

## Article

# Data-Driven Estimation of Time-Varying Stochastic Effects on Building Heat Consumption Related to Human Interactions

Christoffer Rasmussen <sup>1,2,\*</sup> , Niels Lassen <sup>3</sup> , Peder Bacher <sup>1</sup> , Tor Helge Dokka <sup>3</sup> and Henrik Madsen <sup>1</sup> 

<sup>1</sup> Department of Applied Mathematics and Computer Science, The Technical University of Denmark, 2800 Kongens Lyngby, Denmark; pbac@dtu.dk (P.B.); hmad@dtu.dk (H.M.)

<sup>2</sup> Department of Research and Development, Watts, 4600 Køge, Denmark

<sup>3</sup> Skanska Technology, Skanska Norway, 0187 Oslo, Norway; niels.lassen@skanska.no (N.L.); tor.helge.dokka@skanska.no (T.H.D.)

\* Correspondence: christoffer.rasmussen@watts.dk or chrras@dtu.dk

**Abstract:** Within the field of statistical modelling and data-driven characterisation of buildings' energy performance, the focus is typically on parameter estimation of the building envelope and the energy systems. Less focus has been put on the stochastic human effect on energy consumption. We propose a new method for estimating the thermal building properties while, in parallel, estimating time-varying effects caused by the humans' interactions with the building. We do that by combining a smooth, non-linear formulation of the energy signature method known from the literature with a hidden state formulated as a random walk to describe the human interactions with the building. The method is demonstrated on data obtained from autumn 2019 to late spring 2021 from a 900 m<sup>2</sup> newly built school building located south of Oslo, Norway. The demonstration case has shown that the model accuracy increases and the model bias decrease when cross-validated. The estimated hidden state has also been shown to resemble the estimated combined mechanical and natural ventilation pattern controlled by the building users and operational staff. These human interactions have increased the total heat loss expressed in kilowatts per kelvin by around 50% over the course of one year from before the COVID-19 pandemic to after its outbreak.

**Keywords:** data-driven modelling; building physics; hidden state estimation; time-varying estimation; interpretability; occupants' behaviour; building operation



**Citation:** Rasmussen, C.; Lassen, N.; Bacher, P.; Dokka, T.H.; Madsen, H. Data-Driven Estimation of Time-Varying Stochastic Effects on Building Heat Consumption Related to Human Interactions. *Energies* **2023**, *16*, 5991. <https://doi.org/10.3390/en16165991>

Academic Editor: Francesco Nocera

Received: 8 June 2023

Revised: 28 July 2023

Accepted: 5 August 2023

Published: 15 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The interest in data-driven methods for the characterisation of buildings' thermophysical properties is increasing and is today widely represented in the literature. Some of the first studies on the characterisation of buildings' thermophysical properties deal with steady-state characterisation [1–4]. More recently, the dynamic aspects of the building's physical properties are taken into account by utilising either grey- or black-box models [5–8]. Other studies focus specifically on the estimation of building dynamics and buildings' energy flexibility [9]. Related to occupants' behaviour, several studies deal with the estimation of occupants' presence in buildings. One study demonstrated a method for estimating the number of occupants by using Markov-switching models and CO<sub>2</sub> measurements [10]. Yet another study utilises 3D camera data to estimate occupants' presence [11]. However, less focus has been put on data-driven methods where the human interactions with the building are taken explicitly into account when estimating buildings' thermophysical properties. Without methods for estimating building physical properties while taking human interactions into account, documentation of the actual building performance might be difficult to achieve.

In the IEA EBC, Annex 53 report [12], it is stated that the energy consumption of a building is influenced by six main factors: climate, building envelope characteristics, building services and energy systems characteristics, building operation and maintenance, occupant activities and behaviour, and indoor environmental quality provided. A similar

categorisation is found in [13]. However, building designers mainly tend to focus on the first three factors [14]. Likewise, the literature on data-driven methods for the estimation of thermal building properties—as the literature mentioned above—shows a similar tendency. Despite these tendencies, it is found that the latter three of the before mentioned six influential factors have an equally big impact on energy consumption as the first three factors—namely building operation and maintenance, occupant activities and behaviour, and indoor environmental quality provided which all are human-centric effects [15].

It cannot be said for certain that the lack of focus on the human-centric effects is the reason for the frequently observed discrepancy between the actual and estimated energy consumption (the performance gap). However, several studies have shown that the discrepancy is significant [16]. One study showed that the difference between the estimated and the actual energy consumption can exceed 100% due to occupants' behaviour [17], and a difference of 300% was observed between identical buildings in another study [18].

The performance gap is, however, not limited to the occupants' behaviour. Several studies have shown that discrepancies between the prescribed building design and reality can vary drastically. One study from 2011 found that 18 out of 18 (100%) newly built British dwellings had a significantly higher heat loss coefficient than anticipated in the design when it was assessed by co-heating methods on the finished building [19]. Furthermore, energy performance certificates (EPCs) suffer from low accuracy as well. The Danish Energy Agency found that 23% of the Danish energy performance certificate issued during 2018 were misclassified, and 21% and 31% were misclassified during 2017 and 2016, respectively [20].

The reasons for the observed discrepancies can be manifold, as is seen in the literature. This has potentially large economic consequences for the end-user. Especially in times with extraordinarily high energy prices, such as those we experience today. The annual gas inflation in Europe was 51.4% in July 2022, and for energy in general 41.1% [21]. It is, therefore, urgent and important to develop easily scalable methods for disaggregating heat consumption of buildings to gain better insight into how the energy is used. Such insight is of high importance for prioritising future building renovations, optimising the operation of the buildings, and educating building occupants on better utilisation of the energy available.

In this article, we investigate how the advanced steady-state energy signature model—namely the model proposed by Rasmussen et al. [22]—can be altered to estimate unknown disturbances ultimately caused by the user and operational staff. Those previously unobserved effects were caused by changing ventilation rates, indoor temperature, and internal heat gains.

### *Research Objective and Outline*

This study aims to apply statistical modelling techniques to deduct latent effects on the total heat consumption caused by human interaction with a building (e.g., changing ventilation rates, indoor temperature, and internal heat gains) solely from heat consumption and weather data.

The remainder of the paper is structured as follows. In Section 2, the model structure of the energy signature and the implementation of a time-varying state are presented. In Section 3, the demonstration case and the data used are described. In Section 4, the results are presented. First, the model residuals are analysed; second, the parameter estimates are discussed; lastly, the time-varying hidden state is evaluated against other measurements. In Section 5, the method and the modelling technique are discussed, and finally, the conclusions are drawn in Section 6.

## **2. Modelling Approach**

This study is based on the model developed by Rasmussen et al. [22]. In that study, a general framework for estimating thermal building energy performance parameters of single-family houses is presented. The article proposes a method for estimating the influence of the occupants from the systematic model errors, which cannot be led back

to exogenous effects such as weather conditions. The proposed method is a two-stage approach where the model parameters are estimated in stage one, and the residuals are modelled and analysed in stage two. This approach makes the algorithm greedy, meaning that it is locally optimal at each stage rather than globally optimal. To avoid this, a new method is proposed. A hidden state is introduced in the model to account for the latent effects on heat consumption unrelated to the weather conditions.

### 2.1. Model Structure

The basic model structure used is near-identical to the one found in the original paper [22]. A minor variation is, however, made to separate error terms related to the weather-dependent heat load,  $\Phi_{h_1}^{(t)}$ , and the weather-independent heat load,  $\Phi_{h_2}^{(t)}$ . The model structure is

$$\Phi_h^{(t)} = \text{LSE}[\Phi_{h_1}^{(t)}, \Phi_{h_2}^{(t)}], \quad (1)$$

where  $\Phi_h^{(t)}$  is the daily average heating effect with superscript  $(t)$  denoting time. Furthermore, the weather-dependent daily average heating effect is

$$\Phi_{h_1}^{(t)} = (UA_0 + W_s^{(t)} UA_w) (T_b - T_a^{(t)}) - gA I_g^{(t)} + \varepsilon^{(t)} \quad (2)$$

and the weather-independent heating effect is

$$\Phi_{h_2}^{(t)} = \Phi_0 + \zeta^{(t)}. \quad (3)$$

In Equations (2) and (3),  $W_s^{(t)}$  is the wind speed,  $T_a^{(t)}$  is the outdoor ambient air temperature,  $I_g^{(t)}$  is the global solar irradiation. These are all time-dependent and act as the model inputs. Furthermore,  $UA_0$  is the heat loss coefficient under wind-free conditions,  $UA_w$  is the wind dependency,  $T_b$  is the base temperature,  $gA$  is the solar transmittance,  $\Phi_0$  is the background heat load, and  $\varepsilon^{(t)}$  and  $\zeta^{(t)}$  are the error terms. Finally, **LSE** (LogSumExp) is a smooth maximum function defined for two inputs ( $f$  and  $g$ ) as

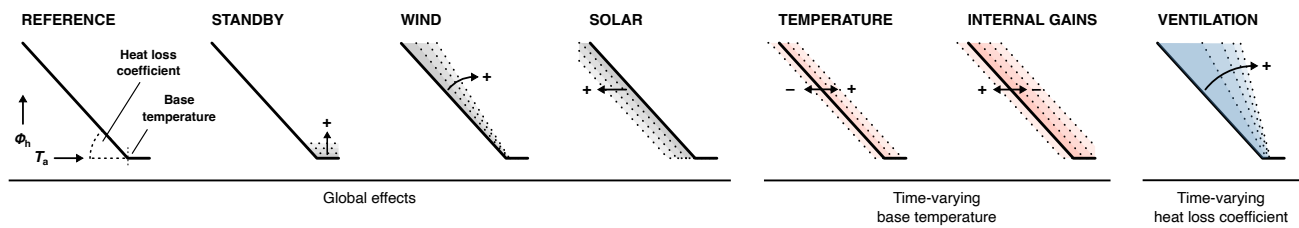
$$\text{LSE}[f(x), g(x)] = \underbrace{\log[\exp(f(x)k) + \exp(g(x)k)]k^{-1}}_{\text{Smooth approximation of } \max[f(x), g(x)]}, \quad (4)$$

where  $k$  is a parameter determining the transition rate between the two functions,  $f(x)$  and  $g(x)$ . Substituting  $f(x)$  with  $\Phi_{h_1}^{(t)}$  and  $g(x)$  with  $\Phi_{h_2}^{(t)}$  as seen in Equation (1), a function for the weather-dependent heating effect which transition smoothly into the weather-independent heating effect is obtained. For a more thorough description of the LogSumExp function and the energy signature, one is referred to [22].

