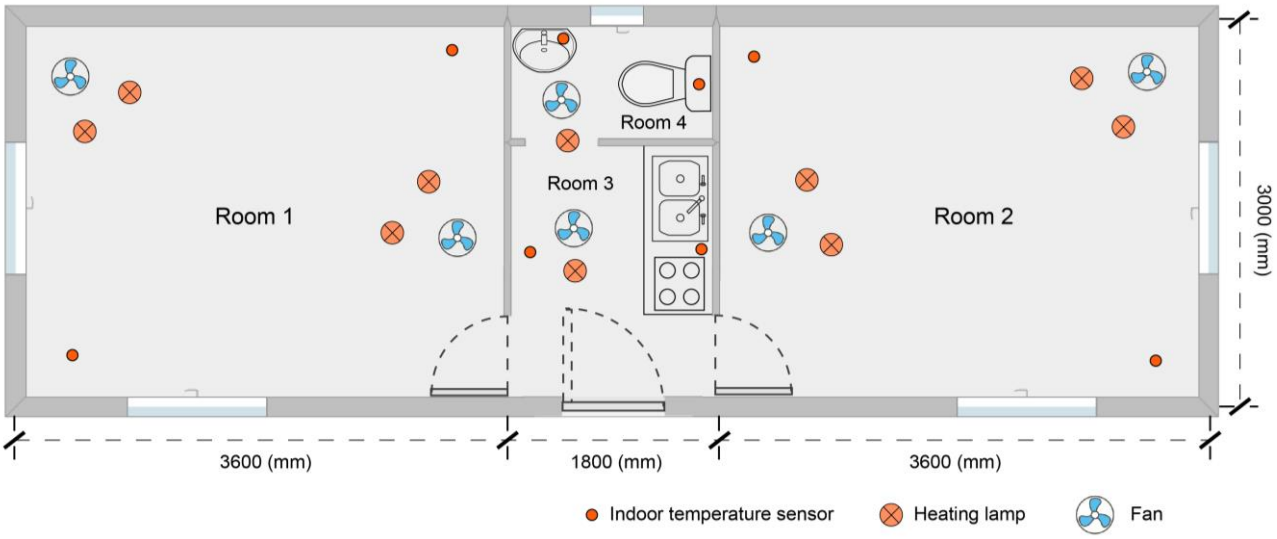


Fig. 7 The layout of the PSO with PRBS scenario 1.

Fig. 8 The parking lot where the PSO (red) is located, surrounded by some small trees and in height varying buildings.

Fig. 9 The shadow pattern from surrounded obstacles around 10 a.m. on 21 Jan. 2020.

Fig. 10 The surrounding environment and shading obstacles of the PSO.

Fig. 11 The image of the PSO under scenario 2.

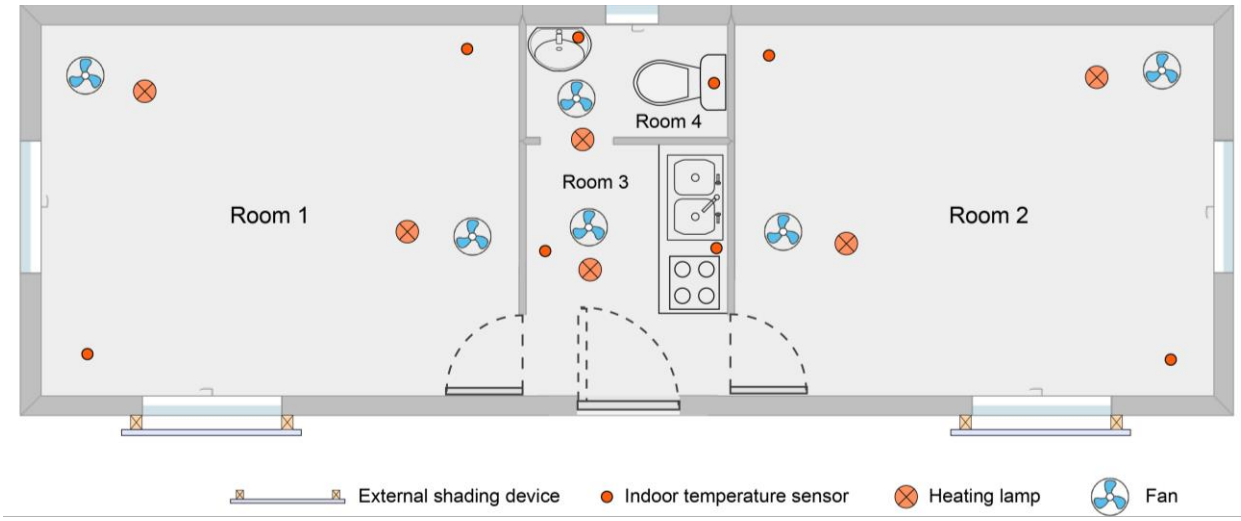


Fig. 12 The PRBS scenario 2 of PSO.

(a) Entrance (b) Test room 1

Fig. 13 The pseudo random binary sequence (PRBS) heating test equipment in the PSO.

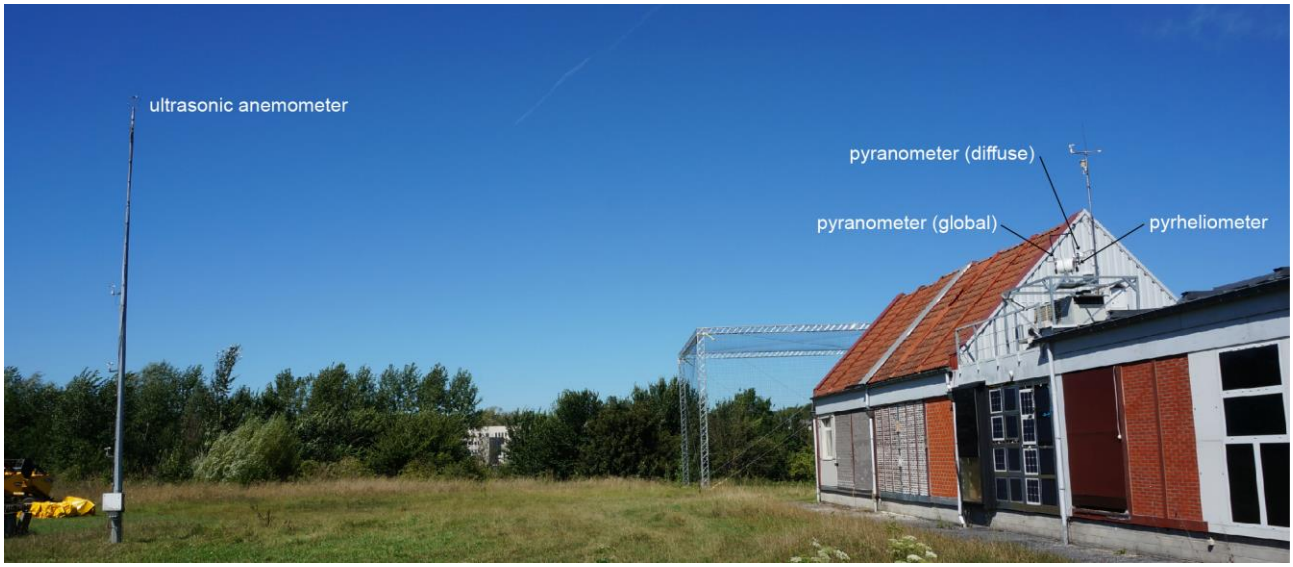


Fig. 14 The weather station at the VLIET test building.

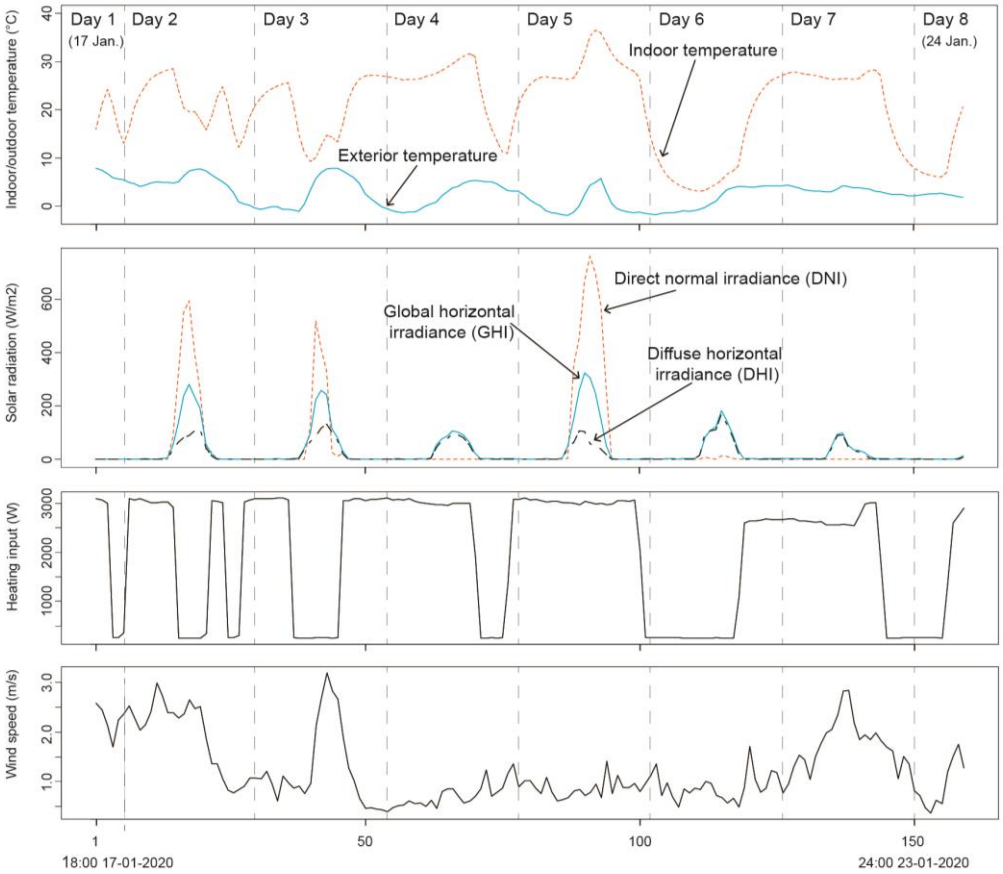


Fig. 15 The plots of DS1 dataset.

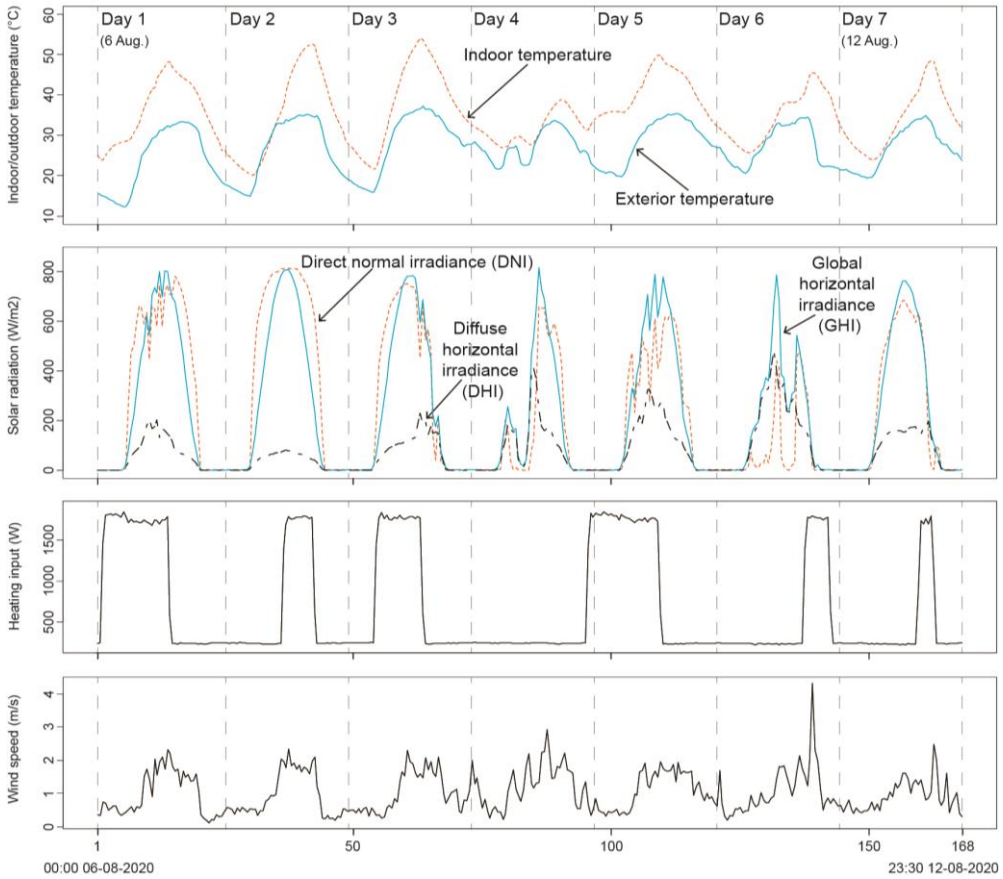


Fig. 16 The plots of DS2 dataset.

