Design of a short perturbation method for on-site estimation of a building envelope thermal performance

Sarah Juricic, Jeanne Goffart, Simon Rouchier, Arnaud Jay, Pierre Oberlć

Highlights

Design of a short perturbation method for on-site estimation of a building envelope thermal performance

Sarah Juricic, Jeanne Goffart, Simon Rouchier, Arnaud Jay, Pierre Oberlé

- Novel 24-hour perturbation method for the estimation of the Heat Transfer Coefficient
- Binary signals with 5^{+}-hour heating and free-floating periods show highest accuracy
- BIC model selection among RC models most accurate for physical interpretation
- Proposed method very accurate in winter conditions
- Mid-season conditions might need slighlty longer measurement durations

Design of a short perturbation method for on-site estimation of a building envelope thermal performance

Sarah Juricic[a], Jeanne Goffart[a], Simon Rouchier[a,*], Arnaud Jay[b], Pierre Oberlé[c]

[a] *Univ. Savoie Mont-Blanc, CNRS, LOCIE, 73000 Chambéry, France*
[b] *CEA/LITEN*
[c] *INES PFE*

Abstract

On-site measurement of the actual thermal performance of a building envelope could considerably benefit to the building sector by securing an effective performance of newly built or retrofitted buildings. Current methods to estimate the Heat Transfer Coefficient, for example, suffer however from either excessive equipment downtime and therefore large costs, or fail to achieve a reasonable accuracy. This paper aims therefore at developing and assessing an original 24-hour perturbation method for estimating the whole-building thermal resistance. Based on a binary heating signal, the proposed method exploits the data collected with stochastic RC models.

The paper first presents a numerical comparison of 398 binary signals along with variable initial indoor temperatures to bring out common characteristics of signals that achieve highly accurate estimations of the thermal resistance. An experimental campaign in a highly insulated wooden-framed house then implemented one particular signal and assessed its reproducibility under variable weather conditions. The paper shows that the proposed 24-hour method indeed achieves a highly accurate estimation when performed in cold weather. In mid-season conditions, the method shows a better robustness from 30- to 36-hour measurements. The results invite to further investigate other building types in variable weather conditions to validate the proposed method.

Keywords: Thermal performance, Heat Transfer Coefficient, Perturbation method, Binary signal, Stochastic RC models

1. Introduction

In a world-wide effort to decrease the carbon footprint of the existing building stock, massive and extensive building retrofit operations are to be expected.

*Corresponding author
Email address: simon.rouchier[at]univ-smb.fr (Simon Rouchier)

Energy performance contracting (EPC) could benefit the investors by securing an effective thermal performance. Achieving a win-win partnership between contractor and client may, however, be jeopardized by a number of risks. Berghorn and Syal [1] identified in particular that the quality of the energy audit is crucial to the success of the EPC. In addition, Mozzo underlined in [2] that the energy audit method must be agreed upon by all parties and must be binding before project implementation. Among the available methods, the author recalls that actual measurements can be a useful, yet maybe costly, method to set a baseline to the energy consumption.

On-site measurements to set a baseline for an extensive building retrofit operation or to assess the performance after retrofit will rely on measurement and verification of the actual intrinsic thermal performance of the building envelope. The estimation of the whole-building Heat Transfer Coefficient (HTC), or its inverse the whole-building thermal resistance R_{eq}, is then an interesting option. For on-site measurements of the HTC to be financially incentive for all parties, a test should fulfil the two following criteria. First it needs to be short as to not be burdensome or costly to carry out. Secondly, it needs to be accurate and robust so that all parties may trust its result.

Therefore, this paper first provides a broad background on data-based methods for whole-building HTC or R_{eq} estimation. The literature review then focuses on so-called active or perturbation methods used for whole-building HTC estimations and more specifically on signal design for system identification.

Background on data-based whole-building HTC or R_{eq} estimation

HTC or R_{eq} estimation from on-site collected data has been a continuous focus for decades now, as summarised in Figure 1. Researchers focussed on either wall-scale or building-scale HTC or R_{eq} estimations. Wall-scale HTC or R_{eq} estimations may however not be representative of the overall thermal intrinsic performance of the building envelope. It does not account for air infiltrations and may overlook heterogeneity of wall composition over the overall envelope as well as thermal bridges. Wall-scale or element-scale methods are therefore not further discussed in this paper. We refer to Soares et al [4] for a recent and broad review of such non-destructive methods.

Whole-building HTC or R_{eq} estimations on the other hand have been documented to rely on data from either occupied or unoccupied buildings. Accurate HTC estimations from data collected in occupied buildings requires several days to several months to converge and be reasonably accurate.

The energy signature method for example, also called change-point regression model, uses several months data. The collected data needs to be aggregated, at least daily [13], or monthly as in Rabl and Rialhe [14]. Even then, both authors underline the difficulty to provide a trustworthy physical interpretation of the estimated parameters. Hammarsten [13] precisely raised criticism against physical interpretation given the unverified assumptions made in such models.

Using RC models, Hollick et al. [15] inferred an HTC estimation of two case studies from respectively 5 and 10 days data: one was occupied the other had

Nomenclature

ΔT	Difference in indoor and outdoor temperature $T_i - T_{ext}$ (K)
η_{HS}	Heating system efficiency (%)
Φ_i	Other indoor heat gains (W)
Φ_{sol}	Solar heat gains (W)
A_{sol}	Solar aperture (m^2)
$I_{10\%}$	10% interpretability score, see section 2.5 (no unit).
I_{sol}	Solar global horizontal radiation (W/m^2)
P_h	Measured heating power (W)
R_{eq}	Overall or whole-building thermal resistance (K/W). R_{eq} is the inverse of the whole-building HTC. Target numerical or experimental value is noted R^*_{eq}
S_{wind}	Wind speed (m/s)
AIC	Akaike Information Criterion
BEM	Building Energy Model
BIC	Bayesian Information Criterion
EPC	Energy performance contracting
HTC	Heat Transfer Coefficient (W/K): overall or whole-building heat transfer of the building envelope towards exterior as defined in the ISO 13789 standard [3]. The HTC is the inverse of the whole-building thermal resistance R_{eq}.
MLS	Maximum Length Sequence
PRBS	Pseudo-Random Binary Signal
ROLBS	Randomly ordered logarithmically distributed binary sequence
SA	Sensitivity analysis, in this paper performed with the EASI RBD-FAST method

occupant-friendly indoor temperature set-points and internal electric gains but no actual occupants. They used a first-order RC model and investigated several models accounting for solar gains. HTC estimations fell under 10% error in winter, although the estimation uncertainties seem overconfident compared to

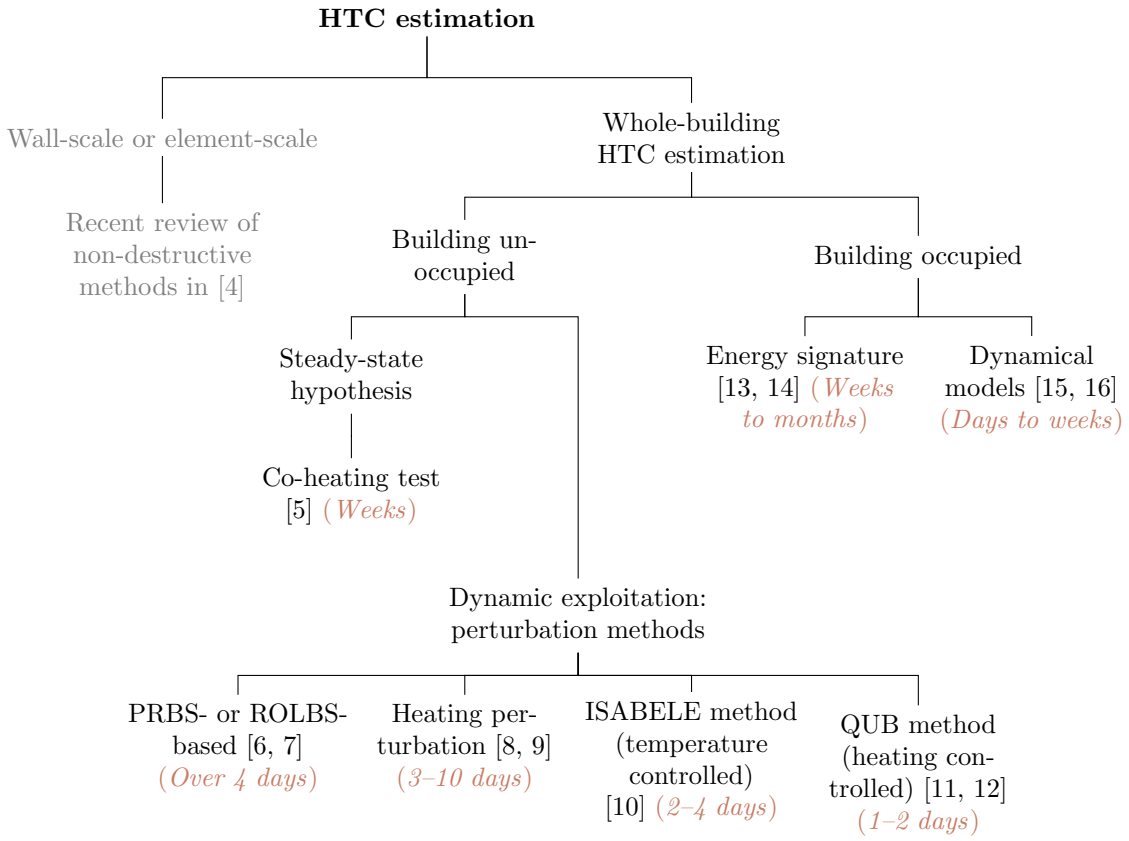


Figure 1: Literature background on whole building HTC estimation. Duration dataset needed is given in orange italic.

the target values. Senave et al [16], in a fully numerical study, trained grey-box ARX models on 26 weeks synthetic winter data. They found that the models tested are unable to provide a satisfactory HTC estimation when indoor and solar heat gains are neglected but give an acceptable estimation when the global solar radiation is taken into account.

More generally, it can concededly be assumed that these methods require a very reasonable amount of sensors which makes such methods easily deployable. However, long equipment downtime may be problematic for practitioners and occupants. In addition, Baasch et al [17] found in a numerically based study that occupancy contributed greatly to the uncertainty of HTC estimations inferred from both grey- and black-box models.

Unoccupied premises allow on the contrary to fully control the indoor ambiance and therefore provide measurement conditions that are favourable for HTC or R_{eq} estimations.

Based on the hypothesis that the building energy balance can reach steady-state under particular conditions, the co-heating test sets a constant indoor temperature with electric heater and fans during several weeks. During the test, the shutters are closed and the mechanical ventilation system, if there is any, is shut. With daily averaged data, the steady-state assumption can be roughly considered to hold. A multiple linear regression model allows then to infer an estimation for the HTC. Physical assumptions and thorough literature review can be found in Bauwens and Roels [5]. Let us underline that, notwithstanding its strong assumptions on building physics, the co-heating test is very often used as a reference value for comparison, such as in [18] or in [10].

To avoid, again, an equipment downtime of several weeks, other methods aim at determining the dynamical characteristics of a building envelope and succeed within a few days. These methods are developed on the basis of fully controlled and optimised, in a loose sense, indoor conditions. They are usually called active or perturbation methods, given that a perturbation is induced in the building by controlling the indoor conditions. The response to this perturbation is then measured and used to infer the thermal characteristics of the envelope. The next paragraphs review a few significant and recent contributions to envelope thermal characterisation by active methods.

Madsen and Schultz presented in [6] an identification method based on a pseudo-random binary (PRB) heating signal. Their idea was to use a highly informative perturbation signal to excite the building envelope system in both large and short frequencies. PRB signals have the advantage of exciting the building envelope within a desired range of frequencies, here chosen to be representative of the expected time constants of a building. PRB signals shift between two constant levels of heating (for example on–off) and have the very desirable property not to be correlated to the outdoor conditions. In [6], the authors calibrated a second-order stochastic RC model and concluded that two different PRB signals were necessary to identify with better accuracy both short time and long time dynamics of the building measured.

