Influence of weather natural variability on the thermal characterisation of a building envelope

Sarah Juricic, Jeanne Goffart, Simon Rouchier, Aurélie Foucquier, Nicolas Cellier, Gilles Fraisse

Highlights

Influence of natural weather variability on the thermal characterisation of a building envelope

Sarah Juricic, Jeanne Goffart, Simon Rouchier, Aurélie Foucquier, Nicolas Cellier, Gilles Fraisse

- Thermal characterisation of a building envelope influenced by weather conditions
- Original methodology assesses minimal measurement duration for robust estimation
- Thermal characterisation by RC models is accurate and faster than steady-state methods
- Necessary 11-day measurements for robust estimation sets new benchmark value in field
- Estimations highly influenced by outdoor temperature and wind speed in case study

Influence of natural weather variability on the thermal characterisation of a building envelope

Sarah Juricic[a,*], Jeanne Goffart[a], Simon Rouchier[a], Aurélie Foucquier[b], Nicolas Cellier[a], Gilles Fraisse[a]

[a] *Univ. Savoie Mont-Blanc, CNRS, LOCIE, 73000 Chambéry, France*
[b] *Univ Grenoble Alpes, CEA, LITEN, INES, 38000 Grenoble, France*

Abstract

The thermal characterisation of a building envelope is usually best performed from on-site measurements with optimised controlled indoor conditions. Conversely, occupant-friendly measurement conditions provide less informative data. Notwithstanding occupancy, the boundary conditions alone contribute to a greater extent to the energy balance, which implies that non-intrusive conditions bring into question the reproducibility and relevance of such measurement. This paper proposes an original numerical methodology to assess the reproducibility and accuracy of the estimation of the overall thermal resistance of an envelope under variable weather conditions. A comprehensive building energy model serves as reference model to produce synthetic data mimicking non-intrusive conditions, each with a different weather dataset. An appropriate model is calibrated from the synthetic data and provides a thermal resistance estimate. The accuracy of the estimates is then assessed in light of the particular weather conditions used for data generation. The originality also lies in the set of weather data that allow for uncertainty and global sensitivity analyses of all estimates with respect to six weather variables. The methodology is applied to a one-storey house reference model, for which thermal resistance is inferred from calibrated RC models. Robust estimations are achieved within 11 days. The outdoor temperature and the wind speed are highly influential because of the large air change rate in the case study.

Keywords: Parameter estimation, Reproducibility, Frequentist calibration, Stochastic RC models, Weather variability, Global sensitivity analysis

*Corresponding author
Email address: sarah.juricic@gmail.com (Sarah Juricic)

Nomenclature

Abbreviations

ARX Auto Regressive Models with eXternal inputs

IWEC International Weather for Energy Calculations

ML Maximum-Likelihood

TMY Typical Meteorological Year

Parameters of interest

HTC Heat Transfer Coefficient: overall heat transfer of the building envelope towards exterior as defined in the ISO 13789 standard [1] (W/K)

R_{eq} Equivalent overall thermal resistance of the envelope, inverse of HTC (K/W)

R^*_{eq} Target thermal resistance: overall theoretical thermal resistance of the reference model (K/W)

U-value Thermal transmittance of a wall (W/m^2K)

C_w, C_i, R_o, R_i, A_w Parameters of the 2^{nd}-order model T_wT_i R_oR_i A_w

Time dependent variables

I_{sol} Solar global horizontal irradiation (W/m^2)

$P_{heating}$ or P_h Heating power delivered in the building (W)

$Q^{in}_{storage}$ and $Q^{out}_{storage}$ Heat transfered in or out the building envelope (W)

Q_{ground} Heat transfers from the indoors towards the ground (W)

Q_{sun} Heat gained indoors by solar irradiation (W)

$Q_{ventilation}$ Heat transfers by air change in the building (W)

T_w Unmeasured temperature of the building envelope ($°C$)

T_{in} Indoor temperature ($°C$)

T_{out} Outdoor temperature ($°C$)

1. Introduction

Background and motivation

The renovation of buildings is a growing concern with respect to the reduction of their global energy consumption, as stated by the European Commission in its strategic long-term policy for a climate neutral economy [2]. As underlined in the contribution *Buildings* to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [3], there is a need for accurate estimations of the thermal performance of building envelopes in order to drive relevant energy conservation measures. Heo et al [4] show for example how the estimation of the actual performance of a building benefits a retrofit analysis under uncertainty. When known, the thermal performance of an investigated building serves the energy retrofit plan by accurately reflecting on the possible energy gains.

On-site monitoring has been shown to be a promising lead for performing an accurate thermal characterisation of the envelope. In particular, the estimation of the Heat Transfer Coefficient (HTC) or its inverse the overall thermal resistance (R_{eq}) has been studied as a result of controlled experiments.

Uncertainty of the estimation of the thermal performance from on-site measurements is indeed usually reduced by means of an optimally designed heating or indoor temperature control in the building. By fully controlling the indoor environment in unoccupied buildings, such experiments may last only a few days while achieving satisfactory estimations. In Thebault and Bouchié [5], the indoor temperature setpoint is set to 35 °C for four days, which suffices for an accurate estimation of the overall thermal resistance. Ghiaus and Alzetto explain the principles of the QUB method in [6]: it relies on a rectangular large heating power excitation signal overnight. By monitoring the indoor temperature and assuming that the thermal response of the building is exponential, the slope of the monitored indoor temperature gives an estimation of the overall HTC of the building in less than 12 h and with errors lower than 10 %. Bacher and Madsen show in [7] how to perform model selection in order to identify the heat dynamics of a building. The authors exploit 6 days of data collected in a house in which heating power had been delivered according to a pseudo-random signal. The parameter estimates the authors obtained from the selected model were found to be satisfactory. Aside from these perturbation methods, co-heating tests also provide a fully controlled indoor environment by setting the indoor temperature to a constant value, as detailed in [8]. This experiment mimics steady-state conditions in the building. By daily averaging of the collected data from measurement durations of between 12 and 18 days, the dynamics driven by the variable outdoor condition are diminished and allow for an accurate estimation of the HTC. The coheating test is indeed regularly used as a benchmark for comparison of other methods, such as in Alzetto et al [9] or in Thébault and Bouchié [5]. In a nutshell, these building-scale characterisation methods are efficient for providing an accurate estimation of the heat transfer coefficient HTC or the overall thermal resistance of the envelope at the expense of a highly uncomfortable indoor environment during the experiment.

In some cases, however, buildings are continuously occupied, such as health-care facilities or nursing homes. Similarly, some other non-residential buildings cannot be left vacant as it would involve a loss of income, such as hotels or restaurants. In these building types energy conservation measures are relevant too. Balaras et al show in [10] that hotels and restaurants as well as health-care facilities have the highest energy use intensities in many European countries as well as in the United States. In addition, among all non-residential buildings, the authors found that heating was unanimously the main source of energy use. Development of an accurate and more importantly of a non-intrusive thermal characterisation method is therefore undisputably relevant.

The thermal characterisation of the envelope of many non-residential buildings cannot be achieved by any of the aforementioned perturbation methods and must instead rely on non-intrusive monitoring and a reduced number of sensors.. This implies that the sensors should not alter the building envelope and that occupants should not be disturbed by the monitoring equipment. A non-intrusive approach then relies on very few sensors in non-optimally controlled and possibly only partially known operating conditions. On the other hand, relying on fewer sensors may positively contribute to providing a test easier to implement and less costly, if ,however, an experiment is found to be feasible and accurate.

Feasibility and accuracy are indeed questioned by the uncontrolled nature of a non-intrusive experiment. The building envelope energy balance as presented in Equation 1 in an uncontrolled experiment shows different dynamics than in controlled experiments. The heating power delivered indoors $P_{heating}$ is less informative than when it is optimally designed. The outdoor boundary conditions, by their contributions through T_{out}, $Q_{ventilation}$ and Q_{sun}, have therefore a proportionally larger influence on the energy balance. This will particularly be the case when shorter datasets are used to infer the HTC or R_{eq}, whereas longer experiments provide de facto a wider natural variability of the boundary conditions which should flatten the effect of uncommon specific outdoor conditions.

$$\begin{aligned} HTC \cdot (T_{in}(t) - T_{out}(t)) + Q^{in}_{storage}(t) + Q^{out}_{storage}(t) \\ +Q_{ventilation}(t) + Q_{ground}(t) - P_{heating}(t) - Q_{sun}(t) \quad &= \quad 0 \end{aligned} \tag{1}$$

In addition, the outdoor weather conditions have correlated frequencies, which are themselves correlated with common indoor temperature setpoint schedules. This may impede the accuracy of the estimation, let alone the identifiability of the heat transfer coefficients.

For these reasons, the estimation of the HTC or R_{eq} coefficients in non-intrusive conditions may be considerably influenced by the boundary conditions. The number of data needed to secure a robust and accurate HTC or R_{eq} estimation could be significantly larger than what is usually considered in literature for tests in a controlled framework. The risk with too short datasets is obtaining significantly different results following identical procedures but carried out a day, a week or a year later. Non-intrusive conditions are indeed poorly

informative for the estimation process: the non-intrusive framework brings into question the uncertainty level and therefore the reproducibility of such estimation.

Estimation of the U-value, HTC or R_{eq} coefficients in non-intrusive conditions: models, experiment duration and influence of weather conditions in the existing literature

Previous work on thermal performance estimation of the envelope in a non-intrusive framework has focused on the thermal characterisation of either a single wall or the entire building envelope. To the best of our knowledge, how variability of the weather conditions influences the accuracy of this estimation has not been extensively investigated to date. A few papers however address the issues of convergence of the estimation and measurement duration with, for some of them, insight on the influence of outdoor conditions on the quality of the results.

At wall scale, Rasooli and Itard [11] showed from numerically generated data how solar irradiation significantly defers stability and convergence of the estimation of the conductive thermal resistance of a wall R_c using heat flow meters following the ISO 9869 standard [12]. Simulated data with and without solar irradiation showed significantly different convergence towards the final estimation value. The authors suggest using flow meters on both sides of the walls to secure a robust and faster estimation of R_c.

Petojevic et al [13] proposed an innovative method to exploit non-intrusive data from heat flux and temperature meters in order to determine the dynamic thermal characteristics of a wall. The use of 12.5 day-data, although not justified by the authors, led to acceptable accuracy in the results.

Gaspar et al [14] studied the minimal duration of a heat flux meter test for estimating the U-value of a wall by the average and dynamic methods of the ISO 9869 standard [12]. The authors compared the stability of the estimation with the criteria given by the ISO 9869 standard [12], which specifies that three conditions must be met simultaneously to end the test:

- The first condition is that the test must last 72 h or longer,
- The second condition is that the U-value obtained at the end of the test must not deviate more than 5 % from the value obtained 24 h earlier,
- The third condition is that the U-value obtained from the first N days and from the last N days must not deviate more than 5 %, with $N = 2/3\times$ total duration.

Gaspar et al [14] concluded that these conditions provide trustworthy estimations and secure convergence towards the final value within 4–5 days for both the average and dynamic methods described by the ISO 9869 standard [12]. The influence of weather variability on the convergence rate was not investigated. A higher actual transmittance of the façades was found to be the main reason for slower convergence.

At wall scale again, Gori and Elwell [15] as well as Gori et al [16] exploited heat flux measurements with RC models and introduced the idea of stabilisation of the estimation: from short datasets, the estimates suffer from the prominent noise in the data. As the dataset grows, the values stabilise towards a final value. Applying the criteria of the ISO 9869 standard [12], they found that a period of up to 10 days was necessary to reach stabilisation in autumn and winter season whereas longer periods were necessary in warmer seasons. The minimum length tested was 3 days, as demanded by the ISO 9869 standard [12], but the authors found that shorter datasets sufficed in some cases with the use of an RC model, implying that the three conditions of the ISO 9869 standard [12] might be too conservative when applied to other methods.

At wall scale too, Rodler et al [17] compared a dynamic model calibrated by Bayesian inference with the average and dynamic methods described in the ISO 9869 standard [12] and found that the temperature difference was more determinant than the length of the dataset, therefore confirming the significant role of boundary conditions in uncontrolled experiments.

Deconinck and Roels [18] applied dynamic grey box modelling in a non-intrusive framework to assess the thermal performance of a single wall based on heat flux measurements. The authors used two different data subsets of 10 days in winter (steady indoor temperature assumed at 20 $°C$) and 9 days in summer (free floating indoor and outdoor temperatures). They found that winter conditions with constant indoor temperatures were not appropriate to identify the parameters of interest, considering that temperatures are the main variables of the differential equations used for the exploitation of the data. Summer free-floating conditions were then found to be more informative and led to identifiable and interpretable parameters. Let us note here that identifiability relates to the unicity of the parameter estimation and interpretability to the ability to give the estimation a physical meaning. Both may be confounded if the model characterises perfectly the system.