In the following section, we will investigate how different effects will alter the heat load,  $\Phi_h^{(t)}$ , in the energy signature. Based on that, suggestions on how the model can be further improved to account for unknown effects, such as changing ventilation rates, temperature setpoints, and internal heat gains, are given.

### 2.2. Driving Forces of Heat Consumption

In Figure 1, a conceptual illustration shows how different external and internal forces drive the energy signature function. For example, the wind speed increases the heat loss caused by the increased air exchange as demonstrated by Rasmussen et al. [23]. Ultimately, this leads to a steeper gradient of the energy signature function, as seen in Figure 1 (Wind). Figure 1 (Solar) shows how the solar gain alters the base temperature while maintaining the gradient of the energy signature function. In both cases, the effects can be modelled as a function of either wind speed or solar irradiation (see Equation (2)) which are easily obtained.



**Figure 1.** Conceptual illustration of different effects on the energy signature function. In the leftmost sub-plot, the heat consumption as a function of the outdoor temperature alone is shown. The heat loss coefficient (the slope) and the base temperature (transition point) are highlighted. In the following three grey sub-plots the effect of the standby heat load, wind speed, and solar irradiation are shown. The two red sub-plots show how the effect of changing indoor temperature and internal heat gains alter the base temperature, and the last blue sub-plot shows how the ventilation rate alters the apparent heat loss coefficient.

For other effects driving heat consumption, such as changing indoor temperatures, internal heat gains, and ventilation rates, measurements of relevant variables are, in many cases, infeasible to obtain. However, by studying the effects of the indoor temperature, internal heat gains, and ventilation rate in the three right-most sub-plots of Figure 1 we might learn how to deal with such effects by indirect methods.

First, a change in the indoor temperature or the internal heat gains, such as electricity loads or metabolism, will offset the heat demand and consequently shift the base temperature (i.e., the intersection with the base heat load). This is seen from Figure 1 (Temperature and Internal gains) and Equations (1)–(3) while realising that

$$T_b^{(t)} = T_i^{(t)} - \frac{\Phi_x^{(t)}}{UA_0 + W_s^{(t)} UA_w}, \quad (5)$$

where  $T_i^{(t)}$  and  $\Phi_x^{(t)}$  are the unobserved indoor air temperature and internal heat gains, respectively. Under conditions with varying indoor temperature and otherwise constant conditions, the heating effect  $\Phi_h^{(t)}$  increases by a factor of  $UA_0 + W_s^{(t)} UA_w$  for every increment of the indoor temperature. Likewise, the heating effect decrease by  $\Phi_x^{(t)}$  for every increment in  $\Phi_x^{(t)}$ .

Second, the heating effect increases as the ventilation heat loss increases. The ventilation heat loss increases linearly with the temperature difference of the supply and return temperature given a constant ventilation rate, and for an increment in the ventilation rate, the slope of the energy signature increase as illustrated in Figure 1 (Ventilation).

Depending on what the hidden driving forces are—either changing indoor temperature, heat gains, or ventilation rates—a natural extension of the energy signature model in Equations (1)–(3) is to alter the base temperature,  $T_b$ , in case of changing indoor temperature or internal heat gains or to alter  $UA_0$  in case of changing ventilation rates.

As we do not want to rely on additional measurements to model  $T_b$  or  $UA_0$ , we can treat them as stochastic processes formulated as random walks. In the following sections, we will show what a random walk is and how it can be used to represent hidden driving forces as those outlined above.

### 2.3. The Random Walk

In discrete time, the time series  $\{S^{(t)}\}$  is a simple random walk defined as

$$S^{(t)} = \sum_{i=1}^t X^{(i)} \quad (6)$$

where  $t$  is the time step and  $X^{(i)}$  is a random Gaussian variable.

For  $X^{(i)}$  being independent and normal distributed with mean zero and variance  $\sigma^2$ —that is  $X^{(i)} \in N(0, \sigma^2)$ —the expected value of  $S^{(t)}$  is

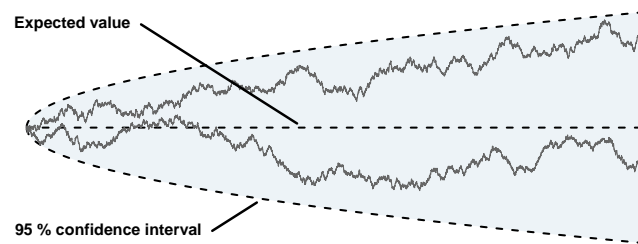
$$\mathbf{E}[S^{(t)}] = \sum_{i=1}^t \mathbf{E}[X^{(i)}] = 0. \quad (7)$$

Further, the variance of  $S^{(t)}$  is

$$\mathbf{Var}[S^{(t)}] = \sum_{i=1}^t \mathbf{Var}[X^{(i)}] = t \sigma^2, \quad (8)$$

which scales linear with time,  $t$ .

In Figure 2, two realisations of the same random walk, the expected value, and 95% confidence interval are shown.



**Figure 2.** Two realisations of a random walk of 10,000 steps each, their expected mean and 95% confidence interval.

Møller et al. [24] has shown how a continuous-time random walk—specifically a Wiener process—can be implemented in grey-box models as a hidden state to identify unobserved system dynamics. A similar discrete-time approach is used in this study and outlined below.

#### 2.4. Random Walk and Energy Signature

Presuming that the indoor temperature, internal heat gains, and ventilation rate are not necessarily constant, they can be treated as hidden states evolving in time. For a model like the energy signature in Equations (1)–(3), that means that the effects of the human interaction with the building are embedded in the measured heat consumption, but the hidden states themselves—for that reason the name—are unknown.

As we have already seen in Section 2.2, the energy signature can be altered in two ways. Neglecting that both changes in the indoor temperature and internal heat gains, and ventilation rates can occur simultaneously or during different periods, the heating effect,  $\Phi_{h1}^{(t)}$  in Equation (2), might either be offset or the slope (2) might change.

To model the former scenario, the base temperature  $T_b$  in Equation (2) is substituted with the random walk

$$T_b^{(t)} = T_b^{(0)} + \sum_{i=1}^t X_{T_b}^{(i)}, \quad (9)$$

where  $T_b^{(0)}$  is the initial condition and  $X_{T_b}^{(i)} \in N(0, \sigma_{T_b}^2)$ .

For the latter scenario,  $UA_0$  could be substituted with a random walk to account for the time-varying slope. Instead, we choose to alter the *total* UA value to simplify the model. That is

$$UA^{(t)} = UA^{(0)} + \sum_{i=1}^t X_{UA}^{(i)}, \quad (10)$$

where  $UA^{(t)}$  is the random walk which substitutes  $UA_0 + W_s^{(t)} UA_w$  in Equation (2),  $UA^{(0)}$  is the initial condition of  $UA^{(t)}$ , and  $X_{UA}^{(i)} \in N(0, \sigma_{UA}^2)$ . Notice that the potential effects on the heating effect caused by the changing wind speeds will now be captured by the random walk.

The full models, including a static baseline model (Model 0) and two time-varying models (Model 1 and Model 2), which are compared in this study are:

*Model 0 (Static)*

$$\Phi_h^{(t)} = \text{LSE} \left[ (UA_0 + W_s^{(t)} UA_w) (T_b - T_a^{(t)}) - gA I_g^{(t)} + \varepsilon^{(t)}, \Phi_0 + \zeta^{(t)} \right] \quad (11)$$

*Model 1 (Time-varying base temperature)*

$$\Phi_h^{(t)} = \text{LSE} \left[ (UA_0 + W_s^{(t)} UA_w) (T_b^{(t)} - T_a^{(t)}) - gA I_g^{(t)} + \varepsilon^{(t)}, \Phi_0 + \zeta^{(t)} \right] \quad (12a)$$

$$T_b^{(t)} = T_b^{(0)} + \sum_{i=1}^t X_{T_b}^{(i)} \quad (12b)$$

*Model 2 (Time-varying heat loss coefficient)*

$$\Phi_h^{(t)} = \text{LSE} \left[ UA^{(t)} (T_b - T_a^{(t)}) - gA I_g^{(t)} + \varepsilon^{(t)}, \Phi_0 + \zeta^{(t)} \right] \quad (13a)$$

$$UA^{(t)} = UA^{(0)} + \sum_{i=1}^t X_{UA}^{(i)} \quad (13b)$$

## 2.5. Estimation

One approach to estimate the model parameters and hidden state is to use maximum likelihood estimation (MLE). The advantage of the MLE method is that the method allows for estimating parameters related to the noise term and, in this setting, the hidden states. The outline of the MLE method is omitted here, but details can be found in *Parameter Estimation in Stochastic Grey-box Models* by Kristensen et al. [25]. In practice, the model parameters are found by maximising the log-likelihood function, and the hidden state is found by Laplace approximation using R [26] and TMB (Template Model Builder) [27,28].