3. B-splines integrated grey-box modelling

3.1 Grey-box model

3.1.1 Forward selection

As stated in section 1.1, a grey-box model consists of system equations, such as Eq. (1-2), based on (super-simplified) physical assumptions. For a specific building, various candidate grey-box models are available to model its thermal dynamics based on different physical assumptions. To identify the most suitable grey-box model among candidates for each case (i.e., DS1-2), a forward model selection method is used in this study. This is a common grey-box model identification technique [3,29,33]. The forward selection starts from the simplest (smallest) feasible model based on physical knowledge [13], and it is step-by-step extended by adding one variable at a time, (statistical) improving the model most, until no candidate variables integration can significantly enhance the fitting of a model to the particular dataset. In general, embedding more parameters in a grey-box model - increasing model complexity - 'always gives rise to an increased value of the maximized log-likelihood' (*LogL*) [54], indicating a better fitting of the grey-box model to a particular dataset, but also increases the overfitting risk of local noise, as shown in Fig. 17. The 'underfitting' and 'overfitting' of a grey-box model are unwanted, and loglikelihood-value (*LogL*), AIC (Akaike information criterion), and BIC (Bayesian information criterion) are used to pinpoint the 'balanced' model (Fig. 17). Referring [54,55] for more details. In this study, a step forward in Fig. 21 and Fig. 22 requires a considerable increase *LogL* and remarkable drops of AIC-value of the grey-box model extension, which is regarded as a pre-set selection criterion of 'balanced' grey-box identification.

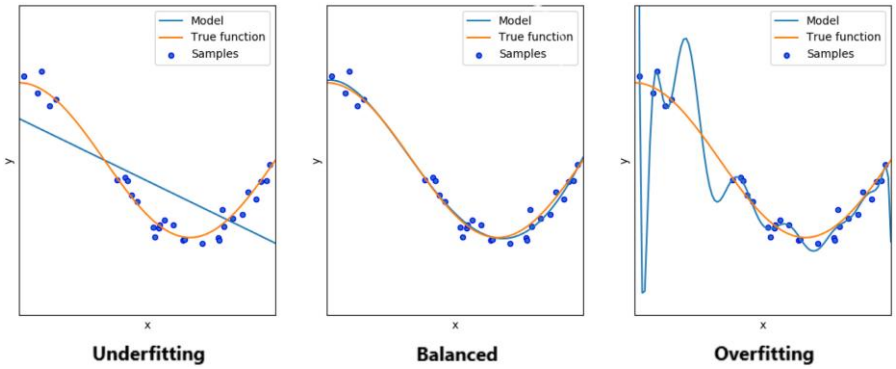


Fig. 17 Underfitting, overfitting and balanced statistical models' performances [56]

3.1.2 Model qualification evaluation

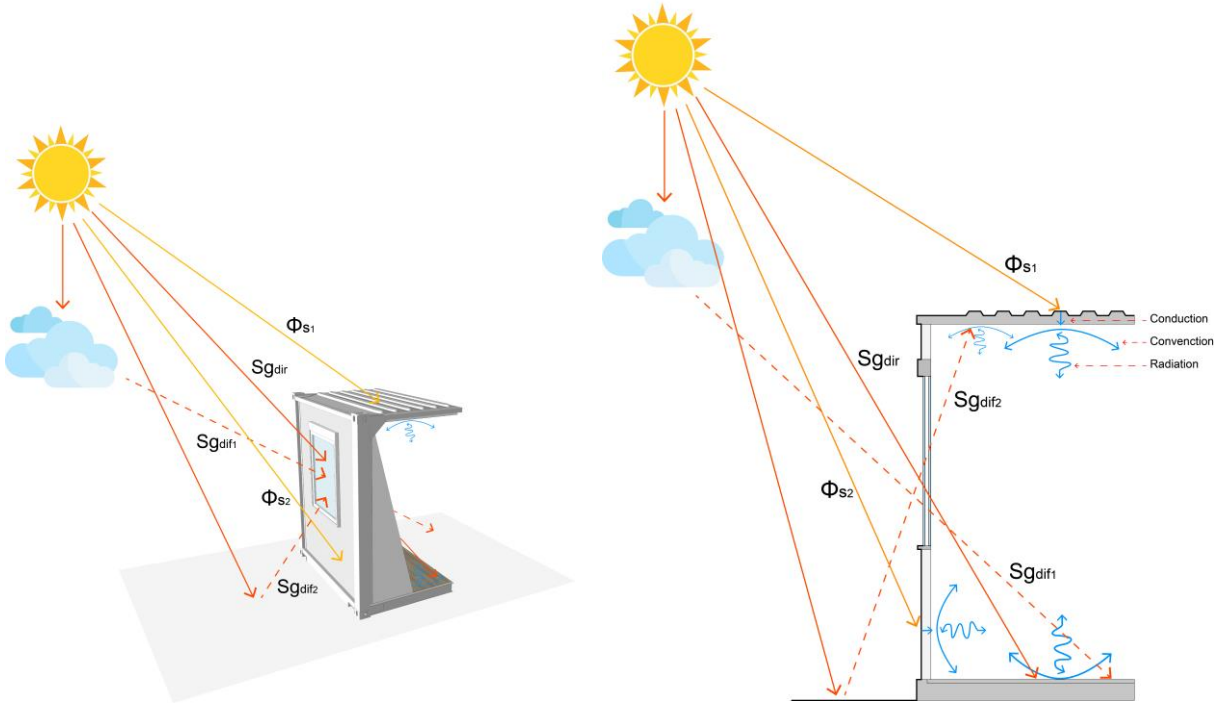
The selected grey-box models in the forward selection also have to meet additional requirements, such as: the residuals of the model are white-noise, and the model gives physically reasonable estimation values [13]. Auto-correlation function (ACF) and the cumulated periodogram (CP) are standard techniques to test the assumption of white noise residuals. In addition, the raw periodogram (RP) and plots of inputs versus model residuals could help to identify potential impacts, which are not properly characterized. In practice, the R package named 'ctsm-r' (Continuous Time Stochastic Modelling for R, version 0.6.17 [57]) is used in this study, which is based on the extended Kalman filter and maximum likelihood method for parameter estimation [21].

3.1.3 Decomposition of the lumped *gA*-value

As mentioned in section 1.2, in almost all grey-box modeling works in literature, the complicated impacts from solar radiation to indoor thermal dynamics (e.g., indoor temperature variation) are simplified as an Eq. (5), which is based on a single lumped gA-value. However, as visualized in Fig. 18, in reality, there are, at least, three main paths of heat transfer from the sun to the indoor environment, resulting in three main types of solar gains: Sg_{dir}, Sg_{dif} and ϕ_s. The explanations of three main solar gain types are tabulated in Table 3, but note that the real physical heat transfer from the sun to the PSO indoor thermal environment is even much more complicated than the diagrams of Fig. 18.

Table 3 Three main type of solar gains based on different heat transfer approaches, as visualized in Fig. 18.

Symbols	Type of solar gains	Explanation
Sg_{dir}	Direct solar gain (penetrating via glazing)	The part of solar gain (thermal radiation) from direct normal irradiance (*DNI*) that passes through the window glazing. As simplified visualized in Fig. 18, the heat transfer of Sg_{dir} is mainly based on two types of heat transfer: radiation and convection.
Sg_{dif}	Diffuse solar gain (penetrating via glazing)	The part of thermal radiation from diffuse horizontal irradiance (*DHI*) passing through the window glazing. As simplified illustrated in Fig. 18, the heat transfer of Sg_{dif} is mainly based on two types of heat transfer: radiation and convection.
ϕ_s	(Total) solar gain via opaque envelope	The part of solar gain from both direct normal irradiance (*DNI*) and diffuse horizontal irradiance (*DHI*), in essence from global horizontal irradiance (*GHI*) based on Eq. (9), via the opaque envelope, including opaque walls, roof, etc. As simplified shown in Fig. 18, the heat transfer of ϕ_s depends on all three types of heat transfer: radiation, conduction, and convection.



(a) 3D model diagram (a) Section diagram

Fig. 18 Diagrams of 'direct' and 'indirect' ways of solar energy-based heating to indoor environment.

With the three mains types of solar gains in mind (Fig. 18), it is not hard to find that modeling solar gain (*Sg*) based on only a single lumped gA_t might lead to a weak interpretation of estimated dynamic gA_t ,e.g., as in [33]. Specifically, in the single lumped gA_t based grey-box model, the part of ϕ_s will be 'forced' by the (super-simplified) physical framework, such as Eq. (1-2), to be incorporated in gA_t. This fact leads to an estimated gA_t (as in [33]), which is not perfectly in line with the gA-definition to merely account for the heat flux through window glazing (see section 1.2). Therefore, in this study, on the basis of [33], three types of solar gains - Sg_{dir}, Sg_{dif} and ϕ_s – will be systematically modelled. Fig. 23 and Fig. 25 could serve as references for systematic modeling the three main parts of solar gains. In addition, the reason of splitting the solar gain penetrating the windows (*Sg*) into two parts Sg_{dir} and Sg_{dif}, referring Eq. (5) and Table 3, is because Sg_{dif} links much less to the sun position, in contrast to Sg_{dir}. Specifically, Sg_{dir} and Sg_{dif} contributed by *DNI* and *DHI* respectively, and the *DNI* is strongly sun position-dependent, but the *DHI* varies limited with sun movement. Hence, the time-dependent (or sun position-dependent) nature of gA_t (section 1.2) is mainly attributed to the sun position dependence of Sg_{dir_t}, instead of Sg_{dif_t}. Hence, modeling the dynamics of the direct solar gain (Sg_{dir_t}) is of key interest, instead of Sg_t, since estimating the (lumped) Sg_t

straightforward could not merely lead to a clearer physical interpretation of the estimated outcomes, but also reduce estimation uncertainties. In this regard, a direct solar gain (Sg_{dir_t}) tailored solar aperture, denoted as gA_{dni_t}, is defined as Eq. (6). Meanwhile, a similar customized solar aperture for Sg_{dif_t} (diffuse solar gain) is proposed as gA_{dhi_t}, formulated as Eq. (7).

$$gA_{dni_t} = \frac{Sg_{dir_t}}{DNI_t} \tag{6}$$

$$gA_{dhi_t} = \frac{Sg_{dif_t}}{DHI_t} \tag{7}$$

Nevertheless, to reduce the grey-box model complexity, in this study, a constant gA_{dhi} assumption is applied to focus on the modeling of daily dynamic gA_{dni_t} curve, which is physically reasonable, since Sg_{dif_t} is much less sun position (time) dependent. The dynamic modeling of gA_{dni_t} curve is based on the customized B-splines, which will be introduced in section 3.2.

3.2 B-splines fitting

To model non-linear relationships, such as the key interest of this study: daily varying gA_{dni_t} curve with time, polynomial regression, spline regression, and generalized additive model (GAM) are commonly used techniques. Polynomial regression models dependent variable as an n^{th} degree polynomial of the independent variable, which requires an increased polynomial degree to fit the more abundant dynamics. However, the increased degree (of the polynomial) generally leads to enlarged uncertainties. Thus, to avoid high degrees required, spline regression split the full range of independent variable into several sub-intervals firstly and fit low degree polynomial in each sub-interval one by one, and connecting them with specific rules to provide a full integral smooth spline later on. Spline regression could lower the polynomial degree needed, but the optimized splitting in the first step is generally case by case, leading to its poor standardization. B-splines are used as flexible 'smoother' in GAM models [58], which meets the need of this study better (e.g., standard flow modeling), since a linear combination of weighted B-splines can represent any spline function in the same degree, and only one combination exists for a particular spline [59]. The pre-construction of B-spline is unified, and based on the B-splines technique; the complex modeling of non-linear relationships (e.g., spline) are transferred as 'analogues' of multiple linear regression, like Eq. (8). The general expression of B-splines fitting is formulated as Eq. (8), where ϕ_i is the weighting (scaling factor) of the i^{th} B-spline in, $Bs_{i,m}(t)$ in m^{th} order, with q basis splines in total.

$$S(t) = \sum_{i=1}^{q} \phi_i Bs_{i,m}(t) \tag{8}$$

The compact definition of B-splines is first reported by Schoenberg [60], and more practical information can be found in [33]. A vital property of B-splines is that the point-wise sum of the infinitely B-spline series is always equal to unity for the entire range of interest [61]. As stated, the B-splines are designated in this study to depict the daily dynamic gA_{dni_t} curve, where gA_{dni_t} is represented by $S(t)$, in Eq. (8). Therefore, the pre-construction of B-splines is restricted by the physical meaning of gA_{dni_t}. Considering the gA_{dni_t} has no physical meaning before sunrise or after sunset, in this study, the pre-constructed B-splines are customized to show in the time interval when a considerable global solar irradiation (GHI_t) appears. Fig. 19 shows that GHI_t generally appears from 8:00 to 17:00 for scenario 1 (January) and 6:00 to 20:00 for scenario 2 (August), at the said location of the PSO. Taking six or eight B-splines as examples, this leads to the customized pre-constructed B-splines for DS1-2 dataset respectively as visualized in Fig. 20. Regarding the B-splines application in R, the B-splines basis functions can be generated by the R-core package 'Splines' (version 3.6.3) [62].