At building scale, Reddy et al [19] identified the issue of data informativeness as well as the influence of weather conditions when assessing the overall heat loss and overall ventilation rate of a large commercial building. To perform the assessment, non-intrusive measurements are averaged and exploited by a steady-state equation. They found that daily averaged data over 1 year combined with a multi-step regression technique, where multiple regressions are performed one after the other to estimate parameters one by one, achieved the best results. Parameter identification over a single season was less accurate: in winter and summer seasons, the combined variability of the outdoor temperature and the relative humidity was narrower than during the spring season. Large variability of these two weather variables yielded parameters estimates with less correlation and more accurate overall parameter identification.

More recently, Senave et al [20] studied the physical interpretation of auto-regressive models with exogenous inputs (ARX), aiming for the estimation of the HTC via on-board monitoring, i.e. in a non-intrusive measurement framework. Four different indoor temperature scenarios were tested twice through 20 days of synthetic data: once for calibration and once for validation. The building

model is a single-zone opaque box and the study focused on the estimation of the HTC in the case of heat losses to the ground. There is no mention of the influence of the measurement duration on the results. In a later paper by Senave et al [21], ARX models and linear regressions were compared to exploit data collected from non-intrusive experiments on datasets of 26 weeks.

Grey-box models such as stochastic RC models have not been used to exploit data at building scale, but showed promising results at wall scale as shown in Deconinck and Roels [18] or in Gori et al [16] with shorter datasets than the average method of the ISO 9869 standard [12]. Stochastic RC models have also been used in controlled experiments and provide a satisfactory HTC estimation from short datasets, as in Thébault and Bouchié [5].

In conclusion, the minimal duration for heat flow measurements to infer the actual thermal transmittance of a single wall has been addressed in the literature and the relevance of the ISO 9869 standard [12] criteria has been discussed. At wall scale, accurate estimations of U-values may be performed within 1 week, sometimes quicker in the case of low-transmittance façades. Solar irradiation seems to influence the convergence rate.

Building-scale HTC estimation aims at a better representation of the overall thermal performance of the envelope. The literature, however, is scarce on non-intrusive experiments at building scale and the methods use much larger datasets, which increases the variability of the weather conditions during the test and therefore reduces the influence of the boundary conditions on the quality of the results. It has also been suggested that large variability in the outdoor temperature and relative humidity lead to a better identification.

Objectives of this work

Accurate estimation of the HTC or R_{eq} coefficients at building scale in non-intrusive experiments would certainly be beneficial as guidance for relevant energy conservation measures. Although the existing literature suggests that such estimation is feasible from datasets of several weeks, shorter experiments are desirable to decrease immobilisation of the measurement devices and the cost of the procedure. The feasibility and accuracy of the HTC or R_{eq} identification from shorter datasets is questionned, however, under naturally variable weather conditions given their larger contribution to the building envelope energy balance.

This paper therefore intends to assess the feasibility, i.e. the accuracy and reproducibility of the R_{eq} estimation from non-intrusive experiments. The underlying hypothesis is that there is a minimal measurement duration after which the R_{eq} estimation keeps steady, regardless of the boundary conditions. Provided that the measurement duration is sufficient, an R_{eq} estimation could then start at any time and on any day under any usual weather conditions, and the estimation should remain robust. The accuracy and feasibility of the estimations are therefore assessed in light of the natural and local variability of weather conditions.

This paper proposes performing the estimation with stochastic RC models in the hope of achieving reasonably fast estimations of the overall thermal resistance of the envelope R_{eq}, as was suggested by the literature review.

The investigation of the feasibility of such estimation in non-intrusive measurement conditions is based on an original numerically based methodology. Indeed, simulated datasets from a reference model are used to calibrate a stochastic RC model and to infer an estimation of R_{eq}. The characteristics of the reference model and the case study that served for data generation are presented in Section 2, which also details how natural weather variability is introduced by the use of multiple weather datasets for the simulations of the reference model. In particular, the weather datasets used are stochastically generated so as to explore to full extent the natural variability that can be expected in a typical January in Geneva. The reproducibility of the estimation of the thermal resistance is then discussed in Section 3.1. Section 3.2 examines the influence of the six stochastically generated weather variables on the thermal resistance estimates. Section 4 finally puts the results into context with the previous literature review and discusses the relevance of using the ISO 9869 standard [12] criteria to assess the convergence of an estimation.

2. Methodology

To assess the accuracy and reproducibility of the thermal resistance R_{eq} estimation of a building envelope while accounting for natural weather variability, the general idea is to study how different yet coherent weather conditions influence such estimations. The methodology developed for this paper is detailed in this section.

2.1. Overview of the applied methodology

The objective is to collect a number of measurement datasets, each obtained under different weather conditions. Each set is expected to provide measurement data of indoor and outdoor conditions so as to enable the calibration of an appropriate model $\mathcal{M}$. The model $\mathcal{M}$ is calibrated from each dataset, providing as many estimations of its parameters θ as there are datasets. From the hence estimated parameters θ, an estimation of R_{eq} is inferred. Measurement duration and variability in the weather conditions themselves will then influence the accuracy and uncertainty of the estimations. In principle, for a certain minimal measurement duration, the final estimates are expected to show robustness and to remain significantly similar regardless of the natural variability in weather conditions.

The essence of the methodology is therefore to use multiple datasets, each in different weather conditions. Datasets from actual on-site measurements during an entire heating season, or better yet during several heating seasons would concededly ensure strong realism. However, the use of real datasets may only deliver an incomplete view of the issue and be difficult to analyse. Indeed, as the initial conditions and the thermal state of the envelope would be different

for each dataset given the previous days, analysing the variability of the R_{eq} estimations could not be attributed only to weather variability.

To avoid this pitfall, the original methodology proposed here relies on a fully numerical four-step procedure, as illustrated in Figure 1:

- A computer-based model serving as *reference model* is implemented in a program for dynamical thermal simulations (here, Energy Plus). Figure 2 illustrates this point and shows how synthetic data is generated from a certain case study, based on appropriate modelling choices. The building energy modelling choices are detailed in section 2.2.1 and the case study that served in this study is described in section 2.2.2. As also described in Figure 2, given that the reference model is purely numerical, a precise value of the target R^*_{eq} can be calculated, as is detailed in section 2.2.3.

- Simulation and output processing (**step I** part 1): a dynamical simulation of the reference model is run under known boundary weather conditions. The choice of a set of synthetic weather datasets to perform the simulation is detailed in section 2.3. The simulation output provides synthetic data of the resulting indoor conditions. White noise is added to mimic actual measurements of indoor and outdoor variables. Step (I) is repeated n times, each time with a different weather dataset. Section 2.3 gives further details on the weather data that are used to perform the energy simulations.

- Data subset extraction (**step I** part 2): Seven subsets of data are extracted from each simulated dataset, all starting on January 2^{nd} with growing lengths from 2 days to 25 days. Simulation parameterisation and data selection are described in section 2.4.

- Calibration (**step II**): Model $\mathcal{M}$ under study, chosen to provide an estimate of R_{eq}, is calibrated on each subset of the simulation output data. Model calibration is performed from a frequentist approach with a BFGS algorithm minimising the negative log-likelihood. The calibration procedure is thoroughly described in 2.5.2. A R_{eq} estimation can be derived from the estimated parameters of each calibrated model $\mathcal{M}$. Overall, step (II), which comprises the parameter estimation of model $\mathcal{M}$ and subsequent R_{eq} inference, is performed $7 \times n$ times. All steps from the choice of model to R_{eq} inference are detailed in section 2.5.

- Accuracy assessment (**step III**): the estimated R_{eq} is compared with the target value R^*_{eq} of the reference model by means of a novel interpretability indicator, described in section 2.6, which reflects on both the uncertainty and the relative error to the target value. The variability and accuracy of the R_{eq} estimation are assessed by studying the evolution with growing measurement duration of the total variance of the maximum-likelihood estimates as well as that of the interpretability indicator.

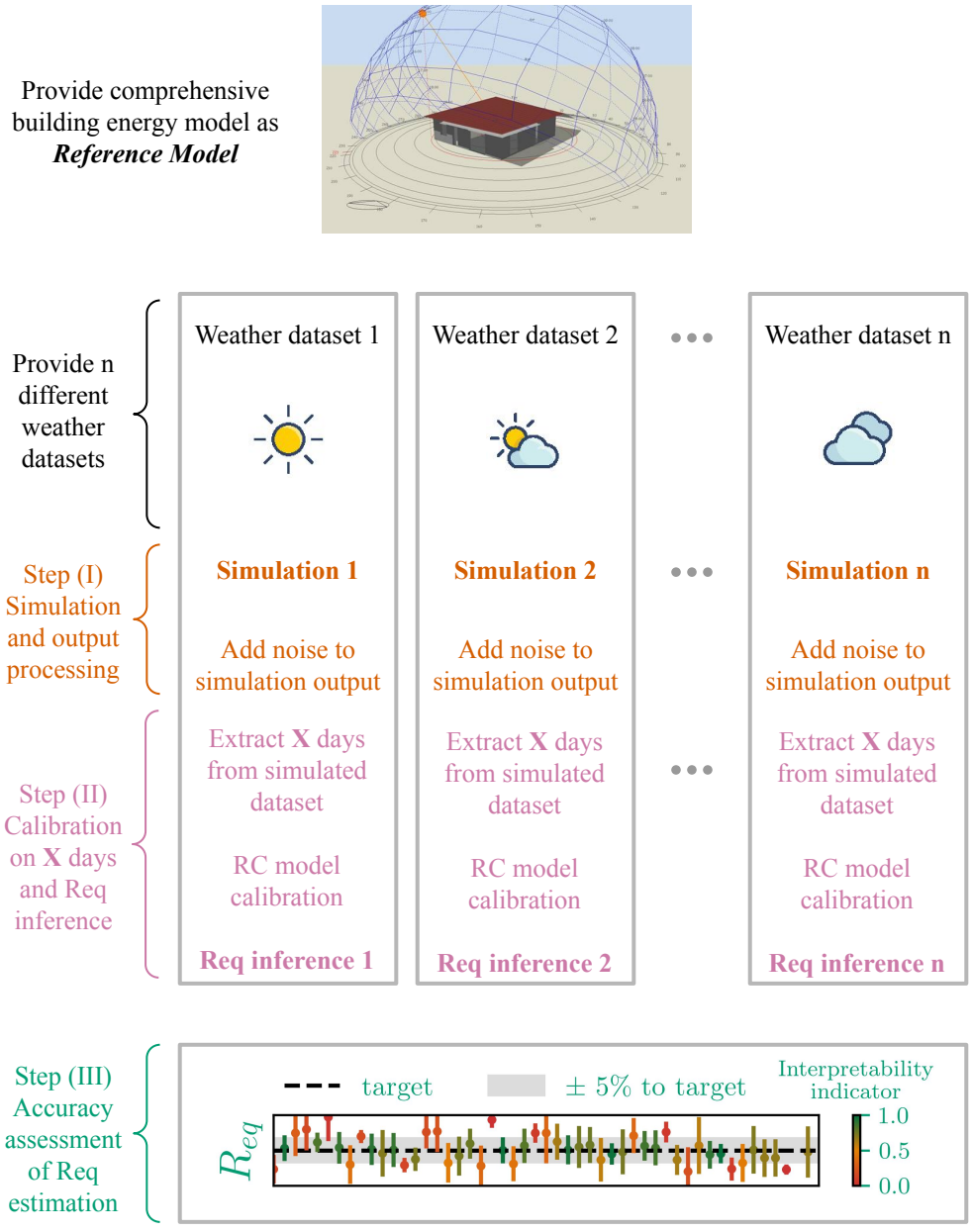


Figure 1: Methodology overview to assess the influence of weather variability on the accuracy of a R_{eq} estimation. In the end, each estimation is compared with the target value by means of a novel interpretability indicator.

