



HAL
open science

The Sereine test: advances towards short and reproducible measurements of a whole building heat transfer coefficient

Sarah Juricic, Mickael Rabouille, Arnaud Challansonnex, Arnaud JAY, Simon Thébault, Simon Rouchier, Rémi Bouchié

► **To cite this version:**

Sarah Juricic, Mickael Rabouille, Arnaud Challansonnex, Arnaud JAY, Simon Thébault, et al.. The Sereine test: advances towards short and reproducible measurements of a whole building heat transfer coefficient. *Energy and Buildings*, 2023, 299, pp.113585. <10.1016/j.enbuild.2023.113585>. <cea-04692055>

HAL Id: cea-04692055

<https://cea.hal.science/cea-04692055v1>

Submitted on 9 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

The Sereine test: advances towards short and reproducible measurements of a whole building heat transfer coefficient

Sarah Juricic^a, Mickaël Rabouille^b, Arnaud Challansonnex^a, Arnaud Jay^c, Simon Thébault^d, Simon Rouchier^e, Rémi Bouchié^a

^aCentre Scientifique et Technique du Bâtiment (CSTB), 84 avenue Jean Jaures, Champs-sur-Marne, 77420, France

^bInstitut National de l'Energie Solaire (INES), Plateforme Formation Evaluation, Savoie Technolac Bat. Helios, Le Bourget du Lac, , France

^cUniv. Grenoble Alpes, CEA, Liten, INES, , Le Bourget-du-Lac, 73375, France

^dCentre Scientifique et Technique du Bâtiment (CSTB), 24 rue Joseph Fourier, Saint-Martin d'Herès, 38400, France

^eLOCIE, UMR 5271 USMB CNRS, Savoie Technolac Bat. Helios, Le Bourget du Lac, 77370, France

Abstract

In a global effort to reduce carbon emissions from the building sector, reliable on-site thermal intrinsic performance measurements could positively contribute to reducing the performance gap in newly constructed or retrofitted buildings. A measurement is considered reliable when it is both accurate and precise, meaning it is, on average, on target with a finite, ideally low level of uncertainty.

This paper aims to assess the reliability of a Heat Transfer Coefficient measurement using the Sereine method. The paper conducts an extensive uncertainty analysis, defining a new expanded coverage interval and thereby introduces an innovative approach to addressing uncertainties related to weather and building type. This approach relies on an extensive set of numerical experiments to quantify the measurement uncertainty under diverse measurement conditions.

An experimental campaign in an externally insulated house demonstrates good agreement among all Sereine tests. It reveals that the expected performance consistently falls within the final widened coverage interval, which suggests reproducibility of the measurement. Given its thorough and transparent uncertainty quantification, the Sereine method could be a cornerstone for reliable performance measurement and, consequently, a valuable tool for future performance contracting protocols or building certification.

Keywords: Inverse problem, Sereine test, Heat Transfer Coefficient, Reproducibility, Uncertainty analysis

1. Introduction

The changing climate raises serious threats for buildings and occupants [1]: discomfort, illness, mortality, financial consequences and further carbon emissions. Key factors for mitigation and adaptation are high performance of new constructions and efficient mass renovation of the existing building stock. One of the strategies to achieve high performance is the implementation of performance contracting and performance guarantees in general. To be effective, it is essential that the energy savings be verifiable by methods agreed upon by all parties involved [2].

In that perspective, the measurement of the actual intrinsic energy performance of the building envelope, such as the overall Heat Transfer Coefficient

(HTC), can significantly contribute to a unified approach to the verification of actual building performance. Deb et al. [3] reported that stakeholders recognize the numerous associated benefits of on-site measurements. In particular, stakeholders expect that an overall HTC measurement should be easy to deploy, completed in a few days or less, and conducted at a reasonable cost. More importantly, accuracy is for stakeholders a major concern as is minimising the HTC associated uncertainty.