In addition, another natural characteristic of B-splines is visualized in Fig. 20: only one to two basic splines dominate the start and end zones, marked as blue in Fig. 20. The feature of B-splines might create difficulties of B-splines fitting when the non-zero values of daily gA_{dni_t} curve starts (sunrise) and ends (sunset), since very limited flexibility of B-splines is available in the above-mentioned 'blue' zones (Fig. 20), leading to enlarged estimation uncertainties. This nature limitation of B-splines will be discussed in section 6.2. Moreover, candidate B-splines integrated grey-box models with 4, 6, 8, or 10 tailor-made B-splines will be explored subsequently in section 3.3, for DS1-2 datasets.

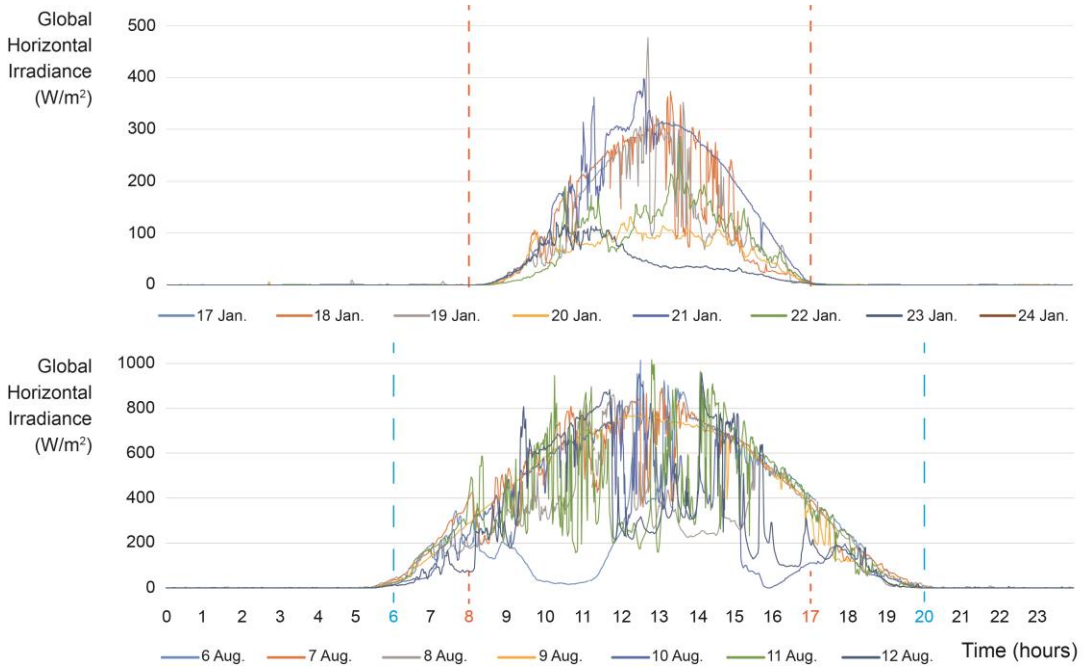


Fig. 19 Daily plots of global horizontal irradiance in one-minute frequency from 17 to 24 Jan. (up) and 6 to 12 Aug. 2020 (down).

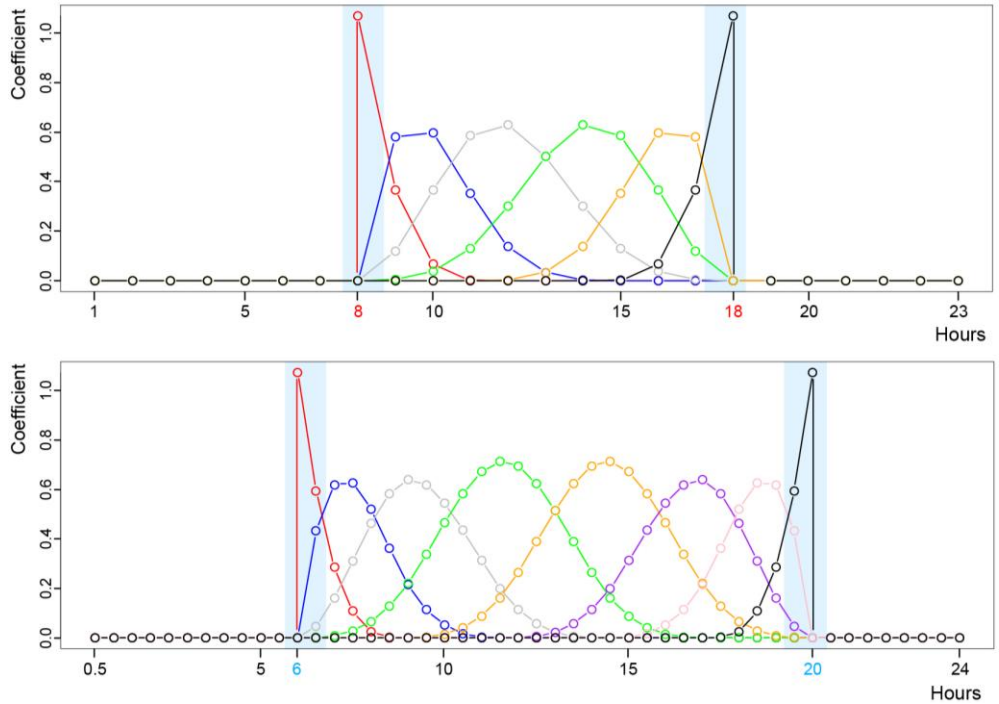


Fig. 20 Examples of customized daily B-splines in line with GHI_t presence in DS1 (up) and DS2 (down), taking six (up) and eight (down) splines as examples.

3.3 Selection of grey-box model

3.3.1 Candidate models and roadmap

A two-steps strategy is used to select the most suitable grey-box model for each of the PSO scenarios, as shown in Fig. 21 and Fig. 22. Firstly, based on a constant gA_{dni} assumption, the best model for the particular dataset is identified. Secondly, the pre-constructed customized B-splines - $S(t)$ in Eq. (8) – will replace the gA_{dni} in each of the selected models, to estimate the varying daily gA_{dni_t} curves for DS1 and DS2 respectively. The *Sn* suffixes (e.g., S6) indicate how many pre-constructed customized B-splines (e.g., Fig. 20) are used in the specific B-splines integrated grey-box model. The forward model selection for both datasets starts from the Ti* model, and iteratively extends step-by-step by increasing the complexity. The star mark (*) indicates the Sg is systematically considered independently, such as Sg_{dir} and Sg_{dif} are modeled separately in the Ti* model, and Sg_{dir}, Sg_{dif} and ϕ_s are constructed individually in TiTe*_aen model. In addition, "*Ria*" reflects the potential direct thermal linkage between indoor and outdoor without the impact of envelope capacity such as heat loss via air infiltration. The "*aen*" and "*ws*" indicate the modeling of solar energy absorbed by the envelope (ϕ_{s_t}) and the impact of wind on thermal resistances. Similar to Eq. (5), Eq (9) is proposed to estimate ϕ_{s_t}, assuming an (unknown) constant absorption coefficient (i.e., aen). In addition, following [63], the effects of wind speed on all thermal resistances are simplified modeled as a coefficient, 1/(1+*co***ws*), to the original thermal resistance (e.g., *Rie* and *Rea*). The unknown coefficients *aen* and *co* will be estimated in the data fitting of grey-box models. Moreover, the third state "*Tm*" accounts for the temperature of the interior thermal mass such as furniture, internal walls, etc.

$$\phi_{s_t} = aen * GHI_t \tag{9}$$

As shown in Fig. 21 and Fig. 22, two roadmaps of model identification for the DS1 and DS2 datasets are presented. The 'balanced' model (Fig. 17) selected in each roadmap generally refers to the one that shows the best overall performance in both *LogL* (high-value desired) and AIC (low-value wanted). In addition, both considerable increase of *LogL* (marked as red) and decline in AIC (signed as blue) indicate a success model extension, visualized as red arrows. Thus, guided by the red arrows, TiTe*_aen_ws_S6 and TiTe*_Ria_aen_ws_S8 models are selected for the PSO scenario 1 and 2 respectively. It is noted that, compared to AIC, BIC imposes a harsher penalty to the model complexity and prefers a model with limited parameters [64], since incorporating B-splines into the grey-box model expands the number of parameters (prm) sharply but might not improve the model significantly enough, to compensate for the penalty of BIC caused by the enlarged *prm*. Thus, the BIC value is only considered as a reference.

3.3.2 Selected two-states models

Specifically, for the DS1 dataset (Fig. 21), following the red arrows, indicating *LogL* increasing and AIC reducing, two constant gA_{dni} based grey-box models are selected in step 1: TiTe*_ws model (low *LogL* and lowest AIC and BIC) and TiTe*_aen_ws model (lowest *LogL* and low AIC and BIC). Extending the two models with customized B-splines in different numbers, both results in 4 B-splines (S4) as optimal, leading to TiTe*_aen_ws_S4 and TiTe*_ws_S4 being selected in step 2. However, in line with the gA_{dni_t} verification purpose stated in section 1.4, ϕ_s should be modelled separately (marked as 'aen'). Therefore, although the TiTe*_aen_ws_S4 model has a slightly higher AIC-value (for comparable *LogL*-value), this model is preferred. Moreover, extending model from TiTe*_aen_ws_S4 to TiTe*_aen_ws_S6, a slight increase of AIC is observed, however, a considerable *LogL* rise is also noticed (from -264.62 to -263.73). This might indicate enhanced flexibility of six B-splines is required, and only four B-splines is not sufficient for capturing the dynamics of gA_{dni_t} in DS1. Therefore, the TiTe*_aen_ws_S6 model is ultimately selected as the optimal one for the DS1 (see Fig. 21). The TiTe*_aen_ws_S6, illustrated in Fig. 23, is a two-state model, which systematically models Sg_{dir}, Sg_{dif}, ϕ_s independently and the wind impacts to indoor thermal dynamics in the system equations.

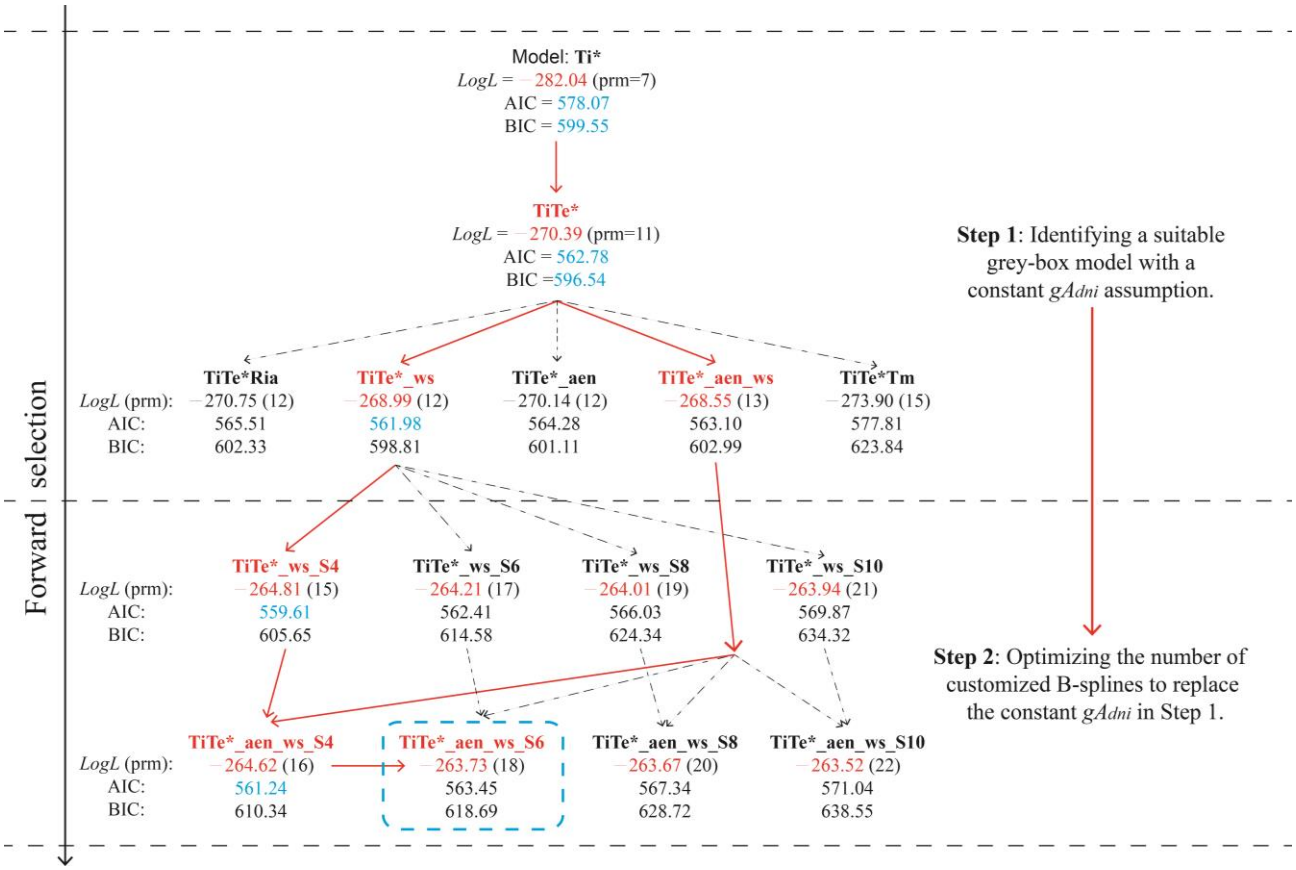


Fig. 21 The forward model selection roadmap for DS1 dataset (scenario 1).