## 2.6. Cross-Validation

Models with hidden and unconstrained states, such as those presented here, are prone to overfitting. To what extent they overfit depends on the data, the estimation approach, and the objective function. To quantify the bias-variance trade-off, the models are cross-validated.

First, each day is categorised according to whether the school was open or closed due to vacations or the COVID-19 lockdown.

Second, 20% of the days where the school was closed and 20% of the days where the school was open were randomly selected as validation data blocks of one to ten continuous observations, i.e., days. The remaining data were used for training. To support the main purpose of cross-validation, only data from the heating season was selected for validation purposes as the hidden states only have meaning under these conditions.

Furthermore, each validation block was chosen such that the block and a three-day burn-in period were contained in a period of the same category, i.e., days where the school were either open or closed. Each block—and the three-day burn-in period—is thereby fully contained within a period of presumed similar building operational conditions.

For each validation block, the heating effect is predicted given the model, the estimated global model parameters, and the expected value of the hidden state variable. The expectation of the hidden state variable is obtained by its prediction throughout the validation block.

The predictor of the random walk  $T_b^{(t)}$  is the naive—or persistent—predictor as it follows from Equation (7). We say that the prediction of  $T_b$  at time  $t + n$  given all information at time  $t$ , is  $T_b$  at time  $t$ . Therefore, we have that

$$\widehat{T_b}^{(t+n|t)} = T_b^{(t)}, \quad (14)$$



where  $n \in \mathbf{Z}^+$ .

The variance of the prediction error is

$$\text{Var}\left[T_b^{(t+n)} - \widehat{T}_b^{(t+n|t)}\right] = n \sigma_{T_b}^2, \quad (15)$$

The predictor and the variance of the prediction error are equivalent for the random walk  $UA^{(t)}$ .

### 3. Demonstration Case

The newly built Montessori school in Drøbak, Norway, was used as the case building. The building is a lower secondary school for 60 pupils with two floors and a heated floor area of approximately 900 m<sup>2</sup>. It has a compact rectangular shape oriented southeast-northwest, and the building volume is intersected by an inclined part of the roof with solar PV panels. The school has been built with the vision to become Norway's most environmentally friendly school. The basis for the energy concept design is a well-insulated building envelope with minimal heat loss, a very efficient lighting system, a high-performance ventilation system, and a ground-source heat pump system that provides low-temperature heating in winter and free cooling in summer. Table 1 contains a few key properties of the building.

**Table 1.** Building data of the Montessori school in Drøbak, Norway.

Building Data	
Net floor area	886 m <sup>2</sup>
Volume	3340 m <sup>3</sup>
Measured infiltration (ACH50)	0.5 h <sup>-1</sup>
Normalised heat capacity	81 Wh/(m <sup>2</sup> K)
Normalised heat loss coefficient for envelope	0.75 W/(m <sup>2</sup> K)

The building has automatic motorised external solar shading for all sun-exposed windows. Heating and cooling are provided by a central air-based system. The air distribution is based on displacement ventilation that varies between fully mechanical and hybrid ventilation depending on the time of the year. Outside the heating season, the exhaust air is directly discharged to the outside through an opening at the top of the atrium in the centre of the building. When heating, the air is mechanically exhausted through the air handling unit after it has passed the heat recovery. The heating, ventilation, and air conditioning (HVAC) system of Drøbak Montessori is unique in the aspect that it is based on the displacement ventilation principle and provides heating through recirculated air outside the occupied hours only. The occupants can open mechanically operable windows in each room. There are no sensors for tracking the opening of windows.

In the winter of 2020, the COVID-19 pandemic hit Europe, and by mid-March 2020, the Montessori school was, as a consequence, closed. The school remained closed and unoccupied until the beginning of May 2020. During this period, the mechanical ventilation rate was turned down. After the lockdown, the mechanical ventilation increased again. From the fourth quarter of 2022, the mechanical ventilation rate increased significantly compared to the pre-pandemic period—likely due to awareness of the continuous presence and spread of COVID-19.

#### 3.1. Data Acquisition

The data foundation is based on sensor data from the building management system and weather data obtained from Copernicus Climate Data Store [29].

Data from the building management system (BMS) has been acquired through an application programming interface (API) for the period 17 October 2019 to 6 April 2021. The data contains measured variables that were used for modelling as well as for model validation. In Figure 3, the hourly pre-processed measurements from the BMS are visualised for the entire measurement period as well as a single week of November 2019.

It has not been possible to acquire either the sensor brand and model or their measurement accuracy. It is, however, presumed that the model noise level is significantly higher than the measurement noise from the sensors, making the specific sensor choice less critical.

The weather data used in this study are from the ERA5-Land data set [30] from the Copernicus Climate Data Store [29]. The ERA5-Land data set provides global historical weather data on the landmasses in a  $0.1^\circ \times 0.1^\circ$  grid. Three weather variables were gathered: the outdoor air temperature 2 m above ground, the global solar irradiation in a horizontal plane, and the wind speed 10 m above ground. Each weather variable was obtained in hourly resolution and represents the estimated mean of the entire spatial grid cell.

### 3.2. Data Pre-Processing

Before any operations were made on the data, the data were re-sampled to obtain hourly equidistant observations. Effectively, the resulting observation at, e.g., 10:00 is the average value from 09:00 to 10:00. The hourly resolution is only used during pre-processing. After the pre-processing steps described below and before modelling, the data were down-sampled to 24 h average values to filter out the majority of the heat dynamics.

#### 3.2.1. Recirculation of Ventilation Air

For validation purposes, mechanical ventilation heat loss was needed. To calculate the mechanical ventilation heat loss, we first estimate if the ventilation air was recirculated or not. Due to the lack of data on the state of recirculation,  $R$ , the following binary assumption was made.

$$R^{(t)} = \begin{cases} 1 \text{ (true)}, & \text{for } Q_r^{(t)} = \eta^{(t)} = 0 \\ 0 \text{ (false)}, & \text{else,} \end{cases} \quad (16)$$

where  $R = 1$  indicates full recirculation and  $R = 0$  indicates no recirculation. Furthermore,  $Q_r^{(t)}$  is the return volumetric airflow, and  $\eta^{(t)}$  is the temperature heat recovery. In Figure 3, the hourly flow rates are illustrated, and one single period with recirculation during night-time is highlighted.

#### 3.2.2. Heating

As all heating and cooling are supplied through the ventilation system, the net heating and cooling supply is calculated as

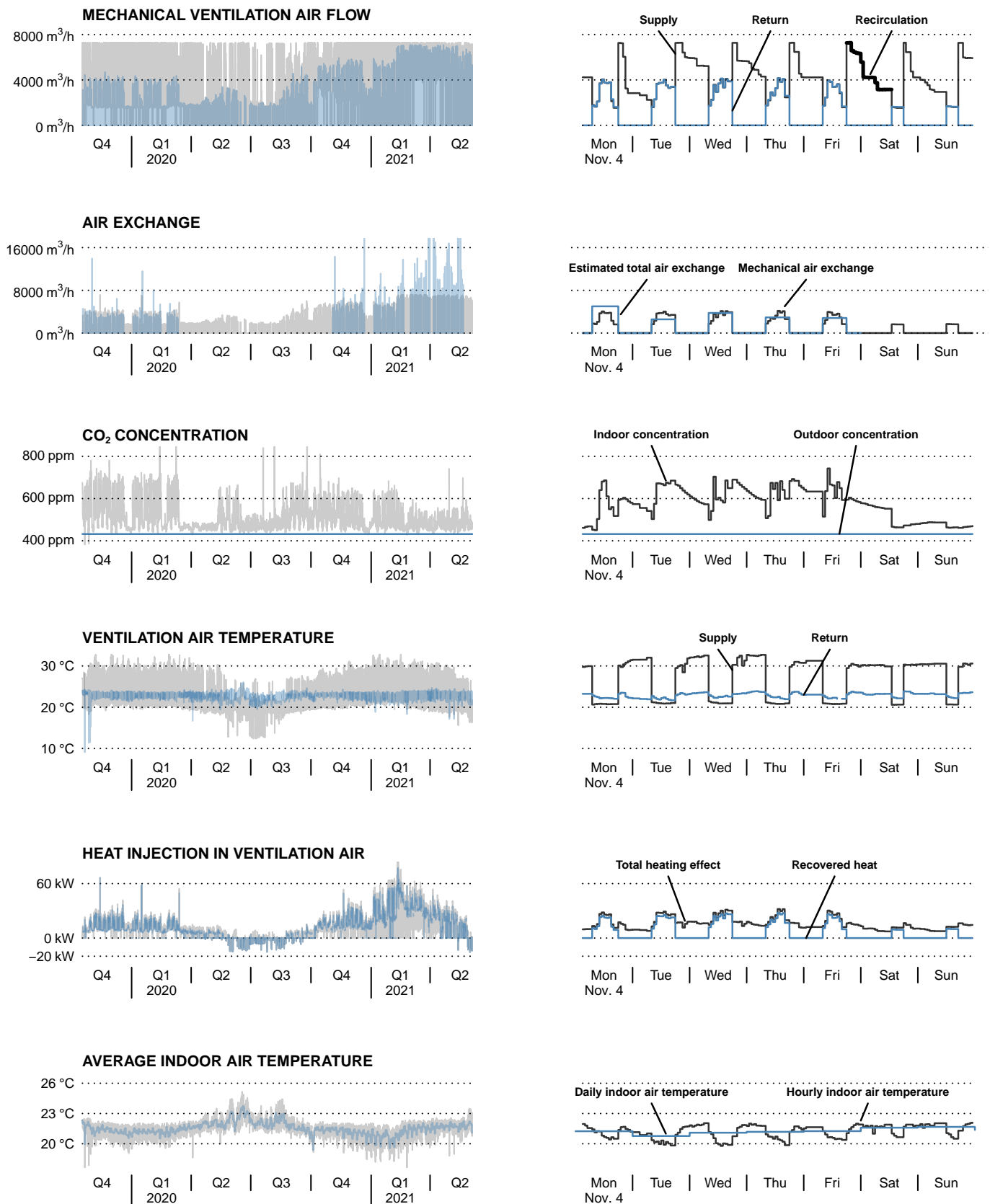
$$\Phi_h^{(t)} = Q_s^{(t)} c_p \rho \Delta T^{(t)} \quad (17)$$