While accuracy of an HTC measurement method is generally tackled in the related literature, there is no common and shared practice for identifying and quantifying the sources of uncertainty of these methods. Fitton et al. [5] outlined indeed that each method has its own uncertainty quantification ap-

| Nomenclature | |
|--------------------|--|
| \hat{X} | (parameter hat) Parameter estimate, i.e. calculated from measured data |
| C_i, C_w | Thermal capacitance parameters of the RC models (J/K), see Appendix A |
| $C_{p,v,air}$ | Volumetric heat capacity of air ($Wh/m^3/K$) |
| H_o, H_i, H_b | Conductance parameters of the RC models (W/K), see Appendix A |
| H_{inf} | Heat transfer coefficient by air infiltration (W/K) as per ISO 13789 standard [4] |
| H_{tr} | Heat transfer coefficient by transmission (W/K) as per ISO 13789 standard [4] |
| 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 [4]. See section 2.3.2 |
| n_{50} | Air change rate at a 50 Pa pressure difference (vol/h) |
| P_h | Measured heating power (W) |
| P_{inf} | Equivalent power lost or gained by air infiltrations during the test (W) |
| Q_{inf} | Air flow rate by infiltration (m^3/h) |
| T_{ext}, T_{int} | External/indoor air temperature ($^{\circ}C$) |
| u_{HTC} | HTC overall uncertainty (W/K) |
| u_{HTC}^{step2} | HTC uncertainty after parameter and variable uncertainty propagation (W/K) (see section 2.3.3) |
| u_{MOB} | \widehat{HTC} estimate uncertainty resulting from the maximum observed bias (W/K) |
| X^* | (starred parameter) True or target value of the measurand/parameter. For example, $HTC_{coheating}^*$ is the target HTC value obtained by a coheating measurement. |
| BIC | Bayesian Information Criterion |
| EWI and IWI | External wall insulation and Internal wall insulation |
| GUM | Guide to the expression of Uncertainty in Measurement (published by the Joint Committee for Guides in Metrology) |
| IWEC | International Weather for Energy Calculation |
| VIM | International Vocabulary for Metrology (published by the Joint Committee for Guides in Metrology) |

proach, with values ranging from 3 % to 30 %. More specifically, without being exhaustive, the literature shows variable levels of maturity regarding uncertainty quantification. Gori et al. [6] or Thébault and Bouchié [7], with earlier work of Stamp [8] and Jack et al. [9], consider model-related, aleatory and systematic sources of error. In contrast, others remain elusive [10, 11, 5], even when repeatability is evidently not achieved.

It is likely that the concept of measurement uncertainty and the estimation thereof is not fully understood by the building energy community. Definitions from the metrology community [12, 13] may be of some help. The expectations of stakeholders as identified by Deb et al. [3], can be translated as follows: an HTC measurement should be *precise* (meaning that when repeated, measurements are close to each other), it should be *accurate* (close

to its true value) and *reproducible* (precision of the same measurement method on the same object under different conditions such as different operators or locations). For the latter to make sense in the case of an HTC measurement, we propose to repute *reproducible* an HTC measurement that is precise when a same building is measured by different operators and under different weather and climate conditions. *Conditions for reproducibility* may also be specifically defined, excluding, for example, summer conditions, to ensure reproducibility under said conditions.

In summary, the lack of common and shared practice for uncertainty quantification poses a serious reliability issue since an overly optimistic uncertainty creates a misleading sense of confidence in a measurement that is thereby neither repeatable nor reproducible. Systemic uncertainty quantifica-

tion combined with accurate methods is, therefore, a key factor in achieving a more reliable HTC measurement.

Objectives, research contribution and paper structure

This paper addresses the issue of HTC measurement reliability by contributing to the accuracy assessment and the uncertainty quantification of the Sereine test method, heir to the ISABELE method [7] and the EPILOG project [14]. The main objectives of the paper are:

- (A) to propose a unified systemic uncertainty estimation of the Sereine HTC measurement;
- thereby (B) to tackle the residual variability identified by Kennedy and O’Hagan [15];
- and (C) to quantify how the residual variability affects the Sereine HTC estimation.

With this work, we aim at raising awareness on the importance of uncertainty analysis and estimation for reproducible HTC measurements and at proposing a novel, systemic, and, most of all, trustworthy uncertainty estimation of an HTC measurement.

The paper begins by detailing the Sereine test and presents an comprehensive uncertainty analysis. Although most of the quantifiable uncertainties are already accounted for, this paper delves further into the residual variability. A computer-based methodology uses dynamic thermal simulations to produce virtual Sereine tests and thereby quantify the residual uncertainty in a controlled environment. The paper identifies and quantifies a maximum bias that depends on the outdoor temperature and building type, which is then added to the HTC estimate uncertainty. Finally, an experimental campaign illustrates how the novel expanded uncertainty of the HTC estimation is consistent and shows reproducibility of the measurement.

