Article

Development, Calibration and Validation of an Internal Air Temperature Model for a Naturally Ventilated Nearly Zero Energy Building: Comparison of Model Types and Calibration Methods

Michael D. Murphy [1,*] **, Paul D. O'Sullivan** [1]**, Guilherme Carrilho da Graça** [2] **and Adam O'Donovan** [1]

[1] Department of Process, Energy and Transport Engineering, Munster Technological University, Rossa Avenue, Bishopstown, T12 P928 Cork, Ireland; Paul.OSullivan@cit.ie (P.D.O.); Adam.ODonovan@cit.ie (A.O.)

[2] Dom Luiz Institute, Faculty of Sciences, University of Lisbon, Campo Grande, 1749-016 Lisbon, Portugal; gcg@fc.ul.pt

[*] Correspondence: michaeld.murphy@cit.ie

Abstract: In this study, a grey box (GB) model for simulating internal air temperatures in a naturally ventilated nearly zero energy building (nZEB) was developed and calibrated, using multiple data configurations for model parameter selection and an automatic calibration algorithm. The GB model was compared to a white box (WB) model for the same application using identical calibration and validation datasets. Calibrating the GB model using only one week of data produced very accurate results for the calibration periods but led to inconsistent and typically inaccurate results for the validation periods (root mean squared error (RMSE) in validation periods was 229% larger than the RMSE in calibration periods). Using three weeks of data from varying seasons for calibration reduced the model accuracy in the calibration period but substantially increased the model accuracy and generalisation abilities for the validation period, reducing the mean RMSE by over 160%. The use of one week of data increased the standard deviation in parameter selections by over 40% when compared with the three-week calibration datasets. Utilising data from multiple seasons for calibration purposes was found to substantially improve generalisation abilities. When compared to the WB model, the GB model produced slightly less accurate results (mean RMSE of the GB model was 1.5% higher). However, the authors found that employing GB modelling with an automatic model calibration technique reduced the human labour input for simulating internal air temperature of a naturally ventilated nZEB by approximately 90%, relative to WB modelling using a manually calibrated approach.

Keywords: model calibration; nearly zero energy building; grey box model; indoor air temperature

1. Introduction

Internal air temperature is a critical parameter for the simulation of the indoor environment in buildings. In order to simulate important metrics such as thermal comfort, occupant productivity and energy consumption, accurate internal air temperature models are required.

White box (WB), black box (BB) and grey box (GB) models have been employed in previous studies for internal air temperature prediction. WB models are physics-based and mechanistic in operation. WB models typically utilise a large number of building descriptive parameters such as material properties, spatial dimensions, fenestration orientations and mechanical/thermal system specifications. BB models are data driven and use regression or machine learning algorithms to map the relationship between system inputs and outputs using large amounts of empirical data without the requirement for static building descriptive parameters [1,2]. GB models, like WB models, are mechanistic in nature, however, the iterative physics-based operations are simplified and aggregated.

Therefore, they are less computationally expensive and require fewer building descriptive parameters [3,4]. The most commonly utilised GB models in thermal engineering applications are resistive-capacitive (RC) models [3–17] (Table 1). An RC model uses an electrical analogy to describe model structure and function, where resistors (R) and capacitors (C) are used to simulate the thermal energy flow and storage in the building. RC models typically simplify WB model parameters by lumping parameters together into single resistors or capacitors. The simplification of RC models is achieved through order-reduction [7,18] and the lumping of parameters [4,14,16].

Table 1. Review of GB models used for predicting internal air temperatures.

Study, Year [Reference]	Model(s)	Building	Data	Method	Duration * (Days)	Season(s)
Fraisse et al., 2002 [10]	1R2C-3R4C	NR	Empirical	NA	0 d/9 d **	Win
Nielsen et al., 2005 [11]	3R2C	NR	Synthetic	NA	0 d/14 d	Win, Sum
Bacher et al., 2011 [12]	1R1C-6R5C	TC	Empirical	Auto	6 d/0 d	Win, Spr
Sturzenegger et al., 2013 [8]	14R14C	NR	Synthetic	Mixed	6 d/6 d	All
Underwood et al., 2014 [4]	3R2C	TC	Synthetic	Auto	365 d/1 d	All
Berthou et al., 2014 [3]	4R2C-7R3C	NR	Synthetic	Auto	14 d/365 d ‡	All
Reynders et al., 2014 [7]	1R1C-8R5C	R	Synthetic †	Auto	135 d/14 d	Win, Spr, Sum
Pavlak et al., 2014 [6]	4R1C	NR	Synthetic †	Auto	21 d/7 d	Spr
Harb et al., 2016 [13]	1R1C-8R3C	R, NR	Empirical	Auto	26 d/168 d	Win, Spr
Oliveira Panao et al., 2016 [14]	5R1C	TC	Empirical	NA	0 d/6 d	Aut
Rodriguez Jara et al., 2016 [15]	2R1C	TC	Synthetic	NA	0 d/365 d	All
Andrade-Cabrera et al., 2017 [16]	9R6C	R	Synthetic	Auto	181 d/0 d	Aut, Win, Spr
Michalak et al., 2017 [17]	5R1C	R	Synthetic	NA	0 d/365 d	All
Cattarin et al., 2018 [5]	11R9C	TC	Empirical	NA	0 d/5 d	Aut
Oliveira Panao et al., 2019 [9]	2R2C	R	Empirical	Auto	NS	NS

Building: TC = test-cell, R = residential building, NR = non-residential. Method: Man = manual calibration, Auto = automated calibration, Mixed = combination of both, NA = method not applicable/no parameter calibration procedure explicitly stated. * Duration: specified in number of days used in calibrating a model and subsequently validating said model (x calibration/y validation), where d = days. NS = Not Specified, where the number of days for calibration/validation were not explicitly stated. The number of days was inferred if no value was explicitly stated. † Noise added to simulate empirical data. ‡ Only the 6R2C model was validated. ** Only the 3R4C model was validated. Season: Win = Winter, Sum = Summer, Spr = Spring, Aut = Autumn, All = All seasons.

Typically, WB models which utilise precise building parameters were found to be reasonably accurate when validated under experimental conditions in building test-cells [19,20]. However, WB models were found to be inaccurate when compared to measured data from real buildings [21]. Applying manual or automated calibration techniques to WB models has been shown to greatly improve accuracy for internal air temperature prediction in real buildings [22–24].

Model calibration often requires parameter tuning or selection. Parameter tuning can be performed using manual [25], automated [26] and semi-automated [27] approaches. The manual calibration approach uses human intuition combined with standardised calibration metrics [28] to minimize prediction error. This approach begins with an initial un-calibrated model populated with design stage building parameters. These parameters are then tuned via a series of revisions to the initial model based on the relative error reduction after each revision. [29,30]. The automated calibration approach uses mathematical or statistical methods for parameter tuning [26]. These methods typically utilise optimisation algorithms to reduce the error between model outputs and measured data [16]. Semi-automated or mixed approaches to calibration combine intuitive and mathematical methods [31]. Typically, an initial model is created manually, parameters are selected for tuning and boundary limits are manually selected for each parameter. The algorithm then selects the final values of these parameters through an iterative error minimization process [8,32]. The difference in time taken to manually or automatically calibrate models has been shown in some instances to be very substantial [26,32]. As manual calibration relies on user intuition, this process can be time intensive (e.g., 40 to 50 h), whereas automated calibration can take over 80% percent less time (e.g., two to seven hours).

In building simulation, WB models are the most widely used models in practice [33]. Typically, WB models are manually calibrated [22,25,34]. However, previous studies have also used semi-automated calibration procedures with WB models [32,35]. BB models are developed solely from empirical data where the internal architecture and coefficients (i.e., number for neurons in the hidden layer of a neural network and the synaptic weightings between neurons) are selected through an automatic procedural method based on globalised error reduction [36–41]. For GB models, several studies use manual [5,11,14] and semi-automated calibration approaches [12]. However, the most common approach is automated calibration [3,4,13,15–17]. The majority of studies using RC models for internal air temperature prediction have either relied on synthetic data generated by WB models or a higher order GB model for calibration and validation [3,6–8,11,15–17]. When empirical data are employed for such purposes, it is most commonly recorded data from unoccupied test-cells or rooms [5,12,14]. The number of RC models that have been calibrated and validated using recorded data from real occupied buildings is limited [10,13].

Nearly zero energy buildings (nZEBs) are now a legislative requirement for new-builds in many parts of the world. A large amount of literature has identified nZEBs and zero energy or net zero energy buildings as the target for buildings in the future [42–45]. Reaching nZEB standards requires improvements in fabric and energy performance when compared to existing buildings. Although energy performance certificates suggest that many nZEBs exist [46], the sample of well documented case study examples in published literature is limited [23,47–50]. Over 80% of the simulation studies that focused on modelling and performance of actual nZEBs, identified occupant behaviour (OB) as the primary inhibiting factor for accurate simulation and predictions of both energy and internal air temperatures [23,49,51–53]. This is due to the near stochastic nature of OB in certain conditions. OB has a strong influence on the internal air temperatures in nZEBs [23], as these buildings are highly insulated and can be thermally decoupled from external climatic conditions. As a result, the internal thermal gains from OB have a more dominant effect on internal air temperature. The observed efficacy of parameter tuning for RC models may be lower if measured empirical data from nZEBs are used (as opposed to synthetic data) due to the included noise from OB. The majority of published research has focused on RC model calibration of air-conditioned buildings [3,6–8,13,16,17]. As these buildings incorporated controlled air-conditioning systems for regulating the internal environment, the influence of OB on air temperature was attenuated, therefore, the model calibration and validation processes are less susceptible to the negative effects of OB noise. Pavlak et al., (2014) trained an RC model on three weeks of synthetic data. Varying levels of noise were added to the calibration dataset to simulate the uncertainty of recorded empirical data. The results of this study showed that least-squares based automatic model calibration error increased relative to noise level and noise type, with high levels of brown noise resulting in a substantial increase in model error. RC models calibrated and validated using automatic calibration algorithms have produced low levels of error with root mean squared error (RMSE), with average of RMSE values of 0.4 °C and a maximum RMSE value of 1.2 °C [6–8,13,16,17]. RC model predictions in naturally ventilated (NV) buildings are reported in the literature as being less accurate, with mean absolute errors of 1.0 °C to 1.1 °C and daily maximum errors of 1.8 °C [14]. Air temperature in NV buildings is much more sensitive to external climatic conditions (external air temperature and wind velocity) than mechanically air-conditioned buildings [54]. OB also has a strong influence on the operation of NV buildings as many NV systems either fully or partially rely on manual occupant-controlled openings.

Table 1 presents findings from a systematic mapping (from peer-reviewed literature) of published RC models that have been calibrated and/or validated for predicting internal air temperatures in different buildings. This table includes the types of RC models that have been used, the building type (i.e., residential, non-residential or test-cell), the types of calibration data used (i.e., synthetic from other software package or empirical from a real building), the calibration method (i.e., manual, automated or mixed methods), the duration

of calibration or validation in days, and the season(s) that were considered during calibration or validation. 40% of the studies identified in Table 1 used a combined calibration and validation approach. Of the studies that calibrate or validate GB models using measured empirical data, the data requirements for calibration purposes were typically between 6 days to 26 days and depended on the application [5,10,12,13] with some studies that used larger datasets broken into smaller periods [7]. Following a review of the RC model studies presented in Table 1, a number of gaps have been identified. Existing literature on GB model calibration has a limited number of validated examples of thermally decoupled environments such as nZEBs, and these examples are in test-cell environments [5]. GB models of nZEBs have yet to be calibrated in occupied conditions. The majority of GB models have used synthetic data for parameter tuning. There are few studies of GB model calibration in NV buildings [5,14]. While some examples have added noise to synthetic data [6,7], calibration or validation with measured empirical data from real buildings is very limited [10,13] and non-existent for NV nZEBs.