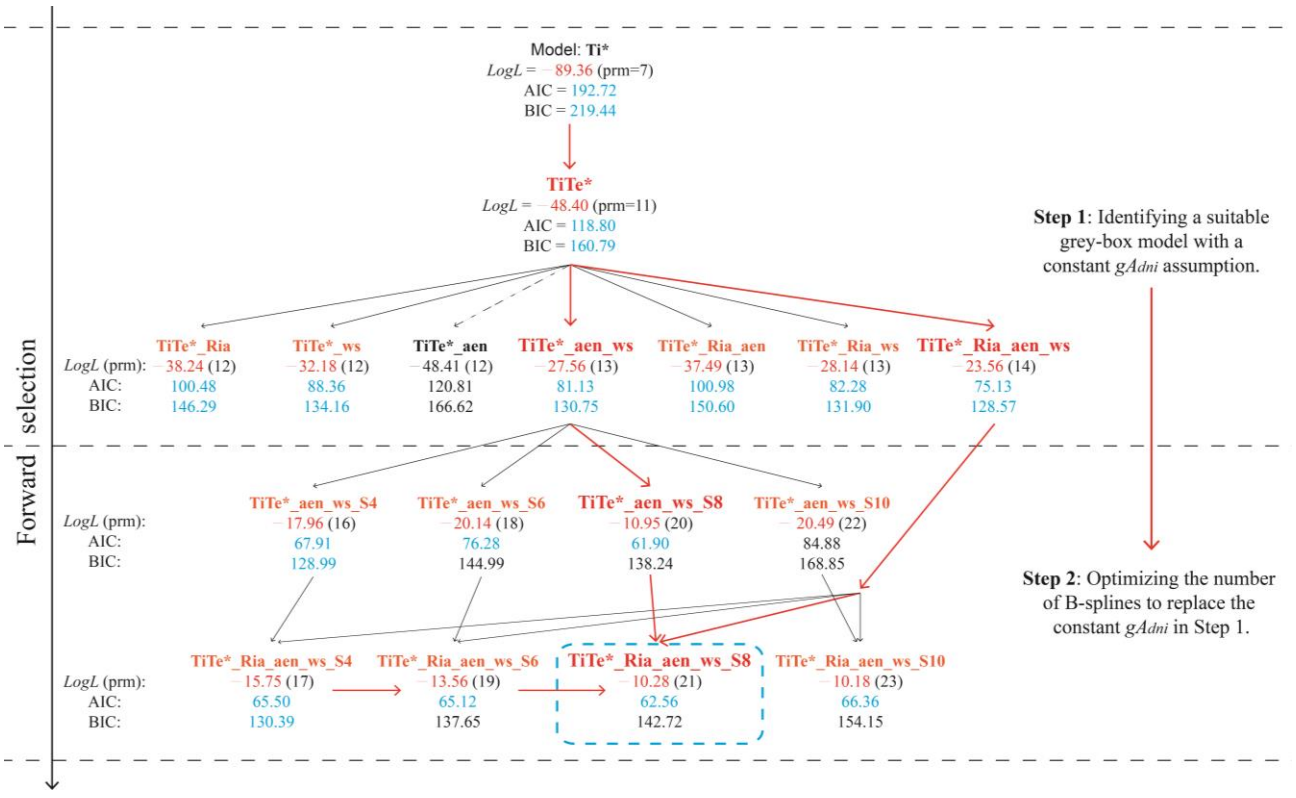


Fig. 22 The forward model selection roadmap for DS2 dataset (scenario 2).

Meanwhile, the red arrows in Fig. 22 point out the TiTe*_Ria_aen_ws_S8 is the selected model, in step 2, for the DS2 dataset (PSO scenario 2). Compared to DS1, the integration of '*Ria*' considerably improved the *LogL* performance in DS2 (Fig. 25), indicating the statistical significance of this model extension, which could be physically attributed to the added extra shading. Fig. 25 illustrates the RC model of TiTe*_Ria_aen_ws_S8 model, with its (white noise) residuals qualified via Fig. 26. It is noted that six and eight B-splines are needed for the PSO scenario 1 and 2 respectively (i.e. TiTe*_aen_ws_S6 for DS1 and TiTe*_Ria_aen_ws_S8 for DS2). This seems physically reasonable since the uneven distribution of windows in scenario 2 trends to result in richer dynamics of both Sg_{dir_t} and gA_{dni_t}. Additionally, for the two final models chosen for DS1-2, their residuals (e_k) between predicted (T_{ik}) and measured (T_{ik}^*) indoor temperature, see Eq. (3) as an example, should correspond to white noise residuals. Hence, Fig. 24 and Fig. 26 checks the characteristics of residuals (e_k) of TiTe*_aen_ws_S6 and TiTe*_Ria_aen_ws_S8, by the plots of model residuals versus time, the auto-correlation function (ACF), and the cumulated periodogram (CP), which indicate two said models meet the white noise residuals assumption. For more specifications, the reader refers to [13].

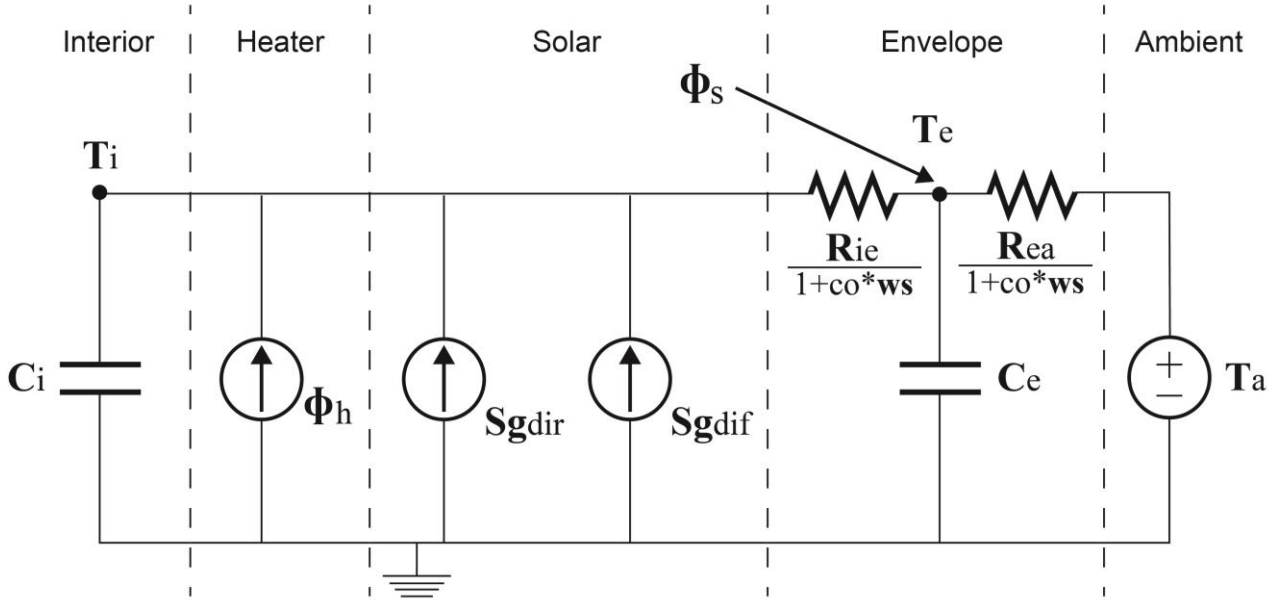


Fig. 23 The selected two-states single zone thermal RC-network model for the DS1 under scenario 1.

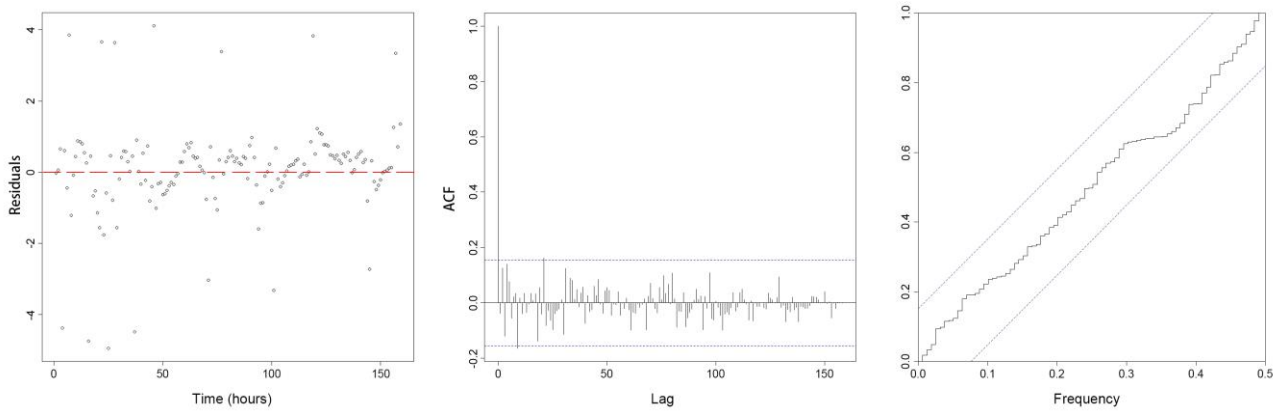


Fig. 24 The plots of residuals versus time, the auto-correlation function (ACF) and the cumulated periodogram (CP) of the residuals for the TiTe*_aen_ws_S6 model of DS1 dataset.

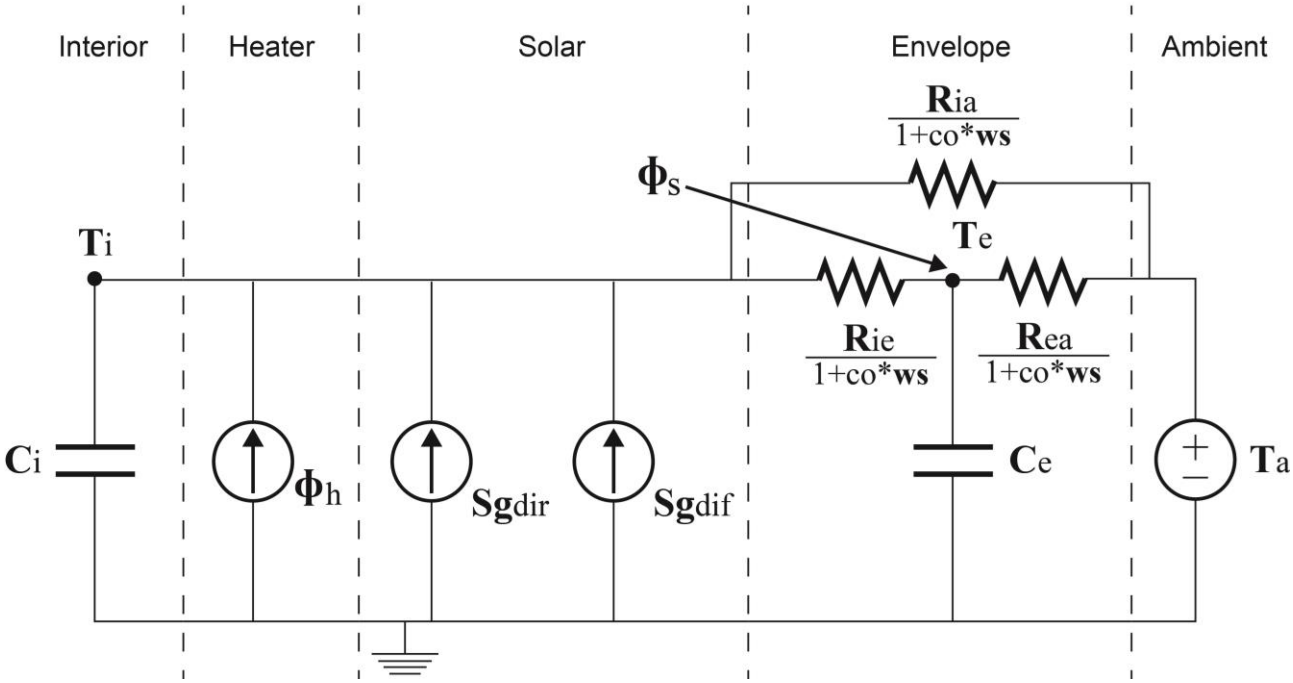


Fig. 25 The selected two-states single zone thermal RC-network model for the DS2 under scenario 2.