2. Overview of the Sereine test and its uncertainty estimation

The Sereine test aims at estimating the overall Heat Transfer Coefficient (HTC) and its intrinsic counterpart H_{tr} , the overall fabric heat transfer coefficient (as defined in the ISO 13789 standard [4] and further explained in section 2.3.2). The Sereine test relies on a specific sensor kit to measure the

building boundary conditions and to control the indoor temperature as to follow an optimised heating scenario. A fully automated procedure analyses the collected data.

This section first conducts a comprehensive analysis of errors and sources of uncertainty that may arise at all steps of the HTC measurement process. The Sereine measurement setup and the data analysis procedure are then detailed in light of the uncertainty analysis. Finally, the section identifies that the residual variability is influenced by weather conditions and building type, that it is poorly accounted for, and that addressing it is necessary for ensuring measurement reproducibility.

2.1. HTC uncertainty budget: components, calculation and combination thereof

To provide a systemic uncertainty assessment of the HTC estimation, an analysis of the possible sources of uncertainty, or uncertainty budget, is performed and summarised in Figure 1.

General perspective of sources of uncertainty in the case of on-site HTC measurement

Many sources of uncertainty play a role in the final estimation of the HTC uncertainty. Uncertainty in thermal and computer inverse problems has been tackled in the literature (among many others [16, 15]) and can be roughly categorised as follows:

- direct calculation error or code uncertainty (due to rough discretization, inappropriate numerical method, initialisation...);
- model error (or inadequacy);
- measurement noise error;
- sensor bias error;
- observation error;
- residual variability.

Residual variability is described in Kennedy and O’Hagan [15] and relates to the possibility that a process takes different values when repeated or reproduced. Residual variability results from either a stochastic error of the process or from a certain form of model error that does not recognise the external factors influencing the process [15]. In the case of an HTC measurement, it could be a significant source of variability because the reproducibility assessment can never be performed under all

imaginable circumstances. Residual variability can therefore be somewhat treacherous as it induces an unknown and non-measurable error onto the HTC estimation.

Application to the Sereine method

The categorization from Maillet et al. [16] and Kennedy and O’Hagan [15], as mentioned earlier, is detailed in Figure 1, completed by the Guide to the expression of Uncertainty in Measurement (GUM) [17] and applied to the case of an HTC estimation.

The overall process is divided into 3 sub-processes: measurement equipment, on-site data acquisition and data analysis. As indicated in Figure 1, some sources of uncertainty are already taken into account in the overall process:

- equipment random uncertainty is accounted for in the data analysis process, where errors are supposed Gaussian;
- equipment systematic uncertainty and spatial variability are likewise propagated in a post-analysis sub-process, as described in section 2.3.3 and Thébault and Bouchié [7];
- informative heating signal: the test conditions are optimised to provide 24- to 48-hour data (or longer if necessary) as informative as possible. The design of the heating signal has been the subject of a previous paper [18];
- model error is accounted for through the use of stochastic modeling, as explained in Section 2.3.1;
- numerical accuracy can be considered established since the algorithmic core relies on the PySIP library for stochastic state-space inference [19].

The human related sources of error will remain out of scope of this work: poorly placed sensors, human error on data acquisition, and human error on algorithm configuration. These sources relate indeed rather to efficient training of the operator and general good practice. They are admittedly important to tackle, but are left outside the scope of this paper.

Then remains uncertainty related to environmental conditions. While solar masks and effect of solar radiation can be considered as accounted for thanks to the Sereine sensor kit, weather variability is yet expected to be a significant source of error given earlier work [18]. Weather-related uncertainty may

also interact with the building-type. Depending on whether the building is externally or internally insulated, or depending on its thermal inertia, the relative effects of weather conditions are likely to differ.

2.2. Sereine measurement equipment and on-site data acquisition setup

The equipment and on-site measurement setup used by the Sereine method are detailed in the next two sections.

The Sereine measurement equipment

The Sereine sensor kit includes equipment for measuring the building’s boundary conditions and for controlling and monitoring the indoor heated volume. While extensive details of the Sereine kit can be found in Juricic et al. [18] and Challanson-nex et al. [20], a brief description is provided here for clarity.