This paper presents the first example of GB model calibration for internal air temperature prediction in an occupied NV nZEB. The five objectives of this study are, (1) investigate GB model performance for predicting internal air temperatures in a thermally decoupled NV building with different sources of empirical calibration and validation datasets, (2) develop a GB model of a naturally ventilated nZEB, (3) apply an automatic calibration algorithm to select optimal GB model parameters using measured internal air temperature data, (4) compare the validated GB model to a calibrated WB model and, finally, (5) investigate the potential for practical implementation of GB models. In Section 3 we present the model theory and the calibration and validation approaches. In Section 4 we analyse the accuracy of the GB model when compared to the WB model and measured empirical data along with an analysis of the parameter selection when using different calibration and validation periods. In Section 5 we discuss the efficacy of different calibration and validation configurations as well as a practical comparison of WB and GB models. Section 6 presents the conclusions of this study.

2. Materials and Methods

2.1. Application

The empirical data used for this study were gathered from a test-bed building known as the zero2020 building. The zero2020 building is a 223 m^2 educational building that functions as a live test-bed known as the National Building Energy Retrofit Test-bed (NBERT) shown in Figure 1. The NBERT is used for research in thermal comfort, ventilative and passive cooling, energy systems, and micro-grid applications [55–59]. The NBERT uses a multi-configuration slotted louvre (MCSL) natural ventilation system for both comfort cooling and air quality. The retrofitted building also has a highly insulated external envelope that has fabric and fenestration u-values that are lower than the legislative requirements for nZEBs in Ireland [60]. The NBERT has one large open plan office area, two small cellular offices, one conference room and one seminar room, with a corridor connecting all occupied zones. For information relating to the thermo-physical properties and the building geometry see [56,61,62], for information regarding the natural ventilation system see [54] and for specification of the instrumentation used in recording the empirical data see [23]. The data used in this paper as well as building information regarding the NBERT test-bed can be found at messo.cit.ie/nbert.

Figure 1. NBERT in the zero2020 building: (**a**) image of exterior of retrofitted building; (**b**) image of open plan office with MCSL ventilation system; (**c**) image of high-performance retrofitted wall detail.

2.2. Grey Box Model Structure

Figure 2 displays the floor plan of the NBERT open plan office, while the bottom-left (b) of Figure 1 displays its interior, which has glazed facades to the west and south, and a roof light. This room was selected as the application in this study for the following reasons: (1) highly insulated external façade, therefore, the internal air temperature is to a higher degree, thermally decoupled from the external air temperature in comparison to a typical building. Occupancy related heat gains therefore, have a stronger influence on internal air temperature. (2) the space has a highly variable occupancy schedule [23], therefore profiling occupancy related gains is more difficult than a conventional open space office. (3) NBERT employs a natural ventilation system with automatic openings (which can be manually overridden by occupants), and manually operated openings, which are controlled exclusively by the occupants. Therefore, the air temperature control in the space relies on both external climatic conditions, (through solar irradiation and ventilation convection exchanges), and internal OB and interactions with the space. For these reasons, the space was deemed an interesting application for a GB model with automatic calibration and validation using measured empirical data.
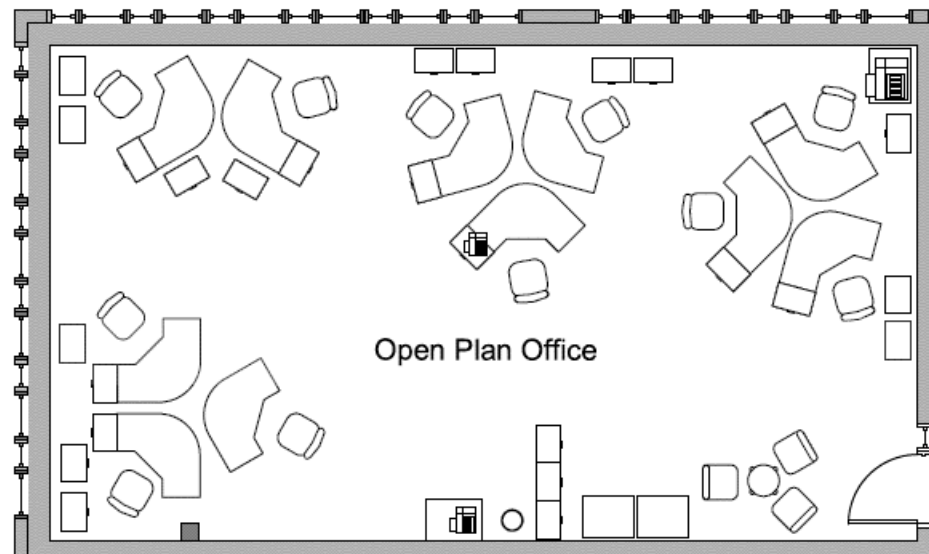
Figure 2. Layout of open plan office in NBERT.

The GB model used in this study was an RC model. Previous studies have found that single capacitor building RC models are too simplistic and do not adequately simulate internal air temperature for high thermal mass buildings [4,11,15,63], while cognate research that utilised multiple capacitors in RC models found that there was a reducing return on increased model accuracy as the number of capacitors was increased beyond two [12–14]. Therefore, a two-capacitor model (2C) was selected for this application. The internal air mass (air capacitor), and the internal material mass (material capacitor), were both modelled as independent capacitors. Three of the six zone boundary surfaces were external (two external walls and roof) while the other three were internal (two internal walls and floor). It was assumed no heat transfer would occur through the internal surfaces. The resistors of the three external surfaces were lumped into one resistor, which was the resistance of the external envelope (U_{Env}). This approach was adopted as the internal air temperature and external air temperature at either side of these three resistors were uniform. A second resistor was applied to the material capacitor representing the materials such as the walls and roof (which was the resistance of the convective heat transfer coefficient (h_{Cap})), therefore, the structure of the model used was 2R2C (for an equivalent RC diagram see [9]). This model structure was found to be very effective in cognate studies [9,12,13], and from the literature review in Section 1, the authors deemed this to be the simplest configuration that would be capable of simulating the dynamic internal air temperature for this application.

3. Theory

3.1. GB Model

The internal air temperature in the room ($T_{Ai(t)}$) at each time-step (t) is described in Equation (1):

$$T_{Ai(t)} = dT_{Ai} + T_{Ai(t-dt)} \tag{1}$$

The change in internal air temperature, T_{Ai}, due to the balance between energy flow in, ($\theta_{A,I}$) and out ($\theta_{A,O}$) of the air in the room is defined in Equation (2):

$$\frac{dT_{Ai}}{dt} = \left(\frac{d\theta_{A,I}}{dt} - \frac{d\theta_{A,O}}{dt} \right) (m_{Ai}C_A)^{-1} \tag{2}$$

where m_{Ai} is the mass of the air in the room and C_A is the specific heat capacity of air.

The sensible energy input to the room is described in Equation (3):

$$\frac{d\theta_{A,I}}{dt} = \beta \left(\frac{d\theta_{Sol}}{dt}\gamma_{Sol} + \frac{d\theta_{Occu}}{dt}\gamma_{IO} + \frac{d\theta_{Int}}{dt}\gamma_{IO} \right) \tag{3}$$

where, θ_{Sol}, $\theta_{Occu,}$ and θ_{Int} are the solar, occupant driven, and internal energy gains added to the room at each time-step. β represents the percentage of thermal gains that enter the air. It was assumed a percentage of all gains would be lost due to latent energy [64]. γ describes the fraction of energy that is convective or radiative. Two independent convective-radiative ratios were employed in this study, one for internal gains and occupant gains (γ_{IO}) and another for solar gains (γ_{Sol}). These ratios represented the percentage of convective gains that enter the air, where the remainder was assumed as radiative and was absorbed by the surface of the material capacitor (i.e., the walls, floor and roof).

Equation (4) describes how the solar gains (θ_{Sol}) were broken into gains entering from the sky-light window on the horizontal, and gains entering from all vertical windows. The g-values for the skylight (g_h) and vertical windows (g_v) were independent as the glazing details varied. The solar radiation incident on the horizontal (I_h) and vertical (I_v) were derived from global horizontal irradiance data and converted to vertical values for the vertical windows using the Perez model [65] swhich was employed from TRNSYS 17 [66] using the integrated weather data and radiation processor (Type 99):

$$\theta_{Sol} = (I_h g_h + I_v g_v) \tag{4}$$

Equation (5) describes totalised outward sensible energy flows from the internal air ($\theta_{A,O}$):

$$\frac{d\theta_{A,O}}{dt} = \left(\frac{d\theta_{Inf}}{dt} + \frac{d\theta_{Vent}}{dt} + \frac{d\theta_{Cond}}{dt} + \frac{d\theta_{Mat}}{dt} \right) \tag{5}$$

where, θ_{Inf}, θ_{Vent}, θ_{Cond} and θ_{Mat} are energy flows due to infiltration, ventilation, conductive losses and material capacitor energy losses (energy flowing from the air to the material capacitor), respectively.

Equation (6) describes the energy losses due to conduction (θ_{Cond}):

$$\frac{d\theta_{Cond}}{dt} = (T_{Ai} - T_{Ae})(U_{Env}A_{Env}) \tag{6}$$

where, T_{Ai} is the internal air temperature of the room, T_{Ae} is the external air temperature, U_{Env} is the u-value of the external envelope and A_{Env} is the area of the external envelope.

Equation (7), where, $\theta_{Mat,I}$ and $\theta_{Mat,O}$ are the energy flows in and out of the material capacitor, m_{Mat} is the mass of the material capacitor, and C_{Mat} is the capacitance of the materials in the room, describes the change in temperature due to the energy flows entering and exiting the material capacitor (walls, floor and roof):

$$\frac{dT_{Mat}}{dt} = \left(\frac{d\theta_{Mat,I}}{dt} - \frac{d\theta_{Mat,O}}{dt} \right) (m_{Mat}C_{Mat})^{-1} \tag{7}$$

The mass of the material capacitor is described in Equation (8) and is a function of its depth (d_{Mat}), area (A_{Mat}) and density (ρ_{Mat}). For this study, the material capacitance for the room was assumed to be concrete from the walls, floor and roof in the room. Therefore, all thermo-physical parameters that were defined for the material capacitor were related to concrete blocks with a maximum thickness of 0.2 m and a maximum utilisable depth of 0.1 m, based on maximum depths for calculations proposed by ISO 13790 [67]. Initial values for the density of the material capacitor were taken from CIBSE Guide A [68]:

$$m_{Mat} = d_{Cap}A_{Mat}\rho_{Mat} \tag{8}$$

Equation (9) describes the energy flowing in to the material capacitor ($\theta_{Mat,I}$):

$$\frac{d\theta_{Mat,I}}{dt} = \left(\frac{d\theta_{Sol}}{dt}(1 - \gamma_{Sol}) + \frac{d\theta_{Occu}}{dt}(1 - \gamma_{IO}) + \frac{d\theta_{Int}}{dt}(1 - \gamma_{IO}) \right) \tag{9}$$

where, θ_{Sol}, θ_{Occu}, and θ_{Int} are the solar, occupant driven and internal energy added to the material capacitor at each time-step. $(1 - \gamma_{Sol})$ and $(1 - \gamma_{IO})$ describe the radiative gains entering the material capacitor.