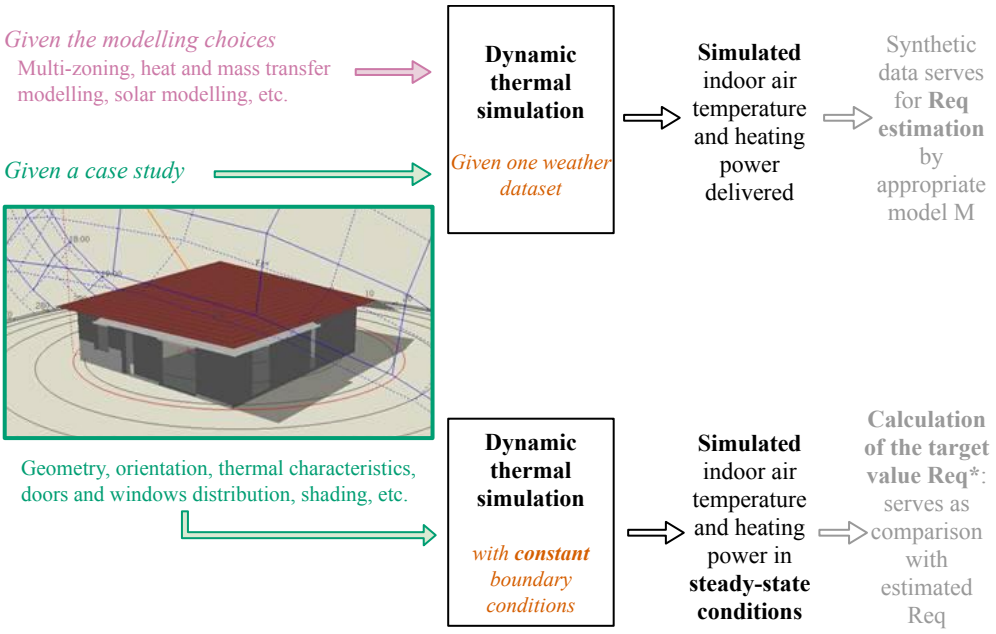


Figure 2: Hypotheses behind the synthetic datasets: given certain modelling choices and given a case study, a dynamical thermal simulation provides synthetic datasets that are used for R_{eq} estimations. The building energy model of the case study is simulated with constant boundary conditions to provide an estimation of its theoretical thermal resistance R^*_{eq}.

2.2. Reference model and description of the case study

The methodology developed and applied in this paper is based on a numerical reference building energy model. Therefore, the objective is, given a case study and certain boundary conditions, to provide synthetic measurements as realistic as possible of the indoor temperature and the heating power delivered in the building. To meet this objective, the modelling choices were carefully made and are detailed in this section. The case study and its thermal characteristics are also presented in this section.

2.2.1. Relevant modelling choices for providing synthetic data

Various choices can be made for the reference model to account for heat and mass transfer modelling, solar irradiation, etc. The choices made for the reference model are therefore driven by the purpose of this study, i.e. thermal behaviour, but also by the need for a reasonable simulation duration since many simulations are planned for the R_{eq} estimation assessment.

Among the different algorithms EnergyPlus has implemented for heat and/or moisture transfers in the building, the *Conduction Transfer Function* seems to be an appropriate option: it is a heat-only algorithm and does not account for moisture storage and diffusion. Using heat and moisture transfer algorithms will only slow down the simulation time without adding significant improvements

to the assessment methodology. The *Conduction Transfer Function* algorithm is fast as it relies on a state space representation with the finite difference wall temperatures as variables [22]. From the state-space representation, it is possible to formulate the model output as a direct function of the input, without calculating storage and temperatures at the discretisation nodes in the wall.

As for heat transfers through ventilation and infiltration, EnergyPlus has two possible options: the *DesignFlowRate* module and the *Airflow Network* model. The *Airflow Network* model, based on pressure and airflow calculations with temperature and humidity calculations, however, has been developed to simulate with accuracy air distribution systems and their performance, such as supply and return duct leaks, multi-zone airflows driven by outdoor wind and mechanical ventilation. Although such detail is without question physically more accurate, it is also much more computationally consuming. The *DesignFlowRate* module, concededly simple, has been chosen instead for the reference model. Infiltration and ventilation flow rates are accounted for by the same module, but in separate inputs so as to enable different values. It relies on equation 2 to calculate the airflow rate at each time step:

$$Q(t) = V_{design} \cdot F_{schedule} \cdot (A + B \cdot |T_{zone} - T_{odb}| + C \cdot WindSpeed + D \cdot WindSpeed^2) \tag{2}$$

where V_{design} the air flow rate (m^3/s)
F_{design} an optional schedule that can vary over time,
A, B, C and D coefficients between 0 and 1,
T_{zone} the zone indoor air temperature ($^\circ C$),
T_{odb} the outdoor dry bulb temperature ($^\circ C$).

A, B, C and D are fixed identically to the default BLAST (EnergyPlus predecessor) [23]:

$$A = 0.606, B = 0.03636, C = 0.1177, D = 0$$

Solar irradiation plays a large role in the building energy balance. In particular as the window blinds are maintained open in the simulations, solar irradiation entering the building through the windows comes from multiple sources: direct beams with time-dependent values for each wall, diffuse irradiation via the environment, reflections of direct irradiation on the environment as well as indoor diffuse reflections. To account for such details, the *Full Interior And Exterior With Reflections* EnergyPlus module is used.

The EnergyPlus simulation run period extends from approximately November 1^{st} to March 31^{st} to cover a winter season. Furthermore, accurate estimations are expected to be larger in the coldest months, from December to February. This is due to the larger temperature difference between indoors and outdoors and lower solar gains. This implies that the energy balance during winter days is more sensitive to the thermal performance of the building envelope. It is, however, important to start the simulation period earlier as to avoid any impact of the warm-up runs performed by EnergyPlus. Indeed,

the software runs multiple times on the same first day, beginning at an indoor temperature of 23 ° C until convergence of the indoor conditions is met. The warm-up allows for credible initial conditions for the rest of the simulation. But, as the first weather day of the simulation period is random, it might place a particular weight on a possibly unusual cold or warm day, thus misleading the first hours and days of simulation, depending on the thermal mass of the building. It is therefore safe to start early in the winter period and then discard the first 15 days of simulation. In the end, the simulation run period starts on November $15^{th}$and finishes on February 15^{th}.

As for the simulation time step, it should not be longer than approximately 15 minutes, and preferably shorter in the case that this reference model has lower characteristic times. Time steps longer than 15 minutes could hide aliasing in the data: short but influential phenomena are not seen. Aliasing in data then leads to potentially dramatically wrong estimations [24]. In this case study, the time step is fixed at 10 minutes.

One of the identified pitfalls for a generalisable model assessment framework is the fact that the simulation outputs are "ideal" measurements: the simulation output is deterministic. There are no systematic errors and no random measurement errors. To maintain the focus on the influence of weather variability on thermal characterisation, systematic errors will not be added to the outputs. Measurement random errors on the other hand are a non-negligible part of the issue of solving inverse problems [25]. White noise is added to the following simulation outputs following normal distributions in agreement with the undermentioned literature sources:

- temperatures: addition of a normal noise $\mathcal{N}(0,\ 0.2\ ^{\circ}C)$, in agreement with Leroy [26],
- heating power: addition of a normal noise $\mathcal{N}(0,\ 20.0\ W)$, in agreement with Sengupta et al [27],
- solar irradiation: addition of a normal noise $\mathcal{N}(0,\ 5.0\ W/m^2)$, in agreement with Stoffel et al [28].

2.2.2. The case study

The case study to which the methodology is applied in this paper is a multi-zone building of a one-storey house, as shown in Figure 3. The heated space is approximately 98 m^2 and has a total volume of approximately 250 m^3. The building is equipped with convective heaters. The air change rate is 1.0 volume per hour.

The building has unheated and unventilated crawlspace and attics. Heat losses towards the crawlspace may be considered insignificant, as the insulation layer under the concrete slab of the ground floor has been set at 30 cm. As for the rest of the building envelope, exterior walls are made up of a 20 cm brick wall, with 10 cm insulation and 1 cm plaster on the interior side whereas the attics and the indoor space are separated by a 1 cm plaster and 30 cm insulation. All windows, frames included, have U-values between 1.3 and 1.6 W/m^2K. Total

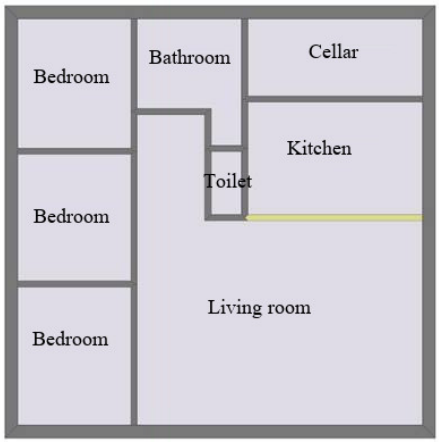


Figure 3: Floor plan of the one storey house serving as reference model

window surfaces add up to 15.9 m^2, among which 0.6 m^2 north, 5.4 m^2 east, 6.7 m^2 south and 3.2 m^2 west. The shading facilities are not activated and therefore allow for solar gains. South- and east-facing facades have a shading overhang, designed to avoid solar irradiation in summer. In winter conditions, most of the irradiation enters the envelope. Table 1 summarises the thermal properties of interest in this case study and Figure 3 shows how the building is configured.

Vertical insulation thickness	10 cm
Attic insulation thickness	30 cm
Ground floor slab insulation thickness	30 cm
Air change rate	1.0 h^{-1}

Table 1: Thermal characteristics of the case study used in this application

The indoor temperature setpoint schedule is designed to mimic occupant-friendly conditions to meet the objective of studying how poorly informative data influences interpretability. Seeing that dynamic models such as RC models cannot adequately learn from data in conditions close to steady state, a realistic temperature setback is therefore scheduled and follows a usual occupant-related schedule. The indoor temperature schedule is set to reach 20 °C in the morning and in the evening for workdays, and all day long during week-ends and on Wednesdays. The rest of the time, the temperature is scheduled to remain at 17 °C. Figure 4 illustrates a week of simulated indoor temperature with such schedule. Noteworthy in this figure is that, as the indoor temperature setpoint is based on the operative temperature, there are slight differences between the simulated indoor air temperature and the temperature setpoint.

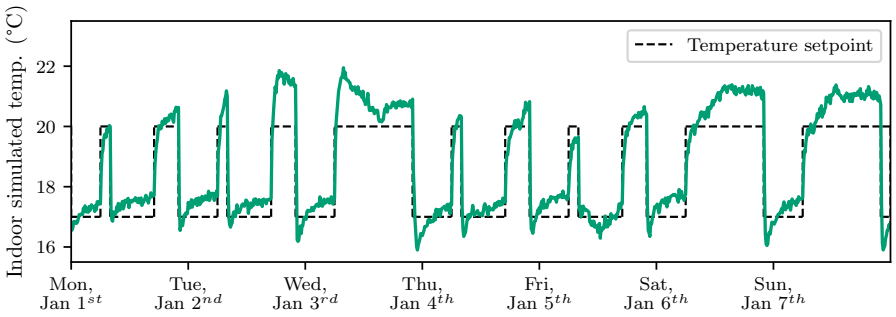


Figure 4: Illustration of 1 week of simulated indoor temperature with added noise: Wednesdays and week-ends have different temperature setpoints.

2.2.3. Thermal performance of the case study

As mentioned earlier, basing the procedure on simulated data offers the advantage that a theoretical thermal resistance of the envelope is known. In our case, it is determined by a simulation run with constant boundary conditions: no solar irradiation, constant indoor and outdoor temperatures. Wind speed, however, is non-null and is kept at the values of the TMY file of Geneva. Removing the wind speed would have decreased the overall heat loss coefficient, by diminishing the heat losses through ventilation as ventilation has been modelled to be wind speed dependent. Since in the synthetic experiment the ventilation is not stopped, the R_{eq} estimations would not have converged towards a target value calculated without wind.

With these steady-state boundary conditions, the dynamic terms $Q^{in}_{storage}$ and $Q^{out}_{storage}$ in the energy balance from Equation 1 become negligible and Equation 3 therefore perfectly describes the linear relationship between the heating power and the indoor-outdoor temperature difference:

$$P_{heating} = HTC \times (T_{in} - T_{out}) = 1/R^*_{eq} \times (T_{in} - T_{out}) \tag{3}$$

A least squares regression is performed on daily averaged data from January, February and March (92 days) and gives $R^*_{eq} = 5.19 \times 10^{-3}\ K/W = 5.19\ K/kW$ with a Pearson (R^2) coefficient of the linear regression of 0.999, showing an excellent fit. This value is from now on called ***target*** R^*_{eq}. Means of comparison of an estimated R_{eq} to the target value R^*_{eq} are described in section 2.6.