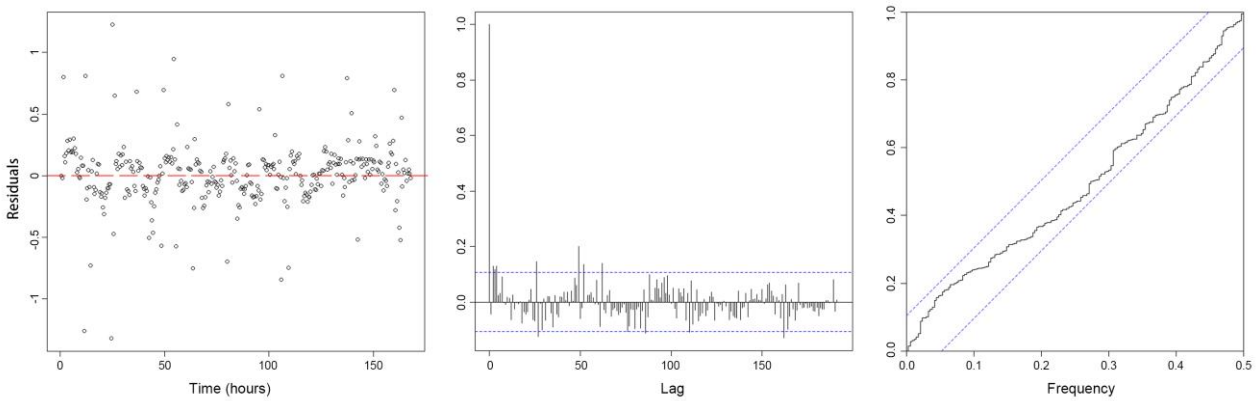


Fig. 26 The plots of residuals versus time, the auto-correlation function (ACF) and the cumulated periodogram (CP) of the residuals for the TiTe*_Ria_aen_ws_S8 model of DS2 dataset.

4. BES-model based simulation

Two classic BES-models are constructed in Modelica (Dymola) for the two PSO scenarios, and each BES-model will be compared with the grey-box model selected for the same scenario, in section 3.3.2. The comparison is aiming to, on the one hand, verify the physical interpretation of estimated outcomes from the enhanced grey-box modeling proposed in this study, and on the other hand, understand the potential limitations of both two techniques (BES- and the enhanced grey-box models) in revealing buidlings' dynamic gA_{dni_t} and Sg_{dir_t}. Appendix B elaborates the construction details of the two BES-models. In brief, the two BES-models, representing only the effective glazing parts of two PSO scenarios (Fig. 7 and Fig. 12), along with the corresponding weather files contribute to the simulated direct solar gains: $Sg_{dir\_sim\_s1_t}$ and $Sg_{dir\_sim\_s2_t}$ for PSO scenarios 1 and 2. As stated in the introduction, for most existing buildings, often not all input data is available, such as the exact layers and material properties of the building envelope, etc. Hence, in conformity with 'reality', only the glass part of the window system and the main shading effect from other buildings and tree groups were constructed in the BES-model in this study.

5. Results and comparison

5.1 Daily dynamic gA_{dni_t} estimations

Based on the selected TiTe*_aen_ws_S6 and TiTe*_Ria_aen_ws_S8 models, two daily dynamic gA_{dni_t} curves, denoted as $gA_{dni\_s1_t}$ and $gA_{dni\_s2_t}$, are evaluated by Eq. (10) and Eq. (11), for PSO scenario 1 and 2 respectively. In Eq. (10-11), *t* expresses daily 1 to 24 hours for DS1, or sequence numbers of 30mins for DS2, and θ is a vector of length *m* containing the spline parameters, and B is a n-by-m matrix containing m basis splines with their corresponding values for each of the n observations. Note that the estimation for each unknown parameter (e.g. ϕ_{s1i} and ϕ_{s2i}) has a mean value with a standard error, thus, based on Eq. (10-11), the estimated $gA_{dni\_s1_t}$ (or $gA_{dni\_s2_t}$) has its 95% confidence interval (zone), which is elaborated in Appendix C. In addition, by Eq. (6), the Modelica-simulated $gA_{dni\_sim\_s1_t}$ and $gA_{dni\_sim\_s2_t}$ for PSO scenario 1 and 2, in one hour or 30 mins sampling time respectively, are calculated based on Eq. (12-13). Note that, in the above-mentioned simulations, the measured DNI_s1_t and DNI_s2_t in the input weather files of the two BES-models is replaced by a constant value of 100 W/m², since the low values of measured direct normal irradiance (DNI_t) in the early morning and late afternoon might lead to simulation errors and this replacement will not impact on the estimations of $gA_{dni\_sim\_s1_t}$ and $gA_{dni\_sim\_s2_t}$.

$$gA_{dni\_s1_t} = \sum_{i=1}^{6}(\phi_{s1i} \cdot Bs_{s1i})_t = B_{s1}\theta_{s1} \quad (10)$$

$$gA_{dni\_s2_t} = \sum_{i=1}^{8}(\phi_{s2i} \cdot Bs_{s2i})_t = B_{s2}\theta_{s2} \quad (11)$$

$$gA_{dni\_sim\_s1_t} = \frac{Sg_{dir\_sim\_s1_t}}{DNI\_s1_t} \quad (12)$$

$$gA_{dni\_sim\_s2_t} = \frac{Sg_{dir\_sim\_s2_t}}{DNI\_s2_t} \quad (13)$$

5.2 Qualitative analysis and quantitative comparison

Prior to the quantitative comparison of the estimated (simulated) outcomes, offered by Eq. (10-13), Fig. 27 firstly visualizes the sun trajectories for the particular scenarios under winter (DS1) or summer (DS2) conditions. This analysis (Fig. 27) shows only one global gA_{dni_t} peak under scenario 1 in January (left) and two gA_{dni_t} 'peaks' under scenario 2 in August (right) are expected. Specifically, in scenario 1 (January) the one gA_{dni_t} peak is expected shortly after sunrise, when the direction of DNI_t is perpendicular to the southern façade of the building with its two windows. In addition, two gA_{dni_t} 'peaks' are expected in scenario 2, of which one appears immediately at sunrise, and the other one, gradual rising and declining, when the direction of DNI_t is perpendicular to the southwestern window. Note that Fig. 27 only visualizes the sun's movement of the first days of both datasets (see Table 4) as representatives, since a small variation in sun trajectory is observed within the seven to eight days (Table 4).

Table 4 Sunrise/sunset orientations in January and August at Leuven [65].

	January DS1 dataset with scenario 1			August DS2 dataset with scenario 2		
No.	Date	Sunrise	Sunset	Date	Sunrise	Sunset
Day 1	17-01-2020	123° ESE	237° WSW	06-08-2020	62° ENE	298° WNW
Day 2	18-01-2020	123° ESE	237° WSW	07-08-2020	62° ENE	297° WNW
Day 3	19-01-2020	122° ESE	238° WSW	08-08-2020	63° ENE	297° WNW
Day 4	20-01-2020	122° ESE	238° WSW	09-08-2020	63° ENE	296° WNW
Day 5	21-01-2020	122° ESE	239° WSW	10-08-2020	64° ENE	296° WNW
Day 6	22-01-2020	121° ESE	239° WSW	11-08-2020	64° ENE	295° WNW
Day 7	23-01-2020	121° ESE	239° WSW	12-08-2020	65° ENE	295° WNW
Day 8	24-01-2020	120° ESE	239° WSW			

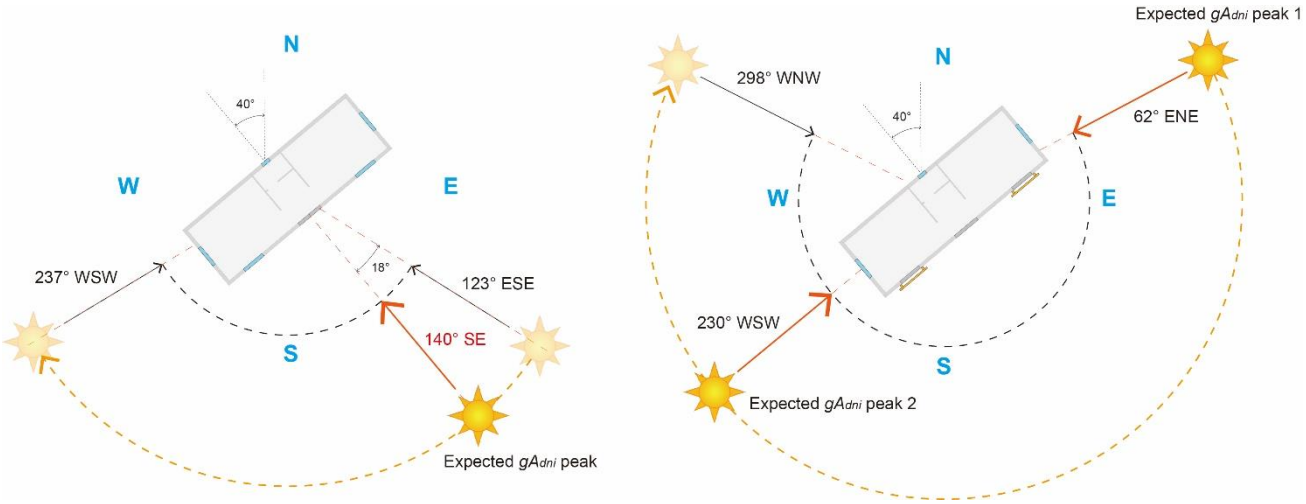


Fig. 27 The expected peak(s) of daily dynamic gA_{dni_t} for January DS1 (left) and August DS2 (right).

Based on the outcomes from Eq. (10-13), Fig. 28 and Fig. 29 further quantitatively compare the estimated (or simulated) gA_{dni_t} for the two scenarios. The expected key dynamic features of both scenario 1 and 2 are clearly unfolded by both approaches, such as the existence of one (two) daily gA_{dni_t} peak(s) for January (August). In addition, more key details are well revealed by both two approaches: for example, in January, the one daily gA_{dni_t} peak is expected to appear around 11 am (Fig. 28), and the two daily gA_{dni_t} peaks for August, of which the first jumps to a high maximum at sunrise and the second one increases in the afternoon and falls back to a lower value gradually over time toward sunset (Fig. 29). Therefore, it is concluded that both the dynamic gA_{dni_t} tailored B-splines integrated grey-box and BES-models, at least in this cases study, are able to reveal the main trend and key characteristics of the daily dynamic gA_{dni_t} curve. Nevertheless, the grey-box fitted and BES-simulated course are not in perfect agreement (Fig. 28-29). At this stage, it is hard to answer which one is 'right', since both two approaches have their limitations, which will be discussed in the next section. The two constant gA_{dni} estimated by TiTe*_aen_ws and TiTe*_Ria_aen_ws models with their corresponding 95% confidence intervals are also shown in Fig. 28-29. It is clear that a constant solar aperture ignores the rich details of gA_{dni_t} dynamics to a large extent, which could increase prediction uncertainties.

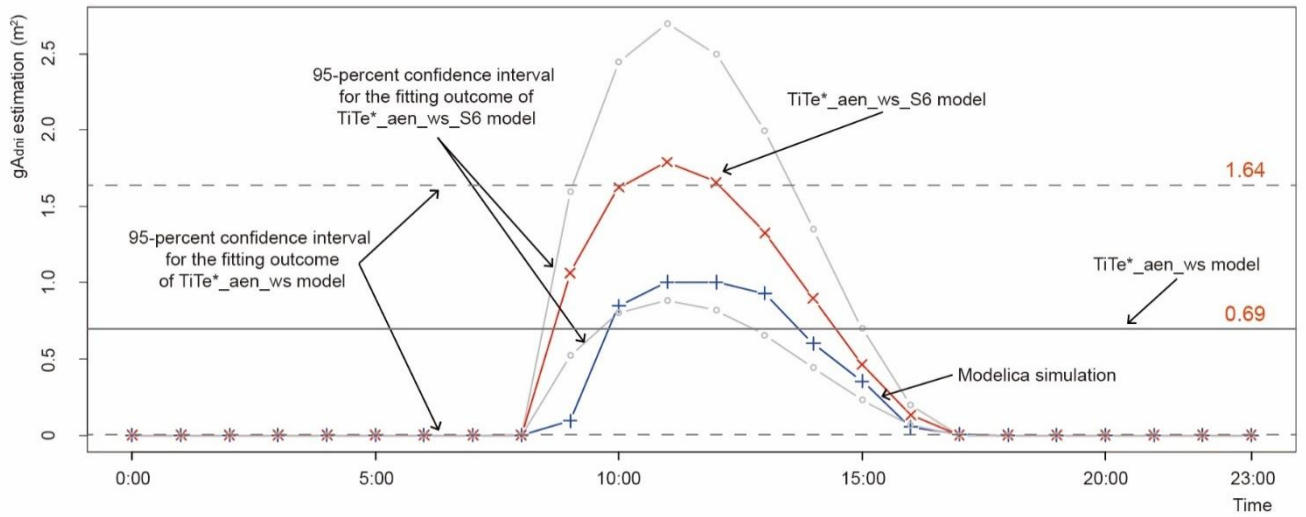


Fig. 28 The hourly dynamic gA_{dni_t} comparison for scenario 1.