Equation (10) describes the energy flowing out the material capacitor ($\theta_{Mat,O}$):

$$\frac{d\theta_{Mat,O}}{dt} = (T_{Ai} - T_{Mat})(h_{Mat}A_{Matex})^{-1} \tag{10}$$

where, T_{Ai} is the temperature of the air in the room, T_{Mat} is the temperature of the material, h_{Mat} is the convective heat transfer co-efficient for the material capacitor, and A_{Matex} is heat exchange area between the material capacitor and the air capacitor.

Equation (11) describes the heat exchange due to infiltration through the building external envelope.

$$\frac{d\theta_{Inf}}{dt} = \frac{dQ_{Ainf}}{dt}(T_{Ai} - T_{Ae})(C_A\rho_A)^{-1} \tag{11}$$

where T_{Ai} is the temperature of the air in the room, T_{Ae} is the external air temperature, C_A is the heat capacity of the air, ρ_A is density of the air in the open plan office, and Q_{Ainf} is the volume of air due to infiltration.

Equation (12) describes the heat loss due to ventilation from the building's natural ventilation system:

$$\frac{d\theta_{Vent}}{dt} = \frac{dQ_{Avent}}{dt}(T_{Ai} - T_{Ae})(C_A\rho_A)^{-1} \tag{12}$$

where Q_{Avent} is the volume of air flowing through the room due to natural ventilation.

The volumetric flowrate of air passing through the room is based on whether the wind driven flow (Q_W) is greater than the buoyancy driven flow (Q_B) or vice versa, as indicated in Equation (13):

$$\frac{dQ_{Avent}}{dt} = \begin{cases} \frac{dQ_B}{dt}, & dQ_W < dQ_B \\ \frac{dQ_W}{dt}, & dQ_W \geq dQ_B \end{cases} \tag{13}$$

Equations (14) and (15) describe the volumetric flowrate for wind driven (Q_W) and buoyancy driven (Q_B) flow according to Warren et al. [69]:

$$\frac{dQ_W}{dt} = (F_r)(\alpha A_{win})(v_w) \tag{14}$$

$$\frac{dQ_B}{dt} = \alpha \left(\frac{A_{win}C_d}{3} \right) \sqrt{\frac{\left| T_{Ai(t)} - T_{Ae(t)} \right| Hg}{(T_{Ai(t)} + T_{Ae(t)})0.5}} \tag{15}$$

where F_R describes the reference flow number, α describes the opening position of openings (the opening scaling factor), A_{win} describes the area of openings in the room, v_w describes the local wind speed, H is the height of the opening and C_d is the discharge co-efficient for all openings.

3.2. Calibration, Testing and Validation Data

The data utilised for calibration, testing and validation in this study were recorded in an occupied open plan office in NBERT (see Section 2.1). In previous work by O'Donovan et al. a WB model was calibrated and validated using the same dataset [23]. Table 2 and Figure 3 display the periods where data were used for calibrating and validating both the GB and the WB model. The WB model was calibrated using data from week 1 to week 3 and validated using data from week 4 to week 6. For more detailed information on these data please see the supplementary data section of O' Donovan et al. [23].

Table 2. Periods used to calibrate or validate the GB model or WB model.

Week Name	(Day)–Date from	(Day)–Date to
Week1	(Wed)–24 December 2014	(Tue)–30 December 2014
Week2	(Mon)–29 June 2015	(Sun)–5 July 2015
Week3	(Mon)–19 October 2015	(Sun)–25 October 2015
Week4	(Mon)–26 October 2015	(Sun)–1 November 2015
Week5	(Mon)–4 May 2015	(Sun)–10 May 2015
Week6	(Mon)–11 May 2015	(Sun)–17 May 2015
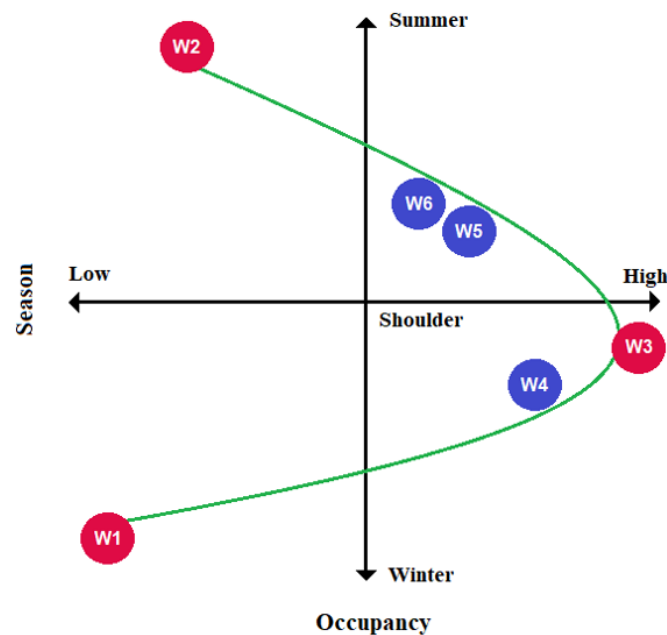
Figure 3. Plot of seasons and occupancy levels considered for calibration in red and validation in blue (Points are illustrative and are representative of typical patterns of occupancy in the case study building).

3.3. Calibration and Validation Approaches and Metrics

Calibration and validation using both WB and GB models were compared. The baseline model geometry for the WB model was developed using the TRNSYS 3D Google Sketchup Plugin [70] and imported into TRNSYS 17 [66]. The WB model was calibrated using a manual approach, where, the baseline model (describing the building fabric and solar heat gains) was calibrated sequentially, first using unoccupied data and later occupied data. Week 1 was used to tune parameters related to the unoccupied condition of the building (i.e., infiltration rate, external shading factor, thermal bridges). Week 2 was used to tune parameters related to the occupied condition of the building (i.e., lighting, appliances, occupant gains, window positions). Week 3 was used to test the WB model performance, while, data from week 4 to week 6 were used to validate the WB model (i.e., no parameters were tuned). Changes from the baseline to the calibrated model were made through a series of iterations that were evidence-based. All changes to the baseline were founded on building information or detailed knowledge of the building performance. Each iteration was, therefore, a manual adjustment of the previous model and did not rely on automation. To assess whether the model was satisfactorily calibrated, standardised calibration metrics with reference to measurement and verification or calibration standards were applied. More information on this approach can be found in O' Donovan et al. [23]. In this study, three GB model validation/calibration approaches were investigated:

(1)   The GB model was calibrated on an individual week, then validated on the corresponding five weeks. This was repeated for all calibration and validation periods.

The purpose of this analysis was to examine the performance of the model when calibrated on a single week with a specific occupancy level during a specific season, then validated on five individual weeks with different occupancy levels over varying seasons. The comparison of calibration and validation metrics was used as an indicator of robustness.

(2)   The GB model was calibrated on three weeks, then validated on the other three weeks. Multiple simulations were carried out where the three calibration weeks and the three corresponding validation weeks contained periods of varying occupancy and different seasons. The efficacy of this calibration and validation method was examined and then compared to the previous method. The model parameters selected by the calibration algorithm for both methods were compared.

(3)   A comparison of both GB and WB models was performed where both models were calibrated and validated using identical data. The GB model was simultaneously calibrated using data from weeks 1, 2 and 3. The WB model was incrementally calibrated for weeks 1 and 2 and tested using data from week 3. Both models were validated using data from week 4 to 6.

To compare the accuracy of each model, the RMSE for a given period, (p), was calculated for each week using Equation (16):

$$\text{RMSE}_\text{P} = \sqrt{\frac{\sum \left(T_{Ai,S} - T_{Ai,\,M}\right)^2}{N}} \tag{16}$$

where, $T_{Ai,S}$ is the simulated internal air temperature, $T_{Ai,M}$ is the measured internal air temperature and N is the number of measurements during period p.

Previous work by O'Donovan et al. noted the need to include a correlation metric when comparing air temperature predictions with empirical data as the range of temperatures in nZEBs can be quite narrow and may not accurately represent the error in a model [23]. Equation (17) shows the calculation of the Pearson correlation co-efficient (r) [71]:

$$r = \frac{\sum \left(T_{Ai,S} - \overline{T}_{Ai,S}\right)\left(T_{Ai,\,M} - \overline{T}_{Ai,\,M}\right)}{\sqrt{\sum \left(T_{Ai,S} - \overline{T}_{Ai,S}\right)^2}\sqrt{\sum \left(T_{Ai,M} - \overline{T}_{Ai,\,M}\right)^2}} \tag{17}$$

where, $T_{Ai,S}$ and $T_{Ai,\,M}$ are values of simulated and measured air temperatures for each instance, and, $\overline{T}_{Ai,S}$ and $\overline{T}_{Ai,\,M}$ are the mean values for the simulated and measured datasets, respectively.

The Pearson correlation co-efficient was used to measure the strength of correlation between the models' outputs and the measured empirical data. This correlation is a direct indication on the models' abilities to profile the transient behaviour of the internal air temperature, and therefore, it provides a good indication on whether the model is accurately capturing the thermodynamic characteristics of the space. When the Pearson correlation co-efficient is greater than 0.5, it can be suggested that the model is representing the dynamic behavior of the building [72].

3.4. GB Model Calibration

The Levenberg-Marquardt algorithm [73] was used in the automatic calibration process. This algorithm was chosen for its computational efficiency and ability to avoid local minima. The Levenberg-Marquardt method has proven to be successful in cognate modelling applications [40,74,75]. The values of the 11 model parameters and their upper and lower limits are shown in Table 3. The algorithm selected the optimal configuration of the 11 parameters through an iterative process. The cost function employed was the sum of squared errors (SSE) (see Equation (18)).

$$\text{SSE} = \sum_{i=1}^{N}(T_{Ai,Mi} - T_{Ai,Si})^2 \tag{18}$$

Table 3. Maximum and minimum limits and initial values for auto-calibration parameters.

Name	Symbol	Eqn.	Units	Minimum	Maximum	Initial Value
Solar convective-radiative ratio	γ_{Sol}	(3), (9)	%	0.1	0.8	0.3
Internal convective radiative ratio	γ_{IO}	(3), (9)	%	0.3	0.7	0.5
Latent loss percentage	β	(3)	%	0.1	0.5	0.3
Horizontal solar transmittance	g_h	(4)	%	0.04	0.3	0.2
Vertical solar transmittance	g_v	(4)	%	0.1	0.5	0.35
External envelope heat loss rate	U_{env}	(6)	W/(m^2 K)	0.17	0.5	0.33
Material capacitor depth	d_{Mat}	(8)	m	0.01	0.1	0.1
Maximum material capacitor area	A_{Mat}	(8)	m^2	150	300	250
Material capacitor exchange area	A_{Matex}	(10)	m^2	100	125	115
Material capacitor convective heat transfer coefficient	h_{Mat}	(10)	W/(m^2 K)	2	13	6
Infiltration rate	Q_{Ainf}	(11)	m^3/h	30	300	150
Opening scale factor	α	(14), (15)	%	0.2	1	0.5

γ_{Sol} is the solar convective-radiative ratio, γ_{IO} is the internal convective-radiative ratio, β is the latent loss percentage, g_h is the horizontal solar transmittance, g_v is the vertical solar transmittance, U_{env} is the external envelope heat loss rate, d_{Mat} is the material capacitor depth, A_{Mat} is the maximum area of the material capacitor, A_{Matex} is the material capacitor exchange area, h_{Mat} is the material capacitor convective heat transfer coefficient, Q_{Ainf} is the infiltration rate, and α is the opening scale factor.

This cost function was selected as it penalizes large residuals, where the error was the difference between simulated air temperature ($T_{Ai,S}$) and the measured air temperature ($T_{Ai,M}$). The value of N is dependent on the length of the calibration dataset. The algorithm stopping criteria were set to a convergence value (1×10^{-3}) and a maximum limit of 1000 iterations.