where  $\Phi_h^{(t)}$  is the heating power (or negative cooling power for  $\Phi_h^{(t)} < 0$ ),  $Q_s^{(t)}$  is the volumetric supply flow rate,  $c_p$  is the specific heat capacity of the air,  $\rho$  is air density, and

$$\Delta T^{(t)} = \begin{cases} T_s^{(t)} - T_a^{(t)}, & \text{for } R^{(t)} = 0 \\ T_s^{(t)} - T_r^{(t)}, & \text{else,} \end{cases} \quad (18)$$

where  $T_a^{(t)}$  is the outdoor ambient air temperature, and  $T_s^{(t)}$  and  $T_r^{(t)}$ , are the supply and return temperature of the ventilation air.





**Figure 3.** Pre-processed hourly observations obtained from the building monitoring system (BMS). The left column shows the entire measurement period, whereas the right column shows one week of data obtained from Monday, 4 November, to Sunday, 10 November, 2019.

The resulting heating power includes the supplied heating and cooling as well as any heat recovered by the heat recovery unit and other heat gains such as heat gain from the ventilators.

In Figure 3, the hourly supply and return temperature of the ventilation airflow are shown, as well as the resulting net heat supply.

### 3.2.3. Total Ventilation Rate

The total ventilation rate is the sum of the mechanical and natural ventilation rates. Assuming that all windows and doors are closed during unoccupied hours, mechanical ventilation is the only potential ventilation during unoccupied hours. During the occupied hours, it is assumed that the total ventilation rate is a combination of mechanical and natural ventilation rates. However, as we only have measurements of the mechanical ventilation, the total ventilation rate must be deducted from an estimate of the CO<sub>2</sub> production from occupants during occupied hours and the outdoor CO<sub>2</sub> concentration.

Several studies have utilised the tracer gas methods for evaluating the ventilation and infiltration rate of buildings (e.g., [31,32]). The tracer gas method utilises the basic principle of mass conservation and is expressed by the continuity equation [33,34]. From monitoring the injection and concentration of tracer gas, the air exchange can be inferred. Sherman has shown several tracer gas techniques for measuring ventilation, including a steady-state method analogue to the method used here [33].

The outdoor ambient CO<sub>2</sub> concentration,  $c_a$ , is approximated by the lower 1%-tile of the measured indoor CO<sub>2</sub> concentrations. It is found to be 430 ppm as shown in Figure 3.

The steady-state CO<sub>2</sub> production from occupants can be calculated as the difference in CO<sub>2</sub> concentration inside and outside multiplied by the ventilation rate. With this, we introduced the reference CO<sub>2</sub> production,  $G_{ref}$ , defined as the average daily CO<sub>2</sub> production during the occupied hours before the COVID-19 lockdown. The occupied hours are assumed to be equal to the hours without recirculation ( $R^{(t)} = 0$ ) and outside weekends, vacations, and the COVID-19 lockdown.

The combined mechanical and natural ventilation rate,  $Q_{total}^{(t)}$ , in the occupied hours can now be approximated by the ventilation rate it requires to maintain the measured average CO<sub>2</sub> concentration during the given day, assuming that the production equals  $G_{ref}$ . For unoccupied hours it is assumed that the natural ventilation rate and the infiltration are negligible, and the ventilation rate during these hours equals the mechanical ventilation rate. Introducing  $\bar{c}$  as the daily average CO<sub>2</sub> concentration only accounting for the occupied hours, we calculated the total ventilation rate as

$$Q_{total}^{(t)} = \begin{cases} G_{ref} (\bar{c}^{(t)} - c_a)^{-1}, & \text{for occupied hours} \\ Q_{mech}^{(t)}, & \text{else.} \end{cases} \quad (19)$$

where the mechanical ventilation rate,  $Q_{mech}^{(t)}$ —i.e., the mechanically forced air exchange between inside and outside, not to be confused with the airflow provided by the air handling unit—is determined by the maximum of the supply and return airflow within each hour, with the addition that the flow is zero for hours with recirculation (i.e., for  $R^{(t)} = 1$ ).

$$Q_{mech}^{(t)} = \begin{cases} \max[Q_s^{(t)}, Q_r^{(t)}], & \text{for } R^{(t)} = 0 \\ 0, & \text{else,} \end{cases} \quad (20)$$

Notice that the increments of  $t$  are 1 h, and the daily average concentration  $\bar{c}$  in Equation (19) remains constant for every hour within the same day (see Figure 3).

It should be noted that this calculation does not explicitly include infiltration. The infiltration is, however, included implicitly for occupied hours as the effect of the infiltration is reflected in the measured CO<sub>2</sub> concentration. The infiltration is, however, neglected for

unoccupied hours, which might introduce some uncertainty. Despite this crude approximation,  $Q_{\text{total}}^{(t)}$  gives us an estimation of what the combined mechanical and natural ventilation rate is during occupied hours.

### 3.2.4. Natural Ventilation Rate

As stated above, the total ventilation rate is the sum of the mechanical,  $Q_{\text{mech}}^{(t)}$ , and natural ventilation rates,  $Q_{\text{nat}}^{(t)}$ . The natural ventilation rate can therefore be approximated by

$$Q_{\text{nat}}^{(t)} = Q_{\text{total}}^{(t)} - Q_{\text{mech}}^{(t)} \quad (21)$$

### 3.2.5. Ventilation Heat Loss

During periods without recirculation, the airflow passing the heat exchanger equals the return airflow. For periods where the ventilation rate is unbalanced ( $Q_r^{(t)} \neq Q_s^{(t)}$ ), the difference between the air flows  $Q_r^{(t)}$  and  $Q_s^{(t)}$  is assumed to be extracted through the hybrid ventilation exhaust louvre or forced through the building envelope. Finally, for any period with recirculation, both the airflow through the heat exchanger and the building envelope is assumed to be zero.

The specific mechanical ventilation heat loss, expressed in watts per kelvin, is calculated as

$$\Phi_{\text{vmech}}^{(t)} = \begin{cases} c_p \rho (Q_{\text{hr}}^{(t)} (1 - \eta^{(t)}) + Q_{\text{env}}^{(t)}) , & \text{for } R^{(t)} = 0 \\ 0 , & \text{else,} \end{cases} \quad (22)$$

where  $Q_{\text{hr}}^{(t)}$  is the airflow passing through the heat exchanger,  $\eta^{(t)}$  is the heat recovery rate between zero and one, and  $Q_{\text{env}}^{(t)}$  is the airflow entering or leaving the building through the envelope. That being

$$Q_{\text{env}}^{(t)} = |Q_s^{(t)} - Q_r^{(t)}|. \quad (23)$$

The specific natural ventilation heat loss is calculated as

$$\Phi_{\text{vnat}}^{(t)} = c_p \rho Q_{\text{nat}}^{(t)}, \quad (24)$$

and the total ventilation heat loss is found by adding the mechanical and natural ventilation loss.

$$\Phi_{\text{vtotal}}^{(t)} = \Phi_{\text{vmech}}^{(t)} + \Phi_{\text{vnat}}^{(t)}, \quad (25)$$

### 3.2.6. Down-Sampling

Before the data are used in the algorithms presented in Section 2, the hourly data are down-sampled to daily values. The down-sampling ensures that the majority of the thermal dynamics (i.e., changing and discharging of the thermal mass) are filtered out [1,35], and the underlying steady-state model assumption is not violated.

The down-sampling also filters out the intraday effects caused by human interaction with the building. Consequently, the emphasis is on the time-varying daily average human effects.