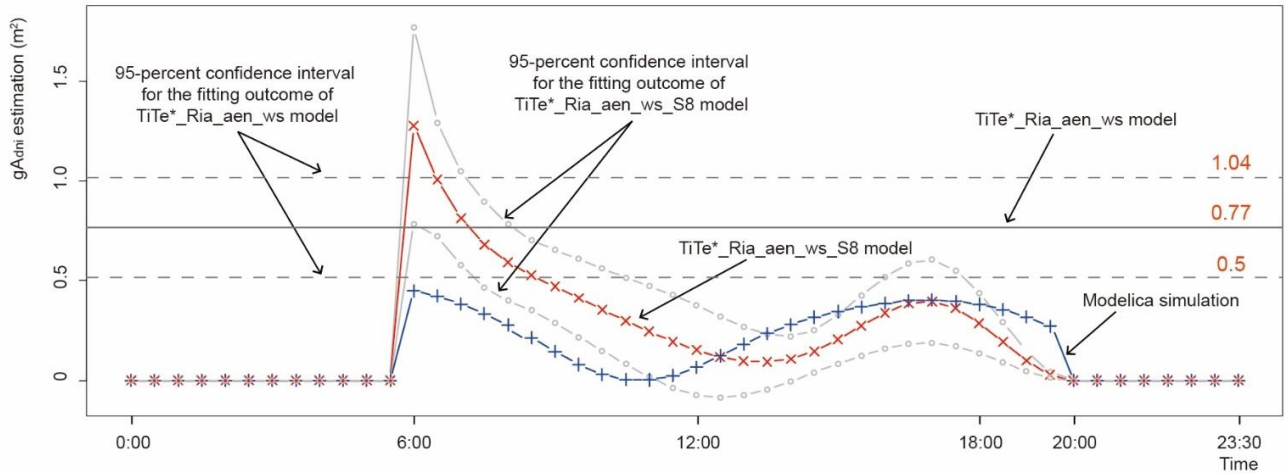


Fig. 29 The hourly dynamic gA_{dni_t} comparison for scenario 2.

For example, Fig. 30 compares the predicted direct solar gains by the three methods (constant gA_{dni} and dynamic gA_{dni_t} based grey-box models and BES-model) for the DS2 dataset (PSO scenario 2). It visualizes how the constant gA_{dni} assumption in TiTe*_Ria_aen_ws model predicts high solar gains around noon, when the solar beam (i.e. DNI_t) is orientated to the PSO main façade of which the two windows are shaded (Fig. 11). It is not hard to understand the prediction uncertainties caused by this wrong Sg_{dir_t} dynamic information. This clearly illustrates why a constant gA_{dni} based Sg_{dir_t} information is rejected in some model predictive control (MPC) studies [38,44].

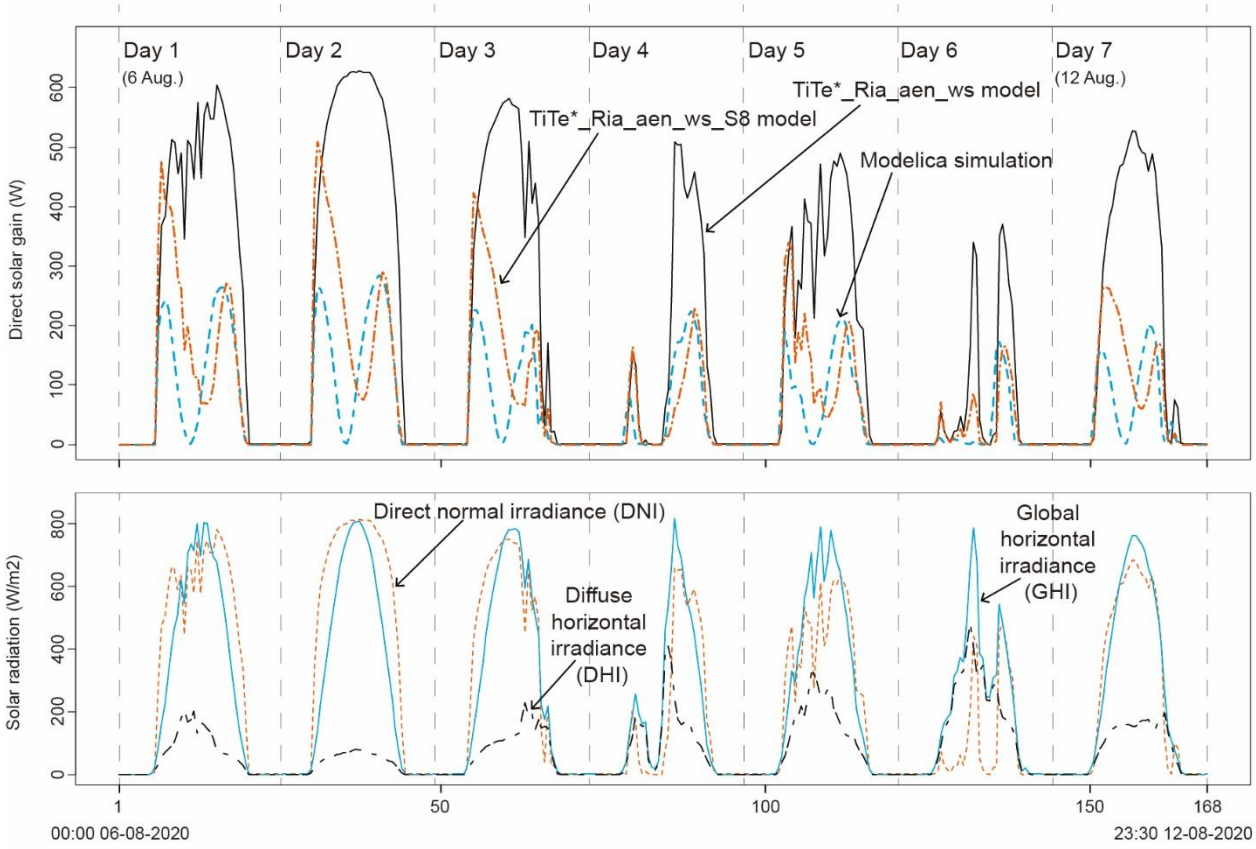


Fig. 30 The comparison of three types of direct solar gain estimations in scenario 2.

6. Limitations of grey-box and white-box approaches

6.1 Limitations of the white-box approach (BES-model)

As stated in section 5.2, discrepancies are observed for the dynamic details between daily gA_{dni_t} curves estimated by proposed grey-box and classic white-box approaches for both scenarios (Fig. 28-29). The differences might be caused by the limitations of both approaches. As mentioned in section 1.1, the key limitation of the BES-model simulation is its massive demand of input data, such as layers and material properties of the building envelope and the shading effect from individual obstacles, which is often unavailable. As reported in section 4, the BES-model in this study, merely considered the windows' glazing and the main shading effect (e.g. from buildings and tree groups), but ignored solar gains through the opaque building envelope and shading from discrete individual trees. Ignoring the opaque building fabric parts could cause an 'ahead-of-time' estimation of the gA_{dni_t} dynamics in BES-simulation. Specifically, without modeling the opaque building fabrics, simulated direct solar gain such as $Sg_{dir\_sim\_s1_t}$ is assumed to influence the indoor air temperature immediately. However, as shown in Fig. 18, the real heat transfer process will take some time, since the solar beam heats interior building surfaces via radiation firstly and the heated surface will gradually warm up the indoor air via convection subsequently. Therefore, the BES-model has a certain risk of an 'ahead-of-time' estimation, referring to Zone C in Fig. 31, in simulated $gA_{dni\_sim_t}$. Moreover, when the beam radiation is blocked, such as the sun facing the covered windows in scenario 2, the simulated $Sg_{dir\_sim_t}$ will drop more sharply than 'reality'. In the real case, the stored part of solar gain in the building envelop will still be gradually released as an alternative 'energy supplier' to the indoor air mass, postponing the drop of indoor temperature. In addition, unmodeled individual trees surrounding the PSO might underestimate potential shading effects, resulting in 'overestimation' in the simulated $Sg_{dir\_sim_t}$, compared to the real case (refers to Zone D in Fig. 31). For instance, Fig. 32 shows a potential significant shading effect from a few individual trees to the southwest window of PSO, which is unmodelled in the simplified BES-models.

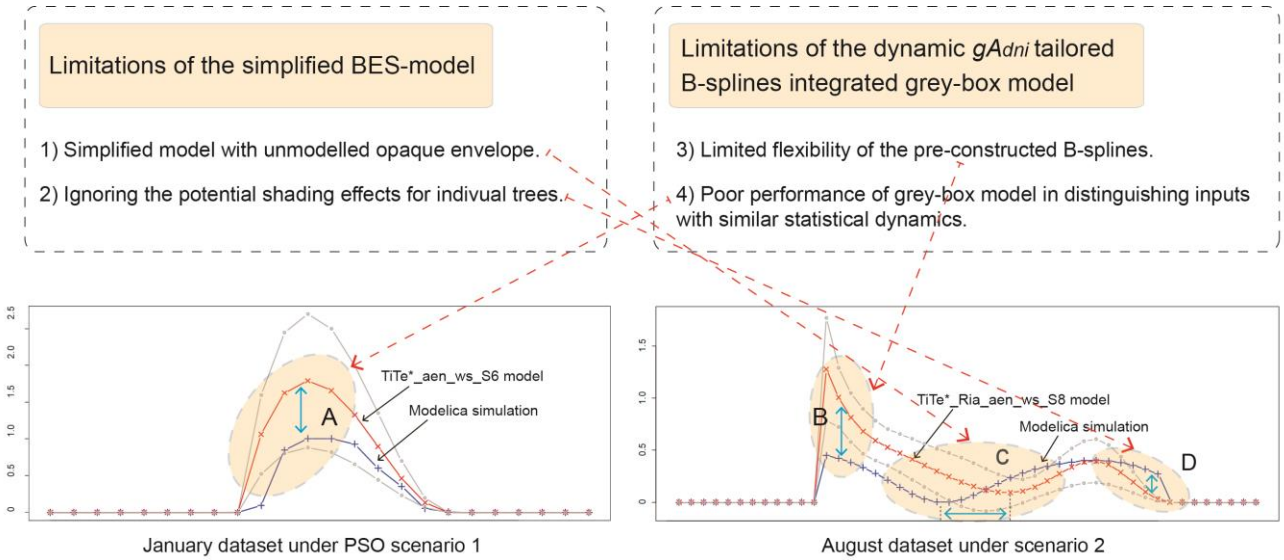


Fig. 31 Summarized limitations of two approaches and possible explanations for deviations between the fitted and simulated gA_{dni_t}.

Fig. 32 Potential shading effects from several individual trees to the southwestern window of PSO.

6.2 Limitations of the B-splines integrated grey-box approach

Meanwhile, the method proposed in this study (i.e. enhanced B-splines integrated grey-box model) also has its limitations 1) limited flexibility of B-splines for particular zones and 2) limitation of grey-box modeling in decoupling lumped physical phenomena. First, for the pre-constructed B-splines, as shown in Fig. 20, around sunrise and sunset (marked as blue zones), there is only one (or two) basis spline with a 'steep' shape available for the data fitting. Close to sunrise and sunset, the direct solar radiation (DNI_t) is very low. Hence, the scaling factor(s) of the particular basis spline(s) have to fit the Ti_t dynamics under a low DNI_t, prohibiting a reliable estimation of gA_{dni_t} dynamics. Especially, in PSO scenario 2, visualized in Fig. 27, the east-northern window is immediately exposed to solar radiation as soon as the sun is above the horizon, leading to a considerable Ti_t increase. With a low DNI_t at that moment, it is statistically 'reasonable' to assign a physically 'unreasonably' high gA_{dni_t} value for the particular moments. As reported in section 2, the glazed area of the north-eastern window is around 0.7 m^2, indicating the gA_{dni_t} value should be lower than 0.7 m^2 in the morning. However, the estimated gA_{dni_t} ranges from 1.3-0.7 m^2 in the morning hours (Fig. 29). Luckily, the exaggerated gA_{dni_t} estimation in the early morning or late afternoon will only impose a limited effect on the grey-box model states, due to the low amount of corresponding Sg_{dir_t}, caused by the low DNI_t-value at that time.