In [7], Baker and van Dijk use a similar type of heating signal called randomly ordered logarithmically distributed binary sequence (ROLBS) to calibrate a RC model of a Passys test facility. A ROLBS sequence is very similar to a PRBS sequence, with ordered frequencies across the test duration, as can be seen in Figure 2b.

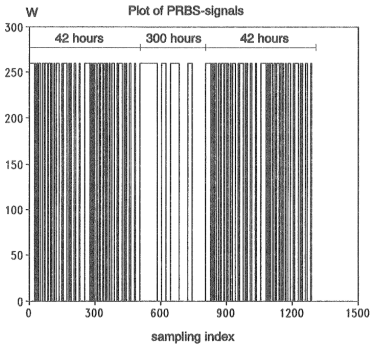
Thébault and Bouchié described and refined in [10] a HTC estimation method, called ISABELE, based on a 2- to 4-day experiment (see figure 2c). During an ISABELE test, the indoor temperature is regulated at a fixed value, for example 30 °C. The authors compared the proposed ISABELE method with other types of signals by means of an experimental campaign in a test cell: constant heating power of various duration, constant temperature setpoint of various duration, pseudo-random heating binary signal and a triangle-shaped heating signal. A set of 20 RC models, from first- to third-order models, is calibrated from the collected data of each experiment and a likelihood ratio test as in Bacher and Madsen [19] is performed to select the best-fitting model. They found that the repeatability was satisfactory for both the temperature and the

heating control strategies for the 2-day and longer tests. The heating controlled tests showed a faster convergence, within 2 days. Shorter tests, such as the nightly tests, seemed to fail to provide repeatable results. Last but not least, a particularity of the ISABELE method is to propagate measurement systematic uncertainties onto the final estimation. The authors found on 2-day tests on an individual house and after uncertainty propagation that the relative error to the target value did not exceed 15–20 %.

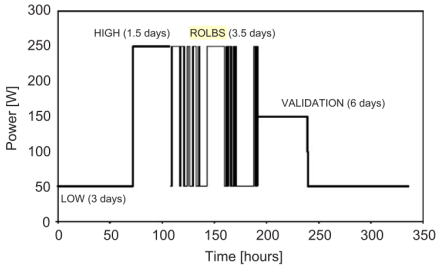
Brastein et al [8] assessed the identifiability and physical interpretation of a second order RC model. The model was calibrated using a 10-day dataset shown in Figure 2d. The heating perturbation signal is a binary signal. Unlike PRB signals, this signal does not cover all frequencies in a given range, but does cover a few short and a few large frequencies. This test duration seems rather long compared to the other perturbation methods. Optimisation of the heating signal was indeed not the objective of the paper. However, the authors investigated what they call the consistency of dynamic information. They calibrated the RC model with a 2- or 4-day data subset and with a given offset. The authors showed how the estimations of four thermal parameters varied with the offset, in particular with the 2-day data subset. Although the authors did not conclude on a preferential signal design for parameter estimation, this result definitely implies that part of the signal is more informative.

In [9], Rouchier et al investigated parameter identification and indoor temperature forecasting from various perturbation signals. The paper aimed at comparing stochastic and deterministic RC models and the target HTC value seems unknown, which is why the authors did not focus on physical interpretation from parameter estimation given each signal. However, signal shape and identified parameters can be discussed. Each sequence had a 3–5-day duration with at some point during the test an imposed 2 kW heating period of either 24 or 48 hours. When not heated, the test box is left in free-floating conditions. A first-, second- and third-order models were used to exploit the data. Noteworthy is the HTC estimation variability : sequence K1 systematically under-estimates the HTC compared to the other sequences whereas sequence D1 tends to over-estimate it. From this paper can be understood that binary signals could be, to a certain extent, optimised for an accurate HTC estimation.

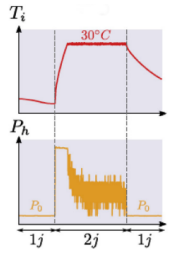
The shortest test found in the literature is the QUB method [11], which is performed overnight, as illustrated in Figure 2e. The night signal is split in two regimes. During the first n_h hours of the night, the heating power is set at power P_1. During the remaining n_h hours, the heating power is set back at power P_2. This signal induces a rise and decay of the indoor temperature. The slopes at the end of the rise period and at the end of the decay period are measured and used to infer a value of the HTC. According to a numerical study in Ahmad et al [12], the accuracy of the HTC estimation is moderately influenced by the initial conditions, by the weather conditions and by the choice of heating power P_1 and P_2. Overall, the error is expected to be bounded within a 20 % range. Unlike the other perturbation methods, based on data exploitation by RC models, the QUB method does, however, not provide any uncertainty range of the estimated HTC.



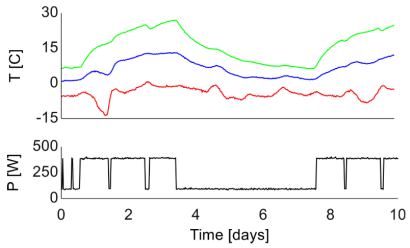
(a) Heating power induced perturbation in Madsen and Schultz (see Fig. 3 in [6]. Note from the authors : "the sampling index corresponds to 5 minutes at the two periods of 42 hours, and to L hour at the long period of 300 hours")



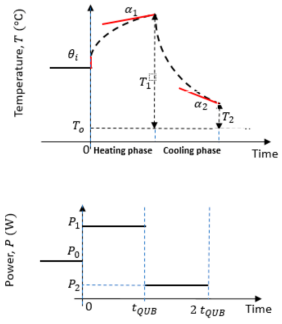
(b) ROLBS heating sequence perturbation in Baker and van Dijk (see Fig. 10 in [7])



(c) Heating power induced perturbation in the ISABELE method (see Fig. 11a in [10])



(d) Heating power induced perturbation in Brastein et al (see Fig. 18 in [8])



(e) Heating power induced perturbation in the QUB method (see Fig. 1 in [12])

Figure 2: Perturbation signals used for thermal characterisation of a building envelope

From general perturbation signal theory to easily deployable whole-house HTC estimation method

More generally, perturbation signals are efficient to make linear system identifications [20] and therefore efficient to provide insight into the thermal dynamics of a building envelope. If the input signal, i.e. the perturbation, is indeed carefully designed, it should maximize the information contained in the collected measurements and therefore allow for highly accurate parameter identification, at least for a given experimental duration. In addition, in the case of building envelopes, a well-designed heating signal is also expected to make the identification less sensitive to the uncontrollable outdoor weather conditions.

The previously surveyed methods all either used a heating power based perturbation binary signal or a temperature regulated perturbation. Other perturbation signals are, however, available in the literature for system iden-

tification. Among many available references, Godfrey et al [20] conveniently reviewed readily accessible perturbation signals whereas Tan and Godfrey [21] proposed a guide to the design and selection of perturbation signals. Signal types and applications are also thoroughly discussed in [22]. Shortly, other than pseudo-random binary signals (PRBS), linear system identification can indeed be performed by the use of pseudo-random multi-level signals (PRMLS), Discrete Interval Binary or Ternary (DIB/DIT) signals, Multilevel Multiharmonic (MLMH) signals, Multisine signals or a hybrid combination of these signals.

These perturbation signals are not binary and can take three or more values. Creating such a signal is concededly feasible, provided that electric portable heaters used to deliver the heating power indoors are equipped with appropriate power regulators. Whether a multi-level power supply and the induced indoor temperature can easily be made homogeneous across all zones during a whole house test is however less certain. In addition, the available amount of heating power to be delivered is also limiting. In France for example, 6 kW electric power is often the upper limit in a house. Therefore, a first reasonable step would be to use a binary signal with constant heating power input up to 6 kW.

Should a binary signal be used, another practical limitation would be the cost. The applications in [6] or in [7] showed that both PRBS and ROLBS tests run over several days, which is prohibitive in terms of equipment downtime. To deploy such methods widely and for a reasonable cost, a 24-hour test would be more easily acceptable among practitioners. The literature review showed that experiments under 48-hours might be achievable, but none of the signals used in the presented methods were optimised in the sense of signal design. It is very likely that richer optimised signals could achieve more accurate results within possibly 24 hours, provided that data are exploited with an adequate model.

Objectives of this paper

Within the scope of the SEREINE project in the PROFEEL program[1], the work presented in this paper aims at developing and assessing the practical limits of a 24-hour test for the estimation of the whole-building thermal resistance. In particular, the focus of this paper is placed on accurate estimations even for highly insulated buildings, which are expected to be difficult to characterise given their higher time constants.

Given the desirable properties of binary signals, this paper intends to design a 24 h binary heating signal that would enhance the accuracy of the estimation of an overall thermal resistance R_{eq} of a building envelope. The paper proposes therefore a numerical methodology to characterise well-suited binary heating signals as well as suitable initial temperature conditions to achieve satisfactory estimations of an overall thermal resistance. Then, an experimental campaign tested one particular binary heating signal on a full-scale experimental house. Results and reproducibility are also presented to discuss the practical feasibility of such short experiment.

[1] programmeprofeel.fr

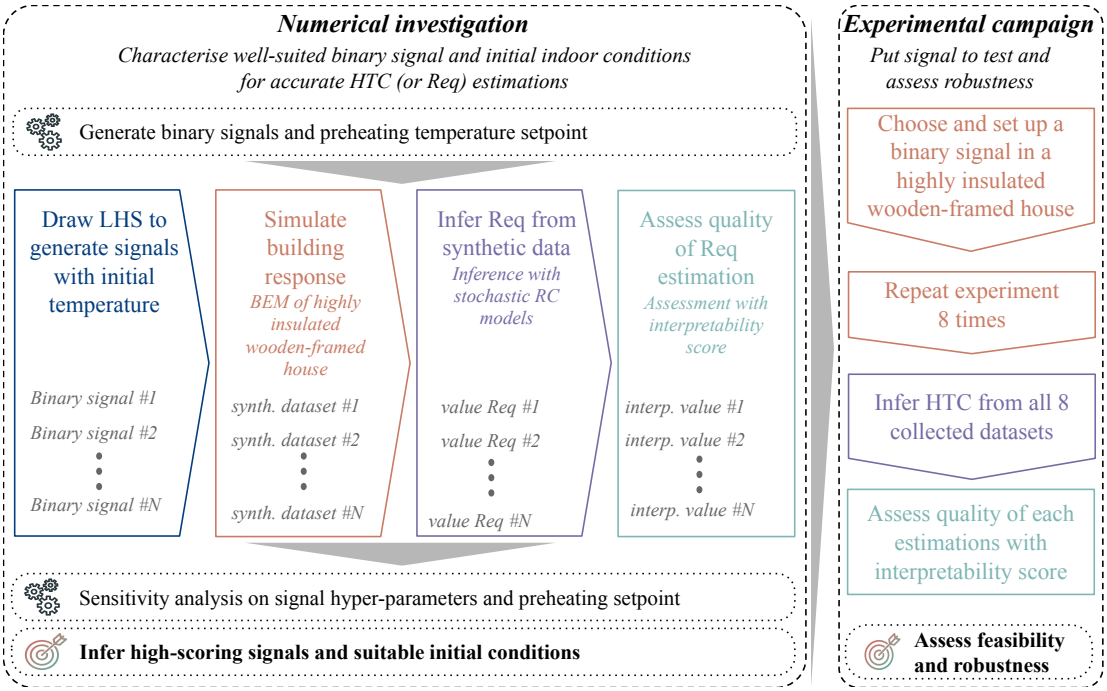


Figure 3: Overview methodology for the design and the experimentation of a 24-h on-site test for accurate R_{eq} (or HTC) estimation based on a binary heating signal

2. Methodology

2.1. Overview of the methodology

The methodology adopted in this paper, as illustrated by Figure 3, is twofold: first the design of satisfactory binary heating signals is performed in a numerical investigation, then a signal is implemented during an experimental campaign in an actual house for validation.