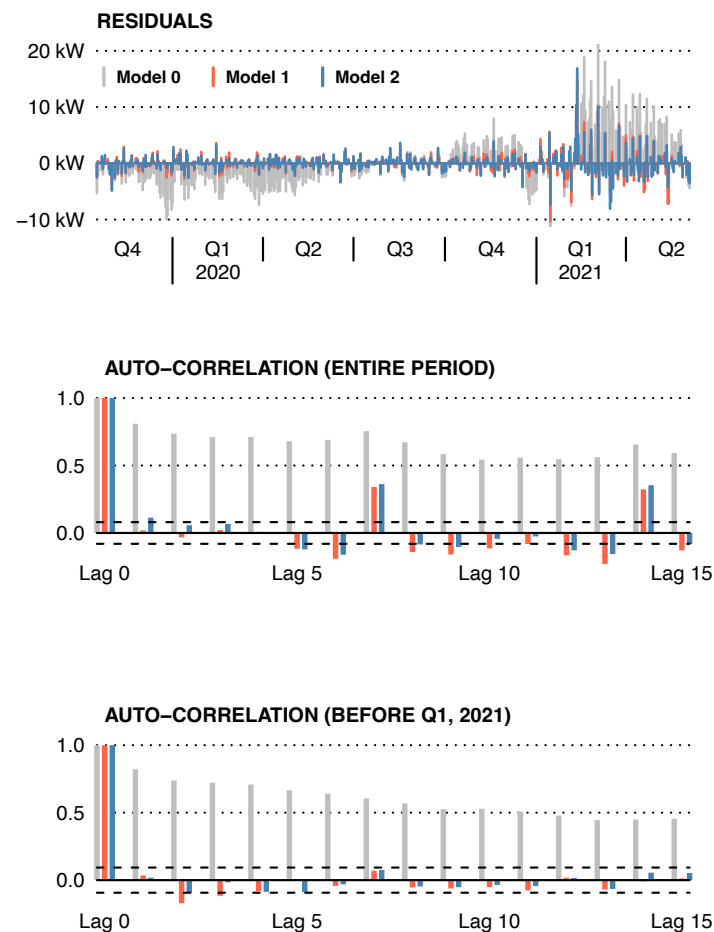
## 4. Results

The results section is divided into three sections. First, it is shown through residual analyses that both bias and variance in the prediction errors are reduced significantly by introducing a random state. Second, we investigate the consistency of the parameter estimates across the three models. Third, the interpretability of the estimated random states is investigated.

#### 4.1. Residual Analyses

The inclusion of a hidden state describing the evolution of the effective heat loss coefficient,  $UA^{(t)}$ , and the base temperature,  $T_b^{(t)}$ , has in both cases reduced the systematic errors compared to the static baseline model in Equation (11).

In the residuals plot in Figure 4 major systematic errors are seen for the static model. For the time-varying models, the residuals resemble white noise during the first five quarters. In the first and second quarters of 2021, all models start to show systematic errors. The errors are, however, far from as systematic as for the static model.

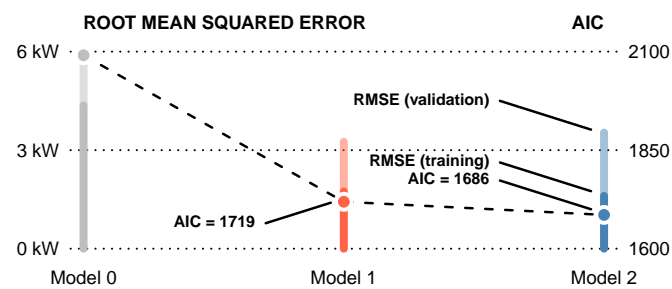


**Figure 4.** Residuals of static and time-varying models (**top**), auto-correlation functions of entire data period (**mid**), and first five quarters of data period (**bottom**).

In the auto-correlation plot for the entire period in Figure 4 (mid), the static baseline model shows significant systematic errors for all of the first 15 lags. For the time-varying models, the auto-correlations are reduced but not eliminated. Specifically, high auto-correlations are found for lag 7 and 14, indicating that some weekly dynamics are not modelled. Looking at the auto-correlation for the first five quarters alone (Figure 4 (bottom)), the time-varying models barely show any signs of systematic errors.

By introducing a state formulated as a random walk to a model, it is expected that the systematic errors will reduce as the degrees of freedom increase. The remaining errors in the last two quarters might indicate that the operation of the building has changed to the extent that a single simple random walk cannot describe.

In Figure 5 the model accuracy for the three models is shown. The dark vertical bars indicate the root mean squared error (RMSE) on the training set, while the bright vertical bars indicate the RMSE on the validation set. The increase in accuracy of the time-varying models remains low even on the validation set.



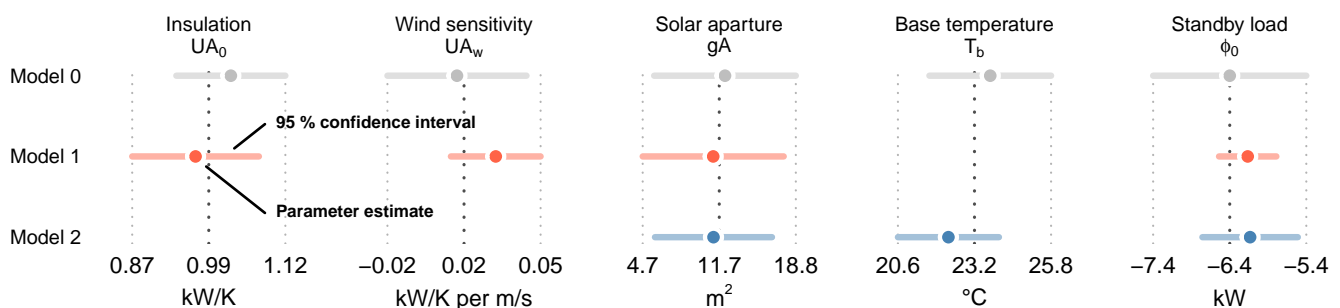
**Figure 5.** Root mean squared error and Akaike information criterion (AIC) (lower is better) of Model 0, 1 and 2. The dark and bright vertical bars indicated the root mean squared error of the training and validation data, respectively.

The coloured circles connected with a black dashed line show the Akaike information criterion (AIC). The AIC is a result of the model likelihood penalised by the number of model parameters. Additionally, here, Models 1 and 2 are in favour of the static model (Model 0).

Both the auto-correlation function and the model accuracy show that the inclusion of a hidden state is a significant improvement over the static baseline model. With the time-varying models, we have, therefore, not only reduced the accuracy, but also the generalisability.

#### 4.2. Parameters Estimates

For each of the three models, a number of parameters are estimated. In Figure 6, the global parameters related to the building's physical properties are shown. Each parameter estimate is indicated by a dot, and their associated 95% confidence intervals are indicated by the horizontal lines to the right and left of the dots.



**Figure 6.** Estimated physical building parameters by Models 0, 1, and 2. The dots indicate the estimate, and the horizontal bars indicate the 95% confidence intervals. For Model 1, the base temperature ( $T_b$ ) is modelled as a random walk and is therefore not shown here. For Model 2, the insulation and wind sensitivity ( $UA_0$  and  $UA_w$ ) are omitted for the same reason.

Across all models, the parameter estimates are consistent, and no model parameters diverge significantly from the others. Comparing the estimated  $UA_0$  values in Figure 6 with the theoretical heat loss coefficient multiplied by the net floor area in Table 1, we see that the estimated  $UA_0$  value is between 0.9 and 1.1 kW/K, while the  $UA$  values obtained from theoretical calculations is 0.7 kW/K. The exact reason for the discrepancy is unknown but is partly related to the fact that the estimated  $UA_0$  value contained the average ventilation heat loss.

#### 4.3. Hidden State Estimates

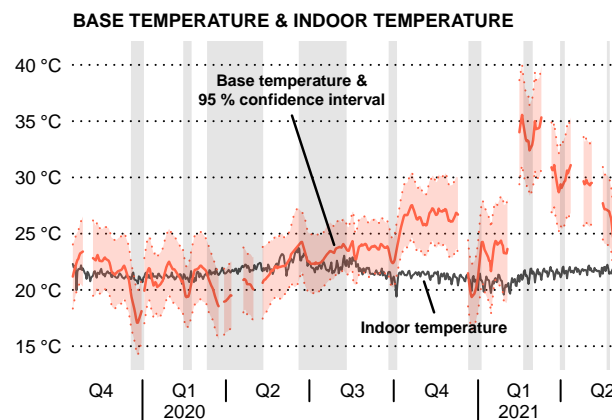
While the parameters estimated across all models are consistent, and the residual analyses in Section 4.1 show that the hidden state models are in favour of the static models, the choice between the two models with either a time-varying heat loss coefficient or base

temperature remains inconclusive as they tend to perform equally well. In the following, an attempt to validate and interpret the hidden states of the models in Equations (12a), (12b), (13a) and (13b) are made. Since the hidden states are truly hidden and, by nature, hard to observe, the validation is done in terms of indirect measures.

#### 4.3.1. Time-Varying Base Temperature

The model in Equations (12a) and (12b) with the implementation of a time-varying estimate of the base temperature,  $T_b^{(t)}$ , has shown significant reductions in both bias and variance.

In Figure 7, the estimated state of  $T_b^{(t)}$  is shown alongside the observed mean indoor temperatures, which do not show many similarities. The estimated state does, however, show drops in the base temperature for longer unoccupied periods (indicated by the grey regions in the figure), which means that the net heat gains increase—or rather, the net heat loss reduces by means of, e.g., reduced ventilation rates—during these periods. Typically, a drop in the base temperature could also indicate a decrease in the indoor temperature. This is, however, not the case as it is seen from the indoor temperature shown in Figure 7.



**Figure 7.** Estimated base temperature and average measured indoor air temperature (Model 1). Grey regions indicate vacations and the COVID-19 lockdown, and the gaps in the red line correspond to the test date blocks.

As stated earlier, the base temperature,  $T_b^{(t)}$ , is the outdoor temperature at which the building is in thermal balance by passive means. The base temperature comprises the indoor temperature and the unknown contributions to the heat balance, such as internal heat gains and ventilation losses. Regardless of what the unknown contributions to  $T_b^{(t)}$  are, it is unreasonable to believe that the building is in thermal balance, by passive means only, at outdoor temperatures of up to 35 °C as the estimated state of  $T_b^{(t)}$  indicates in Figure 7.

The high estimates of the base temperature might be a result of the chosen model structure. In essence, the state of  $T_b^{(t)}$  might very likely capture the changing ventilation heat loss by altering the base temperature and thus lose its direct interpretability.