While many of the standard building parameters (such as the density of concrete blocks) were assumed to be known, the building specific parameters were assumed to be unknown but could be estimated within a set range [76]. The 11 building specific parameters shown in Table 3 could be altered during the auto-calibration process. As these parameters were unknown, the initial parameter values were estimated and the upper and lower limits were either set by the physical constraints of the building or maximum plausible boundary limits determined by the authors.

4. Results and Discussion

4.1. Calibration and Validation of GB Model

In this section, we compared the performance of the GB model for two different calibration and validation datasets. The first dataset contained one week for calibration and five weeks for validation purposes (C1V5), while the second dataset contained three weeks for calibration and three weeks for validation purposes (C3V3). Each dataset was partitioned and folded into multiple calibration and validation subsets (Figure 4) to allow multiple calibration and validation arrangements [77]. The purpose of comparing the calibration and validation of the GB model for these two datasets was to determine the performance gap when the model was calibrated using just one week in a particular season with a particular occupancy level versus being calibrated using multiple weeks over varying seasons and occupancy levels.

4.1.1. GB Model Performance (C1V5)

Table 4 displays the calibration and validation results for the GB model for C1V5. The dataset contained six individual weeks of data. These weeks were recorded over varying seasons and contained multiple occupancy levels (see Section 3.2). The GB model was calibrated using one week and then validated on the other five weeks, this was repeated for each week. For example, when calibrating using week 2, the validation period consisted of weeks 1, 3, 4, 5 and 6 (C2_V13456) (see Figure 4). We can see from the results that the GB model correlated very strongly and produced a very low error for the weeks on which it was calibrated.
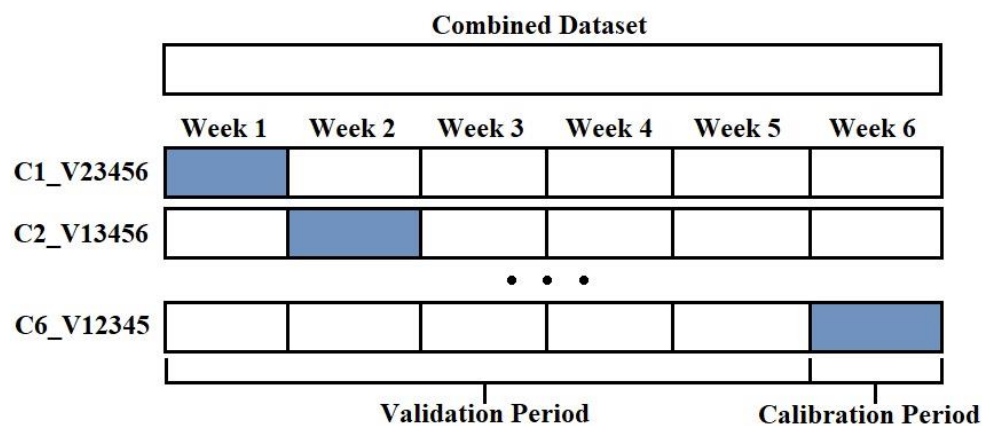
Figure 4. Folded data configurations for validation and calibration periods (C1V5).

Table 4. Accuracy of models calibrated on one week of data. The values in bold indicate the accuracy of the training week. The corresponding line plots for these bold values are in Figure 5.

Name	Metric	Week1	Week2	Week3	Week4	Week5	Week6
C1_V23456	RMSE	**0.25**	0.53	1.24	0.64	2.14	0.87
	r	**0.99**	0.76	0.47	0.92	0.28	0.85
C2_V13456	RMSE	1.94	**0.21**	1.80	1.83	2.11	1.04
	r	0.99	**0.95**	0.80	0.31	0.35	0.91
C3_V12456	RMSE	1.02	1.16	**0.50**	0.60	0.81	0.43
	r	0.99	0.79	**0.86**	0.84	0.60	0.90
C4_V12356	RMSE	0.57	1.46	1.52	**0.22**	1.77	2.39
	r	0.99	0.85	0.45	**0.96**	0.33	0.83
C5_V12346	RMSE	0.35	1.44	0.85	0.38	**0.58**	0.59
	r	0.99	0.76	0.69	0.90	**0.66**	0.86
C6_V12345	RMSE	2.43	0.76	0.79	0.93	0.75	**0.31**
	r	0.99	0.91	0.86	0.80	0.76	**0.95**

Figure 5 shows the calibrated models' performance for each respective calibration period. The models' very tight fit to the empirical data in each calibration period demonstrates the algorithm's ability to finely tune the model parameters to produce the minimum error. However, the correlation substantially decreases and the error substantially increases for the validation periods. For the six simulations of C1_V23456 through to C6_V12345, the mean RMSE was 334%, 739%, 62%, 593%, 25% and 265% higher in the validation period compared to the calibration period, respectively. The mean increase in RMSE from the calibration period to the validation periods on average was 229%.

Figure 6 displays the set of parameters selected by the calibration algorithm for each calibration period (the parameters are displayed as a percentage between the upper and lower limits). We can see there is substantial variation between periods for the majority of parameters. Depending on the week selected for calibration, the tuned parameter values varied substantially. The mean of the standard deviations for the 11 parameters was 19%.

4.1.2. GB Model Performance (C3V3)

Table 5 displays the calibration and validation results for the GB model for C3V3. The GB model was calibrated using three weeks and then validated on the other three weeks. The composition of the three calibration weeks were selected and folded to ensure the model was calibrated using data containing varying occupancy levels during different seasons. The last simulation employs all six weeks for calibration purposes (C123456).

For the four simulations of C123_V456 through to C256_V134, the mean RMSE was 109%, 53%, 65%, and 29% higher in the validation period compared to the calibration period, respectively. The mean increase in RMSE from the calibration periods to the validation periods was 62%. Figure 7 illustrates the final parameter selections that the calibration