The numerical investigation first generates a wide variety of binary signals, each with a different pre-heating temperature. This step is detailed in section 2.2. Then, these signals are implemented in a high-fidelity building energy model (BEM) of a wooden-frame house, which is detailed in section 2.3. The response of the BEM to each signal is then simulated under known outdoor conditions. From this synthetic dataset, an estimation of the overall thermal resistance R_{eq} is inferred. The inference is done with calibrated stochastic RC models and is described in section 2.4. Whether the inferred R_{eq} estimation is satisfactory is determined with an interpretability index which is detailed in section 2.5. In the end, the numerical investigation therefore allows to draw a map of theoretically low- and high-scoring binary signals and to infer favourable initial temperature conditions.

To validate the results, a particular binary signal is selected and implemented in an experimental campaign. The signal is tested in the actual house that

served as case study for the BEM. The signal is tested several times as to assess reproducibility. Information on the house can be found in section 2.3 whereas details on the experimental setup are given in section 4.2.

2.2. Generating and mapping suitable binary heating signals for R_{eq} estimation

2.2.1. Generation of a wide variety of binary heating signals

Since the test is meant to last only 24 hours, the binary heating signal is not, stricto sensu, a pseudo-random signal. However, using pseudo-random signal generation methods allows to cover a very wide variety of binary signals that can be truncated to 24 hours. For this reason, a maximum length sequence (MLS) method is used to generate binary signals.

Originally, the objective of a MLS is to produce a signal which contains all possible frequencies given a specified length, and which has very low auto-correlation. By truncating the signal to 24 hours, it concededly annihilates these original properties. Nevertheless, MLS produces a large variety of signals with very few hyper-parameters. In this paper, only three hyper-parameters were used. To produce as many signals with hyper-parameters directly describing heating and free-floating periods, much more parameters would have been necessary, which makes the MLS solution preferable.

The MLS is generated with linear feedback shift registers. An algorithm of this generator is implemented in the Python library SciPy [23]. The algorithm can be used with physically-scaled hyper-parameters which allows to link their values to the shape of the binary heating signal. The largest hyper-parameter is the first heating period of the signal (in hour), the difference between the largest and second largest hyper-parameter is the first free-floating duration (in hour) and the difference between the second largest and last hyper-parameter is the immediately next heating period (in hour). If this last difference is negative, as can be the case for some signals, it produces a very short second heating period. In any case, it also contributes to creating a wide set of variable binary signals.

The scaling trick is illustrated here in Code 1 with ordered hyper-parameters 14/6/5 that generate a signal that starts with a 14-hour heating period. To be in line with experimental and practical feasibility considerations, the binary signal produced with this code has a time step of 5 minutes. This implies that the duration of the heating or free-floating periods are multiples of 5 minutes.

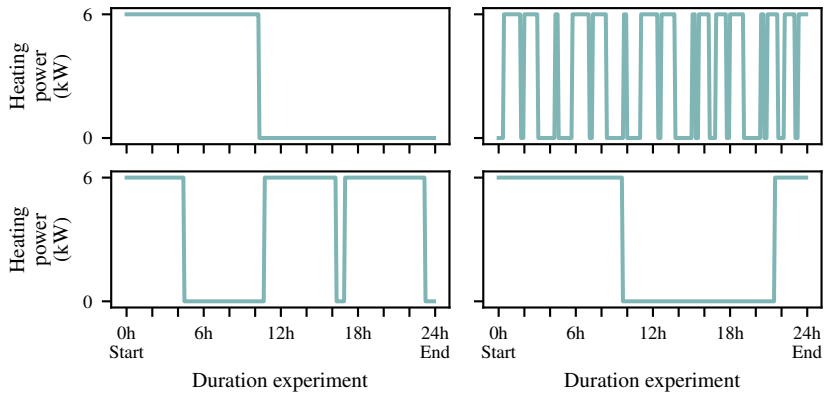


Figure 4: Four binary signals that can be generated from a 3-hyper-parameter MLS generator

Code 1: Generating a binary heating signal of predefined duration with scaling trick

```
# hyperparameters for binary signal generation
hyperparameters = [14, 6, 5]  # order of magnitude of an hour,
    represents heating and free-floating durations
timestep = 5  # minutes
length = 24 * 12  # duration 24h with 12 states/h (time step 5
    minutes)

# inputs for max_len_seq
taps = [int(i * 60 / timestep) for i in hyperparameters]
nbits = max(taps)

# generate pseudo-random binary signal with predefined duration
signal,_ = max_len_seq(nbits=nbits, taps=taps, length=length)
```

Figure 4 shows four examples of 24-hour signals that can be generated with a 3-hyper-parameter MLS-generator. Both single heating periods and multiple heating and free-floating periods can be generated which indeed ensures a wide coverage of pseudo-random signal types over a 24-hour timespan.

2.2.2. Mapping suitable binary heating signals for R_{eq} estimation

Finding a suitable binary signal for an accurate R_{eq} estimation could be performed by optimisation of the MLS hyper-parameters by minimizing a measure of the error between estimated and reference thermal resistance. An optimisation routine on the MLS parameters is, however, not guaranteed to be successful. It would indeed rely on a major and unverified hypothesis that there exists a global maxima for the accuracy indicator assessing the quality of the estimation. Instead, there could be many sets of MLS parameters achieving similar outcomes and/or multiple local maxima could also co-exist.

Exploring the hyper-parameter space and searching for optimal areas is therefore a safer option and allows for a global glance at all signals which should

help uncover common characteristics. To efficiently explore the 3-dimensional hyper-parameter space, a Latin Hypercube Sampling is performed to draw samples under a uniform distribution: $\mathcal{U}(0.5, 24)$ for the first parameter and $\mathcal{U}(0.25, 22)$ for the other two parameters.

If, however, a set of hyper-parameters produces a signal with a heating or a free-floating period shorter than 15 minutes, it is discarded. Periods shorter than 15 minutes are indeed difficult to implement due to limitations of the electric heaters used for the experimental campaign and are in any case less informative.

In addition to varying the binary signal shape, the pre-heating temperature setpoint is also investigated and included in the Latin Hypercube sampling. The objective is indeed to ascertain that the initial indoor temperature, when kept in a usual range, is not influential on the R_{eq} estimation. It can indeed be expected that, regardless of the initial temperature, the dynamic induced by the binary signal alone suffices to provide an accurate R_{eq} inference. The pre-heating temperature setpoint, in °C, follows a uniform distribution $\mathcal{U}(15, 23)$.

In the end, after clearing out the inapplicable signals, a set of 398 hyper-parameter triplets are used in the numerical investigation.

2.3. The case study and its building energy model

2.3.1. Description of the experimental INCAS house

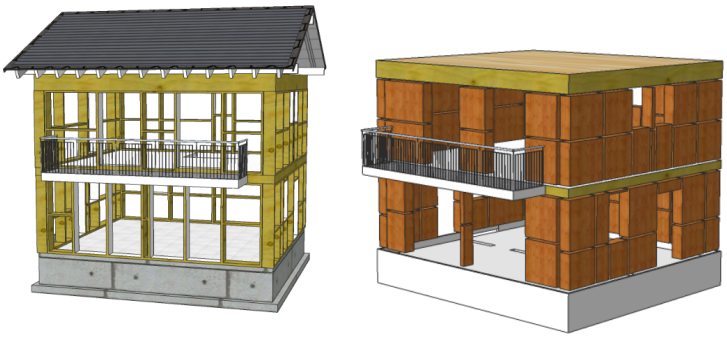
The case study used in this paper for both the numerical investigation and the experimental campaign is a wooden-framed two-storey house in the French Alps (see Fig. 5a). This house is one of the four experimental houses called INCAS, built on the full-scale Building Energy platform at the French National Solar Energy Institute (INES). The platform is located in Le Bourget-du-Lac (N: 45°650, E: 5°867).

Regarding geometry, all four houses are oriented at 15.3° on a North/South axis. Their geometries are simple, compact and similar in shape. Their net usable area is around 95 m^2 with on the first floor an entrance, kitchen, living room, store room as well as lavatory and on the second floor 3 bedrooms and a bathroom. They have been designed to achieve similar low-energy needs and are equipped with approximately 100 sensors for experimental purposes.

In this paper, the wooden-framed house (so-called I-OB visible in Figure 5) is used as case study. This house is a low thermal inertia building built in 2010. As suggested by Figure 5b, house I-OB has unheated but heavily insulated attics and crawlspace. The insulation of the walls and attics is made of wood wool whereas the insulation from the crawlspace is made of polyurethane. Across the envelope, the U-value varies from 0.08 to 0.17 W/m^2K. The windows are equipped with triple-glazing on the South- and West-façades. On the East- and North-façades, two double-glazing windows are installed, with a 15 cm air gap between each window.

A 32-day co-heating test was performed late 2020 and established the overall thermal resistance $R_{coheating}$ at 15.7 K/kW (standard deviation 0.5 K/kW). The airtightness has also been measured with multiple blower-door tests in December

(a) The experimental house is a two-storey wooden-framed house (net floor area $\approx$ 95 m^2, measured overall thermal resistance R^*_{eq} = 15.7 K/kW)



(b) Timber-structure and insulation views of the I-OB house

Figure 5: Case study : wooden-framed INCAS house

2020. Its air change rate at 50 Pa n_{50} is in the range 2.85–2.97 h^{-1} (with a $\pm$5% uncertainty) and can be considered quite efficient. A larger air change rate would have been problematic, as it would contribute greatly to the energy balance with a strong dependence on wind speed and outdoor temperatures. Here, sensitivity of the overall thermal resistance estimation to wind speed and outdoor temperatures should be limited.

2.3.2. Building energy model of the INCAS house

For the numerical investigation, a 5-zone energy model of the wooden-framed INCAS house was used and implemented in the simulation software EnergyPlus. Ventilation is not modelled as to mimic experimental conditions, where air vents would be sealed and mechanical ventilation switched off. In addition, in order to keep the focus on the thermal behaviour of the envelope, air infiltrations are assumed null. The time characteristics of the envelope alone are indeed expected to be significantly larger than for air infiltration or ventilation. In

the case of buildings with large thermal inertia, it can be expected that a 24-hour test might fail to provide an accurate estimation of R_{eq}, precisely because the largest time characteristics in such buildings are too high. Air infiltration on the other hand might concededly affect reproducibility of a 24-hour test, in particular under windy conditions. Wind speed will be kept in mind when exploiting the experimental campaign results.

The overall thermal resistance of the building energy model is $R^*_{eq} = 23$ K/kW. It is calculated by an appropriate simulation of the building energy model under constant outdoor temperature and no solar radiation. The indoor temperature is set at a constant setpoint. In these conditions, when averaged daily, the heating power delivered to maintain the indoor temperature setpoint is a direct function of the indoor outdoor temperature difference with $\overline{P}_h = \Delta\overline{T}/R^*_{eq}$, from which a target value of R^*_{eq} is calculated.

This numerical reference value is higher than the measured thermal resistance of the actual building ($R_{co-heating} = 15.7$ K/kW). The absence of air infiltrations, modelling simplifications and highly possible discrepancies between model and actual envelope composition in terms of thermal bridges and heterogeneity of a wooden-framed envelope probably explain this difference, which should be kept in mind for later discussion of the results.

The building energy model is equipped with electric convective heaters in each zone. The total heating power available is 3.1 kW and is split among the five zones as to allow reasonably close indoor temperatures across all zones. The 3.1 kW heating power available in the building energy model is lower than what can be installed in an actual experiment, which is expected to be up to 6 kW. However, given the larger thermal resistance of the model, delivering more heating power would be followed by a indoor temperature rise so steep that the 35°C limit would be attained within a few hours, which in turn will jeopardise the identification process. In practice, the number of heaters during an actual experiment would also be adapted to the expected thermal performance of the house.