#### 4.3.2. Time-Varying Heat Loss Coefficient

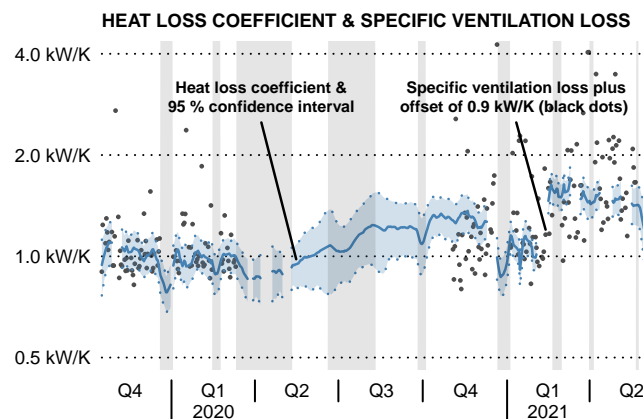
As was the case for the previous model, the model in Equations (13a) and (13b) with a time-varying estimate of the heat loss coefficient,  $UA^{(t)}$ , improved in terms of bias and variance compared to the baseline model with globally estimated model parameters.

The hypothesis of implementing the  $UA^{(t)}$  state is that it will describe the building's thermal insulation level and any changes in the ventilation heat loss. The reason is that a change in the ventilation flow rate effectively means that the heat loss coefficient is changing as outlined in Section 2.2.

To evaluate if the estimated state of  $UA^{(t)}$  replicates the ventilation heat loss, the specific ventilation loss (i.e., the ventilation loss in watts per kelvin) is calculated based on



the estimated total ventilation loss in Section 3.2.3. As the estimated heat loss coefficient,  $UA^{(t)}$  (blue line in Figure 8), includes both the building envelope's thermal resistance and ventilation losses, the estimated specific ventilation loss (black dots) is offset by a constant of 0.9 kW/K (found by optimisation) such that it can be compared to the state of  $UA^{(t)}$ .



**Figure 8.** Estimated heat loss coefficient and specific ventilation loss corresponding to the ventilation rate required to maintain observed indoor CO<sub>2</sub> concentrations (plus a constant) (Model 2). Grey regions indicate vacations and the COVID-19 lockdown, and the gaps in the blue line correspond to the test date blocks. The specific ventilation loss (black dots) is offset by 0.9 kW/K to compensate for the fact that the estimated heat loss coefficient (blue line) is the combined effect of thermal insulation and ventilation loss. Notice that the y-axis is logarithmic.

Despite the high variance of the estimated specific ventilation loss (black dots in Figure 8), it shares many characteristics of the estimated heat loss coefficient (blue line). Before the COVID-19 lockdown, starting at the end of the first quarter of 2020, the estimated heat loss coefficient does not deviate significantly from a value of around 1.0 kW/K, except for the Christmas holidays and winter vacation. Likewise, the specific ventilation heat loss is distributed around the same level. The distinct drops during vacations and the COVID-19 lockdown possibly indicate a lower level of ventilation.

In the fourth quarter of 2020, a slight increase in the specific ventilation heat loss and a significant increase in the estimated heat loss coefficient is seen. However, for the following two last quarters, the pattern seems to follow the trend to a greater extent.

During the warmest period around the summer of 2022, it has not been possible to estimate the specific ventilation heat loss reliably, and it can therefore not serve as validation.

To quantify if the state of  $UA^{(t)}$  is a result of pure chance, 10,000 realisations of the estimated random walk with an incremental variance of  $0.0024 \text{ (kW/K)}^2$  and initial condition of 1.0 kW/K were simulated. Of these, only around 2% of the realisations give rise to a better fit in terms of RMSE to the specific ventilation loss (black dots) in Figure 8. So, given the estimated parameters of the random walk in Equation (13b), there is a probability of less than 2% that the estimated state of  $UA^{(t)}$  resembles the specific ventilation loss as good or better by pure chance. Additionally, a constant heat loss coefficient will, at best, result in a fit to the specific ventilation loss, which is 85% as good in terms of RMSE.

Conclusively, the random walk presumably reveals the evolution of the total ventilation rate while performing better than two natural benchmarks—the realisations of the random walk itself and a constant UA value.

## 5. Discussion

The classical methods for estimating building physical parameters, such as the energy signature found in [1,2], as well as more recent developments of the same method [22,35,36], omit the fact that human interactions with the building can have significant effects on the observed energy consumption [16–18]. Instead, they estimate a global set of parameters for the entire data set, which results in a 'best-on-average' estimate, and, therefore, does

not reflect the potential variations in the heat consumption caused by changing temperature setpoints, ventilation rates, etc. Even though static models—such as Model 0 in Equation (11)—do not explain these human effects, no apparent variations in the estimated building physical parameters are evident from the three models investigated (see Figure 6). It should, however, be noted that the static model produces highly auto-correlated prediction errors (see Figure 4), which violates the underlying model assumptions, and proves that more advanced techniques like those in Model 1 and 2 are needed to accommodate humans' interactions with the building.

According to Hammarsten and Eriksson et al. [1,35], the dynamics due to charging and discharging of the thermal mass are negligible for daily average values which are used in these models. Consequently, it appears reasonable to assume that the auto-correlation is most likely a result of human interactions (i.e., changing ventilation rate).

Model 1 showed high accuracy and low auto-correlation in the residuals relative to Model 0. Despite that, Model 1 gave a hardly interpretable estimate of the time-varying base temperature. Reformulating the model with a constant base temperature and a time-varying total heat loss coefficient (Model 2), the estimated time-varying heat loss coefficient showed a high resemblance with the combined heat loss through the building envelope and the ventilation heat loss. From the results in Figure 8 it is shown that the total heat loss has increased by approximately 50% from the five months before the COVID-19 lockdown in March 2020 to late winter of 2021. Based on all quantitative evidence—such as a constant indoor temperature, measured mechanical ventilation rates, measured CO<sub>2</sub> concentrations and from that, the deducted natural ventilation rates—the changes are very likely a result of human interactions with the building (i.e., increased mechanical and natural ventilation). For the specific case study, a natural explanation might be an increased awareness of air hygiene as a result of the prevailing COVID-19 pandemic. To back up this reasoning, Navas-Martín and Cuervo-Vilches have concluded from surveys from Spanish homes that that people have increased the level of natural ventilation during the COVID-19 pandemic compared to pre-pandemic [37]. The conjecture that the users of the building have increased the natural ventilation rate during the COVID-19 pandemic has not been confirmed by the staff (one person was asked), but might be a result of unconscious actions.

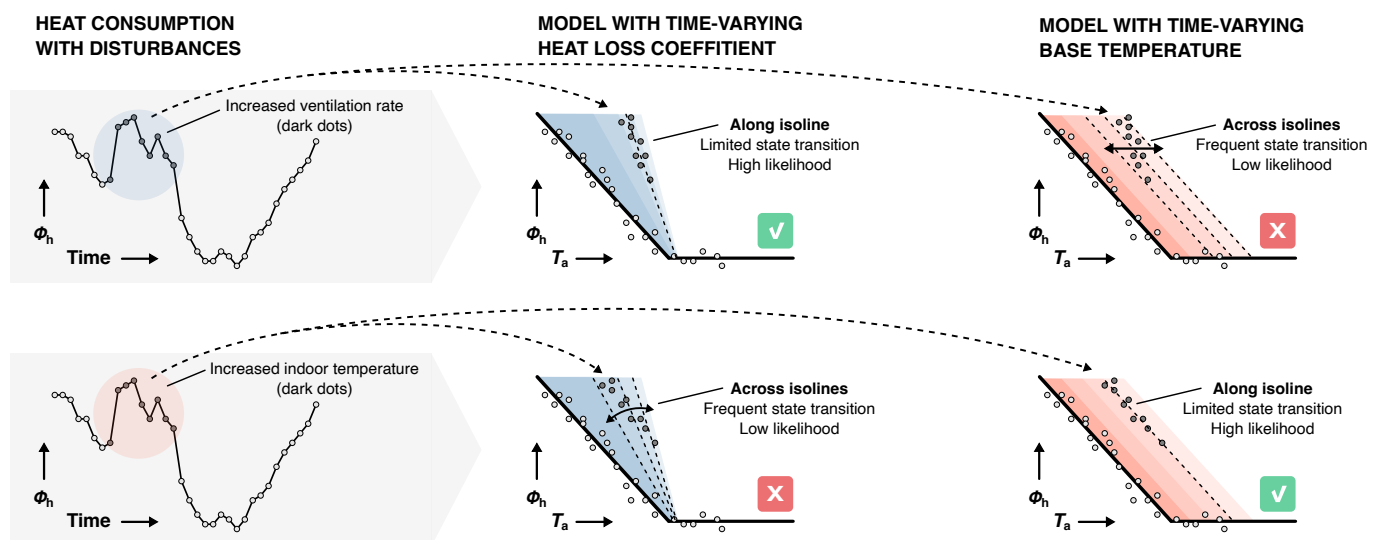
### 5.1. Choosing the Appropriate Model Structure

The choice of which model parameter to describe by a time-varying state ( $UA^{(t)}$  or  $T_b^{(t)}$ ) is not trivial, as one model is a transformation of the other. Specifically, this means that prediction errors and accuracy are similar for the two suggested models, but the interpretability differs.