algorithm made for each three-week combination of calibration and validation datasets. As can be seen, the parameter values selected by the algorithm can vary depending on the three weeks used for the calibration dataset. However, the variance was substantially lower in comparison to the C1V5 results (Figure 6). The mean of the standard deviations for the 11 parameters was 13%.
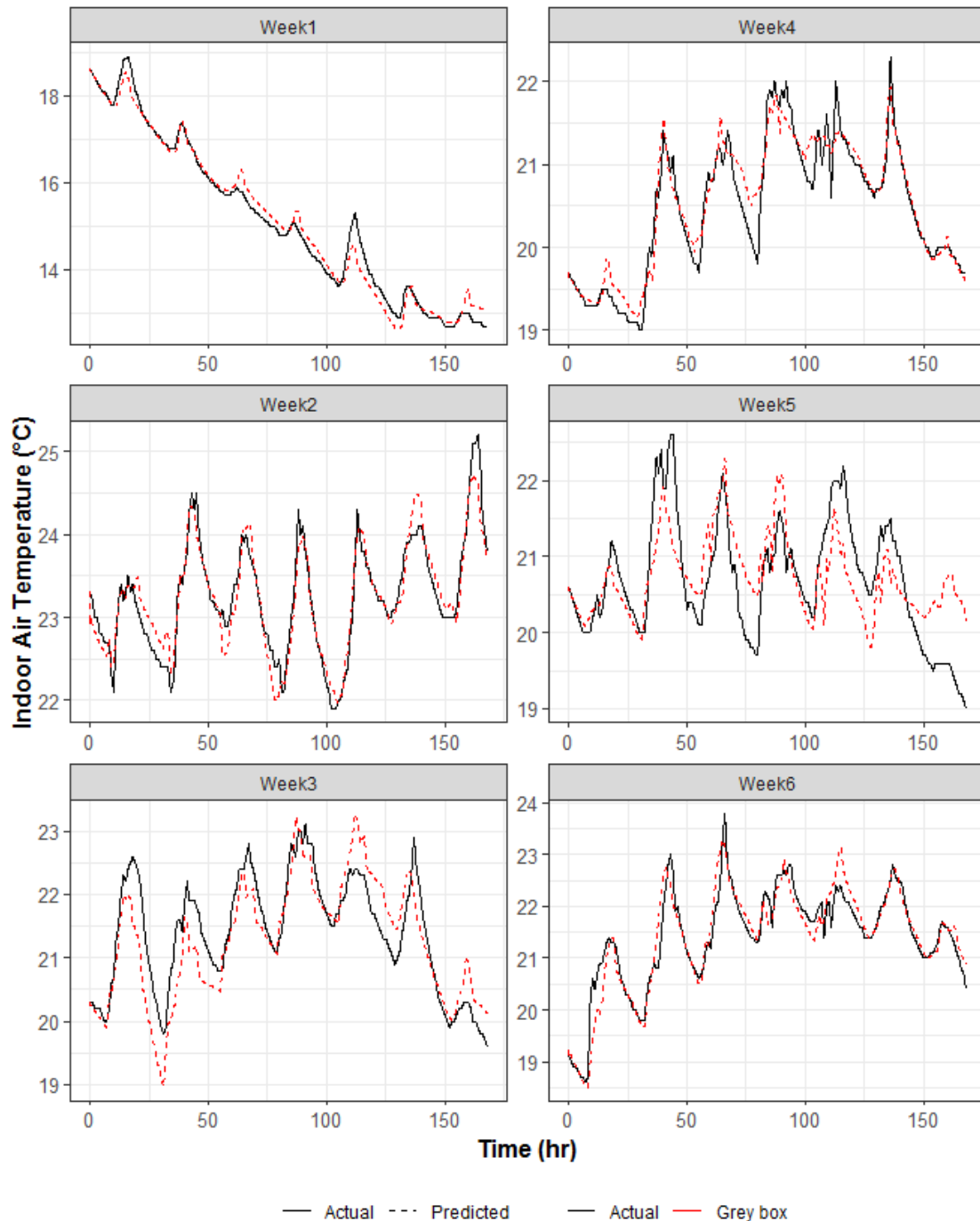
Figure 5. Line graphs of GB model performance showing empirical and predicted air temperatures using C1V5 calibration approach for calibration weeks only.
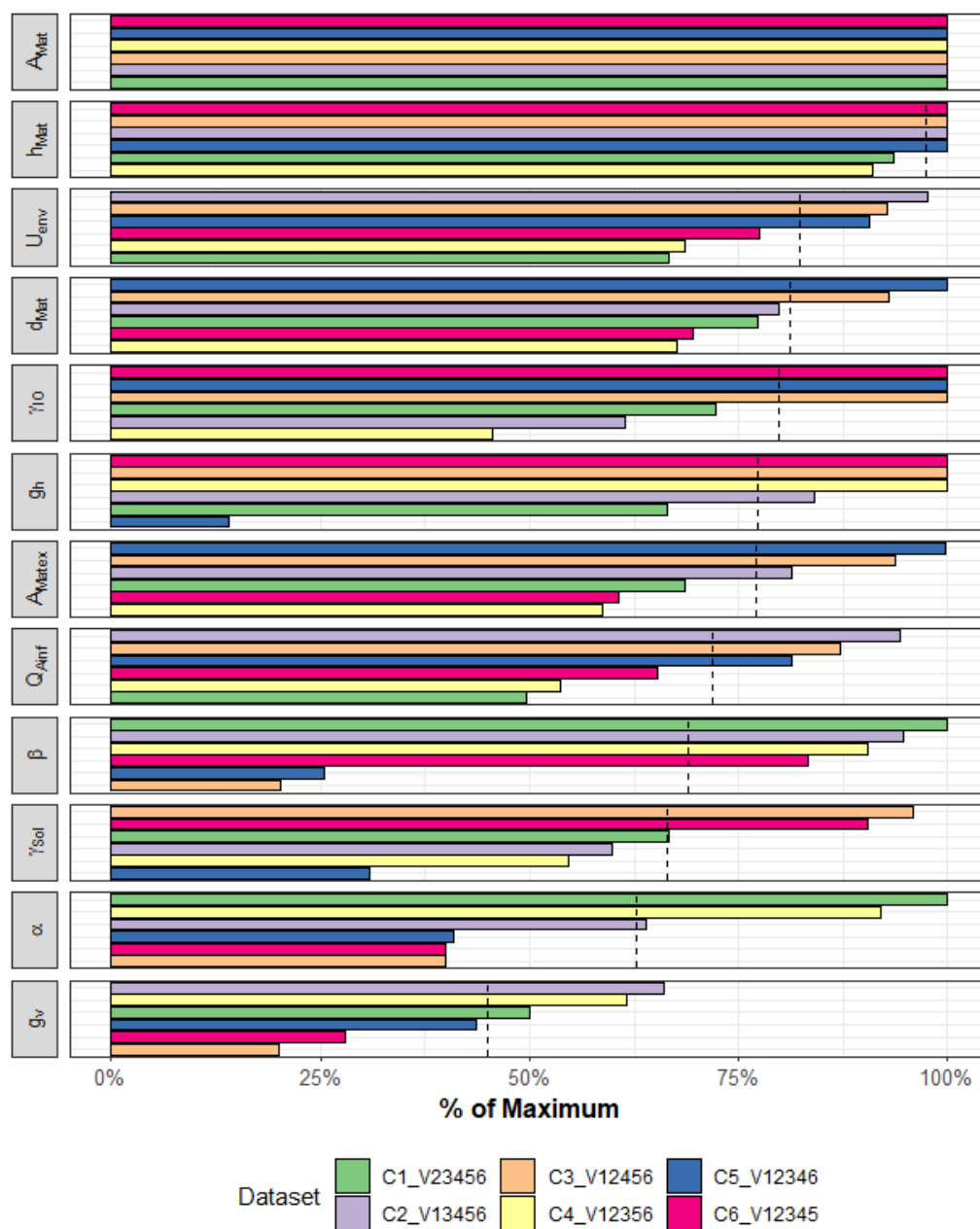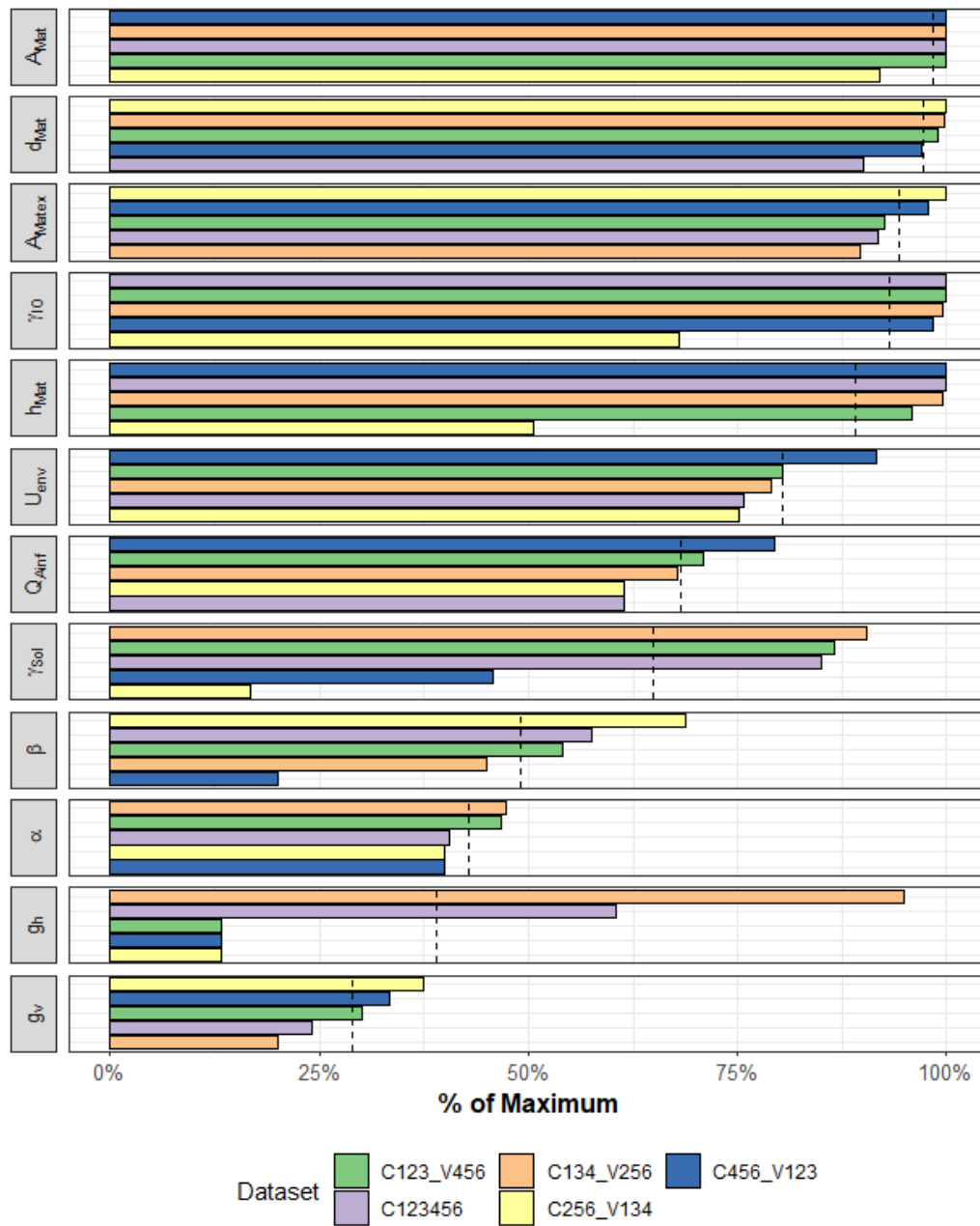
Figure 6. Parameter selecting based on GB models that were calibrated using one week (C1V5). Coloured bars indicate parameters selected for each calibration set, dashed line indicates mean value.

Table 5. Accuracy of models calibrated using three weeks of data. The values in bold indicate the accuracy of the calibration weeks.

Name	Metric	Week1	Week2	Week3	Week4	Week5	Week6
C123	RMSE	**0.27**	**0.46**	**0.59**	0.67	1.18	0.91
V456	r	**0.99**	**0.82**	**0.81**	0.83	0.49	0.90
C456	RMSE	0.29	1.10	0.78	**0.37**	**0.58**	**0.46**
V123	r	0.99	0.78	0.73	**0.90**	**0.65**	**0.89**
C134	RMSE	**0.27**	0.89	**0.58**	**0.47**	0.81	0.49
V256	r	**0.99**	0.85	**0.77**	**0.88**	0.55	0.88
C256	RMSE	0.84	**0.49**	0.67	0.62	**0.71**	**0.46**
V134	r	0.98	**0.83**	0.72	0.84	**0.49**	**0.88**
C123456	RMSE	**0.27**	**0.82**	**0.64**	**0.45**	**0.76**	**0.52**
	r	**0.99**	**0.84**	**0.73**	**0.90**	**0.58**	**0.88**

Figure 7. Parameter selecting based on GB models that were calibrated using three-week datasets (C3V3). Coloured bars indicate parameters selected for each calibration set, dashed line indicates mean value. Calibration is also indicated (C123456).

4.1.3. Comparison of GB Model Performance for Both Calibration Methods

Figure 8 displays the performance of the GB model when different sources of data were used for calibration and validation. The GB model calibrated on three weeks of data (C3V3), produced results with a relatively consistent level of accuracy, with mean RMSE values of between 0.58 °C and 0.67 °C, maximum RMSE values of between 0.82 °C and 1.18 °C, Pearson correlation coefficients ranged from 0.49 to 0.99, for calibration and validation periods, respectively.
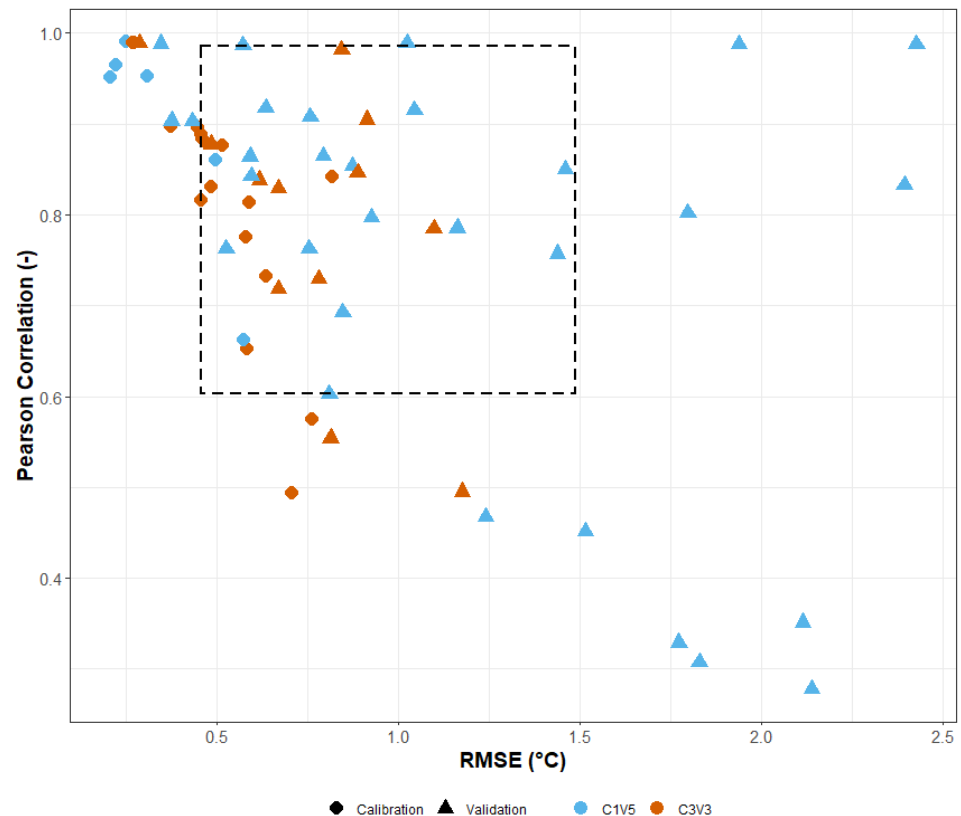
Figure 8. Scatterplots of Pearson correlations with respect to RMSE. Data grouped according to calibration approaches, i.e., for one week (C1V5) and three week (C3V3), using colour scheme and according to stage (blue for one week and orange for three weeks) and using data point shape for calibration or validation (circle for calibration period and triangle for validation period). Dashed box indicates the accuracy of WB model (see Section 4.2 for comparison with GB model).