2.4. Data exploitation : inference of R_{eq} with stochastic RC model calibration

Whether it be with synthetically generated data or with experimental data, an adequate model is needed to exploit the data and infer the thermal resistance R_{eq}.

The literature review previously revealed that stochastic RC models are almost always chosen to exploit dynamic data, with the exception of a few papers that used auto-regressive models. This tendency can be understood in the light of review [24] on techniques and tools for building energy performance prediction. The authors found that hybrid or grey-box models combine the advantage of being physically interpretable while needing a reasonable amount of data for training. RC models therefore are a sensible choice for exploiting 24-hour dynamic data.

Ideally, data exploitation by RC models is done by a very large number of different models, among which, for each dataset, a best-fitting model is chosen. However, a very large model bank will be computationally costly, which is

Model name	Model order	Number of parameters	Calculation R_{eq}
T_wT_i R_oR_i in Fig. 6a	2	8	$R_o + R_i$
T_wT_i R_oR_i R_p in Fig. 6b	2	9	$1/(R_o + R_i) + 1/R_p$
$T_wT_iT_m$ R_oR_i in Fig. 6c	3	11	$R_o + R_i$
$T_wT_iT_m$ R_oR_i R_p in Fig. 6d	3	12	$1/(R_o + R_i) + 1/R_p$

Table 1: Description of the four RC models calibrated in this study

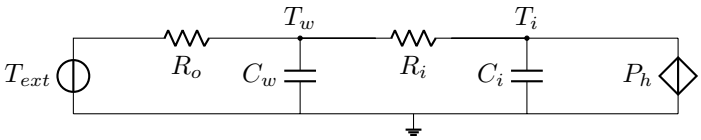
precisely the problem encountered by Baasch et al [17]. To alleviate this issue, Baasch et al chose to narrow their study down to only one first- and one second-order model, without any model selection process, which makes interpretation of the results all the more difficult.

This paper adopts instead the following compromise : pre-select a set of likely models and then apply model selection. The models calibrated in this paper are described in Table 1 and represented in Figure 6. A preliminary model selection process had indeed found that first order models fit very poorly the data whereas fourth order models seemed to be constantly overfitting and yielded extremely correlated parameters. Furthermore, solar aperture coefficients were systematically found non significant which is why there are none in the chosen models. This is barely surprising since the shutters are simulated closed during the numerical experiment and are also closed during the experimental campaign.

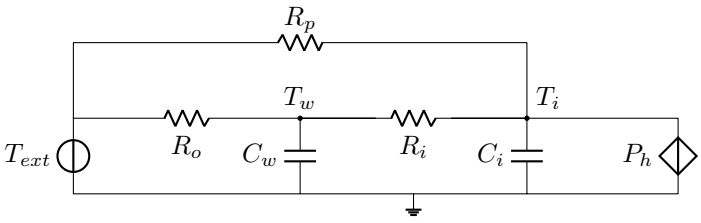
A stochastic formulation of the RC models is used in agreement with [6], to account for model prediction discrepancy. Calibration of each model is performed in a frequentist approach with the pySIP python package [25], using the BFGS algorithm.

After all models are calibrated, it is necessary to select the best-fitting model to infer a single thermal resistance. How the overall thermal resistances R_{eq} are calculated is described in Table 1. It is indeed possible to find a disagreement between all model-inferred resistances. Model selection procedures for physical interpretation is under discussion in the literature. Deconinck and Roels [26] as well as Thébault and Bouchié [10] proceed to stochastic RC model selection by the use of the likelihood ratio statistical test as described in Bacher and Madsen [19]. Other non hierarchical selection have also been used in literature, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC). Rasmussen compares both criteria for ARX model selection in [27] and underlines that BIC is the most likely to avoid overfitting although the selected model might show auto-correlated residuals which makes model validation difficult. Hastie et al [28] also underlined that, with non-finite samples, the BIC often selects too simple models, compared to the AIC. Too simple models are indeed likely to show auto-correlated residuals, when they do not capture all important thermal dynamics.

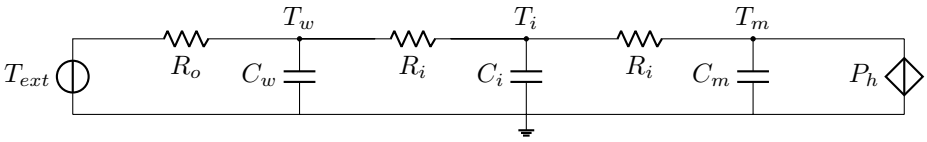
As part of the study, these three model selection procedures, likelihood ratio



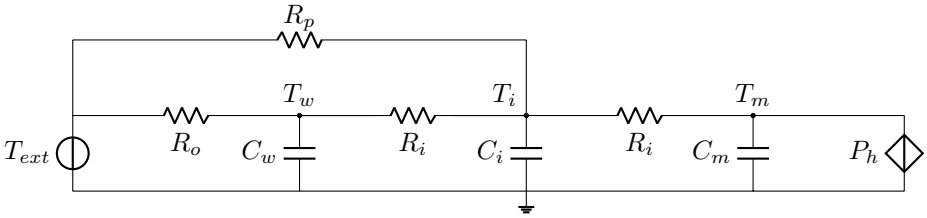
(a) Model $T_w T_i\ R_o R_i$



(b) Model $T_w T_i\ R_o R_i\ R_p$



(c) Model $T_w T_i T_m\ R_o R_i$



(d) Model $T_w T_i T_m\ R_o R_i\ R_p$

Figure 6: Four RC models used in the model selection procedure

test, AIC and BIC, will therefore be compared in the results section and their individual performance discussed in the light of physical interpretation of the calibrated parameters.

2.5. Thermal resistance R_{eq} interpretation and assessment

Each dataset provides an estimation of the overall thermal resistance R_{eq} which can be compared to a target value R^*_{eq}. Let us recall that the target value for the synthetically generated data is the numerically calculated resistance from the building energy model whereas the experimental target value is the estimated thermal resistance from the co-heating test.

A mere relative error between the estimated and the target values only gives partial knowledge for interpretation: it does not account for uncertainty. Instead

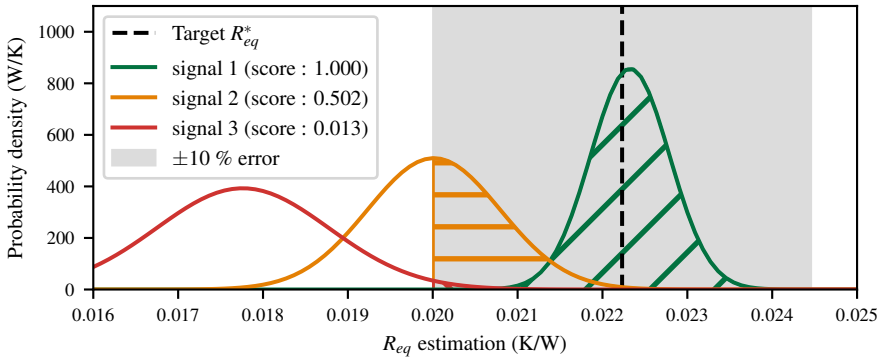


Figure 7: Illustration of the interpretability score : it takes values between 0 and 1. Scores close to 1 are considered very satisfactory estimations whereas scores close to 0 are unacceptably far from the target value.

of a relative error based comparison, an interpretability score is defined and used. This score is defined as the area under the R_{eq} estimation curve that is within a $\pm x$ % error range to the target value R^*_{eq}. In the numerical investigation, a 10 % error range is set here to allow for satisfactory discrimination among all estimations.

As illustrated in Figure 7, the interpretability score takes values between 0 and 1. A 0-score means that the estimation is completely outside the ±10 % error range whereas a score close to 1 means that the estimation and its uncertainty are entirely within the ±10 % bounds. Scores higher than 0.5 can be considered as acceptable as it means that half of the area under the curve is within the ±10 % error range.

3. Results of the numerical investigation of binary signals

This section presents the results of the numerical part of the study which aims at characterising the type of binary signal that yields the best estimations of the overall thermal resistance R_{eq}. This section first describes the numerical experiment as has been implemented in Energy Plus. Before analysis, section 3.2 compares three model selection procedures to determine which one is more likely to give interpretable results. Then in section 3.3, the 398 binary signals are examined by means of the interpretability score. From there, common characteristics of high-scoring binary signals are inferred. Influence of the pre-heating temperature is also assessed in section 3.3.

3.1. Numerical test setup

The numerical simulations are performed in January using actual weather data measured in Le Bourget-du-Lac (France) in 2019, on the INCAS platform. The simulation starts on January 1^{st}. The test is then implemented as to

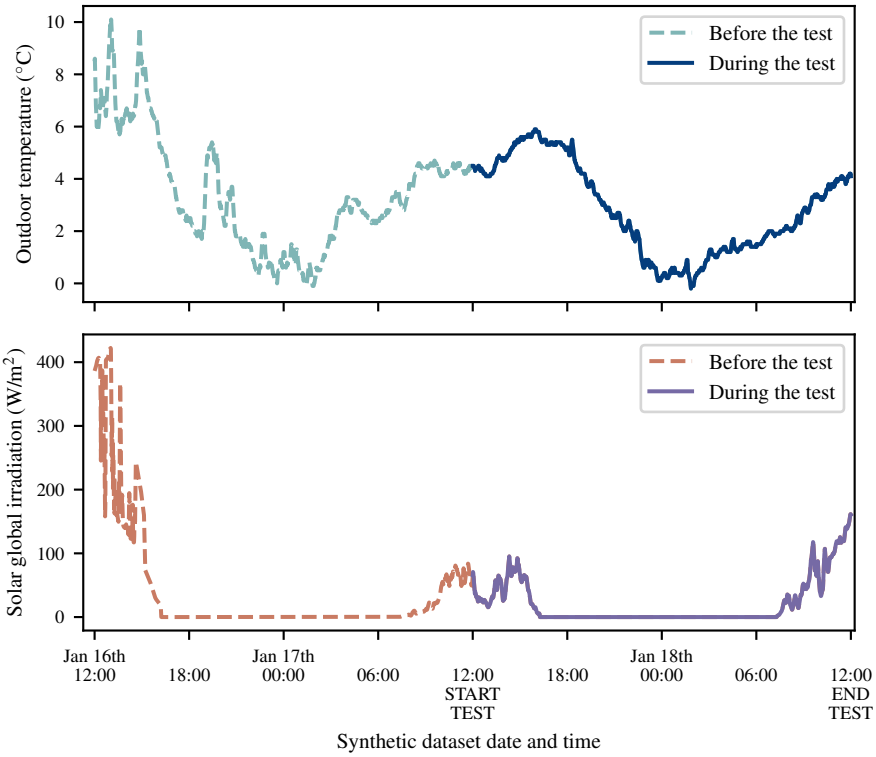


Figure 8: Outdoor temperature and global horizontal solar radiation used as boundary conditions in the building energy model simulation, the day before and during the test.

start on January 17^{th}. This allows for over two weeks of dynamic thermal simulation and should avoid any influence of the warm-up iterations performed by EnergyPlus on the first day of the weather dataset. The boundary conditions before and during the numerical experiment are given in Figure 8. It can be noticed that the outdoor temperatures are reasonably cold and that the solar radiation is very low. This needs to be kept in mind for the analysis of the experimental campaign.

In the simulation, the building is ideally heated at a constant temperature $T_{preheating}$ during 2 weeks before the test begins. As mentioned in the methodology section, temperature $T_{preheating}$ varies with each simulation as to assess whether it influences the accuracy of the R_{eq} estimation.

The numerical test starts on January 17^{th} at noon, with the hypothesis that in actual measurement conditions, a practitioner would set up the test in the morning and collect the sensors on the morrow afternoon.