A general recommendation is, therefore, that the highest level of interpretability is achieved by selecting the model as a result of which dominating driving forces that make the heat consumption deviate from the building's average heat consumption. If the dominating force is changing ventilation rates—as was the case in this study—Model 2 in Equations (13a) and (13b) with time-varying heat loss coefficient is recommended. If, on the other hand, the dominating forces are changing indoor temperatures and internal heat gains, Model 1 in Equations (12a) and (12b) is recommended. Recommendations like these are, however, not directly applicable as they require additional data sources to determine what the ventilation rate, indoor temperature or internal heat gains are. One alternative is to base the decision on a preliminary data analysis. Following the conceptual illustration of the effects on the heat consumption in Figure 1, a constant variance of the heat consumption for the heating period indicates that the main driving forces for the variation are related to changing indoor temperatures and internal heat gains as these typically are independent of the outdoor temperature (see sub-plot of temperature and internal gains). If the variance increases with decreasing outdoor temperature, the plausible dominating driving force causing the variation is the ventilation rate, as this effectively alters the heat loss coefficient (see Figure 1, sub-plot of ventilation).

Taking the arguments for how to choose between the time-varying heat loss coefficient or base temperature one step further, one can lean towards Occam's razor. Occam's razor states that *plurality should not be posited without necessity*, or put in simple terms that *of two competing theories, the simpler explanation of an entity is to be preferred* [38]. This principle is analogue to the likelihood, which is the underlying objective of the model estimation.

In Figure 9, two identical time series of observed heat consumption are shown. In both cases, nine of the observations are assigned with an—for the observer—unknown disturbance. In the first row of sub-plots in Figure 9, the disturbance is caused by an increased ventilation rate and hence an increased heat consumption (dark dots). In the second row of the figure, the disturbance is caused by an increased indoor temperature and hence an increased heat consumption (dark dots). In both cases, the modeller can choose to model the heat consumption by including either a time-varying heat loss coefficient (second column of sub-plots in Figure 9) or a time-varying base temperature (third column of sub-plots). In the scenario where the disturbances are caused by an increased ventilation rate (first row of sub-plots), the model with a time-varying heat loss coefficient is capable of explaining the disturbance by a single increment in the ventilation rate (i.e., a one-time change in the heat loss coefficient as shown in the first row, second column of the sub-plots in Figure 9). On the other hand, choosing the model with a time-varying base temperature (first row, third column of the sub-plots in Figure 9), the disturbances can only be explained equally well by several subsequent changes in the base temperature.



**Figure 9.** Conceptual illustration of how two fundamental types of disturbances (changing ventilation rate and indoor temperature or internal heat gain) affect the energy signature, and consequently how they dictate which model to use.

To summarise, for the first scenario (first row of sub-plot in Figure 9), we can either explain the disturbances by a single state transition related to the ventilation rate (i.e., a single state transition of the heat loss coefficient), or as several sequential state transitions of the base temperature. Following Occam's razor, we chose the model with the least state transitions (i.e., the model that can explain the data in the simplest possible manner) and followed by that, the model with the highest likelihood—or, analogue to that, the lowest AIC as we see in Figure 5.

In case the disturbances are caused by an increment in the indoor temperature and/or internal heat gain, the model with time-varying base temperature is preferred, as shown in the second row of sub-plot in Figure 9.

## 5.2. Limitations

The validation of the models involved a comparison between the estimated total heat loss coefficient and the total ventilation rate. As a consequence of the fact that we did not have data on the natural ventilation rate, this was estimated from the measured CO<sub>2</sub> concentrations and an assumed constant CO<sub>2</sub> production throughout the data period. This is, as stated earlier, a crude estimate which might affect the validations of the models.

The method presented utilises daily average values for heat consumption and weather conditions; this will naturally omit the models' capabilities to describe intraday variations of human interactions. To reach that level, high-resolution data are needed (e.g., hourly resolution), and the model needs to be reformulated as a dynamical model which can describe the charging and discharging of the thermal mass.

In Figure 6, the physical parameters estimates and their 95% confidence intervals were presented. Although the  $UA_0$ ,  $T_b$ , and  $\Phi_0$  have a relative uncertainty of 10 to 15%, the  $UA_w$  and  $gA$  have an uncertainty of around 50%. For the  $UA_w$  and  $gA$ , the relatively high uncertainty might be related to the fact that the building might be more wind sensitive from certain angles of attack and receives higher levels of solar gain for certain sun positions. Describing such effects by simple constants might give rise to higher uncertainty in the parameter estimates. One solution might be to describe the solar gain by a non-parametric function—such as splines—of the sun's position. One example of this can be found in [39], and a similar approach could be used for wind direction-dependent wind sensitivity.

The models demonstrated only handle one type of human-driven interaction each—either changing internal heat gain and indoor temperature or changing ventilation rates. A limitation of the models is that both types of human interactions can (and in many cases will) occur over a period of time. Further research is needed to develop methods which can handle both types of interactions.

In this study, part of the heat consumption is assigned to the building properties, and part of it is assigned to the human interactions with the building. The human interactions are understood in a broad sense, meaning that they can be related to occupants' behaviour and operational staff's changes to the mechanical systems (e.g., increasing the ventilation rate). A third reason for the changes which might be captured by the estimated time-varying states is system faults causing changes in systems efficiencies. This has not been the scope of this study but is naturally a limitation of the methods in terms of determining the cause of the estimated variations.

## 6. Conclusions

The study has shown, first, that disturbances in a building's heat consumption caused by human interactions can be deducted by time-varying parameter estimates with high confidence. Second, for the demonstration of the method on data from a school south of Oslo, Norway, the results have shown that the total heat loss, including mechanical and natural ventilation, has increased by approximately 50% over the course of one year dominated by the COVID-19 pandemic.

We have shown that two different advanced energy signature model representations, using either a time-varying base temperature or heat loss coefficient, fit and predict with almost identical accuracy. Moreover, the inclusion of a time-varying hidden state has been shown to improve in terms of bias and variance compared to the static energy signature model (Model 0) found in previous studies. However, the interpretation of the two time-varying models has shown to differ significantly. To address this, two methods for selecting the most interpretable model have been proposed. Models can be selected either by a graphical inspection of the data before selecting the model structure or by making the model selection based on likelihood (with emphasis on the likelihood of the hidden state transition) rather than the prediction errors on the validation data (e.g., root mean squared error).

The method has shown promising results on real data, although better data for validation would have strengthened the support of the method. The work opens up the possibility for additional research in the proposed or similar methods. Solid continuous research on methods like the proposed will make it possible to separate the effects of human interactions with buildings from effects caused by the weather conditions and the building's physical properties. Ultimately this might enable better documentation of existing buildings' energy performance, such as energy performance certificates.

To support further research and development of the proposed method, the following future work could be considered.

1. Test the proposed model on data obtained from a building with ground truth data on total ventilation rate, indoor temperature and internal heat gains.
2. Combining the effects of time-varying heat loss coefficient and base temperature into one single model.
3. Formulate a dynamic model to capture intraday variations and estimate time constants.
4. Assess the models' capabilities of medium to long-term probabilistic simulation and forecasting of heat consumption.

**Author Contributions:** Conceptualization, C.R.; methodology, C.R., N.L. and T.H.D.; software, C.R.; validation, C.R., N.L., P.B., T.H.D. and H.M.; formal analysis, C.R.; investigation, C.R., N.L., P.B., T.H.D. and H.M.; resources, T.H.D. and H.M.; data curation, C.R., N.L. and T.H.D.; writing—original draft preparation, C.R.; writing—review and editing, C.R., N.L., P.B., T.H.D. and H.M.; visualization, C.R.; supervision, T.H.D. and H.M.; project administration, T.H.D.; funding acquisition, T.H.D. and H.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the project 'SmartTune' founded by The Research Council of Norway [296345]; the project Smart Cities Accelerator+ (SCA+) founded by The European Regional Development Fund, Interreg Öresund-Kattegat-Skagerrak (European Union) [20293290]; and the project 'Investigation of savings potential of better installation and continued adjustment of heat pumps, including better financing of heat pumps for Danish homeowners' founded by Elforsk [353-026].

**Data Availability Statement:** Restrictions apply to the availability of these data. Data was obtained from Piscada AS. Data availability for other than the authors is unknown.

**Conflicts of Interest:** The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results

## Nomenclature

All symbols are specified in the text immediately after their first appearance. Some general conventions are used to distinguish model parameters (constants) and variables. Model parameters are written in upright letters (e.g.,  $g_A$  which is the effective solar transmittance), while variables such as model output and inputs are written with italic letters (e.g.,  $W_s$  denoting the wind speed).