The second limitation is mainly attributed to the data-driven feature of the grey-box model. As stated in section 1.1, the grey-box model estimation is based on statistical correlations of dynamics between different input data, such as indoor temperature (Ti_t) and direct normal irradiance (DNI_t). This mechanism of estimation makes it hard to separate each physical phenomenon (or decouple the lumped physical phenomena), when two separated physical processes are highly statistically correlated. Given an example, in PSO scenario 2, when the sun orientates to the two covered windows in Fig. 27, direct solar gains are blocked, but the heat release from the solar gain through the opaque envelope (ϕ_{s_t}; unconsidered in the BES-models in section 4) might still be statistically attributed to DNI_t, owing to considerable statistical correlation between Ti_t and DNI_t during the noon hours. This might explain why the dynamic gA_{dni_t} never drops to zero and keeps a value above 0.1 m^2 (Zone C in Fig. 31) during the noon period, which is again physically 'unreasonable' in fact, but 'reasonable' in statistics. Given another example, in PSO scenario 1, the total area of two windows on the PSO southern façade is 1.4 m^2, indicating the maximum of $gA_{dni\_s1_t}$. However, the peak of estimated $gA_{dni\_s1_t}$ is beyond 1.5 m^2 (Fig. 28). Keeping the decoupling difficulty of the grey-box model in mind, this could explain why the estimated gA_{dni_t} of grey-box model trends to be higher than simulated $gA_{dni\_sim_t}$ most of the time, since the estimated gA_{dni_t} is not purely contributed by solar gains through the windows and might incorporate partial ϕ_{s_t}. To sum up, the limitations of both data-driven and numerical approaches are summarized in Fig. 31, and potential main explanations for deviations are specified. In addition, let us stress again, that none of the blue or red curves in Fig. 28-29 are in line with the 'truth', due to the said limitations of both approaches.

7. Conclusion

In this study, an enhanced B-splines integrated grey-box modeling technique is proposed to gauge gA_{dni_t} and Sg_{dir_t}, as an efficient alternative to the classic building energy simulation (BES-) method. The significance of precise gA_{dni_t} (or *gA*) information, instead of the commonly used constant one, to prediction works, such as in model predictive control (MPC), is also demonstrated. It was shown that a constant gA_{dni} assumption can lead to high solar gain predictions, when there is in fact no direct solar gain. Meanwhile, the method proposed in this study is compared with the classic BES-model simulation method, to test both its robustness and physical interpretation of estimation outcomes. The comparison shows both two approaches (i.e. proposed advanced grey-box and classic BES-model) can estimate (or simulate) the key dynamic features of daily gA_{dni_t} curve (Fig. 28 and Fig. 29), but both have certain limitations on quantifying dynamic details. However, considering the key dynamics of gA_{dni_t} and the variation between the estimated dynamic gA_{dni_t} curves and 'true' ones is small, the proposed method is still regarded as a promising technique, to fix the problem caused by the common constant solar aperture assumption, referring to Fig. 4 and Fig. 30. Most importantly, this proposed method cooperates with in-situ data and only requires a very limited amount of input data (e.g. 30 mins or hourly data of 7 parameters for around one week), showing its merits. Its excellence could be understood as two-fold: 1) The data measured on-site is generally regarded as more reliable, since it embeds 'everything' into data, including 'known' and 'unknown' impacts, where the 'unknown' part, such as unexpected shading effects is almost impossible to be incorporated into BES-models; 2) the required data (i.e. *Ti*, *Te*, ϕ_h, *GHI*, *DNI*, *DHI*, *ws*) are generally easy to access, and the low demands of data size (e.g. only one week date sampling in 30 mins) could significantly reduce the cost of measurement. To sum up, the technique (gA_{dni_t} tailored B-splines integrated grey-box model) proposed in this study could work much more efficiently than the traditional numerical (BES-model) approach, and the estimated gA_{dni_t} curves could significantly reduce the prediction uncertainties caused by the invariant solar aperture assumption, which is widely used in current grey-box modeling works. Therefore, the technique can considerably contribute to multiple fields, such as building thermal performance assessment, model predictive control, fault detection and diagnosis, etc. In future work, more cases studies with increased complexity, such as multiple thermal zones with complex occupancy profiles could be studied to tests the robustness of this technique.

Acknowledgment

This research was funded by the Research Foundation Flanders (FWO), application number G0D2519N, and by KU Leuven, grant C24/18/040. This support is gratefully acknowledged.

Appendix A

In practice, the indoor temperature (*Ti*) in DS 1-2 datasets is processed in two steps. First, the 5 mins indoor temperature data from 8 horizontally and vertically distributed sensors (Fig. 7 and Fig. 12) are averaged to one representative indoor temperature. Then, the resulting 5 mins (representative) indoor temperature is further averaged to hourly values for DS1 or 30 mins ones for DS2 and used for the subsequent statistical modellings. For the processing of the remaining data in DS 1-2 (i.e. outdoor temperature (*Te*), heating input (*Ph*), direct normal irradiance (*DNI*), etc.), the values are averaged directly from the 5 mins (or 1 min) data to 60 or 30 mins frequency in DS 1-2 respectively. For the measurement accuracies and frequencies of original data used for DS 1-2 datasets construction, they are tabulated in Table 5.

Table 5 The basic information of data measurement.

Data type	Instrument	Manufacturer	Type	Measuring range	Accuracy	Frequency
Indoor temperature	--	Eltek	GC-10	+5 / 40°C	± 0.4 %	5 mins
Heating input	power meter	Elster	A100C	100A	± 1 %	5 mins

Ambient temperature	Air Temperature Sensor (HMP155A)	Vaisala	--	-40 / +60°C	±0.2°C	1 min
Global radiation	pyranometer	Kipp & Zonen	CMP 3	--	±3%	1 min
Diffuse radiation	pyranometer	Kipp & Zonen	CMP11-V	--	±1.5%	1 min
Direct normal radiation	pyrheliometer	Kipp & Zonen	SHP1-V	--	±1.1%	1 min
Wind Speed	ultrasonic anemometer	Thies	3D No. 4.3830.20.340	0-85 m/s	0-5 m/s: ±0.1 m/s; > 5 m/s: ±2%	1 min

Appendix B

Modelica is an open-source object-orientated modeling language supporting multi-domain physical system modeling, which is increasingly used in the dynamic simulation of thermal-hydraulic systems [66–68] and building energy systems [66,69–72]. In this study, in cooperation with the Dymola tool, the IDEAS (Integrated District Energy Assessment by Simulation) library version 2.1 is used. The simulation reliability of the IDEAS library has been verified in several approaches [73]. As stated, the two BES-models are minimal models, indicating that only the transparent part of window system (without considering window frames or opaque elements) is considered and constructed, in line with the difficulties of data-collecting in reality. Thus, limited data is required, which is: 1) the volume of the indoor air mass, 2) the size and properties of the glazed envelope, 3) the total size of ceiling and floor (the construction details can be set as default), and 4) information about the surrounding shading obstacles (e.g. heights and distances to the PSO). Besides, the glass property for all PSO windows is set as uncoated double glazing with a U-value of 2.9 W/m²K, 4mm (glass)/ 12mm (spacer)/ 4mm(glass). In addition to the BES-models, weather files are needed. Two weather files, exactly in line with DS1 and DS2 datasets (i.e. same in time span, data value, and frequency) are constructed according to the guidance of [74]. Orientated by the Sg_{dir_t} and gA_{dni_t} estimation in this study, only the global horizontal radiation (W/m^2), direct normal radiation (W/m^2), and diffuse horizontal radiation (W/m^2) are required here, and the remaining data are set to zero in the two weather files. Finally, based on the constructed BES-models, modelling the glazed enveloped of two PSO scenarios as shown in Fig. 7 and Fig. 12, along with the corresponding pre-established weather files, dynamic Sg_{dir_t} outcomes for the two PSO scenarios (denoted as $Sg_{dir\_sim\_s1_t}$ and $Sg_{dir\_sim\_s2_t}$) are simulated in a pre-set 5 mins frequency. Then, the original 5mins simulated dynamic Sg_{dir} outcomes are averaged to hourly $Sg_{dir\_sim\_s1_t}$ and $Sg_{dir\_sim\_s2_t}$ in 30 mins frequency ultimately, to ensure the consistency of frequencies in both estimated dynamic Sg_{dir_t} and gA_{dni_t} outcomes from BES- and grey-box models.

Appendix C

As reported in section 5.1, The mean values ($B\theta$) of dynamic $gA_{dni\_s1_t}$ and $gA_{dni\_s2_t}$ can be. Meanwhile, the standard errors for the estimations of two dynamic gA_{dni_t} can be calculated in a general from - $\sqrt{\text{diag}(B\Sigma B^{\mathrm{T}})}$, where diag($B\Sigma B^{\mathrm{T}}$) is the diagonal of the n-by-n matrix $B\,\Sigma\,B^{\mathrm{T}}$ and Σ is the covariance matrix of the spline parameters. Additionally, the covariance matrix (Σ) can be calculated based on Eq. (16) with the standard errors of the spline parameters σ_s and the correlation matrix of the spline parameters ρ obtained directly from ctsm-r, where $\circ$ is the Hadamard operator representing elementwise multiplications of the two m-by-m matrices. Finally, assuming that the parameter estimates are normally distributed, the 95% confidence interval of the estimated spline of dynamic $gA_{dni\_s1_t}$ and $gA_{dni\_s2_t}$ (CI_{95}) can be calculated via Eq. (16-17), where Φ is the cumulative distribution function of the normal distribution. It refers to the zones between two grey lines (visualized in Fig. 28-29) as examples for the 95-percentage confidence interval (CI_{95}) of gA_{dni_t}.

$$\Sigma = \sigma_s {\sigma_s}^{\mathrm{T}} \circ \rho \tag{14}$$

$$CI_{95} = B\theta \pm \Phi\left(0.5\left(1 - \frac{95}{100}\right)\right)\sqrt{\text{diag}(B\Sigma B^{\mathrm{T}})} \tag{15}$$

Reference

[1] IEA(International Energy Agency). World Energy Outlook 2021. 2021. https://doi.org/https://iea.blob.core.windows.net/assets/88dec0c7-3a11-4d3b-99dc-8323ebfb388b/WorldEnergyOutlook2021.pdf.

[2] Zhang X, Lau S-K, Lau SSY, Zhao Y. Photovoltaic integrated shading devices (PVSDs): A review. Sol Energy 2018;170. https://doi.org/10.1016/j.solener.2018.05.067.

[3] Geert B, Katia R, Staf R, Jade D, Gabrielle M, Marieline S, et al. EBC Annex 71: Building Energy Performance Assessment Based on In-situ Measurements. Leuven: 2021.

[4] Chaudhuri T, Soh YC, Li H, Xie L. A feedforward neural network based indoor-climate control framework for thermal comfort and energy saving in buildings. Appl Energy 2019;248:44–53. https://doi.org/https://doi.org/10.1016/j.apenergy.2019.04.065.

[5] Fang T, Lahdelma R. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. Appl Energy 2016;179:544–52. https://doi.org/https://doi.org/10.1016/j.apenergy.2016.06.133.

[6] Roels S, Bacher P, Bauwens G, Castaño S, Jiménez MJ, Madsen H. On site characterisation of the overall heat loss coefficient: Comparison of different assessment methods by a blind validation exercise on a round robin test box. Energy Build 2017. https://doi.org/10.1016/j.enbuild.2017.08.006.

[7] Bauwens G. In Situ Testing of a Building's Overall Heat Loss Coefficient - Embedding Quasi-stationary and Dynamic Tests in a Building Physical and Statistical Framework. KU Leuven, 2015.

[8] Deconinck AH, Roels S. Comparison of characterisation methods determining the thermal resistance of building components from onsite measurements. Energy Build 2016. https://doi.org/10.1016/j.enbuild.2016.08.061.

[9] De Coninck R, Magnusson F, Åkesson J, Helsen L. Toolbox for development and validation of grey-box building models for forecasting and control. J Build Perform Simul 2016;9:288–303. https://doi.org/10.1080/19401493.2015.1046933.

[10] Killian M, Mayer B, Kozek M. Cooperative fuzzy model predictive control for heating and cooling of buildings. Energy Build 2016;112:130–40. https://doi.org/https://doi.org/10.1016/j.enbuild.2015.12.017.

[11] Andriamamonjy A, Klein R, Saelens D. Automated grey box model implementation using BIM and Modelica. Energy Build 2019;188–189:209–25. https://doi.org/https://doi.org/10.1016/j.enbuild.2019.01.046.

[12] Reynders G, Nuytten T, Saelens D. Robustness of reduced-order models for prediction and simulation of the thermal behavior of dwellings. Proc. BS2013 13th Conf. Int. Build. Perform. Simul. Assoc. Chambéry, Fr., 2013.