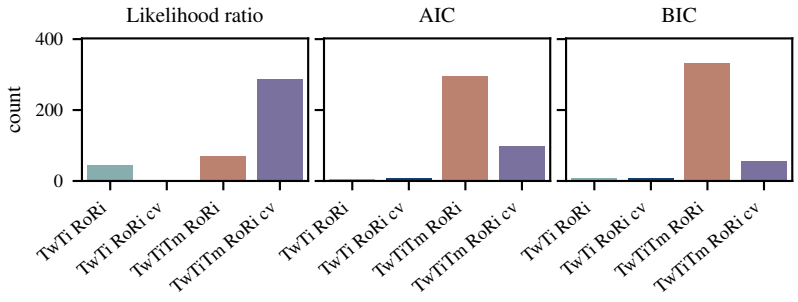


Figure 9: RC model choice for each model selection procedure: the likelihood ratio tends to favour the model with highest complexity, whereas AIC and especially BIC favour the simpler third order model.

3.2. Model selection for physical interpretation

For each of the 398 simulated datasets, estimations of the overall thermal resistance R_{eq} have been inferred with all four stochastic RC models. To move forward with the analysis of the results, a single model among the four needs to be selected. As mentioned in section 2.4, several procedures exist but have originally been designed for model prediction rather than for physical interpretation. Likelihood ratio, AIC and BIC procedures are compared in this section.

The three selection procedures tend indeed to favour different models, as shown in Figure 9. The likelihood ratio favours in overall the third order model with higher complexity $T_wT_iT_m$ R_oR_i R_p (as in fig. 6d). On the contrary, AIC and in particular BIC tend to favour the third order model $T_wT_iT_m$ R_oR_i (as in fig. 6c), which is slightly simpler. Second order model are almost never selected. This difference is completely in agreement with the idea that AIC and BIC tend to select simpler models ([28, 27]).

Physical interpretability for each selection procedure is compared in Figure 10 by means of the cumulated interpretability scores calculated for 5, 10 and 15 % error ranges. The lower the cumulated interpretability score line, the more accurate the estimation of the thermal resistance R_{eq}. The BIC procedure is for all error ranges the lowest curve. This means that in overall, the interpretability is better with a BIC-estimation than with the other selection procedures. Around 70% (resp. 84%) of the BIC-estimations score indeed higher than 0.5 in the error range 10% (resp. 15%). The interpretability score within the 5% error range is much more discriminant, but the BIC shows again in overall a better R_{eq} interpretability than the likelihood ratio and the AIC test.

Noteworthy is that although model $T_wT_iT_m$ R_oR_i R_p is often chosen by the likelihood ratio test, and sometimes by the AI- and BI- criteria, calibrated parameter R_p is actually often not considered significant. In 90% of all cases, the p-value of the estimated R_p is indeed higher than 0.05. There is therefore in 90% of all estimations no solid evidence that parameter R_p is necessary to the model to explain the dynamics in the data as its value could very well be null.

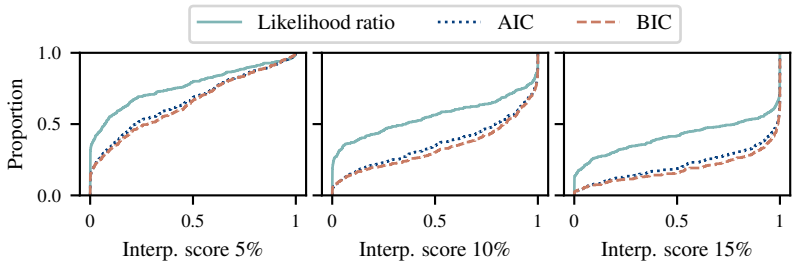


Figure 10: Comparison of three model selection procedures: likelihood ratio, AIC and BIC scores. The lower the cumulated interpretability score line, the closer the estimation to the target 5, 10 or 15 % error area and therefore the more satisfactory the estimation.

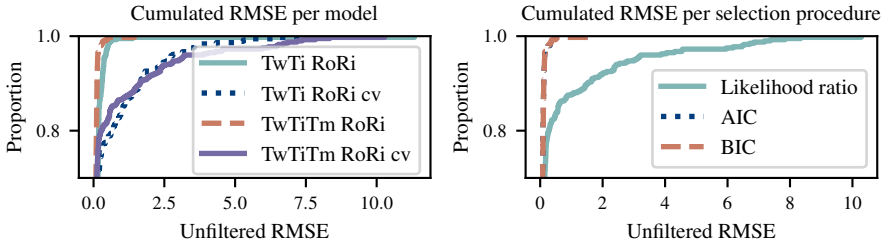


Figure 11: Root Mean Square Error of the unfiltered predictions for each model on the left hand side and for each model selection procedure on the right hand side (zoom on the y-axis). Model $T_w T_i T_m\ R_o R_i$ shows in overall the best fit to the data without Kalman filtering, meaning it catches satisfactorily the thermal dynamics.

Cases where the p-value of R_p is higher than 0.05 should therefore be discarded from any selection procedure.

To assess whether the BIC-chosen models carry a risk of correlated residuals as suggested in [27], a comparison of the root mean square error (RMSE) of the unfiltered prediction of all four models is shown in Figure 11. The idea is to perform a prediction on the same set of data without the automatic Kalman update. If the parameters of the model without filtering describes satisfactorily the thermal dynamics in the data, the RMSE should be very low. A validation analysis would concededly be best performed on a different validation dataset, but this allows a primary analysis without the need of more data. Figure 11 suggests that the best-fitting model is model $T_w T_i T_m\ R_o R_i$. Both models with a R_p parameter tend to have much higher prediction errors. On the right hand side of Figure 11 can be seen that AIC and BIC perform very similarly and in any case much better than the likelihood ratio procedure. In this case study, it will therefore be rather unlikely to find autocorrelated residuals with the calibration results selected by the AIC or BIC procedure.

To sum it up, the BIC model selection procedure seems preferable when physical interpretation of the overall thermal resistance is the objective. For

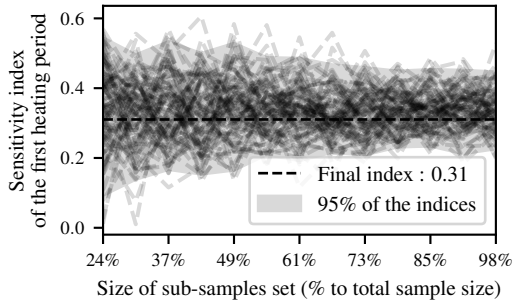


Figure 12: Convergence check of the global sensitivity analysis of the R_{eq} estimations under variable binary signal hyper-parameters and pre-heating temperature:

this reason, the following results are given for R_{eq} estimation given by the BIC.

3.3. Characterising signals and initial indoor conditions that provide satisfactory estimations

The numerical investigation is based on the one hand on the study of high-scoring signal shapes and on the other hand on suitable initial indoor conditions through pre-heating temperature setpoints. This result section first presents the results of the EASI RBD-FAST sensitivity analysis to establish relevant characteristics to an accurate test. Common characteristics of high-scoring signals and influence of the pre-heating temperature setpoint are then produced.

3.3.1. Common characteristics of high scoring signals

To establish what characteristics are relevant to an accurate estimation of the thermal resistance R_{eq}, a global sensitivity analysis (SA) is first performed. Given that the four input parameters are sampled by Latin-Hypercube, it is possible to calculate first order sensitivity indices with the EASI [29] RBD-FAST [30] method (with bias correction according to [31]). Provided convergence is verified and uncertainty considered, it will give a sense of what characteristics are influential on the R_{eq} estimation considered as output of the SA.

The global sensitivity analysis reveals one single significant first-order index: the first heating period duration. Its index is 0.31, which means that it directly explains about 30% of the variability of R_{eq}. The other input parameters, pre-heating temperature and the other two signal hyper-parameters, are non-significant. This means that the remaining variability of the R_{eq} estimations is explained by interactions between some or all four input parameters [32]. Convergence of the index related to the first heating period, the only one significant, is calculated by bootstrap on sub-samples of growing size with replacement and is shown in Figure 12, proving convergence. Noteworthy is the relatively high uncertainty (± 0.1), which incidently confirms that the other three indices are non-significant.

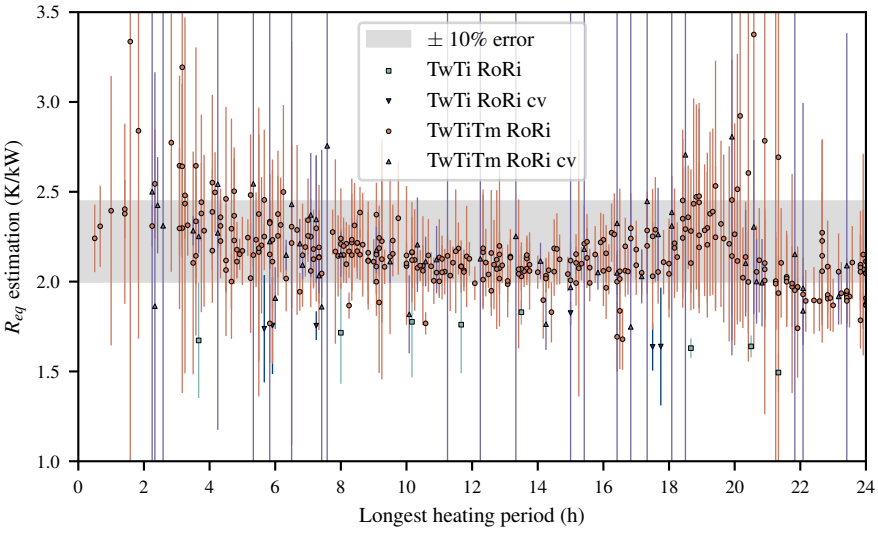


Figure 13: Variability of the thermal resistance R_{eq} estimation as a function of the longest heating period of the binary heating signal.

To further investigate how the signal shape is influential on the accuracy of the R_{eq} estimation, it makes therefore sense to study how the heating periods affect accuracy. Given that relevant interactions are likely to be found to explain the variability in R_{eq} estimations, free-floating durations too are further studied.

Figure 13 shows how the longest heating period of each binary signal influences the R_{eq} estimation. The longest heating period varies between 1/2 h (mostly free-floating periods) and 24 h (heating only during 24 h). A significant variability is encountered when the longest heating duration is less than around 8 h or over 18 h. On the contrary, in the range 8–18 h, the R_{eq} estimations show an overall steadiness and their maximum-likelihood estimators (the dots in the figure) have relative errors below 10%.

The interaction between longest heating and longest free-floating durations further explains the variability in R_{eq} estimations, as shown in Figure 14. In this figure, the signals are assessed by their 10% interpretability score, calculated with the BIC-estimation of R_{eq}. The high-scoring signals are coloured in deep green whereas low-scoring signals turn to dark red.

All signals in Figure 14 are naturally found in the bottom left part of the figure, given that heating and free-floating durations cannot sum up higher than 24 h. The good scoring signals with a 8–18 h of heating duration are again visible as a much greener area in the center of the figure. Within this range, the bottom signals, with the shortest free-floating durations, score in overall worse. Signals with a 5–8 h heating duration and free-floating periods larger than 8 h also score highly. From this figure can therefore be understood that most of the high-scoring signals have a 5–18 h longest heating duration and a longest

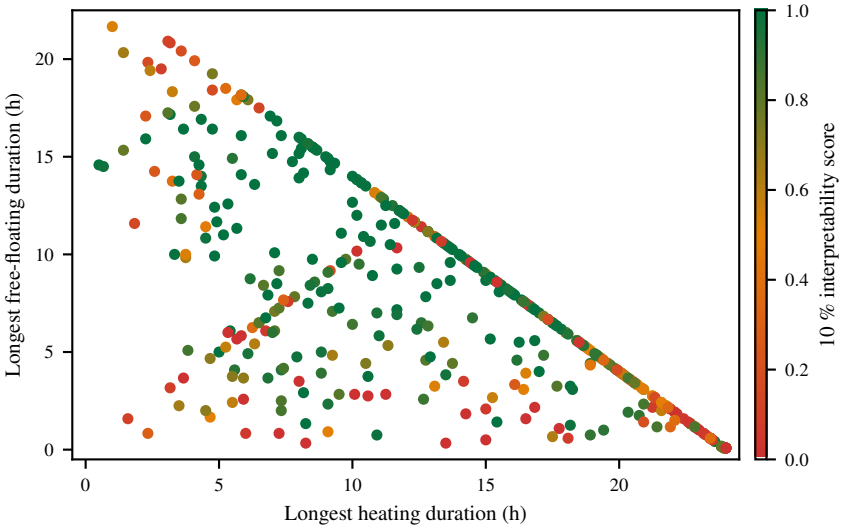


Figure 14: Assessment of each binary signal by means of the 10% interpretability score: dispersion of the coloured dot signal uncovers common favourable characteristics in terms of heating and free-floating duration.

free-floating duration higher than 5 h but lower than 18 h.