### Mathematical operators and notation

<b>E</b>	Expectation
<b>Var</b>	Variance
<b>LSE</b>	The smooth maximum function, LogSumExp
$N(0, \sigma^2)$	Normal distribution with mean zero and variance $\sigma^2$
$\mathbf{Z}^+$	Set of positive integers (1, 2, 3, ...)
$\in$	Denotes set membership, and is read "in" or "belongs to"

### Superscripts

$(t)$	Discrete time index
$(i)$	Discrete index

*Physical quantities*

$\Phi_h$	Heating effect
$\Phi_x$	Secondary net heat gain. Sum of metabolism, ventilation loss, etc.
$\Phi_{v_{\text{mech}}}$	Specific mechanical ventilation loss
$\Phi_{v_{\text{nat}}}$	Specific natural ventilation loss
$\Phi_{v_{\text{total}}}$	Specific total ventilation loss
$\eta$	Sensible heat recover of ventilation
$\rho$	Density of air
$\bar{c}$	Daily average CO <sub>2</sub> concentration for occupied hours
$c_a$	Outdoor ambient CO <sub>2</sub> concentration
$c_p$	Specific heat capacity of air
$G_{\text{ref}}$	Reference CO <sub>2</sub> production only accounting for occupied hours
$I_g$	Global solar irradiation
$Q_s, Q_r$	Mechanical volumetric ventilation supply/return airflow
$Q_{\text{total}}$	Total natural and mechanical volumetric ventilation airflow
$Q_{\text{mech}}$	Mechanical volumetric ventilation airflow
$Q_{\text{nat}}$	Natural volumetric ventilation airflow
$Q_{\text{env}}$	Forced volumetric ventilation airflow through building envelope
$Q_{\text{hr}}$	Volumetric ventilation airflow over heat recovery unit
$R$	State of recirculation (binary)
$T_a$	Outdoor ambient air temperature
$T_i$	Indoor air temperature
$T_s, T_r$	Ventilation supply/return temperature
$\Delta T$	Temperature difference over heating/cooling coil in ventilation system
$W_s$	Wind speed

*Model parameters*

$\varepsilon, \zeta$	Noise terms
$\Phi_0$	Background heating effect
$g_A$	Effective solar transmittance
$k$	Transition rate of LogSumExp function (LSE)
$T_b$	Base temperature
$UA_0$	Heat loss coefficient under wind-free conditions (i.e., $W_s = 0$ )
$UA_w$	Wind dependency
$T_b^{(t)}$	Time-varying base temperature, i.e., base temperature at time $t$
$UA^{(t)}$	Time-varying heat loss coefficient (including wind effects), i.e., heat loss coefficient at time $t$

**References**

1. Hammarsten, S. A critical appraisal of energy-signature models. *Appl. Energy* **1987**, *26*, 97–110. [\[CrossRef\]](#)
2. Hammarsten, S. *Estimation of Energy Balances for Houses*; National Swedish Institute for Building Research: Gävle, Sweden, 1984.
3. Dick, J.B.; Thomas, D. Ventilation Research in Occupied Houses. *J. Inst. Heat. Vent. Eng.* **1951**, *19*, 306–326.
4. Fels, M.F. PRISM: An introduction. *Energy Build.* **1986**, *9*, 5–18. [\[CrossRef\]](#)
5. Bacher, P.; Madsen, H. Identifying suitable models for the heat dynamics of buildings. *Energy Build.* **2011**, *43*, 1511–1522. [\[CrossRef\]](#)
6. Madsen, H.; Bacher, P.; Bauwens, G.; Deconinck, A.H.; Reynders, G.; Roels, S.; Himpe, E.; Lethé, G. *IEA-EBC Annex 58—Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements. Report of Subtask 3, Part 2: Thermal Performance Characterisation Using Time Series Data—Statistical Guidelines*; Technical report; The International Energy Agency: Paris, France, 2016.
7. Jiménez, M.J.; Madsen, H.; Andersen, K.K. Identification of the Main Thermal Characteristics of Building Components Using MATLAB. *Build. Environ.* **2008**, *43*, 170–180. [\[CrossRef\]](#)
8. Bauwens, G.; Ritoša, K.; Roels, S. *IEA EBC Annex 71—Building Energy Performance Assessment Based on In-Situ Measurements: Physical Parameter Identification*; Technical report; The International Energy Agency: Paris, France, 2021.



9. Junker, R.G.; Azar, A.G.; Lopes, R.A.; Lindberg, K.B.; Reynders, G.; Relan, R.; Madsen, H. Characterizing the Energy Flexibility of Buildings and Districts. *Appl. Energy* **2018**, *225*, 175–182. [CrossRef]
10. Wolf, S.; Møller, J.K.; Bitsch, M.A.; Krogstie, J.; Madsen, H. A Markov-Switching model for building occupant activity estimation. *Energy Build.* **2019**, *183*, 672–683. [CrossRef]
11. Sangoboye, F.C.; Kjærgaard, M.B. Plcount: A probabilistic fusion algorithm for accurately estimating occupancy from 3d camera counts. In Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, Palo Alto, CA, USA, 16–17 November 2016; pp. 147–156.
12. Yoshino, H.; Hong, T.; Nord, N. IEA EBC Annex 53: Total energy use in buildings—Analysis and evaluation methods. *Energy Build.* **2017**, *152*, 124–136. [CrossRef]
13. Yu, Z.; Fung, B.C.; Haghighat, F.; Yoshino, H.; Morofsky, E. A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy Build.* **2011**, *43*, 1409–1417. [CrossRef]
14. Fabi, V.; Buso, T.; Andersen, R.K.; Corgnati, S.P.; Olesen, B.W. Robustness of building design with respect to energy related occupant behaviour. In Proceedings of the 13th Conference of International Building Performance Simulation Association (IBPSA), Chambery, France, 25–28 August 2013.
15. Schweiker, M.; Shukuya, M. Comparative effects of building envelope improvements and occupant behavioural changes on the energy consumption for heating and cooling. *Energy Policy* **2010**, *38*, 2976–2986. [CrossRef]
16. Haldi, F.; Robinson, D. The impact of occupants' behaviour on building energy demand. *J. Build. Perform. Simul.* **2011**, *4*, 323–338. [CrossRef]
17. Brohus, H.; Heiselberg, P.; Simonsen, A.; Sørensen, K.C. *Influence of Occupants' Behaviour on the Energy Consumption of Domestic Buildings*; Aalborg University: Aalborg East, Denmark, 2010.
18. Socolow, R.H. The Twin Rivers Program on Energy Conservation in Housing: Highlights and Conclusions\*. *Energy Build.* **1977**, *1*, 207–242. [CrossRef]
19. Wingfield, J.; Miles-Shenton, D.; Bell, M. *Comparison of Measured Versus Predicted Heat Loss for New Build UK Dwellings*; Unpublished Data; Leeds Metropolitan University: Leeds, UK, 2011.
20. The Danish Energy Agency. *Status for Energimærkningsordning for Bygninger*; Technical report; Energistyrelsen (The Danish Energy Agency): Copenhagen, Denmark, 2018.
21. Fleck, A. Energy Crisis—Energy Prices Keep Climbing in the EU. 2022. Available online: <https://www.statista.com/chart/28004/eu-energy-inflation-rates/> (accessed on 11 October 2022).
22. Rasmussen, C.; Bacher, P.; Cali, D.; Nielsen, H.A.; Madsen, H. Method for Scalable and Automatised Thermal Building Performance Documentation and Screening. *Energies* **2020**, *13*, 3866. [CrossRef]
23. Rasmussen, C.; Hviid, C.A.; Bacher, P.; Cali, D.; Madsen, H. Estimating Building Airtightness from Data—A Case Study. In Proceedings of the Cold Climate HVAC & Energy 2021, Tallinn, Estonia, 18–21 April 2021; Volume 246.
24. Møller, J.K. Stochastic State Space Modelling of Nonlinear systems—With application to Marine Ecosystems. Ph.D. Thesis IMM-PHD-2010 No. 246, Technical University of Denmark: Lyngby, Denmark, 2011.
25. Kristensen, N.R.; Madsen, H.; Jørgensen, S.B. Parameter estimation in stochastic grey-box models. *Automatica* **2004**, *40*, 225–237. [CrossRef]
26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
27. Kristensen, K.; Nielsen, A.; Berg, C.W.; Skaug, H.; Bell, B.M. TMB: Automatic Differentiation and Laplace Approximation. *J. Stat. Softw.* **2016**, *70*, 1–21. [CrossRef]
28. Thygesen, U.H.; Albertsen, C.M.; Berg, C.W.; Kristensen, K.; Nielsen, A. Validation of ecological state space models using the Laplace approximation. *Environ. Ecol. Stat.* **2017**, *24*, 317–339. [CrossRef]
29. Copernicus Climate Data Store. Available online: <https://cds.climate.copernicus.eu> (accessed on 14 March 2022).
30. ERA5-Land Hourly Data from 1950 to Present. 2022. Available online: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview> (accessed on 14 March 2022).
31. Duarte, R.; Gomes, M.; Rodrigues, A. Estimating ventilation rates in a window-aided room using Kalman filtering and considering uncertain measurements of occupancy and CO<sub>2</sub> concentration. *Build. Environ.* **2018**, *143*, 691–700.
32. Penman, J. An experimental determination of ventilation rate in occupied rooms using atmospheric carbon dioxide concentration. *Build. Environ.* **1980**, *15*, 45–47. [CrossRef]
33. Sherman, M. Tracer-gas techniques for measuring ventilation in a single zone. *Build. Environ.* **1990**, *25*, 365–374. [CrossRef]
34. Awbi, H. *Ventilation of Buildings*; Spon Press: London, UK, 2005.
35. Eriksson, M.; Akander, J.; Moshfegh, B. Development and validation of energy signature method Case study on a multi-family building in Sweden before and after deep renovation. *Energy Build.* **2020**, *210*, 109756. [CrossRef]
36. Rohdin, P.; Milic, V.; Wahlqvist, M.; Moshfegh, B. On the use of change-point models to describe the energy performance of historic buildings. In Proceedings of the 3rd International Conference on Energy Efficiency in Historic Buildings (EEHB2018), Visby, Sweden, 26–27 September 2018; Uppsala University: Uppsala, Sweden, 2018; pp. 512–520.
37. Navas-Martín, M.; Cuervo-Vilches, T. Natural ventilation as a healthy habit during the first wave of the COVID-19 pandemic: An analysis of the frequency of window opening in Spanish homes. *J. Build. Eng.* **2023**, *65*, 105649. [CrossRef]

38. Occam's Razor. 2022. Available online: <https://www.britannica.com/topic/Occams-razor> (accessed on 6 April 2022).
39. 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*, 249–258. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.