[13] Bacher P, Madsen H. Identifying suitable models for the heat dynamics of buildings. Energy Build 2011. https://doi.org/10.1016/j.enbuild.2011.02.005.

[14] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 2018;81:1192–205. https://doi.org/https://doi.org/10.1016/j.rser.2017.04.095.

[15] Pan Y, Zhang L. Data-driven estimation of building energy consumption with multi-source heterogeneous data. Appl Energy 2020;268:114965. https://doi.org/https://doi.org/10.1016/j.apenergy.2020.114965.

[16] Ciulla G, D'Amico A. Building energy performance forecasting: A multiple linear regression approach. Appl Energy 2019;253:113500. https://doi.org/https://doi.org/10.1016/j.apenergy.2019.113500.

[17] Afram A, Janabi-Sharifi F. Review of modeling methods for HVAC systems. Appl Therm Eng 2014;67:507–19. https://doi.org/10.1016/j.applthermaleng.2014.03.055.

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

[19] Crabb JA, Murdoch N, Penman JM. A simplified thermal response model. Build Serv Eng Res Technol 1987;8:13–9. https://doi.org/10.1177/014362448700800104.

[20] Howell JR, Menguc MP, Siegel R. Thermal radiation heat transfer. CRC press; 2010.

[21] Kristensen NR, Madsen H, Jørgensen SB. Parameter estimation in stochastic grey-box models. Automatica 2004;40:225–37. https://doi.org/https://doi.org/10.1016/j.automatica.2003.10.001.

[22] Madsen H. Time series analysis. CRC Press; 2007.

[23] Madsen H, Bacher P, Bauwens G, Deconinck AH, Reynders G, Roels S, et al. IEA EBC Annex 58: Report of Subtask 3, part 2: Thermal performance characterisation using time series data – statistical guidelines. 2016.

[24] Yu X, Georges L, Imsland L. Data pre-processing and optimization techniques for stochastic and deterministic low-order grey-box models of residential buildings. Energy Build 2021;236:110775. https://doi.org/https://doi.org/10.1016/j.enbuild.2021.110775.

[25] Yu X, Skeie KS, Knudsen MD, Ren Z, Imsland L, Georges L. Influence of data pre-processing and sensor dynamics on grey-box models for space-heating: Analysis using field measurements. Build Environ 2022:108832. https://doi.org/https://doi.org/10.1016/j.buildenv.2022.108832.

[26] 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. https://doi.org/https://doi.org/10.1016/j.enbuild.2014.01.038.

[27] Hu J, Karava P. A state-space modeling approach and multi-level optimization algorithm for predictive control of multi-zone buildings with mixed-mode cooling. Build Environ 2014;80:259–73. https://doi.org/https://doi.org/10.1016/j.buildenv.2014.05.003.

[28] Reynders G, Diriken J, Saelens D. Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals. Energy Build 2014;82:263–74. https://doi.org/https://doi.org/10.1016/j.enbuild.2014.07.025.

[29] Zhang X, Ritosa K, Saelens D, Roels S. Comparing statistical modeling techniques for heat loss coefficient estimation using in-situ data. J Phys Conf Ser 2021;2069:12101. https://doi.org/10.1088/1742-6596/2069/1/012101.

[30] Oliveti G, Arcuri N, Bruno R, De Simone M. An accurate calculation model of solar heat gain through glazed surfaces. Energy Build 2011;43:269–74. https://doi.org/https://doi.org/10.1016/j.enbuild.2010.11.009.

[31] Kontoleon KJ. Glazing solar heat gain analysis and optimization at varying orientations and placements in aspect of distributed radiation at the interior surfaces. Appl Energy 2015;144:152–64. https://doi.org/https://doi.org/10.1016/j.apenergy.2015.01.087.

[32] Evola G, Marletta L. The Solar Response Factor to calculate the cooling load induced by solar gains. Appl Energy 2015;160:431–41. https://doi.org/https://doi.org/10.1016/j.apenergy.2015.09.072.

[33] Rasmussen C, Frölke L, Bacher P, Madsen H, Rode C. Semi-parametric modelling of sun position dependent solar gain using B-splines in grey-box models. Sol Energy 2020;195:249–58. https://doi.org/10.1016/J.SOLENER.2019.11.023.

[34] Hens H. Heat transfer. Build Phys - Heat, Air Moisture 2017:15–124. https://doi.org/doi:10.1002/9783433608548.ch1.

[35] Jiménez MJ, Madsen H. Models for describing the thermal characteristics of building components. Build Environ 2008;43:152–62. https://doi.org/https://doi.org/10.1016/j.buildenv.2006.10.029.

[36] Olazo-Gómez Y, Herrada H, Castaño S, Arce J, Xamán JP, Jiménez MJ. Data-Based RC Dynamic Modelling to Assessing the In-Situ Thermal Performance of Buildings. Analysis of Several Key Aspects in a Simplified Reference Case toward the Application at On-Board Monitoring Level. Energies 2020;13. https://doi.org/10.3390/en13184800.

[37] Díaz-Hernández HP, Torres-Hernández PR, Aguilar-Castro KM, Macias-Melo E V, Jiménez MJ. Data-Based RC Dynamic Modelling Incorporating Physical Criteria to Obtain the HLC of In-Use Buildings: Application to a Case Study. Energies 2020;13. https://doi.org/10.3390/en13020313.

[38] Sourbron M, Verhelst C, Helsen L. Building models for model predictive control of office buildings with concrete core activation. J Build Perform Simul 2013;6:175–98. https://doi.org/10.1080/19401493.2012.680497.

[39] Erfani Beyzaee A, Yu X, Kull TM, Bacher P, Jafarinejad T, Roels staf, et al. Analysis of the impact of predictive models on the quality of the model predictive control for an experimental building. Proc Build Simul 2021 2021.

[40] Saelens D. Energy performance assessment of single storey multiple-skin facades. KU Leuven, 2002.

[41] Karlsson J, Roos A. Modelling the angular behaviour of the total solar energy transmittance of windows. Sol Energy 2000;69:321–9. https://doi.org/https://doi.org/10.1016/S0038-092X(00)00083-9.

[42] Hollick FP, Gori V, Elwell CA. Thermal performance of occupied homes: A dynamic grey-box method accounting for solar gains. Energy Build 2020;208:109669. https://doi.org/https://doi.org/10.1016/j.enbuild.2019.109669.

[43] Thilker CA, Bacher P, Bergsteinsson HG, Junker RG, Cali D, Madsen H. Non-linear grey-box modelling for heat dynamics of buildings. Energy Build 2021;252:111457. https://doi.org/https://doi.org/10.1016/j.enbuild.2021.111457.

[44] Picard D, Drgoňa J, Kvasnica M, Helsen L. Impact of the controller model complexity on model predictive control performance for buildings. Energy Build 2017;152:739–51. https://doi.org/https://doi.org/10.1016/j.enbuild.2017.07.027.

[45] Yu Z, Fung BCM, Haghighat F, Yoshino H, Morofsky E. A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy Build 2011;43:1409–17. https://doi.org/https://doi.org/10.1016/j.enbuild.2011.02.002.

[46] Cattarin G, Causone F, Kindinis A, Pagliano L. Outdoor test cells for building envelope experimental characterisation – A literature review. Renew Sustain Energy Rev 2016;54:606–25. https://doi.org/https://doi.org/10.1016/j.rser.2015.10.012.

[47] Deb C, Gelder L V, Spiekman M, Pandraud G, Jack R, Fitton R. Measuring the heat transfer coefficient (HTC) in buildings: A stakeholder's survey. Renew Sustain Energy Rev 2021;144:111008. https://doi.org/https://doi.org/10.1016/j.rser.2021.111008.

[48] Godfrey KR. Correlation methods. Automatica 1980;16:527–34. https://doi.org/https://doi.org/10.1016/0005-1098(80)90076-X.

[49] Madsen H, Holst J. Estimation of continuous-time models for the heat dynamics of a building. Energy Build 1995;22:67–79. https://doi.org/https://doi.org/10.1016/0378-7788(94)00904-X.

[50] Roels S, Deurinck M. The effect of a reflective underlay on the global thermal behaviour of pitched roofs. Build Environ 2011;46:134–43. https://doi.org/https://doi.org/10.1016/j.buildenv.2010.07.005.

[51] Saelens D, Roels S, Hens H. The inlet temperature as a boundary condition for multiple-skin facade modelling. Energy Build 2004;36:825–35. https://doi.org/https://doi.org/10.1016/j.enbuild.2004.01.005.

[52] Senave M, Roels S, Verbeke S, Lambie E, Saelens D. Sensitivity of characterizing the heat loss coefficient through on-board monitoring: A case study analysis. Energies 2019. https://doi.org/10.3390/en12173322.

[53] Senave M, Roels S, Reynders G, Verbeke S, Saelens D. Assessment of data analysis methods to identify the heat loss coefficient from on-board monitoring data. Energy Build 2020;209:109706. https://doi.org/https://doi.org/10.1016/j.enbuild.2019.109706.

[54] Claeskens G, Hjort NL. Model selection and model averaging. Cambridge Books 2008.

[55] Chatterjee S, Hadi AS. Regression analysis by example. John Wiley & Sons; 2015.

[56] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.

[57] Juhl R. ctsmr package-Continuous Time Stochastic Modelling in R 2013. http://ctsm.info/.

[58] Ross N. Generalized Additive Models in R n.d. https://noamross.github.io/gams-in-r-course/chapter1.

[59] Prautzsch H, Boehm W, Paluszny M. B-spline representation BT - Bézier and B-Spline Techniques. In:

Prautzsch H, Boehm W, Paluszny M, editors., Berlin, Heidelberg: Springer Berlin Heidelberg; 2002, p. 59–75. https://doi.org/10.1007/978-3-662-04919-8_5.

[60] Schoenberg IJ. Contributions to the problem of approximation of equidistant data by analytic functions. Part B. On the problem of osculatory interpolation. A second class of analytic approximation formulae. Q Appl Math 1946;4:112–41. https://doi.org/10.1090/qam/16705.

[61] Christensen O. B-splines BT - Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering. In: Christensen O, editor., Boston: Birkhäuser Boston; 2010, p. 203–14. https://doi.org/10.1007/978-0-8176-4980-7_10.

[62] R Core Team. R: A language and environment for statistical computing 2020. http://www.r-project.org/.

[63] Janssens A, Hens H. Effects of wind on the transmission heat loss in duo-pitched insulated roofs: A field study. Energy Build 2007;39:1047–54. https://doi.org/https://doi.org/10.1016/j.enbuild.2006.10.016.

[64] Dayton CM. Model comparisons using information measures. J Mod Appl Stat Methods 2003;2:2.

[65] Leuven, Flemish Brabant, Belgium — Sunrise, Sunset, and Daylength, January 2020 n.d. https://www.timeanddate.com/sun/belgium/leuven?month=1&year=2020(accessed on 10 July 2020).

[66] Perera DWU, Winkler D, Skeie N-O. Multi-floor building heating models in MATLAB and Modelica environments. Appl Energy 2016;171:46–57. https://doi.org/https://doi.org/10.1016/j.apenergy.2016.02.143.

[67] Bünning F, Sangi R, Müller D. A Modelica library for the agent-based control of building energy systems. Appl Energy 2017;193:52–9. https://doi.org/https://doi.org/10.1016/j.apenergy.2017.01.053.

[68] Wetter M, Zuo W, Nouidui TS, Pang X. Modelica Buildings library. J Build Perform Simul 2014;7:253–70. https://doi.org/10.1080/19401493.2013.765506.

[69] Sangi R, Baranski M, Oltmanns J, Streblow R, Müller D. Modeling and simulation of the heating circuit of a multi-functional building. Energy Build 2016;110:13–22. https://doi.org/https://doi.org/10.1016/j.enbuild.2015.10.027.

[70] Spiliotis K, Gonçalves JE, Saelens D, Baert K, Driesen J. Electrical system architectures for building-integrated photovoltaics: A comparative analysis using a modelling framework in Modelica. Appl Energy 2020;261:114247. https://doi.org/https://doi.org/10.1016/j.apenergy.2019.114247.

[71] Fuchs M, Teichmann J, Lauster M, Remmen P, Streblow R, Müller D. Workflow automation for combined modeling of buildings and district energy systems. Energy 2016;117:478–84. https://doi.org/https://doi.org/10.1016/j.energy.2016.04.023.

[72] Ali M, Vukovic V, Sahir MH, Fontanella G. Energy analysis of chilled water system configurations using simulation-based optimization. Energy Build 2013;59:111–22. https://doi.org/https://doi.org/10.1016/j.enbuild.2012.12.011.

[73] Jorissen F, Reynders G, Baetens R, Picard D, Saelens D, Helsen L. Implementation and verification of the IDEAS building energy simulation library. J Build Perform Simul 2018;11:669–88. https://doi.org/10.1080/19401493.2018.1428361.

[74] IDEAS. IDEAS weatherdata template 2018. https://github.com/open-ideas/IDEAS/blob/master/IDEAS/Resources/weatherdata/template.TMY (accessed on 1 July 2020).