2.3. Providing weather data to the simulation step (I)

2.3.1. Using actual weather data: a limited insight

As a first attempt to understand the influence of weather conditions, the reference model was run with 10 years of historical winter weather data in Geneva (Switzerland), from 1990 to 1999 included. From each simulation, subsets of data of variable durations are extracted: 2, 3, 5, 8, 11, 15 and 25 days. Model $T_wT_i\ R_oR_i\ A_w$ as in Equation 4 is then calibrated. An estimation

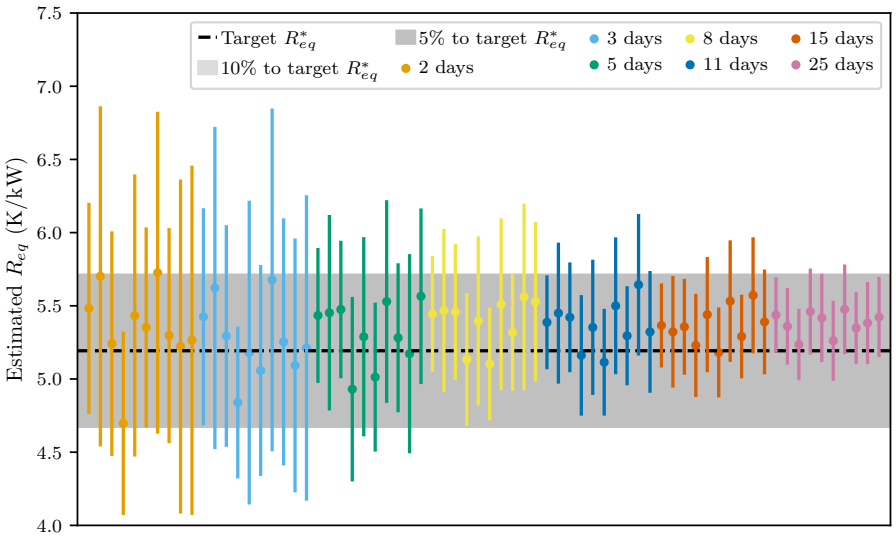


Figure 5: Evolution of the variability in the estimation of R_{eq} from 2-, 3-, 5-, 8-, 11-, 15- and 25-day data subsets: data simulated with historical weather data for

of R_{eq} is inferred for each subset and shown in Figure 5: each dot represents the maximum likelihood (ML-) estimate and the bars represent the confidence intervals. As will be specified in 2.5.2, the ML-estimates are the most likely values for the parameters given the collected data.

Preliminary results can be inferred from Figure 5:

- Short datasets provides R_{eq} estimations with a large variability and high uncertainty. There seems indeed to be no agreement in the ML-estimates (visible as dots) and the confidence intervals are large. This suggests in particular that 2 or 3 days are insufficient for a robust estimation of R_{eq}.
- Variability of the ML-estimates decreases as the measurement duration increases. All ML-estimates converge within a 5 % error band around the target R^*_{eq} with 25-day datasets.
- Regardless of the measurement duration, the significant variability of the ML-estimates can only be attributed to weather variability induced by the different weather datasets used for simulation. However, the particular conditions that cause under- or overestimation can however not be inferred from this first application.

These preliminary results suggest a decrease in variability with growing measurement duration. However, seeing that this first application is performed on only 10 weather datasets, concluding on a minimal length would be statistically weak. In addition, the datasets do not allow to attribute the estimation variability to one or more specific weather variables.

2.3.2. Stochastic weather data to perform global sensitivity analysis

To perform a more exhaustive assessment of the R_{eq} estimations under variable weather conditions, the proposed methodology is now applied to a set of 2000 synthetic weather datasets with which a variance-based sensitivity analysis is possible [29].

A total of six weather variables are stochastically generated to be representative of the usual weather conditions in Geneva in winter, following the methodology described in [29], as a time series constructed by a combination of statistical and deterministic features. The characteristics are extracted on the basis of the $IWEC$ weather data file (International Weather for Energy Calculations) [30] from Geneva. $IWEC$ files are built like TMY weather files [31] for locations outside the United States and Canada. The TMY file, standing for Typical Meteorological Years, is built by concatenation of typical months. Each month is chosen from 30 years of actual data: each monthly dataset is weighted as a sum of 13 Finkelstein–Schafer statistics [32] from the temperature, wind and solar radiation data. In the end, the monthly dataset chosen is the one that shows statistics closest to the mean, median of the 30-year data distribution, after having discarded years with exceptionally long periods of consecutive warm, cold or low radiation days. The stochastic generation [29] thus contains as much variability than in the TMY file: if the TMY has for one particular variable a lower variability than the rest of the 30-year actual weather data, it will reflect this in the synthetic data.

From the TMY file, Goffart et al [29] selected six weather variables to stochastically generate 2000 weather files; the rest of the variables were left unchanged. The generated variables are exterior dry bulb temperature, relative humidity, direct normal solar irradiation, horizontal diffuse solar irradiation, wind speed and wind direction.

Finally, the weather data are generated to calculate sensitivity indices through a Sobol variance method able to cope with groups of time-dependent inputs, like the time dependency of each weather variable here. Sensitivity indices by groups estimate the effect of the entire time series of the meteorological variable under study. The sensitivity indices are therefore scalars even though the variables are time series. The indices are calculated from two sets of 1000 samples, each sample of the first 1000 being defined by the characteristic features extracted from the TMY file of each weather variable, the second 1000 samples being a rearrangement of the first.

In this study, the output of interest for the sensitivity analysis is the R_{eq} estimation and in particular the weather conditions leading to an increased or decreased ML-estimation and accuracy.

In order to check the representativeness of the generated weather data, Figure 6 compares the synthetic data with the actual historical data from Geneva and with the TMY data. The figure shows the empirical cumulative distributions of the six weather variables for the month of January of the historical weather data in black thin lines and in orange the TMY data. The grey areas represent the 50 %, 75 % and 95 % quantiles of the synthetic data.

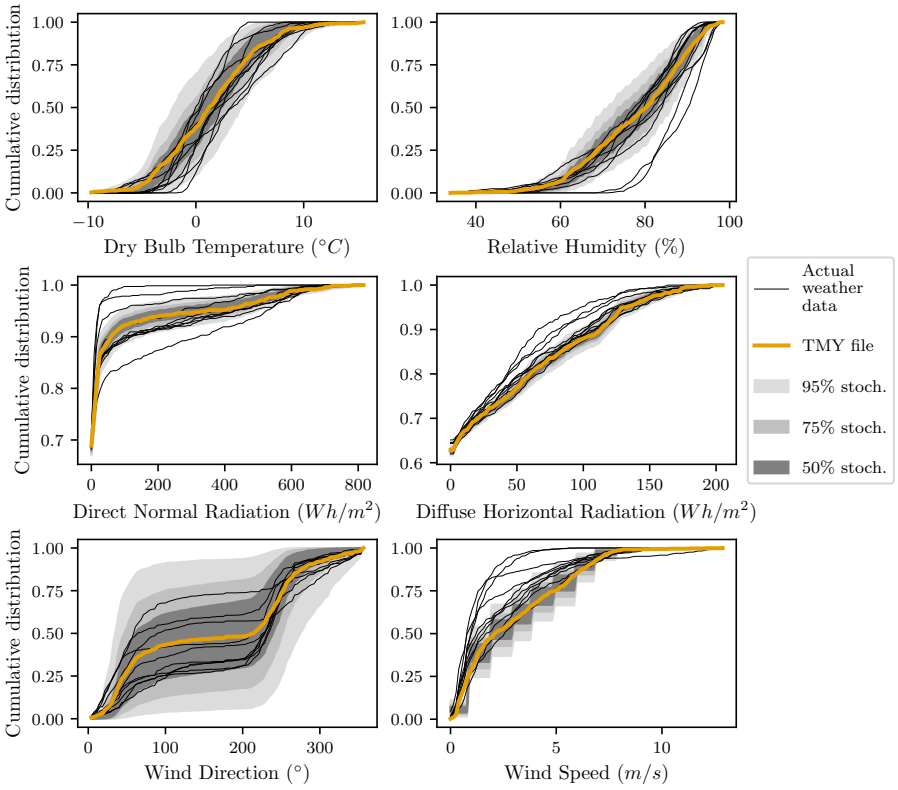


Figure 6: Representativeness of the stochastic weather data with respect to real weather data: the lines represent the cumulative distributions of all six weather variables. The generated outdoor temperature and direct normal irradiation data in grey are representative of historical data whereas wind speed is instead in the generated data than in the historical measurements in Geneva.

The lines represent the cumulative distributions of all six weather variables. The higher the line, the lower the values of its time series. For example, the wind speed in the TMY file, in orange, is lower than any other historical weather data, which means that the TMY file is in overall more windy in January than the 10 years of historical weather data.

From Figure 6, it can be inferred that:

- Synthetic outdoor dry bulb temperatures seem to be representative of the historical measurements. Synthetic wind direction is in good agreement with the historical measurements as well.
- Synthetic relative humidity seems to be lower than some of the historical measurements. The synthetic diffuse radiation on the contrary seems to be slightly overestimated, as does the wind speed.

- The direct normal radiation data generated do not cover a range as wide as the actual data: some of the real data may have much higher or lower direct radiation. This might have an impact on the following results and will be discussed later.

2.4. Step (I): Simulation, output processing and data subset selection

For the purpose of the study, the thermal simulation of the multi-zone reference model needs to be performed so as to deliver energy consumption and temperature. Let us also shortly recall that the reference model is run on a winter season, from November 15th to February 15th. The time step of the output is set at 10 minutes in order to catch higher-frequency phenomena and improve the accuracy of the estimation [33]. Winter season secures outdoor temperatures below 15 °C, which creates a significant temperature difference with indoors and enhances the practical identifiability of the parameters of interest.

Aiming at a comprehensive comparison, the studied model is calibrated on several subsets of each dataset: 2, 3, 5, 8, 11, 15 days. As shown in Figure 7, all subsets start on January 2nd, i.e. far from the warm-up period of Energy Plus, which might have affected the realism of the data. January 2nd is also at the beginning of the month of interest, representative of winter conditions.

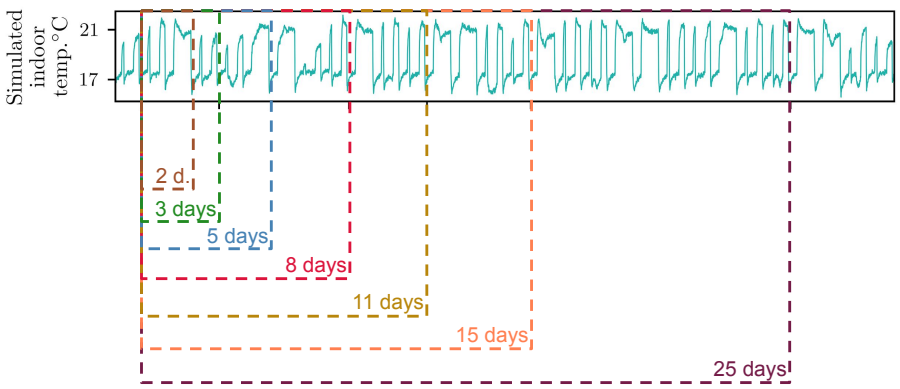


Figure 7: Selection of the data subsets from the stochastic data: 2, 3, 5, 8, 11 and 15 days starting January 2nd

2.5. Step (II): R_{eq} inference

Inferring a physical property from data means solving an inverse problem. To do so, as detailed in this section, an appropriate model is chosen and its parameters estimated so that the model prediction fits the data. The physical property of interest, here the overall thermal resistance, R_{eq}, is inferred from the estimated model parameters.

2.5.1. The choice of stochastic RC models

The limited number of sensors encourages to consider grey-box models as a data analysis tool. Indeed, Foucquier et al [34] distinguish three levels of thermal modelling of buildings: white-, grey- and black-box modelling. Comprehensive thermal dynamic models, i.e. white-box models, rely on an extensive number of parameters as they accurately describe the physical behaviour of a building. Therefore, white-box models calibrated from poorly informative data will most certainly lead to overfitting and non-interpretable parameter values. By contrast, black-box models rely exclusively on statistics and cannot be physically interpreted. In between, grey-box models are a combination of both, with physically inspired mathematical structures and statistical modelling to achieve reliable simulation results [35]. In particular, grey-box models can use a naive description of the building physics to limit the number of parameters [36] and still achieve satisfactory modelling through a stochastic diffusion term in the model. In a comprehensive review of energy modelling of buildings, Li and Wen [37] also underline that grey-box models cut computational costs while maintaining physical meaning. Grey-box models therefore offer a good compromise in a non-intrusive framework.