However, when the GB model was calibrated using one week of data (C1V5), the accuracy of the model was far less consistent in comparison to the C3V3 results, with mean RMSE values of between 0.21 °C and 1.52 °C, and maximum RMSE values of between 1.46 °C and 2.43 °C. Pearson correlation coefficients ranged from 0.28 to 0.99, for calibration and validation periods, respectively. From Figure 8 we can see how the calibration periods for C1V5 (see Figure 5 for line graphs), are highly accurate. However, the corresponding validation periods are typically far less accurate, with lower Pearson correlation and higher RMSE. In contrast, the calibration periods for C3V3 (see Figure 9 for line graph of C123V456), were not as accurate in comparison to C1V5, but the validation periods were far more accurate. We can see the number of weeks used for calibration of parameters in GB models effects the accuracy and repeatability of results (for a breakdown of individual weeks see Figure 10). The inconsistency in the results for C1V5 is likely a result of over-fitting for the conditions present in the individual calibration week. This is reflected in the mean RMSE values for each calibration week or series of weeks and in the difference in standard deviation between C1V5 and C3V3. The use of three weeks of data (C3V3) as opposed to one week of data (C1V5) for calibration, reduces the average RMSE in the validation period by over 160%. The use one week of data (C1V5) increases the standard deviation in final parameter selections by 46%.
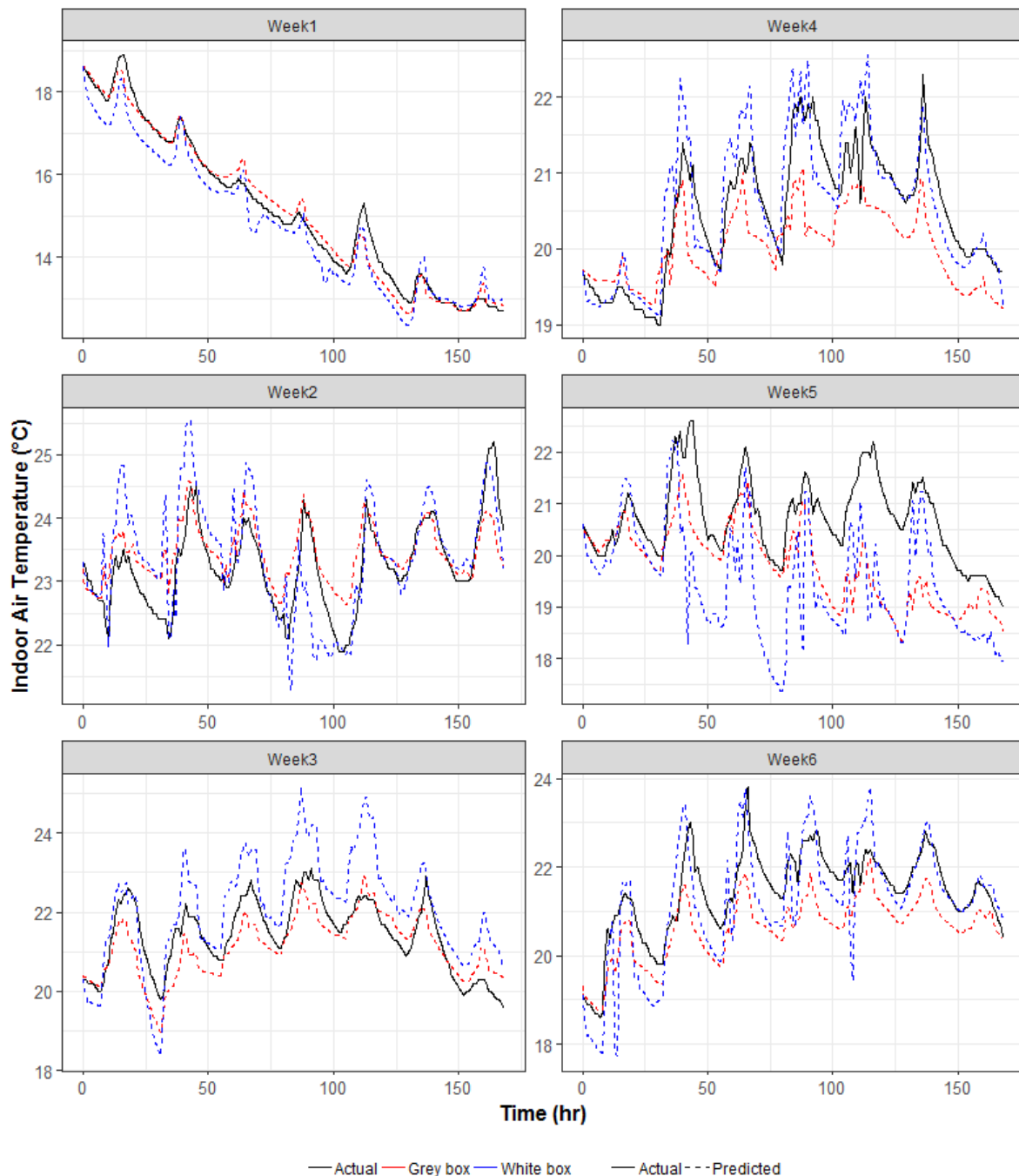
Figure 9. Line graphs of internal air temperature for calibration on weeks 1 to 3 (on the left) and validation on weeks 4 to 6 (on the right), comparing empirical data, GB model predictions (C123_V456) and the WB model predictions.

4.2. Comparison of WB and GB Models

Figure 9 displays the difference observed between recorded air temperature data, GB model predictions and WB model predictions in the NBERT open plan office for both calibration and validation periods. Table 6 displays the Pearson coefficient and the RMSE for both models for the calibration period (weeks 1–3) and the validation period (weeks 4–6). During the calibration period the GB model produced a higher Pearson coefficient for week 2 (0.82), while the WB performed better in week 3 (0.86). In week 1 (non-occupied) both models produced an equivalent Pearson coefficient (0.99). The RMSE in the calibration period for GB was considerably lower in all three weeks (0.27, 0.46, and 0.59). During the validation period the GB model produced a higher Pearson coefficient for week 3 (0.9),

while the WB model performed better in weeks 2 and 3 (0.88 and 0.6). The RMSE values of the WB model over the validation period were lower than the GB model for weeks 1 and 3 (0.46 °C and 0.77 °C), while the GB model produced a lower RMSE in week 2 (1.18 °C).
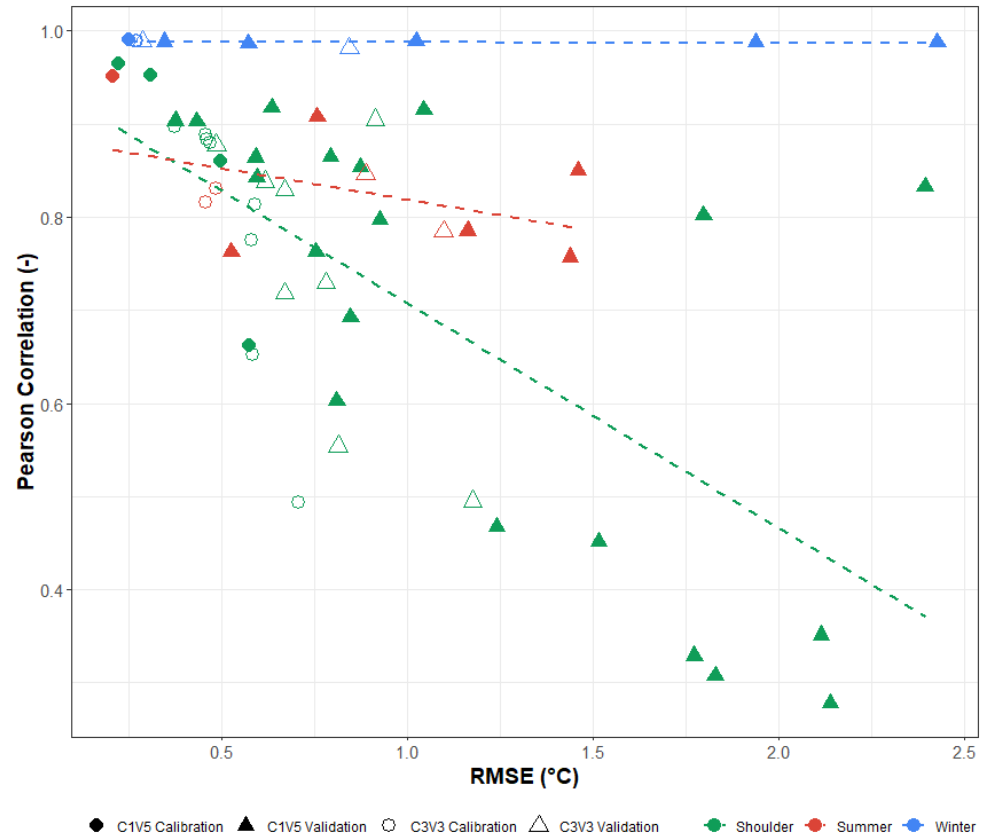
Figure 10. Scatterplots of Pearson correlation co-efficient and RMSE for both GB calibration methods (C3V3 and C1V5). Points indicate the accuracy for a single specific week for the GB model calibrated using different data sources, where, the colour indicates the season and the shape indicates the calibration method (C3V3 or C1V5) and whether it was a calibration or validation week. Dashed line indicates the line of best fit according to season.

The mean RMSE of the GB model during the validation period increased by 109% when compared to mean RMSE of the calibration period. However, the mean RMSE of the WB model only increased by 28% between the calibration and validation period. The drop in GB model accuracy between the calibration and validation periods may be due to the overfitting of the tuning parameters to the calibration data as discussed in Section 3.2 (for the selected parameters please see Figure 7, C123_V456).

Table 6. Calculated error for both GB and WB models during calibration (weeks 1–3) and validation (weeks 4–6).

Week	*r*		RMSE (°C)	
	GB	**WB**	**GB**	**WB**
1	0.99	0.99	0.27	0.50
2	0.82	0.74	0.46	0.62
3	0.81	0.86	0.59	1.00
4	0.83	0.88	0.67	0.46
5	0.49	0.60	1.18	1.49
6	0.90	0.86	0.91	0.77

When directly comparing the performance of the models, we can see the GB model substantially outperformed the WB model over the calibration period. The mean RMSE of the GB model (0.44 °C) was 38% less than the WB model (0.71 °C) over the same period. However, during the validation period the mean RMSE of the GB model (0.92 °C) was slightly higher than the WB model (0.91 °C) and the correlation between the measured empirical data and model predictions were stronger for the WB model during the validation period. The GB model was far more accurate during the calibration period, but this can be attributed to the automatic calibration method, which employed an optimisation algorithm, which was able to converge on a solution. The WB model relied on a manual calibration method, which used a piecemeal evidence-based approach for a larger number of parameters. It must also be noted that WB model parameters were calibrated using data from weeks 1 and 2, while week 3 was used for testing purposes (as described in Section 3.3). While the WB model is only slightly more accurate than the GB model for the validation period, this comparison suggests the WB modelling technique produced better generalisation abilities. Figure 8 provides a breakdown of the accuracy of the individual weeks used for calibration and validation used in C1V5 and C3V3 for the GB model. The accuracy of the calibration and validation periods for the WB model are represented by the dashed box. From this, we can see that the GB model out-performs the WB model for the calibration periods, especially for the C1V5 configuration. For the validation periods, the GB model is more likely to produce less accurate results than the WB model, with the C1V5 configuration producing the least accurate results.

Figure 10 displays the calibration and validation results for individual weeks contained in the C1V5 and C3V3 configurations. We can see calibration periods for both C1V5 and C3V3 fit the winter period very well with low RMSE values (≤ 0.27 °C). However, when the GB model was validated on the winter season, the validation RMSE rose to 0.84 °C for the C3V3 configuration and to 2.43 °C for the C1V5 configuration, while the correlation remained very high ($r \geq 0.98$). The summer calibration periods produced RMSE values between 0.21 °C and 0.49 °C while the validation periods resulted in RMSE values between 0.53 °C and 1.46 °C, with r values between 0.76 and 0.95. The shoulder season produced the widest spread in RSMEs with calibration values between 0.22 °C and 0.71 °C and validation values between 0.38 °C and 2.39 °C. The r values ranged between 0.28 and 0.96. From these results we can see that calibrating the GB model using one week of data from the winter or summer period results in high errors and poor correlations when the model is then validated on shoulder season periods. Likewise, when the GB model is calibrated using one week from the shoulder season periods and validated on the summer or winter period, it generally performs poorly. The occupancy level in the winter period is zero, the summer period is low and the shoulder season is high (Figure 3). These results indicate the OB noise in the shoulder season has a negative effect on model calibration and validation. While the RMSE values for validation on the winter period vary considerably for GB models that have been trained using one week of data from another period, the correlation coefficient remains consistently very high. This may be due to the lack of occupants during this period and therefore the elimination of OB noise. When the C3V3 configuration is employed, the magnitude of validation error for all seasons reduced substantially as the calibration period generally contains a varied mix of data from multiple seasons and occupancy levels.