3.3.2. Pre-heating temperatures in the range 15–23 °C found non influential

To make sure the results are not highly dependent on the indoor temperature conditions before the test, the pre-heating temperature varied in the range 15–23 °C. The global sensitivity analysis shown previously in section 3.3 already indicated that the pre-heating temperature had no first-order influence on the R_{eq} estimation, given that its sensitivity index was non-significant.

Figure 15 shows all R_{eq} estimations according to the pre-heating temperature. The R_{eq} estimations show a certain vertical dispersion which is mainly due to the variability of the binary signals. As suggested by the sensitivity analysis, there is, however, no visible trend along the x-axis that would indicate that the pre-heating temperature influences the accuracy of the R_{eq} estimation.

More importantly, variable pre-heating temperatures also address the issue of temperature difference between indoors and outdoors during the test. Given that the outdoor conditions remain identical in the numerical study, starting at 15°C or at 23°C imply more or less heat flux through the envelope during the test. This could therefore mean that the test can be performed in slightly warmer or slightly colder outdoor temperatures without significant change in accuracy, therefore implying a certain robustness under variable winter weather conditions. This point will be further investigated in the experimental campaign.

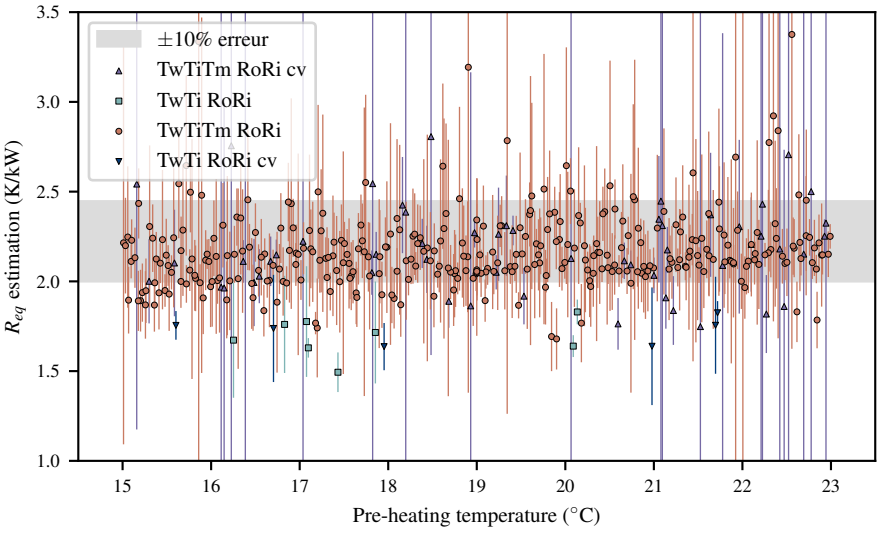


Figure 15: Influence of the pre-heating temperature on the R_{eq} estimation.

4. Experimental results

The numerical investigation assessed the 24-hour estimations of a whole-building thermal resistance R_{eq} from 398 binary heating signals. Common characteristics of high-scoring signals have been found.

An experimental campaign has then been set up to assess the feasibility of an actual on-site 24-hour estimation. One signal has been implemented and repeated 8 times as to assess its reproducibility under variable weather conditions.

This section presents the binary signal that has been selected and tested, describes in detail the experimental setup, then shows the estimation results for a 24-hour, revealing a significant variability. The section finally discusses the advantage of slightly extending the test to 30 or 36 hours to mitigate this variability.

4.1. A binary signal of interest to be further investigated

Preliminary experiments on a first binary signal, with a 7-hour heating period, showed poor results [33]. One plausible reason is that the actual time constants of the INCAS house are larger than what had been modelled in the building energy model. However, minimal heating and free-floating durations have been found necessary to achieve accuracy and are intrinsically related to the building time constants. Then, the 7-hour longest heating duration might be the reason why this original signal did not achieve the expected results. A signal with longer heating and free-floating durations should therefore be explored in the experimental campaign.

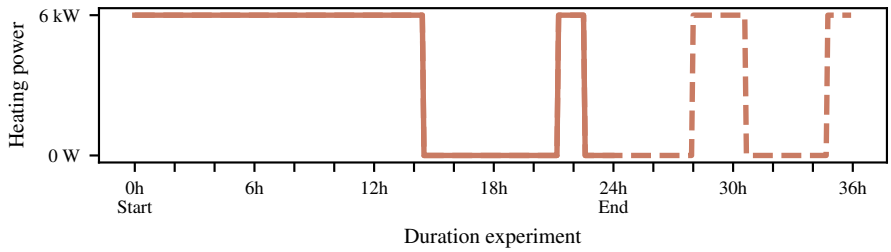


Figure 16: The binary signal investigated in the experimental campaign (details in Table 2)

Signal id.	**MLS parameters**	**Theoretical 10% interpretability**	**Tested on**
BS-1	14.5 - 7.8 - 6.49	$I_{10\%} = 0.89$	26–27 Dec
–			*02–03 Jan**
BS-2			01–02 Mar
BS-3			05–06 Mar
BS-4			08–09 Mar
BS-5			12–13 Mar
BS-6			18–19 May
BS-7			21–22 May
BS-8			30–31 May

Table 2: Summary of all tests performed on the wooden-framed INCAS house
* The test planned on January 2^{nd} failed due to an equipment issue. BS stands for binary signal.

A particular binary signals, shown in Figure 16, has therefore been selected to be tested in the experimental campaign. The longest heating period lasts over 14 hours. This signal is expected to achieve high interpretability scores. Indeed, according to the numerical results presented in section 3.3, its expected 10%-interpretability score $I_{10\%}$ is is expected at around 0.9.

Table 2 summarises the characteristics of the signal tested, its theoretical interpretability score according to the numerical investigation and specifies what days the signal was tested. Noteworthy, the signal has been tested 8 times as to assess its robustness under variable weather conditions.

4.2. Experimental setup in house I-OB

The house used for the experimental campaign is the INCAS I-OB real-size house previously described in section 2.3. The house was left unoccupied and was set up following the protocol defined in the SEREINE project, henceforth called the SEREINE protocol. In this protocol, the heating, domestic hot water and the mechanical ventilation systems are turned off. The air vents and water

evacuation pipes are sealed, identically to what is requested for a blower-door test. Roller shutters are closed as to avoid direct solar radiation into the house.

A set of sensors, called the SEREINE kit, is deployed in and around the house to monitor its thermal behaviour and its boundary conditions during the tests. The SEREINE kit consists indeed of both indoor and outdoor equipment.

Before installing the kit, the test must be configured via the online SEREINE user-interface. This webpage is first used to declare the number, the serial identification and the locations of the indoor and outdoor equipment. The binary heating signal used during the test is also defined via the interface and later fetched through the mobile phone network by the central communication box of the kit to control the indoor conditions.

The indoor equipment of the SEREINE kit includes 3-level power fan coils (adjustable to deliver 650, 1300 and 2000 W) temperature sensors, additional fans to homogenise the air temperature in all equipped room, controllers and a central communication box. The central box connects the SEREINE web-interface with the controllers in order to get the binary signal used during the experiment. It later sends the data collected, again through the mobile phone network. As for the controllers, they can each operate one fan coil and one additional fan depending on the signal played (ON/OFF). The controllers also monitor the indoor temperature in the room and the energy consumption of both the fan coil and the additional fan.

In addition to indoor equipment, the outdoor equipment is used to monitor the boundary conditions during the test. The outdoor kit includes outdoor air temperature sensors, and SENS probes. The SENS probes, shown in Figure 17, have been specifically designed to estimate an equivalent outdoor surface temperature that accounts for both outdoor air temperature and surface temperature rise due to solar radiation. A full description of the probe is given in Bouchié et al [34] . The equivalent surface temperature is indeed estimated with two temperature sensors set up behind a white face for the first one and a black face for the second one. An estimation of the building surface emissivity on each building face allows an estimation of the outdoor surface temperature on each measured wall. In the end, a weighted mean of all measured surfaces allows to infer an overall equivalent outdoor surface temperature of the building.

In the experimental campaign presented in this paper, 7 controllers have been used. Each controlled one 650 W-fan coil, one additional fan and a temperature probe. Three controllers were deployed on the first floor (entrance, living room and cellar) and four on the second floor (one per bedroom and one in the bathroom). To monitor the boundary conditions, 2 air temperature sensors and 7 SENS probes have been used. One air temperature sensor was placed outdoors near the north façade of the house to specifically measure the outdoor air temperature. The other temperature sensor was placed in the crawlspace. The SENS probes were placed vertically in front of each vertical wall, horizontally in the attics and sideways on both sides of the roof, as to cover both roof slopes.

The temperature and electrical consumption data collected are saved at a 1-minute time step during each test. The data were then averaged (or summed

Figure 17: SENS probe used to measure the equivalent outdoor surface temperature of the building

for the energy consumption) at a 5-minute time step in order to speed up data exploitation.

Although designed over 24 hours, all 8 tests were carried out over 48 hours. The proposed binary signal has indeed been generated for 48 hours. For data analysis, the data collected are then truncated to the desired duration, namely to 24, 30 and 36 hours, as shown in Figure 16.

The collected data from all tests of the experimental campaign have been exploited by an extended set of RC models. Unlike the numerical investigation, there are much less estimations to perform. It is therefore less computationally expansive to calibrate a larger set of RC models. In addition, data collected in an actual measurement campaign might present more complex dynamics which implies that an appropriate model is more likely to be found with a larger set of models.

Unlike the numerical investigation, the set of RC models is used to infer a Heat Transfer Coefficient (HTC) value instead of a thermal resistance value R_{eq}. The HTC has indeed been found to be more easily perceived by practitioners given its order of magnitude. All results in this section are therefore given through HTC estimations.

Finally, in agreement with the results given in section 3.2 about model selection, the selected HTC estimations are obtained from the best-fitting model according to the BIC procedure.

4.3. Results of the experimental campaign for 24-hour tests

As mentioned earlier, the tests have been repeated several times as to assess their reproducibility under variable weather conditions. The tests were programmed to cover winter, mid-season and early summer conditions. All test results and average weather indicators are later given in Table 3.

As for the interpretability assessment, from now on, the score is calculated within 15% error bounds. A 10% interpretability score is slightly too discriminant to clearly visualise the variability of the estimations.