The numerical procedure is applied to a lumped capacitance model: $T_w T_i$ $R_o R_i$ A_w. It is a low-order model in the sense that it relies on a system of two differential equations. It incorporates two thermal resistance parameters R_o and R_i, two thermal capacitance parameters C_w and C_i and a single coefficient

for solar aperture hereafter named A_w. RC models are indeed simplified lumped models of the otherwise non-linear thermal exchanges of the building envelope and have physical meaning: thermal capacitances or thermal resistances can be proven to be the lumped capacitances or, respectively, lumped resistances of each layer of the envelope [38, 39].

As with any simplified model, RC state-space models have an intrinsic model error, which can be taken into account as an ad hoc term in the model formulation [40]. If not, Brynjarsdottir and O'Hagan [41] showed that disregarding model discrepancy may lead to biased and over-confident parameter estimation. Therefore, stochastic differential equations [42] are chosen to formulate the RC model, as described in Equation 4.

$$\begin{cases} \begin{bmatrix} \dot{T_w} \\ \dot{T_{in}} \end{bmatrix} &= \begin{bmatrix} -\frac{1}{C_w R_o} & \frac{1}{C_w R_i} \\ \frac{1}{C_i R_i} & -\frac{1}{C_i R_i} \end{bmatrix} \begin{bmatrix} T_w \\ T_{in} \end{bmatrix} + \begin{bmatrix} \frac{1}{C_w R_o} & \frac{A_w}{C_w} & 0 \\ 0 & 0 & \frac{1}{C_i} \end{bmatrix} \begin{bmatrix} T_{out} \\ I_{\text{sol}} \\ P_h \end{bmatrix} + \sigma d\omega \\ y &= \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} T_w \\ T_{in} \end{bmatrix} + \epsilon \end{cases} \tag{4}$$

First and foremost, the structural identifiability needs to be proven, meaning that in theory, given hypothetical ideally informative data, calibrating the model will result in a unique solution. The structural global identifiability of the model is derived by a differential algebra algorithm [43] implemented by Bellu et al [44] in the tool DAISY.

From the estimation of the thermal resistance parameters R_o and R_i of the RC model, R_{eq} can be derived as shown in Equation 5. Equivalent standard deviations are obtained from Equation 6. For readability, the R_{eq} estimations will be given in K/kW, which is equivalent to the order 10^3 K/W.

$$R_{eq} = R_o + R_i \tag{5}$$

$$\sigma_{R_{eq}} = \sqrt{\sigma_{R_o}^2 + \sigma_{R_i}^2 + 2 \times \sigma_{R_o} \sigma_{R_i}} \tag{6}$$

2.5.2. Model calibration

To infer an estimation of the overall thermal resistance, the parameters of the model of interest need to be estimated so as to fit the data. This calibration process is performed by a quasi-Newton optimisation using the BFGS algorithm, operated in the pySIP python library [45]. In particular, the BFGS algorithm minimises the negative log-likelihood of the model prediction which resumes in finding the most likely set of parameters that fit the selected dataset.

The model calibration is performed through a frequentist approach, where all information about the parameters is hypothesised to be acquired in the collected data and where the estimation is supposed to have a Gaussian error. Another option, yet more computationally costly, would be to use a Bayesian approach where prior information about the parameter values is

incorporated in the calibration process as a prior probability density. However, this paper investigates how natural weather variability influences the amount of information in the data itself and how that reflects on the accuracy and robustness of the estimation of R_{eqs}. From this perspective, a frequentist approach makes perfect sense.

As a result, the estimation with the BFGS optimisation provides maximum-likelihood (ML-) estimates of the parameters of the selected RC model, with a Gaussian error uncertainty. Both ML-estimates and their uncertainty need to be considered when assessing the outcome of an estimation.

2.5.3. Model selection and validation

Good practice in model calibration demands that model selection be made so as to infer results from one best fitting the data. The best-fitting model might be different between short- and long-duration datasets. For this reason, even if model selection is performed on a 3-day dataset, the residuals of prediction of a 15-day dataset are verified as well, in order to ensure the selected model still performs well on larger datasets.

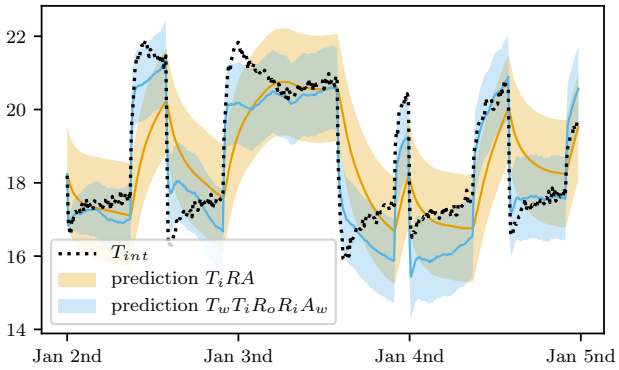
First-order RC models may be quickly discarded as they visibly fail to capture the physics compared with a second-order model, as can be seen in Figure 8a. The residuals of a first-order model are highly auto-correlated, see Figure 8b, which again is proof that such a model does not correctly fit the data. The residuals of a second-order model are indeed much closer to white noise for a 3-day calibration. When a 15-day calibration is performed, the residuals show that the first-order model is still highly auto-correlated and that the second-order model still performs well. Higher-order models achieved highly correlated parameter estimations, which is very undesirable when the parameters need to be physically interpreted.

2.6. Interpretability: a novel indicator to assess the R_{eq} estimations (step III)

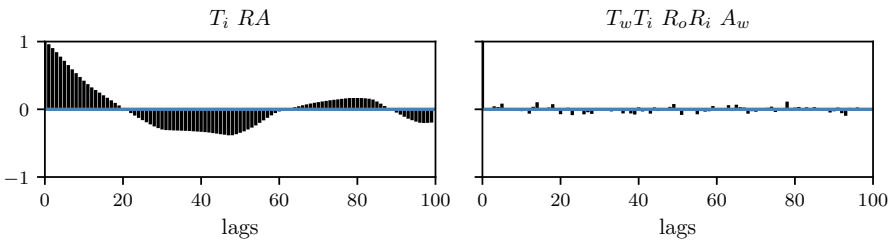
This section proposes a novel scalar indicator, called the interpretability indicator, to assess the closeness of the estimation to the target value.

The need for a novel indicator comes from the observation that a straightforward relative error of the ML-estimator to the target value is not representative of the uncertainty of such estimation. A relative error cannot in fact discriminate between accurate and uncertain ML-estimations. For example, in Figure 9, the case 1 estimation is less desirable than the case 2 estimation: both are equally inaccurate from a relative error point of view, but the case 2 reflects the inaccuracy appropriately through a large confidence interval.

In order to better discriminate between the less desirable estimations from the others, an interpretability indicator is proposed. This indicator represents the area under the bell curve that is ± 5 % of the target R^*_{eq}. The interpretability indicator therefore takes values between 0 and 1. For example, in Figure 9, the case 1 estimation scores close to 0 whereas the case 2 estimation scores at 0.24. Estimations may be considered satisfactory if they score above 0.5, such as case 3 estimation in Figure 9.



(a) Non-filtered prediction on a 3-day dataset



(b) Autocorrelation of the residuals of filtered predictions of models $T_i\ RA$ and $T_w T_i\ R_o R_i\ A_w$ for 3-day data. Basically, this test compares the similarity of the residuals with itself modulo a certain delay, i.e. a lag. Model $T_i\ RA$ shows a significant autocorrelation of its residuals, which implies systematic errors in the prediction. Significant physical phenomena are missing in the model.

Figure 8: Graphical visualisation of how models $T_i\ RA$ and $T_w T_i\ R_o R_i\ A_w$ fit the data

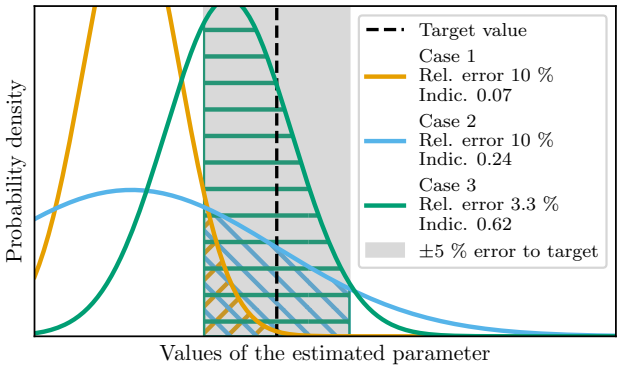


Figure 9: Relative error and interpretability indicator as two accuracy indicators: the relative error reflects poorly on the importance of uncertainty whereas the interpretability indicator accounts simultaneously for accuracy and uncertainty.

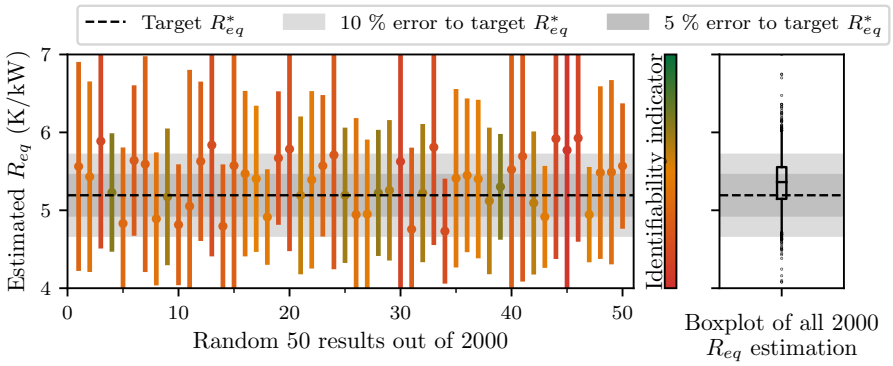


Figure 10: Variability of the R^*_{eq} estimation with 2-days calibration

3. Results

3.1. Decreasing variability of the R_{eq} estimation with experiment duration

As described in the previous section, the assessment methodology was applied to generate 2000 simulations from 2000 different weather datasets. Each simulation provides seven synthetic datasets of growing duration used for the calibration of model T_wT_i R_oR_i A_w. In section 3.1.1, we show how the natural weather variability influences a 2-day calibration and in section 3.1.2, we establish a minimal duration for a robust estimation of R_{eq}.

3.1.1. Variability with a 2-day calibration

For each of the 2000 data sets and for each subset, the stochastic RC model T_wT_i R_oR_i A_w is calibrated. In each case, R_{eq} is inferred as the sum of the resistive parameters estimations. Figure 10 shows on the left hand side 50 randomly picked ML- estimates of R_{eq} with their confidence interval, in order to illustrate the variability of the estimations. The estimations are coloured according to the previously defined interpretability indicator. As a short reminder, it takes values between 0 and 1: the closer to 1, the greener and the closer the estimation to the target value.

Looking at these individual results, three cases can be distinguished:

- The R_{eq} estimation is close to the target R^*_{eq} value: the estimation is accurate and the confidence interval includes the target R_{eq}. This case, visible in the greener colours, is the most desirable case.

- The R_{eq} estimation is far from the target R^*_{eq} value but the confidence interval includes the target R^*_{eq} or the ML-estimate is accurate but with very large uncertainty: the estimation is not accurate but the credible interval relates to this inaccuracy, which keeps the result trustworthy. This case is rendered in orange.

- The R_{eq} estimation is far from the target R^*_{eq} value and the confidence interval does not include the target R^*_{eq}: not only is the result inaccurate but it also give a false sense of confidence in an inaccurate result. These results are visible in red.

The last case is the least desirable one but occurs with many estimations. The right hand side of Figure 10 displays a boxplot af all R_{eq} ML-estimates (dots only) and shows a wide variability: the median of the 2000 ML-estimates falls at 5.36 K/kW with a standard deviation of 0.35 K/kW (5^{th} quantile 4.82 K/kW and 95^{th} quantile 5.98 K/kW). The outlier estimates show absolute errors beyond 20 % of the target R^*_{eq}. This variability confirms the preliminary outcomes by suggesting that the influence of weather conditions on the ML-estimates of R_{eq} is not negligible. A data subset longer than 2 days is certainly needed to decrease this variability.