5. Discussion

From the results in Section 4.1.1 we can see the simple GB model employed in this study was capable of accurately capturing the dynamic characteristics of the internal air temperature profile when calibrated over one week using an automatic algorithm (Figure 5). However, depending on the week selected as the calibration period (varying season and occupancy levels), the parameters selected by the algorithm varied substantially (Figure 6), which resulted in varied model performance (generally poor) when validated (Table 5). Many previous studies utilised only one week or less of empirical data (single season) for model calibration purposes. Using a similar sized single-season calibration period (C1V5)

for the RC model and calibration algorithm used in this study did not produce consistently accurate results. When additional weeks (across varying seasons) where added to the calibration period, the consistency of the model's performance substantially improved. However, the generalisation abilities of the C3V3 GB model where still below that of the WB model. The C3V3 GB model did however produce accuracy levels comparable to that of the WB model (mean RMSE of the GB model was 1.5% higher than mean RMSE of the WB model).

From a practitioner's perspective, the GB modelling and automatic model calibration techniques were found to be far more straightforward and time efficient methods of simulating internal air temperature in an nZEB, in comparison to a WB model manually calibrated using a piecemeal evidence-based approach. The WB model had a complex structure and accounted for many other factors which interact with internal air temperature (such as relative humidity), while the GB model had a simple RC structure, less interactions and fewer parameters. The manual evidence-based WB calibration method was entirely depended on lengthy human interaction, while the automatic calibration method required minimal human interaction. The mean time for the calibration algorithm to converge on a final set of parameters for the full C3V3 configuration (six individual calibration periods made up of three weeks each) was 148 s using a six core Intel i7 3930 3.2 GHz processor with parallel computing enabled. The authors of this study estimated that by employing the GB modelling method with an automatic model calibration technique, the human labour input to simulating internal air temperature was reduced by approximately 90% relative to WB modelling using a manually calibrated evidence-based approach. The labour reduction applies only to the model development, calibration and validation time and does not include the time required to record and process the empirical data, which would be the same for both methods. This results in a significant decrease in human labour input, albeit with a slight decrease in accuracy and drop in generalisation abilities for this application. In this study, both WB and GB modelling techniques and their corresponding calibration methods were found to possess independent attributes and both styles had unique merits.

6. Conclusions

In this study, a GB model for a naturally ventilated nZEB was developed and calibrated using multiple data configurations using an automatic calibration algorithm. Then the GB model was compared to a WB model for the same application with identical calibration and validation datasets. The following conclusions were drawn from the results:

- The GB modelling method used in this study was capable of simulating the dynamic internal air temperature profile of a naturally ventilated nZEB.
- Utilising only one week for the GB model calibration dataset resulted in overfitting. When three weeks of data from varying seasons were used, the GB model was able to consistently produce more accurate results for the validation periods.
- The season and level of occupancy in the calibration and validation data had a strong influence on the GB model's accuracy levels.
- When calibrated and validated using identical data, the WB model produced slightly more accurate results than the GB model and displayed better generalisation abilities.
- Although the GB model was slightly less accurate than the WB model (mean RMSE 1.5% higher), the authors found the development time to be significantly lower for the GB automatic calibration method in comparison to the WB manual calibration method (approx. 90% reduction in human time input).

Abbreviations

Symbols

T	temperature (°C)
m	mass (kg)
C_A	specific heat capacity of air (J/kg K)
θ	energy (J)
U	u-value (W/(m^2 K))
A	area (m^2)
h	heat transfer co-efficient (W/(m^2 K))
γ	convective-radiative ratio (%)
p	density (kg/m^3)
V	volume (m^3)
v	wind velocity (m/s)
α	scaling coefficient
F_r	reference flow number
C_d	discharge coefficient
H	opening height (m)
g	acceleration due to gravity (m/s^2)
r	Pearson coefficient
Q	volumetric flow rate (m^3/s)

Subscripts

Int	internal
e	external dry bulb air
Sol	solar
$Occu$	occupancy
O	out
I	in
h	horizontal
t	time
v	vertical
inf	infiltration
$vent$	ventilation
$Cond$	conduction
Mat	material
A	air
B	buoyancy driven
W	wind driven

Abbreviations

nZEB	nearly zero energy building
MCSL	multi-configuration slotted louvre
NBERT	national building energy retro test-bed
RMSE	root mean squared error
RC	resistance capacitance
NV	naturally ventilated
GB	grey box
WB	white box
BB	black box
SSE	sum of squared errors

References

1. Ji, Y.; Xu, P.; Duan, P.; Lu, X. Estimating hourly cooling load in commercial buildings using a thermal network model and electricity submetering data. *Appl. Energy* **2016**, *169*, 309–323. [CrossRef]
2. Manfren, M.; Aste, N.; Moshksar, R. Calibration and uncertainty analysis for computer models—A meta-model based approach for integrated building energy simulation. *Appl. Energy* **2013**, *103*, 627–641. [CrossRef]
3. 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. [CrossRef]
4. Underwood, C.P. An improved lumped parameter method for building thermal modelling. *Energy Build.* **2014**, *79*, 191–201. [CrossRef]
5. Cattarin, G.; Pagliano, L.; Causone, F.; Kindinis, A.; Goia, F.; Carlucci, S.; Schlemminger, C. Empirical validation and local sensitivity analysis of a lumped-parameter thermal model of an outdoor test cell. *Build. Environ.* **2018**, *130*, 151–161. [CrossRef]
6. Pavlak, G.S.; Florita, A.R.; Henze, G.P.; Rajagopalan, B. Comparison of traditional and bayesian calibration techniques for gray-box modeling. *J. Archit. Eng.* **2014**, *20*. [CrossRef]
7. 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*, 263–274. [CrossRef]
8. Sturzenegger, D.; Gyalistras, D.; Morari, M.; Smith, R.S. Semi-automated modular modeling of buildings for model predictive control. In Proceedings of the 10th ACM Conference on Embedded Networked Sensor Systems (Sensys'12), Toronto, ON, Canada, 6–9 November 2012; p. 99. [CrossRef]
9. Oliveira Panão, M.J.N.; Mateus, N.M.; Carrilho da Graça, G. Measured and modeled performance of internal mass as a thermal energy battery for energy flexible residential buildings. *Appl. Energy* **2019**, *239*, 252–267. [CrossRef]
10. Fraisse, G.; Viardot, C.; Lafabrie, O.; Achard, G. Development of a simplified and accurate building model based on electrical analogy. *Energy Build.* **2002**, *34*, 1017–1031. [CrossRef]
11. Nielsen, T.R. Simple tool to evaluate energy demand and indoor environment in the early stages of building design. *Sol. Energy* **2005**, *78*, 73–83. [CrossRef]
12. Bacher, P.; Madsen, H. Identifying suitable models for the heat dynamics of buildings. *Energy Build.* **2011**, *43*, 1511–1522. [CrossRef]
13. Harb, H.; Boyanov, N.; Hernandez, L.; Streblow, R.; Müller, D. Development and validation of grey-box models for forecasting the thermal response of occupied buildings. *Energy Build.* **2016**, *117*, 199–207. [CrossRef]
14. Oliveira Panão, M.J.N.; Santos, C.A.P.; Mateus, N.M.; Carrilho Da Graça, G. Validation of a lumped RC model for thermal simulation of a double skin natural and mechanical ventilated test cell. *Energy Build.* **2016**, *121*, 92–103. [CrossRef]
15. Rodríguez Jara, E.; Sánchez de la Flor, F.J.; Álvarez Domínguez, S.; Molina Félix, J.L.; Salmerón Lissén, J.M. A new analytical approach for simplified thermal modelling of buildings: Self-Adjusting RC-network model. *Energy Build.* **2016**, *130*, 85–97. [CrossRef]
16. Andrade-Cabrera, C.; Burke, D.; Turner, W.J.N.; Finn, D.P. Ensemble Calibration of lumped parameter retrofit building models using Particle Swarm Optimization. *Energy Build.* **2017**, *155*, 513–532. [CrossRef]
17. Michalak, P. The development and validation of the linear time varying Simulink-based model for the dynamic simulation of the thermal performance of buildings. *Energy Build.* **2017**, *141*, 333–340. [CrossRef]
18. Robertson, J.J.; Polly, B.J.; Collis, J.M. Reduced-order modeling and simulated annealing optimization for efficient residential building utility bill calibration. *Appl. Energy* **2015**, *148*, 169–177. [CrossRef]
19. Lomas, K.J.; Eppel, H.; Martin, C.J.; Bloomfield, D.P. Empirical validation of building energy simulation programs. *Energy Build.* **1997**, *26*, 253–275. [CrossRef]
20. Mateus, N.M.; Pinto, A.; Da Graca, G.C. Validation of EnergyPlus thermal simulation of a double skin naturally and mechanically ventilated test cell. *Energy Build.* **2014**, *75*, 511–522. [CrossRef]
21. Ahmad, M.; Culp, C.H. Uncalibrated Building Energy Simulation Modeling Results. *HVAC&R Res.* **2006**, *12*, 1141–1155. [CrossRef]
22. Royapoor, M.; Roskilly, T. Building model calibration using energy and environmental data. *Energy Build.* **2015**, *94*, 109–120. [CrossRef]
23. O' Donovan, A.; O' Sullivan, P.D.; Murphy, M.D. Predicting air temperatures in a naturally ventilated nearly zero energy building: Calibration, validation, analysis and approaches. *Appl. Energy* **2019**, *250*, 991–1010. [CrossRef]
24. Roberti, F.; Oberegger, U.F.; Gasparella, A. Calibrating historic building energy models to hourly indoor air and surface temperatures: Methodology and case study. *Energy Build.* **2015**, *108*, 236–243. [CrossRef]
25. Snyder, S.C.; Maor, I. Calibrated Building Energy Simulation in Practice: Issues, Approaches, and Case Study Example. *ASHRAE Trans.* **2015**, *121*, 1–8.
26. Chaudhary, G.; New, J.; Sanyal, J.; Im, P.; O'Neill, Z.; Garg, V. Evaluation of "Autotune" calibration against manual calibration of building energy models. *Appl. Energy* **2016**, *182*, 115–134. [CrossRef]
27. Cacabelos, A.; Eguía, P.; Luís Míguez, J.; Granada, E.; Arce, M.E. Calibrated simulation of a public library HVAC system with a ground-source heat pump and a radiant floor using TRNSYS and GenOpt. *Energy Build.* **2015**, *108*, 114–126. [CrossRef]
28. ANSI/ASHRAE ASHRAE Guideline 14-2002 Measurement of Energy and Demand Savings. *Ashrae* **2002**, *8400*, 170.

29. Bertagnolio, S.; Randaxhe, F.; Lemort, V. Evidence-based calibration of a building energy simulation model: Application to an office building in Belgium. In Proceedings of the 12th International Conference for Enhanced Building Operations (ICEBO), Manchester, UK, 23–26 October 2012.