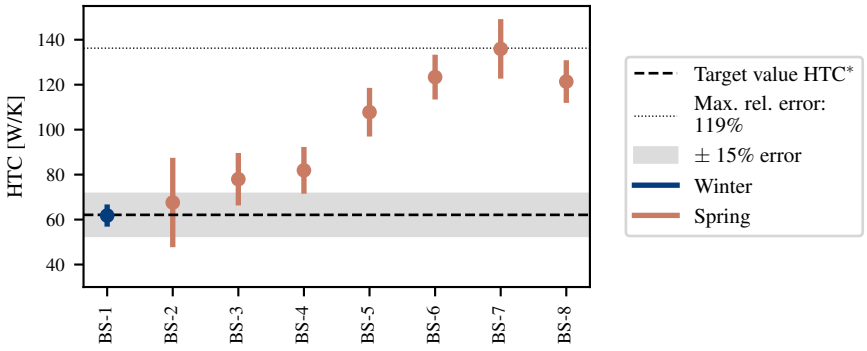


Figure 18: HTC estimations from all 24-hour tests of the experimental campaign

Figure 18 presents all 24-hour HTC estimations from the experimental campaign. The tests have been repeated between late December and May 2020. Test BS-1 achieves a very accurate estimation ($I_{15\%}(BS\text{-}1) = 0.99$) and test BS-2 is fairly satisfactory but has a wide uncertainty ($I_{15\%}(BS\text{-}2) = 0.57$). All other tests are significantly inaccurate.

From Figure 18 could be inferred that there is a significant variability across all test results. The most likely explanation is weather variability from one test to the other. The next section explores therefore in detail the influence of weather conditions on the HTC estimations.

4.4. Reducing weather induced variability : results from 30- and 36-hour tests

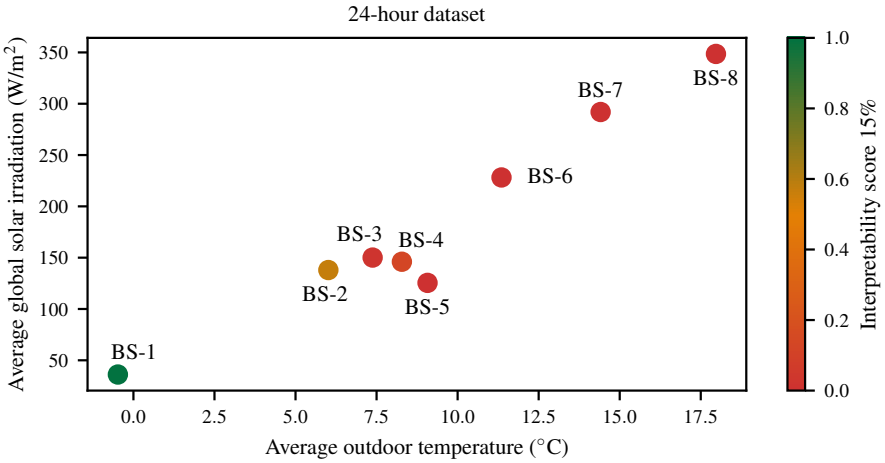
The experimental campaign, performed between late December and May 2021, revealed a significant variability in the accuracy of the HTC estimations. Weather variability is therefore likely to be influential on the estimations. Test results and averaged weather conditions can be found in Table 3 below.

Figure 19a shows the 15% interpretability scores as a function of the average outdoor temperature and the average global solar radiation during the 24-hour test. The variability of the outdoor conditions is clear : average outdoor temperatures vary from below 0 °C to 17.5 °C whereas the average global solar radiation shows that tests have been performed during both cloudy and very sunny days. The mid-season weather conditions are particularly evident on the right side of the figure, where the outdoor temperature and the solar radiation are highly correlated.

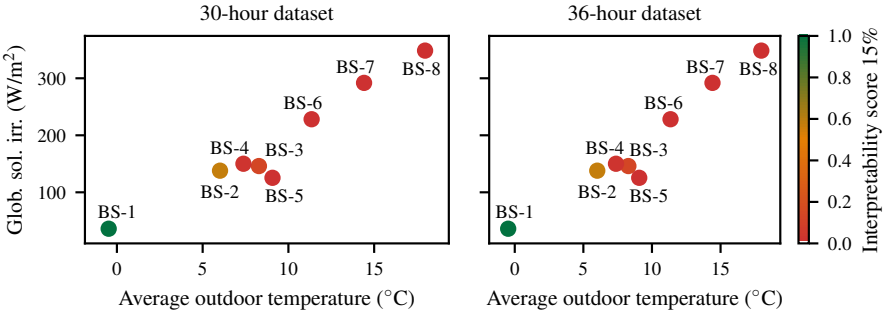
In Figure 19a, the accurate estimations have high 15% interpretability scores, in deep green. Orange coloured estimations can be considered fairly satisfactory.

The first and only very accurate test BS-1 has been performed during very cold and cloudy days. BS-2 is the test with the second coldest average outdoor temperatures and was indeed considered fairly satisfactory. On the other hand, tests BS-7 and BS-8, on the extreme right side of the figure, have been performed in May in early summer conditions and are both severely inaccurate,

with a maximum relative error of 119%.



(a) Illustration of the correlation between the average outdoor temperature, the average global solar radiation and the 15% interpretability scores



(b) Evolution of the 15% interpretability scores $I_{15\%}$ with the 30-hour and 36-hour experiments : the correlation between the average outdoor temperature, the average global solar radiation and the interpretability scores weakens significantly, compared with a 24-hour test.

Figure 19: Correlation between the average outdoor temperature, solar radiation and the HTC estimations

Weather conditions seem then to have, from a 24-hour test, a significant influence. It could be expected that longer tests allow on the contrary a more accurate HTC (or R_{eq}) estimation thanks to the additional information brought by a larger dataset. Let us briefly recall that all tests in the experimental campaign had been performed over 48 hours with 48-hour truncated binary signals. As can be seen from the dotted line in Figure 16, the first 24 hours remain identical to the signals tested in the numerical investigation. The additional 24

Test id	**BS-1**			**BS-2**			**BS-3**			**BS-4**		
Duration	24h	30h	36h	24h	30h	36h	24h	30h	36h	24h	30h	36h
HTC (W/K)	61.8	60.9	60.5	67.6	64.7	41.1	78.0	76.9	70.5	81.9	72.3	51.5
$I_{15\%}$	1.00	1.00	1.00	0.57	0.97	0.00	0.13	0.00	0.59	0.02	0.36	0.32
$\overline{\overline{T}}_{ext}$ **(°C)**	-0.5	1.1	2.2	6.0	6.3	5.9	8.3	8.2	7.8	7.4	7.3	6.1
$\overline{I}_{sol}$ **(W/m^2)**	36	29	24.	138	111	92	146	117	98	150	120	100
$\overline{S}_{wind}$ **(m/s)**	4.3	4.9	5.3	1.0	1.0	0.9	6.0	5.8	5.0	3.2	3.3	3.0
Test id	**BS-5**			**BS-6**			**BS-7**			**BS-8**		
Duration	24h	30h	36h	24h	30h	36h	24h	30h	36h	24h	30h	36h
HTC (W/K)	107.8	55.2	50.9	123.4	62.1	57.9	135.9	62.6	57.9	121.4	60.8	51.5
$I_{15\%}$	0.00	0.51	0.37	0.00	0.72	0.82	0.00	0.46	0.76	0.0	0.77	0.41
$\overline{\overline{T}}_{ext}$ **(°C)**	9.1	9.2	8.9	11.4	11.0	10.3	14.4	15.0	14.4	18.0	17.6	16.8
$\overline{I}_{sol}$ **(W/m^2)**	125	101	84	228	195	163	292	262	219	349	280	245
$\overline{S}_{wind}$ **(m/s)**	4.9	5.3	5.5	2.0	2.1	1.9	3.1	3.0	2.7	2.0	1.8	1.5

Table 3: Summary of all results and averaged weather conditions from the experimental campaign. The red color font underlines the highest wind speeds.

hours are concededly not optimally designed, but add information to the signal in any case. The 24-hour experimental results are simply truncated from the 48-hour datasets.

Figure 19b now shows the 15% interpretability scores from 30 and 36-hour datasets, again as a function of the average outdoor temperatures and average global solar radiations. The overall weather variability observed during the 8 tests is fairly high, with average temperatures in the range 1–17 °C and again both cloudy and sunny couple days. In overall, 30-hour and 36-hour tests seem to achieve more accurate estimations, with, however, differences among tests. Test BS-2 performs significantly worse with the 36-hour dataset than with the 30-hour dataset which is counter-intuitive. Yet, the weather conditions during the test is remarkably similar to the conditions during tests BS-3, -4 and -5. The estimations from BS-3, -4 and -5 are not very satisfactory but their $I_{15\%}$ score is higher than 0.3. The three latest tests, although performed during warmer and sunnier days, gain a considerable accuracy with a longer test.

What the 15% interpretability indicator in Figure 19b fails however to reveal is that the relative error drops significantly with these longer datasets : from a maximum of 119% error from a 24-hour dataset to 24% with a 30-hour dataset and then a maximum of 18% with a 36-hour dataset (except for the counter-performance of BS-2 respectively with 33.7% error). This seems to indicate that a slightly longer test improves significantly the robustness of the HTC estimation.

5. Discussion

The numerical investigation in section 3 found common characteristics to binary heating signals that achieve, to the least numerically, accurate estimations of the whole-house thermal resistance R_{eq} (or HTC). One particular binary signal has then been put to the test in an extensive experimental campaign, shown in section 4. This section will now confront both numerical and experimental results, discuss the lessons learnt from the experiments and elaborate on valuable research prospects.

The comparison of the numerical investigation with the experimental results is undoubtedly challenging, mainly because the experimental campaign could not be achieved in perfect agreement with the conditions simulated in the numerical investigation. First, let us recall that the actual thermal resistance of the I-OB house, estimated at 15.7 k/kW , is significantly different than that modelled in the BEM, calculated at 23 K/kW. In addition, as established in Table 4, the heating power delivered in the INCAS house was variable, in a range between 3200 W and 5500 W. Variable heating power delivered during different outdoor conditions means variable indoor temperatures achieved and therefore variable indoor-outdoor temperature differences. This makes the induced heat flux through the envelope different from one experiment to the other and makes any comparison all the more difficult. As also underlined in Table 4, the actual start time of each test depended on other experiments performed during the week in the INCAS house and made it impossible to start at noon, except once. Let us here shortly recall that a noon-starting test is a recommendation from the practitioners that were contacted in the framework of the funding project.

Test	**BS3-1**	**BS-2**	**BS-3**	**BS-4**
Max heating power (W)	3516	3211	3917	3932
Actual start time	17:30	18:00	18:30	18:30
Test	**BS-5**	**BS-6**	**BS-7**	**BS-8**
Max heating power (W)	4278	4644	5592	4896
Actual start time	17:40	17:40	17:30	12:00

Table 4: Actual test conditions of the experimental campaign : actual maximum heating power delivered and actual start time of each test. These values are to be compared to the constant 3.1 kW delivered in the numerical building energy simulations and the theoretical start at 12:00.

In addition, given that the experimental results have shown a significant correlation between weather conditions and accuracy of each signal, the boundary conditions must be taken into account to compare the expectations from the numerical investigation with the experiments. Figure 8 had shown that, during the numerical simulations , the average outdoor temperature was around 2 °C and that the solar radiation was very low, with a maximum around 150 W/m^2.

With these boundary conditions in mind, the only test in the experimental campaign that can compare is the test BS-1. This test has indeed been performed late December 2020 in cold and cloudy winter conditions.

With these elements in mind, the 24-hour signal BS-1 scores well and in agreement with the findings of the numerical investigation, which predicted an interpretability of $I_{10\%} = 0.89$. In a few words, with this strict comparison between numerical and experimental results, the proposed 24-hour binary heating signal seems to be validated for similar weather conditions.

What the experimental campaign however reveals about the proposed signal is a lack of robustness with respect to weather conditions. With outdoor temperatures warmer than 7°C and correlated solar radiations larger than 100 W/m^2, the interpretability score decreases significantly.

From the point of view of a practitioner, this can be considered as highly problematic. It means indeed that the validity of the signal is limited, especially if the limit is lower than 5°C, which does not even cover the entire winter season. Concededly, other better suited signals for warmer temperatures could easily be determined with another numerical investigation. But in practice, it would mean having to consult weather forecast and then decide what signal is the most appropriate. It would naturally be feasible, but is rather inconvenient.