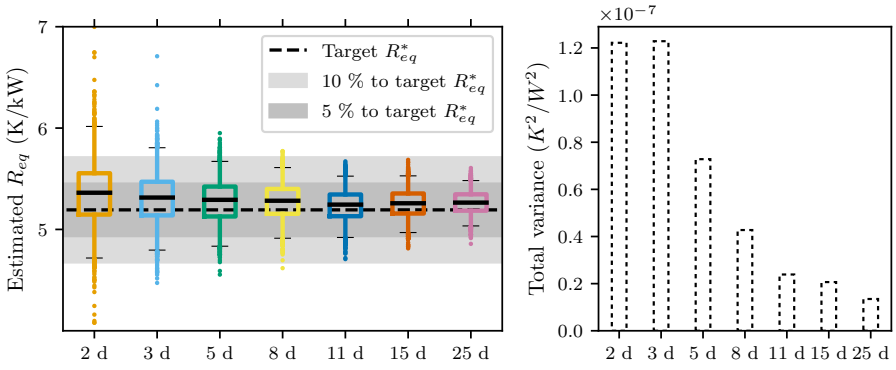
3.1.2. Minimal measurement duration for robust model calibration

The results so far suggest that a period of more than 2 days is necessary to achieve a robust R_{eq} estimation. Figure 11a shows how the variability of all 2000 R_{eq} ML-estimates varies with the seven data subsets: model calibration from 2-, 3-, 5-, 8-, 11-, 15- and 25-day data. From the figure, it can be inferred that the longer the data subset, the lower the variability. There is distinctively a decrease in total variance towards a median value slightly above the target value R^*_{eq}. Calibrations from 11-day data and more show all estimated R_{eq} values within 10 % of their median value, hence ensuring low variability in the R_{eq} estimation with respect to weather influence.

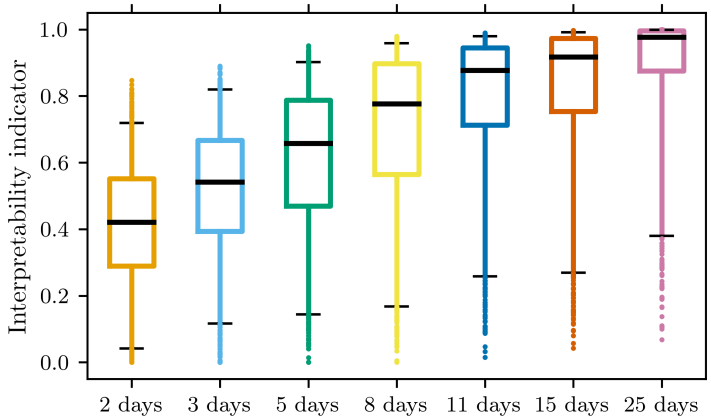
To validate the impression of a decrease in variability on the left hand side in Figure 11a, the right hand side shows for each data subset the evolution of the total variance of ML-estimates. From the 2-day and 3-day data subsets, the R_{eq} ML-estimates have a total variance of approximately 1.2×10^{-7} K^2/W^2. With 8-, 11- and 25-day data subsets, the total variance decreases respectively by a factor 3, 4 and 6. The evolution of the partial variances will be discussed further in the next section.

Studying the variability and variance of the ML-estimates, however, does not reflect properly on the accuracy of the estimation when considering its uncertainty. Figure 11b thus shows the evolution of the interpretability indicator described in section 2.6 with growing datasets. Longer datasets provide increasing interpretability indicator scores. Considering the minimal score of 0.5 as satisfactory, 80 % of the 8-day estimations score higher than 0.5, 90 % of the 11-day estimations and 95 % of the 25-day estimations. Estimations from a period of 8 days and more datasets can therefore be considered as accurate in overall, with a low uncertainty.

As a partial conclusion, 11-day datasets suffice to reduce the error below ± 10 % and provide in 90 % of all cases an acceptable interpretability score. Longer datasets still significantly reduce the overall variance. However, from a practitioner's point of view, longer experiments might be unnecessary, as this



(a) 2000 R_{eq} ML-estimates for datasets with growing duration: datasets over 11 days are all within ±10 % error to the target R^*_{eq}. Total variances indeed decrease with longer calibration datasets.



(b) Evolution of the 2000 interpretability indicators with growing measurement duration: 80 % of the 8-day estimations score higher than 0.5, as do 90 % of the 11-day estimations and 95 %of the 25-day estimations.

Figure 11: Assessing the quality of the R_{eq} estimations through convergence of the ML-estimates within a 10 % error band and through the interpretability indicator.

would immobilise the experimental setup almost twice as long for a relatively small decrease in uncertainty.

3.2. Influential weather variables on an R_{eq} estimation

As mentioned in section 2.3.2, the synthetic weather files allow for a global sensitivity analysis with respect to six weather variables. The variability of the R_{eq} ML-estimates can be attributed to the natural variability of these weather variables. Figure 12a shows the sensitivity indices of the estimations of some parameters with respect to the weather variables: R_{eq}. The sensitivity indices are calculated for all seven data subsets. The indices shown in Figure 12a are the first-order indices, meaning that they only show the direct influence of each weather variable. If the sum of each first-order index is close to 1, it would imply that there were almost no second-order effects, i.e. combined effects of the weather variables. Let us also finally recall that values of sensitivity indices are always simply estimated. The indices given in Figure 12a should mainly be interpreted as order of magnitudes. Indices below 0.1 may be considered insignificant, given the uncertainty of their estimation.

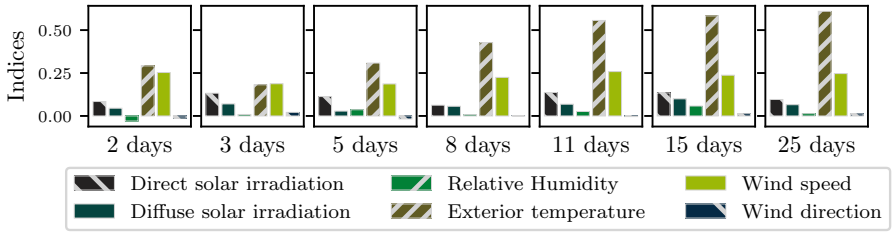
From Figure 12a, it can be seen that the variability of the R_{eq} ML-estimates is mainly influenced by the outdoor temperature and the wind speed. With shorter datasets, the sum of the first-order indices is significantly inferior to 1. This means that the variability is also explained by interactions of weather variables. Variability with longer datasets is by contrast almost only explained by the variability of outdoor temperature and wind speed, seeing that the indices add up to 1. Let us also note that neither the relative humidity nor the wind direction were expected to have an influence on the estimations as they are not used in the infiltration and ventilation model of EnergyPlus. Their sensitivity indices are indeed insignificant.

The influence of outdoor temperature and wind speed is also evident in the evolution of the partial variances of each weather variable shown in Figure 12b. Let us recall that the total variance is the sum of the first-, second- and all higher-order partial variances. The figure shows the first-order partial variances, i.e. the partial variances due to the effect of each weather variable individually. With these elements in mind, it is evident that the total variance of the 11-, 15- and 25-day datasets is solely explained by first-order effects of the weather variables, mainly outdoor temperature and wind speed.

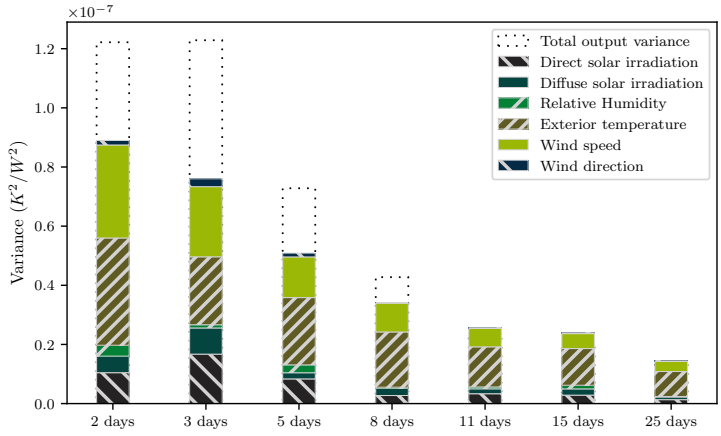
Let us now examine how outdoor temperature and wind speed influence the R_{eq} ML-estimates. Figure 13 shows how the R_{eq} ML-estimates from 11-day datasets vary with the average outdoor temperature on the abscissa and the average wind speed on the ordinate. Warmer periods tend towards over-estimations and colder days towards under-estimations. At the same time, non-windy days produce overall over-estimations, windy days under-estimations.

At the same time, an interaction can also be seen in Figure 13: calibration from warm and unwindy days results in over-estimation, cool and windy days in under-estimations.

This outcome is in agreement with the hypothesis that the large air change rates in the reference model are a cause of inaccuracy in the estimation of the overall thermal resistance. As the ventilation-related heat losses have been modelled in the EnergyPlus simulation environment, there is a direct



(a) First-order sensitivity indices of the R_{eq} ML-estimates: outdoor temperature and wind speed are the main factors in variability



(b) Decrease in the total and partial variances with longer calibration datasets

Figure 12: Individual effect of weather inputs on the estimations of R_{eq}

relationship between the difference in indoor and outdoor temperature and wind speed. Ventilation-related heat losses are greater with cold outdoor temperatures and with high wind speed and on the contrary smaller with warmer and/or non-windy days. It could therefore be expected that acceptable and robust estimations be achieved in less than 11 days in buildings with lower air change rates.

Finally, Figure 14 shows more clearly how the influence of outdoor temperature and wind speed on the R_{eq} ML-estimates evolves from short to longer datasets.

As seen earlier, this figure, too, shows the decrease in total variance of the R_{eq} ML-estimates with longer calibration sets: the vertical spread of all estimations are narrower with the 11-day calibration. Interestingly, while the total variability does decrease, the angle representative of the correlation remains relatively similar whatever the measurement duration. Longer datasets produce averages that are less horizontally spread, but the relationship between temperature and R_{eq} estimation is almost unaltered.

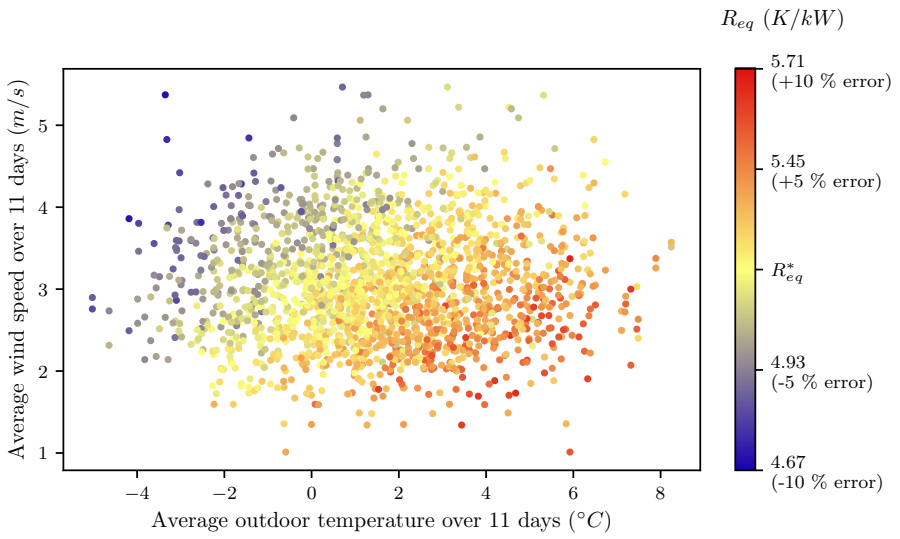


Figure 13: Variability in the R_{eq} ML-estimates from 11-day calibration with respect to outdoor temperature and wind speed. Colours refer to ± 10 % errors to target R^*_{eq}.

A natural assumption would have been to consider that colder days lead to more accurate estimations than warmer days, as colder days increase the heat losses and thus the heating power needed to keep up with the indoor temperature setpoint. This assumption does not seem to hold here. If it were, the variance would be significantly narrower during cold days than during warm days. Here, there is no significant difference in vertical spread between cold and warm days, nor is there any between windy and non-windy days.