30. Raftery, P.; Keane, M.; O'Donnell, J. Calibrating whole building energy models: An evidence-based methodology. *Energy Build.* **2011**, *43*, 2356–2364. [CrossRef]

31. Bertagnolio, S.; Andre, P. Simulation Assisted Audit & Evidence Based Calibration Methodology. Available online: http://www.labeee.ufsc.br/pos-graduacao/ecv_4202/Nxx_D_calibration_methodology_SB_20100413.pdf%5Cnhttp://www.labeee.ufsc.br/sites/default/files/disciplinas/ECV4202_calibration_methodology_SB_20100413.pdf (accessed on 31 January 2021).

32. Ogando, A.; Cid, N.; Fernández, M. Energy modelling and automated calibrations of ancient building simulations: A case study of a school in the northwest of Spain. *Energies* **2017**, *10*, 807. [CrossRef]

33. Gandhi, P.; Brager, G.; Dutton, S. A Comparative Study of Mixed Mode Simulation Methods: Approaches in Research and Practice. In Proceedings of the Symposium on Simulation for Architecture & Urban Design, Alexandria, VA, USA, 12–15 April 2015; pp. 197–204.

34. Raftery, P.; Keane, M.; Costa, A. Calibrating whole building energy models: Detailed case study using hourly measured data. *Energy Build.* **2011**, *43*, 3666–3679. [CrossRef]

35. Monetti, V.; Davin, E.; Fabrizio, E.; André, P.; Filippi, M. Calibration of building energy simulation models based on optimization: A case study. *Energy Procedia* **2015**, *78*, 2971–2976. [CrossRef]

36. Mustafaraj, G.; Lowry, G.; Chen, J. Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office. *Energy Build.* **2011**, *43*, 1452–1460. [CrossRef]

37. Wu, S.; Sun, J.-Q. A physics-based linear parametric model of room temperature in office buildings. *Build. Environ.* **2012**, *50*, 1–9. [CrossRef]

38. Ngo, N.T. Early predicting cooling loads for energy-efficient design in office buildings by machine learning. *Energy Build.* **2019**, *182*, 264–273. [CrossRef]

39. Lu, T.; Lü, X.; Kibert, C. A hybrid numerical-neural-network model for building simulation: A case study for the simulation of unheated and uncooled indoor temperature. *Energy Build.* **2015**, *86*, 723–734. [CrossRef]

40. Mba, L.; Meukam, P.; Kemajou, A. Application of artificial neural network for predicting hourly indoor air temperature and relative humidity in modern building in humid region. *Energy Build.* **2016**, *121*, 32–42. [CrossRef]

41. Pandey, S.; Hindoliya, D.A.; Mod, R. Artificial neural networks for predicting indoor temperature using roof passive cooling techniques in buildings in different climatic conditions. *Appl. Soft Comput. J.* **2012**, *12*, 1214–1226. [CrossRef]

42. Liu, Z.; Liu, Y.; He, B.J.; Xu, W.; Jin, G.; Zhang, X. Application and suitability analysis of the key technologies in nearly zero energy buildings in China. *Renew. Sustain. Energy Rev.* **2019**, *101*, 329–345. [CrossRef]

43. Feng, W.; Zhang, Q.; Ji, H.; Wang, R.; Zhou, N.; Ye, Q.; Hao, B.; Li, Y.; Luo, D.; Lau, S.S.Y. A review of net zero energy buildings in hot and humid climates: Experience learned from 34 case study buildings. *Renew. Sustain. Energy Rev.* **2019**, *114*, 109303. [CrossRef]

44. Al-Saadi, S.N.; Shaaban, A.K. Zero energy building (ZEB) in a cooling dominated climate of Oman: Design and energy performance analysis. *Renew. Sustain. Energy Rev.* **2019**, *112*, 299–316. [CrossRef]

45. Bandeiras, F.; Gomes, M.; Coelho, P.; Fernandes, J. Towards net zero energy in industrial and commercial buildings in Portugal. *Renew. Sustain. Energy Rev.* **2020**, *119*, 109580. [CrossRef]

46. Paoletti, G.; Pascual Pascuas, R.; Pernetti, R.; Lollini, R. Nearly Zero Energy Buildings: An Overview of the Main Construction Features across Europe. *Buildings* **2017**, *7*, 43. [CrossRef]

47. Brambilla, A.; Salvalai, G.; Imperadori, M.; Sesana, M.M. Nearly zero energy building renovation: From energy efficiency to environmental efficiency, a pilot case study. *Energy Build.* **2018**, *166*, 271–283. [CrossRef]

48. Colclough, S.; Kinnane, O.; Hewitt, N.; Griffiths, P. Investigation of nZEB social housing built to the Passive House standard. *Energy Build.* **2018**, *179*, 344–359. [CrossRef]

49. Ascione, F.; Bianco, N.; Böttcher, O.; Kaltenbrunner, R.; Vanoli, G.P. Net zero-energy buildings in Germany: Design, model calibration and lessons learned from a case-study in Berlin. *Energy Build.* **2016**, *133*, 688–710. [CrossRef]

50. Wells, L.; Rismanchi, B.; Aye, L. A review of Net Zero Energy Buildings with reflections on the Australian context. *Energy Build.* **2018**, *158*, 616–628. [CrossRef]

51. Bonte, M.; Thellier, F.; Lartigue, B. Impact of occupant's actions on energy building performance and thermal sensation. *Energy Build.* **2014**, *76*, 219–227. [CrossRef]

52. Paliouras, P.; Matzaflaras, N.; Peuhkuri, R.H.; Kolarik, J. Using measured indoor environment parameters for calibration of building simulation model—A passive house case study. *Energy Procedia* **2015**, *78*, 1227–1232. [CrossRef]

53. Kampelis, N.; Gobakis, K.; Vagias, V.; Kolokotsa, D.; Standardi, L.; Isidori, D.; Cristalli, C.; Montagnino, F.M.; Paredes, F.; Muratore, P.; et al. Evaluation of the performance gap in industrial, residential & tertiary near-Zero energy buildings. *Energy Build.* **2017**, *148*, 58–73. [CrossRef]

54. Carrilho da Graça, G.; Linden, P. Ten questions about natural ventilation of non-domestic buildings. *Build. Environ.* **2016**, *107*, 263–273. [CrossRef]

55. Asaleye, D.A.; Breen, M.; Murphy, M.D. A decision support tool for building integrated renewable energy microgrids connected to a smart grid. *Energies* **2017**, *10*, 1765. [CrossRef]

56. O'Donovan, A.; O'Sullivan, P.D.; Murphy, M.D. A field study of thermal comfort performance for a slotted louvre ventilation system in a low energy retrofit. *Energy Build.* **2017**, *135*, 312–323. [CrossRef]

57. O'Sullivan, P.D.; Kolokotroni, M. A field study of wind dominant single sided ventilation through a narrow slotted architectural louvre system. *Energy Build.* **2017**, *138*, 733–747. [CrossRef]

58. Phan, Q.A.; Scully, T.; Breen, M.; Murphy, M.D. Determination of optimal battery utilization to minimize operating costs for a grid-connected building with renewable energy sources. *Energy Convers. Manag.* **2018**, *174*, 157–174. [CrossRef]

59. O' Donovan, A.; Murphy, M.D.; O'Sullivan, P.D. Passive control strategies for cooling a non-residential nearly zero energy office: Simulated comfort resilience now and in the future. *Energy Build.* **2020**, 110607. [CrossRef]

60. Department of Housing, Local Government and Heritage. *Conservation of Fuel and Energy—Buildings Other than Dwellings—TGDL*; Government of Ireland: Dublin, Ireland, 2017.

61. O' Sullivan, P.D.; Delaney, F.; O' Riain, M.; Clancy, T.; O' Connell, J.; Fallon, D. Design and Performance of and External Building Envelope Retrofit Solution for a Grid Optimised Contcrete Structure: A Case Study. In Proceedings of the 30th International Manufacturing Conference, Dublin, Ireland, 30 August–1 September 2011; pp. 1–12.

62. O'Sullivan, P.D.; O' Donovan, A.; Murphy, M.D. The influence of thermal mass on the predicted climate cooling potential in low energy buildings. In Proceedings of the AIVC, Antibes Juan-Les-Pins, France, 18–19 September 2018; p. 10.

63. Gouda, M.M.; Danaher, S.; Underwood, C.P. Building thermal model reduction using nonlinear constrained optimization. *Build. Environ.* **2002**, *37*, 1255–1265. [CrossRef]

64. Roberti, F. Energy retrofit and conservation of built heritage using multi-objective optimization: Demonstration on a medieval building. In Proceedings of the Building Simulation Applications, Bolzano, Italy, 4–6 February 2015; pp. 189–197.

65. Perez, R.; Ineichen, P.; Seals, R.; Michalsky, J.; Stewart, R. Modeling daylight availability and irradiance components from direct and global irradiance. *Sol. Energy* **1990**, *44*, 271–289. [CrossRef]

66. Laboratory, S.E. *TRNSYS 17: A Transient System Simulation Program*; Solar Energy Laboratory, University of Wisconsin: Madison, WI, USA, 2017.

67. ISO 13790:2008. *Energy Performance of Buildings—Calculation of Energy Use for Space Heating and Cooling*; ISO: Geneva, Switzerland, 2008.

68. Chartered Institution of Building Services Engineers. *CIBSE Guide A: Environmental Design*; Chartered Institution of Building Services Engineers: Norwich, UK, 2006.

69. Warren, P.R. Ventilation Through Openings on one wall only. *Int. Conf. Heat Mass Transf. Build.* **1977**. [CrossRef]

70. TRANSSOLAR & Energietechnik GmbH. *TRNSYS3d*; TRANSSOLAR & Energietechnik GmbH: Stuggart, Germany, 2011.

71. Shine, P.; Scully, T.; Upton, J.; Shalloo, L.; Murphy, M.D. Electricity & direct water consumption on Irish pasture based dairy farms: A statistical analysis. *Appl. Energy* **2018**, *210*, 529–537. [CrossRef]

72. Huerto-Cardenas, H.E.; Leonforte, F.; Aste, N.; Del Pero, C.; Evola, G.; Costanzo, V.; Lucchi, E. Validation of dynamic hygrothermal simulation models for historical buildings: State of the art, research challenges and recommendations. *Build. Environ.* **2020**, *180*, 107081. [CrossRef]

73. Moré, J.J. The Levenberg-Marquardt algorithm: Implementation and theory. In *Numerical Analysis*; Watson, G.A., Ed.; Lecture Notes in Mathematics; Springer: Berlin/Heidelberg, Germany, 1978; Volume 630.

74. Braun, J.E.; Chaturvedi, N. An Inverse Gray-Box Model for Transient Building Load Prediction. *HVAC&R Res.* **2002**, *8*, 73–99. [CrossRef]

75. Murphy, M.D.; O'Mahony, M.J.; Upton, J. Comparison of control systems for the optimisation of ice storage in a dynamic real time electricity pricing environment. *Appl. Energy* **2015**, *149*, 392–403. [CrossRef]

76. Ilaria, P.; Prada, A.; Cappelletti, F.; Gasparella, A. Multi-Stage Multi-Level Calibration of a School Building Energy Model. In Proceedings of the Building Simulation Appliocations (BSA) 2019, Bolzano, Italy, 19–21 June 2019; p. 10.

77. Dirutigliano, D.; Brüntjen, M.-A.; Fliegner, C.; Frisch, J.; Corrado, V.; Treeck, C. van Case Study for Energy Efficiency Measures of Buildings on an Urban Scale. In Proceedings of the Building Simulation Applications, San Francisco, CA, USA, 7–9 August 2017; pp. 403–410.