Although it was not the original objective, a solution that arises from the results of section 4.4 is to extend the duration of the test. In practice, a 24-hour was concededly convenient to perform tests between 12:00 and 12:00 (D+1). Similarly, a 36-hour test remains convenient, in particular if it starts at the end of the afternoon and stops early in the morning two days later.

Again, it must be underlined that the proposed binary signal had not been designed to be optimal in a 30-hour or 36-hour test. Yet, the R_{eq} estimations scored significantly better over longer datasets. This means that an adequate numerical investigation on a 36-hour signal could be performed and would very likely point to other more informative signals. With a longer and better suited 36-hour binary heating signal, weather conditions can be expected to be less influential. Higher overall accuracy can also be expected.

In any case, a focus on the influence of weather conditions and seasonal variability on the accuracy of R_{eq} estimations cannot be overlooked. The experimental campaign indeed revealed a significant influence of seasonal variability. Natural variability of solar radiation, wind speed and outdoor temperature in a single season could, however, not be assessed through the campaign but could nonetheless have a significant influence on accuracy. Robustness of the R_{eq} estimations under natural weather variability is then a first research prospect that should be investigated.

On another note, summer is likely to be the less convenient season for a short test because of low indoor-outdoor temperature differences and therefore low heat flows through the envelope. Given the results of the experimental campaign, a 24-hour test is unlikely to be found suitable. Whether a 36-hour, or longer, test can be found suitable and robust with respect to summer

weather variability is left to be established and also constitutes a major prospect.

The numerical investigation has also been the opportunity to discuss model selection procedures with the particular objective of physical interpretation of the model resistive parameters. The numerical investigation indicated that the BIC procedure is the most promising for physical interpretability. An indisputable limitation is that for computational reasons, the set of RC models in the numerical investigation had been limited to four different models. Although the four models chosen were the most promising in terms of data fitting, the conclusions would be stronger with a larger set of models. Then, the conclusions drawn here might be case-related, i.e. related to a house with moderate thermal inertia and high insulation. Finally, other procedures for model selection could be investigated. This highlights the necessity to further investigate RC model selection for physical interpretation by considering a larger set of RC models and comparing selection procedures on different building types.

Although the experimental campaign showed a decreased interpretability with warmer temperatures, these results are likely to be, to some extent, case-related. Let us indeed underline that the I-OB house is a very highly insulated wooden-framed house. A 24-hour test might achieve much more robust HTC estimations in many other building types with different thermal characteristics.

Comparing the accuracy achieved with the proposed method on this case study to literature is also tricky. In the aforementioned whole-building scale methods, the experiments were carried out on houses with HTC values at least twice as large as in this case study: house I-OB has an estimated HTC around 62.1 W/K against $HTC \approx 125$ W/K in [10], $HTC \approx 185$ W/K in [19] or $HTC_1 \approx 120$ W/K and $HTC_2 \approx 260$ W/K in [11], who tested two houses. This again underlines the interest of testing the proposed method on other building types and on moderately insulated houses. This also puts in perspective the results of the experimental campaign presented in this paper: achieving a high accuracy with the proposed 24-hour signal in test BS-1 is comparatively to the literature case studies an remarkable result given that I-OB house has extreme high insulation levels.

The prospect of a 24-hour accurate test remains valuable in particular for houses before and after retrofit. Renovated buildings are indeed unlikely to either present a wooden-framed structure nor achieve, overall, extreme high thermal performances. Instead, indoor or outdoor insulated houses with moderate to good insulation levels and possibly untreated thermal bridges can be expected. As such, a 24-hour test has higher chances to yield accurate results while having an enhanced robustness under variable weather conditions than this case study.

6. Conclusion

On-site measurement of the thermal performance of a building envelope can be a powerful leverage for energy performance contracting. To do so, accuracy and robustness of measurement protocols need to be developed while maintaining costs at an acceptable level. In this prospect, this paper investigated the feasibility, accuracy and robustness of a 24-hour test based on a binary heating perturbation signal. This 24-hour test procedure relies on data exploitation with stochastic RC models.

The paper first proposed a numerical investigation of binary signal shapes and initial indoor temperatures through a pre-heating temperature setpoint. Overall, 398 signals were implemented in a building energy model of a wooden-framed house, each with a pre-heating temperature in the range 15–23°C. Each synthetic dataset simulated served to infer a whole-building thermal resistance. Given their respective accuracy, common characteristics of high-scoring signal shapes could be determined.

The paper then presented the results of an experimental campaign where a numerically highly accurate signal was tested. The experiment was repeated 8 times as to assess reproducibility, and henceforth robustness under variable weather conditions.

Take-home messages

- The numerical investigation showed that, in theory, high-scoring binary heating signals have both heating and free-floating periods of at least 5 hours each. Pre-heating temperatures between 15–23°C do theoretically not seem to be influential on the estimation accuracy, which makes the method applicable in many indoor conditions regardless of the initial temperature.

- The proposed perturbation method is therefore based on a high-scoring binary heating signal. The data is exploited by a BIC-selected RC model, according to findings presented in the numerical investigation suggesting that BIC model selection achieves higher physical interpretability.

- The experimental campaign showed that, in weather conditions close to those used for the numerical simulations, the proposed perturbation method indeed achieves a very satisfactory accuracy.

- The results also revealed that an accurate test in such a highly insulated house in mid-season and early summer should maybe preferably rely on a slightly longer test, for example over 30- to 36-hours. A longer test should ensure a better robustness of the HTC or whole-house thermal resistance estimation under variable weather conditions.

Research prospects

The results presented in this paper suggest to further use the numerical methodology on other buildings types and under variable weather and season conditions. Characterising binary heating signals suitable for other building types would indeed be valuable to propose adequate signals to any building. In addition, it would be highly relevant to compare the characteristics of suitable signals among different building types to maybe determine common characteristics. From a practitioner's point of view, if such tests were to be deployed at large scale, it would be very convenient to reduce the number of suitable signals to a minimum, as to simplify the procedure. A wider numerical investigation could also quantify the advantage of slightly extending the duration of the test, from a 24-hour test to, for example, a 30- or 36-hour test. An optimum between equipment downtime, inconvenience for occupants and robust accuracy of the HTC estimation may then be found.

The results also invite to further pursue the experimental campaign, not only in other building types but also in buildings with a wider variety of insulation level. In particular, 24-hour tests in the frame of energy retrofit validation remains a promising prospect and need to be thoroughly examined by an appropriate measurement campaign. More on-site experiments may also provide a clearer insight on the model selection procedure by the use of non-synthetic data and therefore further assess the suitability of the Bayesian information criterium, compared to other selection criteria.

Acknowledgments

Research reported in this paper was supported by the PROFEEL program. The content is solely the responsibility of the authors and does not necessarily represent the official views of the PROFEEL program funders.

CRediT author statement

Sarah Juricic: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft **Jeanne Goffart**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - review **Simon Rouchier**: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing - review **Arnaud Jay** Data curation, Writing - contribution to original draft and review **Pierre Oberlé** Data curation, Writing - review

References

[1] G. H. Berghorn, M. M. Syal, RISK FRAMEWORK FOR ENERGY PERFORMANCE CONTRACTING BUILDING RETROFITS, Journal of Green Building 11 (2016) 93–115.

[2] M. A. Mozzo, Setting the Energy Baseline For Performance Contracts, Strategic Planning for Energy and the Environment 21 (2001) 12–19.

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

[4] N. Soares, C. Martins, M. Gonçalves, P. Santos, L. S. da Silva, J. J. Costa, Laboratory and in-situ non-destructive methods to evaluate the thermal transmittance and behavior of walls, windows, and construction elements with innovative materials: A review, Energy and Buildings 182 (2019) 88–110.

[5] G. Bauwens, S. Roels, Co-heating test: A state-of-the-art, Energy and Buildings 82 (2014) 163–172.

[6] H. Madsen, J. M. Schultz, Short time determination of the heat dynamics of buildings, Technical Report 243, Technical University of Denmark, Department of Civil Engineering, 1993.

[7] P. Baker, H. van Dijk, PASLINK and dynamic outdoor testing of building components, Building and Environment 43 (2008) 143–151.

[8] O. M. Brastein, B. Lie, C. Pfeiffer, N. O. Skeie, Estimating uncertainty of model parameters obtained using numerical optimisation, Modeling, Identification and Control 40 (2019) 213–243.

[9] S. Rouchier, M. Rabouille, P. Oberlé, Calibration of simplified building energy models for parameter estimation and forecasting: Stochastic versus deterministic modelling, Building and Environment 134 (2018) 181–190.

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

[11] F. Alzetto, G. Pandraud, R. Fitton, I. Heusler, H. Sinnesbichler, QUB: A fast dynamic method for in-situ measurement of the whole building heat loss, Energy and Buildings 174 (2018) 124–133.

[12] N. Ahmad, C. Ghiaus, T. Thiery, Influence of initial and boundary conditions on the accuracy of the QUB method to determine the overall heat loss coefficient of a building, Energies 13 (2020).

[13] S. Hammarsten, A critical appraisal of energy-signature models, Applied Energy 26 (1987) 97–110.

[14] A. Rabl, A. Rialhe, Energy signature models for commercial buildings: test with measured data and interpretation, Energy and Buildings 19 (1992) 143–154.

[15] F. P. Hollick, V. Gori, C. A. Elwell, Thermal performance of occupied homes: A dynamic grey-box method accounting for solar gains, Energy and Buildings 208 (2020) 109669.

[16] 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) 109706.

[17] G. Baasch, P. Westermann, R. Evins, Identifying whole-building heat loss coefficient from heterogeneous sensor data: An empirical survey of gray and black box approaches, Energy and Buildings 241 (2021) 110889.

[18] F. Alzetto, J. Meulemans, G. Pandraud, D. Roux, A perturbation method to estimate building thermal performance, Comptes Rendus Chimie 21 (2018) 938–942.

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

[20] K. Godfrey, A. Tan, H. Barker, B. Chong, A survey of readily accessible perturbation signals for system identification in the frequency domain, Control Engineering Practice 13 (2005) 1391–1402.

[21] A. H. Tan, K. R. Godfrey, A guide to the design and selection of perturbation signals, in: Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference, January 2010, IEEE, 2009, pp. 464–469. doi:10.1109/CDC.2009.5400077.

[22] A. H. Tan, K. R. Godfrey, Industrial Process Identification, Advances in Industrial Control, Springer International Publishing, Cham, 2019. doi:10.1007/978-3-030-03661-4.

[23] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, l. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nature Methods 17 (2020) 261–272.

[24] 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.

[25] 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.

[26] 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.

[27] C. Rasmussen, Data-driven Methods for Reliable Energy Performance Characterisation of Occupied Buildings, Ph.D. thesis, Technical University of Denmark, 2020.

[28] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, volume 173 of *Springer Series in Statistics*, Springer New York, New York, NY, 2009. doi:10.1007/978-0-387-84858-7.

[29] E. Plischke, An effective algorithm for computing global sensitivity indices (EASI), Reliability Engineering and System Safety 95 (2010) 354–360.

[30] S. Tarantola, D. Gatelli, T. Mara, Random balance designs for the estimation of first order global sensitivity indices, Reliability Engineering and System Safety 91 (2006) 717–727.

[31] J. Y. Tissot, C. Prieur, Bias correction for the estimation of sensitivity indices based on random balance designs, Reliability Engineering and System Safety 107 (2012) 205–213.

[32] I. Sobol', Sensitivity estimates for nonlinear mathematical models, Mathematical Modelling and Computational Experiment 1 (1993) 407–414.

[33] S. Juricic, J. Goffart, S. Rouchier, Designing a 24-hour perturbation method for the estimation of a building heat transfer coefficient, in: International Building Physics Conference, Copenhagen, 2021.

[34] R. Bouchié, C. Abele, S. Derouineau, J.-R. Millet, Conception et validation d'un capteur de mesurage de la température extérieure équivalente d'une paroi opaque d'un bâtiment, in: IBPSA France, Arras, 2014.