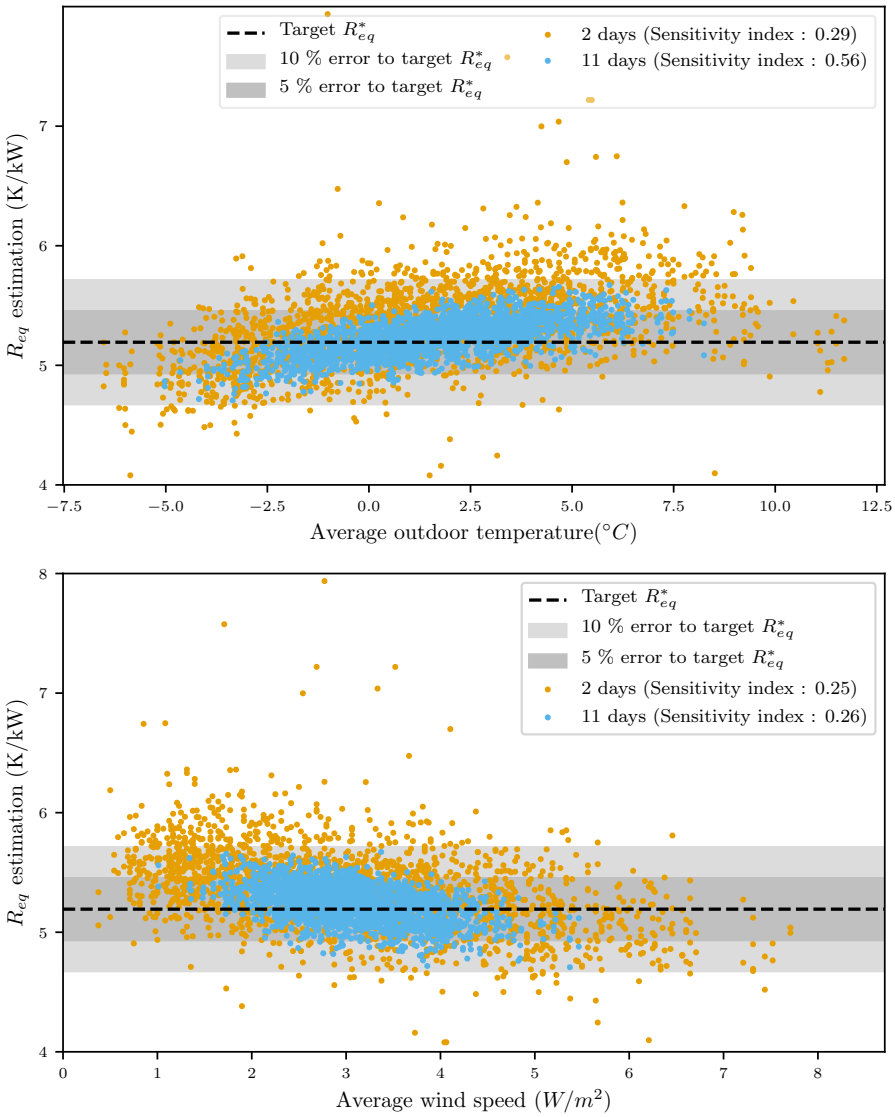


Figure 14: Variability in the R_{eq} ML-estimates with respect to averaged outdoor temperatures and wind speeds, for 2-day and 11-day calibration.

4. Discussion

The results show that, in the particular location, climate and season conditions of this study and in the specific case study, the calibration needs to be based on at least 11 days to ensure convergence within ±10 % of the target value R^*_{eq}. They also show that the longer the dataset, the smaller the

total variance of separate estimations. Calibration from shorter datasets will lead to uncertain results and the variability of the estimations will mainly be due to the variability of the weather conditions. Compared with controlled tests with optimised heating or temperature patterns for which a measurement of a few days is sufficient, 11-day measurements is significantly longer. But considering that optimised heating or temperature patterns create richer and less correlated data, it is quite consistent to find 11 days as a minimum in non-intrusive conditions.

In addition, as non-intrusive measurement design is considered here, 11 days or more is not a prohibitive duration: as long as the test remains user-friendly, leaving data loggers for a few weeks is most probably neither burdensome for the building occupants nor for the expert carrying out the diagnosis. All the more so, compared with data exploitation at building scale by other low-order models such as auto-regressive models as suggested in [20] or with energy signature methods, stochastic RC models have been shown to be a faster approach for exploiting the data, as long as temperature is not kept constant by the occupants. All-day-long constant temperatures would necessitate choosing models which have the heating power as output to exploit the data.

On another note, the results are certainly specific to this case study. Let us indeed recall that the study was conducted in winter conditions, concededly typical, and on a particular building type in Geneva which has a temperate oceanic climate [46]. Whether it is safe to extrapolate the results to other seasons, climates or building types is debatable and needs to be discussed.

Regarding the seasonal and climate-related variability of the results, it can be inferred from the results that larger solar irradiations, lower outdoor temperatures or higher wind speeds will affect the outcomes. This might be the case when this experiment is performed in autumn or spring weather, or in colder or more windy climates. Yet, let us also recall that the same building in different locations will have a different target value R^*_{eq} as it includes heat losses by infiltration. In the end, the results will probably slightly change in much colder or more windy locations, but the extrapolation would still be feasible. It could then be expected that the order of magnitude of the calibration duration would be approximately similar. Such extrapolation would, however, be more risky if the natural variability of one of the weather variables is significantly larger or narrower than that in Geneva. In a colder climate, for example, but with little variability in a winter month, lower outdoor temperatures will certainly have an influence on the amount of heat losses through air change in the building, but will also affect the target value through its air change rate component. In the end, regardless of how cold it might be, low natural variability in outdoor temperatures will barely affect the variability of the R_{eq} estimations. For these reasons, the results of this paper can still be viewed as a benchmark in the field of non-intrusive measurements exploitation. The 11-day minimal duration can then serve as a comparison for other locations, as long as the natural variability of the weather variables is considered.

Summer conditions, as well as dry or tropical climates where active cooling

is needed, are, however, completely out of the scope of the conditions tested in this paper because the proposed experiment uses a heating power signal as model input. If a cooling power-based appropriate methodology were to be developed, the present study may suggest that the variability of the thermal characterisation estimation could be influenced by much larger solar irradiation, correlated with high outdoor temperatures. A minimal measurement duration could not, however, be safely determined from the outcomes of this study and would need further dedicated investigations.

Another distinctive feature of this study is the particularity of the case study: a one-storey internally insulated house, with high insulation in the attics and under the ground floor. Although already discussed, its large air change rate is also a particularity. Whether the outcomes are valid for building with heated or unheated neighbouring zones, for other levels of insulation or for larger buildings such as apartment blocks remains uncertain, not only because the effect of weather variability on the thermal resistance estimation would be different, but also because these conditions raise questions of the feasibility of such non-intrusive experiment in the first place. Further work will be necessary to confirm the feasibility of non-intrusive thermal characterisation in other building types.

A future major development would also be to exploit an actual measurement campaign and for that purpose, an important indicator would be to consider the convergence of the estimation. Let us therefore take the opportunity to make a distinction between convergence of a single estimation and reproducibility of such an experiment. At the scale of a specific measurement campaign, the stop factor would be convergence of the estimation results: continuing the measurements does not significantly change the results.

Although tools for assessing the convergence are not the purpose of this study, one may extrapolate the well-established ISO 9869 standard [12] criteria for wall-scale characterisation to building scale and for results from RC models. First, the R_{eq} estimation should not deviate more than 5% from the 24 h earlier estimation. Secondly, with N the total duration of measurements, the R_{eq} value inferred from the last $2/3N$ days is within 5 % of the first $2/3N$ days.

The second ISO 9869 standard criterion is roughly applied with the available data in Figure 15. The estimation from a 3-day dataset is compared with that from a 2-day dataset by calculating a deviation percentage: $\delta_i = (R_{eq}^{3\ days} - R_{eq}^{2days})/R_{eq}^{2days} * 100$. Then, the estimation from a 5-day dataset is compared with that from a 3-day dataset and so forth. The deviation calculated from the previous estimation is considered satisfactory when it scores below 5 %.

All 2000 deviations calculations are represented as grey shaded areas with 50 %, 75 % and 95 % quantiles. Interestingly, a large majority of cases show a convergence in the sense of the second ISO 9869 standard criteria within 5 days. Eight-day datasets are sufficient for convergence in more than 95 % of all cases. Yet at the same time, reproducibility as defined previously is not quite achieved: the variance caused by weather conditions is still significant.

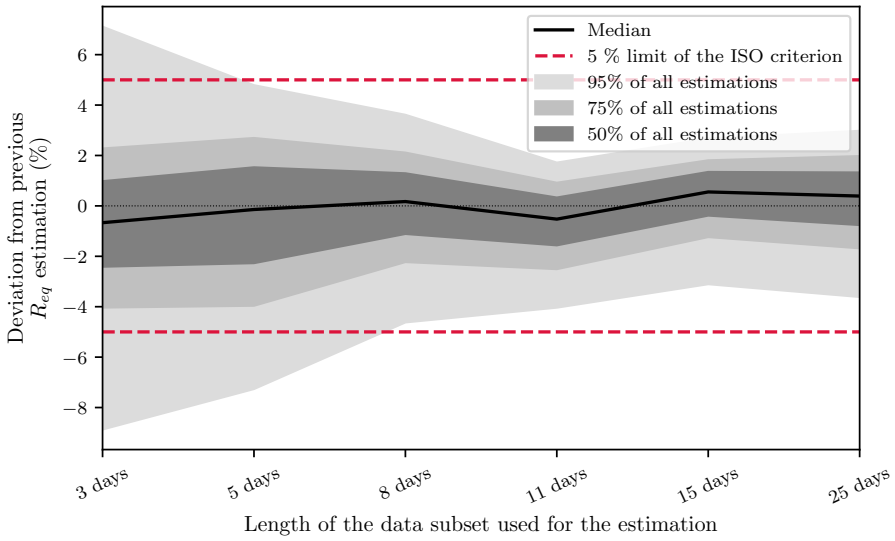


Figure 15: Second ISO 9869-1 convergence criteria: the estimations from N-day measurements should not deviate from the previous (N-1)-day estimations by more than 5 %. Applied here, the second ISO 9869-1 convergence criteria would indicate that a period of 8 days is sufficient.

The third criterion is not directly applicable to the data from this paper, but one may extrapolate that there could be up to 10 % deviation between the first $2/3 \cdot N$ days inference and the last $2/3 \cdot N$ days inference, for example when the first days are particularly cold and the last particularly warm or vice versa. However, with similar consecutive weather conditions, convergence would be considered "achieved" rather quickly.

To summarise the convergence topic, according to these criteria, convergence could in some cases be considered as achieved with fewer measurements than the results from this paper would suggest to reach reproducibility.

The question is then perhaps not to look at deviation in the specific R_{eq} result but instead quantify the information learnt from the data and their evolution, bearing in mind the representativeness of the weather conditions. In this regard, it is a call for a Bayesian perspective on the results: the important outcome to consider is the complete posterior distribution and not the single most probable value of interest. If upon significant variation of weather there is no more information gained from the data, i.e. there is no further change in the posterior distribution, then the measurements may stop.

Figure 16 illustrates for one case how the posterior distribution of a R_{eq} estimation varies with a longer dataset. As a comparison, a 5 % error area around the target value is represented in grey. An estimation from a 2- or 3-day dataset does not provide a satisfactory estimation, seeing that the uncertainties are large. This means that the amount of useful information in short datasets is insufficient for trustworthy estimation. A 5-day calibration provides an under-

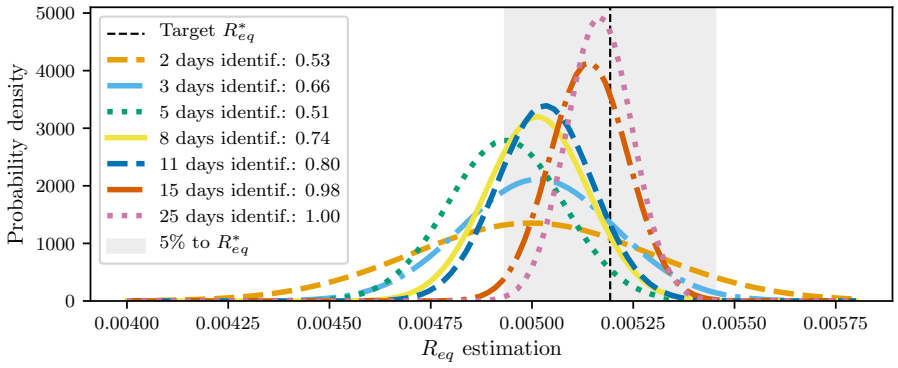


Figure 16: Illustration of the interpretability indicator: the most important is that most of the posterior distribution is satisfactorily close to the target value. Although not perfectly accurate at peak, the posterior distribution is overall within the boundaries of the grey target area, provided there are sufficient data for calibration.

estimation, which is probably related to particular weather conditions. Estimations from 8- or 11- day datasets are relatively similar and could be considered satisfactory. This suggests that data from 8 or 11 days are sufficiently rich to provide a R_{eq} estimation. Finally, estimations from 15 or 25 days are very accurate as their interpretability indicator scores at almost 1.0.

Thus, Figure 16 illustrates how a posterior distribution provides a wider perspective on the convergence of the estimation than the only ML-estimator as suggested in the ISO 9869 standard [12]. The interpretability indicator as defined in relation to the target and henceforth known value can obviously not be a metric for convergence. However, it supports the idea of judging convergence through posterior distribution, by the use of a divergence metric such as the Kullback Leibler divergence. Furthermore, it would make sense to exploit data in a Bayesian approach and, for example, use in-line calibration algorithms such as Sequential Monte Carlo (see [47]).

5. Conclusion

Establishing reliable methods for estimating the thermal performance of buildings remains a challenging issue under the constraint of a non-intrusive measurement framework: the data are collected in non-intrusive conditions, where the indoor air temperature is controlled so as to provide occupant-friendly conditions and are thus less informative. Such data may therefore lead to errors in the thermal diagnosis and the outcome may be uncommonly influenced by the boundary conditions, i.e. the weather conditions.

In this context, this paper has developed an original model assessment framework to investigate the influence of weather conditions on the reproducibility and hence feasibility of the thermal characterisation of a building envelope from measurements in non-intrusive conditions. The methodology

relies on a sensitivity and uncertainty analysis of the overall thermal resistance R_{eq} estimation with respect to six weather variables.

The proposed methodology has proven to be effective for assessing the robustness of the overall thermal resistance estimation. Through the analysis of the variability of all estimations over different measurement durations and through the analysis of their partial variances and sensitivity indices, the minimal measurement duration can be assessed and the main influential weather variables identified.

The paper shows how the methodology is applied to a case study and how a stochastic RC model is used to exploit the data generated by the model assessment framework. After comparing 2, 3, 5, 8, 11, 15 and 25 days of model calibration, it was found that 11 days and longer provide reproducible results regardless of the outdoor conditions.

The variability of the overall thermal resistances R_{eq} estimations from 11 days and longer observed in the outcomes is in the present case study exclusively due to the variability of outdoor temperature and wind speed. This case study has indeed large air change rates which would emphasise the effect of these two weather variables on the overall heat transfers.

Although the 11-day duration is strictly speaking specific to the particular climate conditions and the particular case study, the strength of the uncertainty and sensitivity analysis of this methodology allows us to prudently extend the validity of the results to other cases, as long as similar weather variability remains. Indeed, the robustness of the overall thermal resistance estimation is not simply studied in relationship to the weather conditions seen as absolute values, but rather to weather variability itself. The minimal duration of 11 days found in this application thus gives a sense of the order of magnitude of duration that can be expected from the exploitation of data with stochastic RC models and may serve as a benchmark for future investigations. In addition, the results call for further efforts in establishing reliable tools for the assessment of convergence from a given dataset to later exploit actual on-site experiments.

Acknowledgements

The authors would like to thank the French National Research Agency (ANR) for funding this work through the BAYREB research project (ANR-15-CE22-0003).

CRediT authorship contribution statement

All authors have contributed equally to the conceptualisation and methodology developed in this study. Sarah Juricic and Jeanne Goffart provided the simulated data, with the support of Nicolas Cellier for the numerical implementation. Sarah Juricic conducted the formal analysis, validation and visualisation, with the careful supervision of all authors. The original draft was written by Sarah Juricic, with thorough review by all authors. Funding acquisition was performed by Simon Rouchier.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

[1] ISO 13789, Thermal performance of buildings - Transmission and ventilation heat transfer coefficients - Calculation method, 2017.

[2] European Commission, IN-DEPTH ANALYSIS IN SUPPORT OF THE COMMISSION COMMUNICATIONCOM(2018) 773, Technical Report, European Commission, Brussels, 2018. URL: https://ec.europa.eu/clima/sites/clima/files/docs/pages/com{_}2018{_}733{_}analysis{_}in{_}support{_}en{_}0.pdf.

[3] O. Lucon, A. Ürge-Vorsatz, A. Ahmed Zain, H. Akbari, P. Bertoldi, L. F. Cabeza, et al, Buildings, in: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 2014, pp. 671–738.

[4] Y. Heo, G. Augenbroe, R. Choudhary, Quantitative risk management for energy retrofit projects, Journal of Building Performance Simulation 6 (2013) 257–268.

[5] S. Thébault, R. Bouchié, Refinement of the ISABELE method regarding uncertainty quantification and thermal dynamics modelling, Energy and Buildings 178 (2018) 182–205.

[6] C. Ghiaus, F. Alzetto, Design of experiments for Quick U-building method for building energy performance measurement, Journal of Building Performance Simulation 12 (2019) 465–479.

[7] P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings, Energy and Buildings 43 (2011) 1511–1522.

[8] R. Jack, D. Loveday, D. Allinson, K. Lomas, First evidence for the reliability of building co-heating tests, Building Research and Information 46 (2018) 383–401.

[9] F. Alzetto, D. Farmer, R. Fitton, T. Hughes, W. Swan, Comparison of whole house heat loss test methods under controlled conditions in six distinct retrofit scenarios, Energy and Buildings 168 (2018) 35–41.

[10] C. Balaras, E. Dascalaki, K. Droutsa, M. Micha, S. Kontyiannidis, A. Argiriou, Energy use Intensities for Non-Residential Buildings, in: Proceedings of the 48th International HVAC&R Congress,

December, Union of Mechanical and Electrotechnical Engineers and Technicians of Serbia (SMEITS), 2017, pp. 369–389. URL: https://izdanja.smeits.rs/index.php/kghk/article/view/3315. doi:10.24094/kghk.017.48.1.369.

[11] A. Rasooli, L. Itard, In-situ characterization of walls' thermal resistance: An extension to the ISO 9869 standard method, Energy and Buildings 179 (2018) 374–383.

[12] ISO 9869-1, ISO 9869 Thermal insulation Building elements In-situ measurement of thermal resistance and thermal transmittance Part 1: Heat flow meter method, 2014.

[13] Z. Petojević, R. Gospavić, G. Todorović, Estimation of thermal impulse response of a multi-layer building wall through in-situ experimental measurements in a dynamic regime with applications, Applied Energy 228 (2018) 468–486.

[14] K. Gaspar, M. Casals, M. Gangolells, Energy & Buildings Review of criteria for determining HFM minimum test duration, Energy & Buildings 176 (2018) 360–370.

[15] V. Gori, C. A. Elwell, Estimation of thermophysical properties from in-situ measurements in all seasons: Quantifying and reducing errors using dynamic grey-box methods, Energy and Buildings 167 (2018) 290–300.

[16] V. Gori, P. Biddulph, C. A. Elwell, A Bayesian dynamic method to estimate the thermophysical properties of building elements in all seasons, orientations and with reduced error, Energies 11 (2018).

[17] A. Rodler, S. Guernouti, M. Musy, Bayesian inference method for in situ thermal conductivity and heat capacity identification: Comparison to ISO standard, Construction and Building Materials 196 (2019) 574–593.

[18] A.-H. Deconinck, S. Roels, Is stochastic grey-box modelling suited for physical properties estimation of building components from on-site measurements?, Journal of Building Physics 40 (2017) 444–471.

[19] T. A. Reddy, S. Deng, D. E. Claridge, Development of an inverse method to estimate overall building and ventilation parameters of large commercial buildings, Journal of Solar Energy Engineering, Transactions of the ASME 121 (1999) 40–46.

[20] M. Senave, G. Reynders, P. Bacher, S. Roels, S. Verbeke, D. Saelens, Towards the Characterization of the Heat Loss Coefficient via On-Board Monitoring: Physical Interpretation of ARX Model Coefficients, Energy and Buildings (2019).

[21] M. Senave, S. Roels, G. Reynders, S. Verbeke, D. Saelens, Assessment of data analysis methods to identify the heat loss coefficient from on-board monitoring data, Energy and Buildings 209 (2020).

[22] J. Seem, Modeling of heat transfer in buildings, Ph.D. thesis, University of Wisconsin-Madison, 1989.

[23] Big Ladder Software, EnergyPlus 8.6 Engineering reference Infiltration/Ventilation, 2016. URL: https://bigladdersoftware.com/epx/docs/8-6/input-output-reference/group-airflow.html{#}zoneventilationdesignflowratehttps://bigladdersoftware.com/epx/docs/8-6/engineering-reference/infiltration-ventilation.html{#}infiltrationventilation.

[24] H. Madsen, Time series analysis, Chapman & Hall / CRC, 2008.

[25] D. Maillet, Y. Jarny, D. Petit, Problèmes inverses en diffusion thermique Outils spécifiques de conduction inverse et de régularisation, Techniques de l'ingénieur (2011) 1–2.

[26] M. Leroy, Note technique num. 37 Classification de performance maintenue, Technical Report, METEO FRANCE, 2010.

[27] M. Sengupta, A. Habte, S. Kurtz, A. Dobos, S. Wilbert, E. Lorenz, et al., Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications (www.nrel.gov/publications), Technical Report February, National Renewable Energy Laboratory, 2015. doi:10.1016/j.solener.2003.12.003.

[28] T. L. Stoffel, I. Reda, D. R. Myers, D. Renne, S. Wilcox, J. Treadwell, Current issues in terrestrial solar radiation instrumentation for energy, climate, and space applications, Metrologia 37 (2000) 399–402.

[29] J. Goffart, T. Mara, E. Wurtz, Generation of stochastic weather data for uncertainty and sensitivity analysis of a low-energy building, Journal of Building Physics 41 (2017) 41–57.

[30] ASHRAE, International Weather for Energy Calculations (IWEC Weather Files) Users Manual and CD-ROM, 2001.

[31] G. Pernigotto, A. Prada, A. Gasparella, J. L. Hensen, Analysis and improvement of the representativeness of EN ISO 15927-4 reference years for building energy simulation, Journal of Building Performance Simulation 7 (2014) 391–410.

[32] J. M. Finkelstein, R. E. Schafer, Improved goodness-of-fit tests, Biometrika 58 (1971) 641–645.

[33] G. Ramos Ruiz, C. Fernández Bandera, Analysis of uncertainty indices used for building envelope calibration, Applied Energy 185 (2017) 82–94.

[34] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: A review, Renewable and Sustainable Energy Reviews 23 (2013) 272–288.

[35] T. Bohlin, S. F. Graebe, Issues in nonlinear stochastic grey box identification, International Journal of Adaptive Control and Signal Processing 9 (1995) 465–490.

[36] O. Brastein, B. Lie, R. Sharma, N.-O. Skeie, Parameter estimation for externally simulated thermal network models, Energy and Buildings (2019).

[37] X. Li, J. Wen, Review of building energy modeling for control and operation, Renewable and Sustainable Energy Reviews 37 (2014) 517–537.

[38] R. Kramer, J. van Schijndel, H. Schellen, Simplified thermal and hygric building models: A literature review, Frontiers of Architectural Research 1 (2012) 318–325.

[39] G. Fraisse, C. Viardot, O. Lafabrie, G. Achard, Development of a simplified and accurate building model based on electrical analogy, Energy and Buildings 34 (2002) 1017–1031.

[40] M. C. Kennedy, A. O'Hagan, Bayesian Calibration of Computer Models, Journal of the Royal Statistical Society. Series B (Statistical Methodology) 63 (2001) 425–464.

[41] J. Brynjarsdottir, A. O'Hagan, Learning about physical parameters: the importance of model discrepancy, Inverse Problems 30 (2014).

[42] H. Madsen, J. Holst, Estimation of continuous-time models for the heat dynamics of a building, Energy and Buildings 22 (1995) 67–79.

[43] S. Audoly, G. Bellu, L. D'Angiò, M. P. Saccomani, C. Cobelli, Global identifiability of nonlinear models of biological systems, IEEE Transactions on Biomedical Engineering 48 (2001) 55–65.

[44] G. Bellu, M. P. Saccomani, S. Audoly, L. D'Angiò, DAISY: A new software tool to test global identifiability of biological and physiological systems, Computer Methods and Programs in Biomedicine 88 (2007) 52–61.

[45] L. Raillon, S. Rouchier, S. Juricic, pySIP : an open-source tool for Bayesian inference and prediction of heat transfer in buildings, in: Congrès français de thermique, Nantes, 2019.

[46] M. C. Peel, B. L. Finlayson, T. A. McMahon, Updated world map of the Köppen-Geiger climate classification, Hydrology and Earth System Sciences 11 (2007) 1633–1644.

[47] S. Rouchier, M. J. Jiménez, S. Castaño, Sequential Monte Carlo for on-line parameter estimation of a lumped building energy model, Energy and Buildings 187 (2019) 86–94.