The outdoor equipment aims at measuring an overall equivalent temperature as defined in Bouchié et al. [21]. To do so, SENS probes are deployed in front of each wall and over each roof surface with matching inclination and orientation. The probes, see Figure 2, measure a so-called equivalent temperature. The equivalent temperature represents the apparent temperature experienced by the building envelope surface due to factors such as outdoor air temperature, global irradiation, snow, wind-driven rain and convection. A weighted average from all deployed SENS probes gives an overall equivalent temperature used as external boundary temperature.

Indoors, the Sereine kit utilises portable heating devices (Figure 3), which are simultaneously controlled by a central logger-controller. The devices are equipped with fans that ensure homogeneous temperatures across the building.

The Sereine protocol

The Sereine test relies on a perturbation method where the building envelope and the indoor environment are fully controlled. Solar gains are limited when possible by closing all blinds. If necessary, glazed surfaces that have no blinds are covered with cardboard. Ventilation is shut down and air inlets are sealed similarly to an airtightness measurement. However, airtightness defects, if reported, should be left as they are.

The indoor air temperature is controlled by portable heating devices of the Sereine kit. The

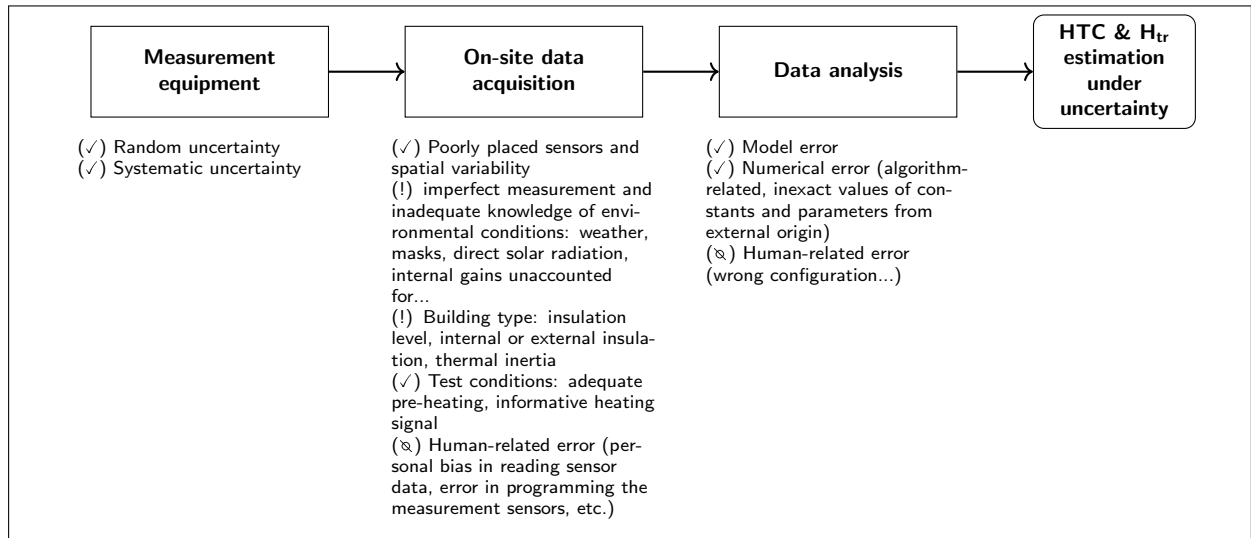


Figure 1: Uncertainty analysis and sources of error of an HTC estimation – application to the Sereine test:

(✓) accounted for (⊗) out of scope (!) untackled or partially tackled and influential



Figure 2: SENS sensor : measurement of the equivalent temperature of a surface (source picture: INES)



Figure 3: Sereine portable heating device (source picture: AQC)

temperature setpoint follows a pseudo-random signal, as defined in Juricic et al. [18] and depicted in Figure 4.

Data acquisition may last one, two days or more, depending on the building type and the insulation level. Regarding the required duration, it is yet to be determined.

2.3. Sereine data analysis

2.3.1. Appropriate models for data analysis and automation

To analyse the collected data on-site, an appropriate model must be used. A model is deemed

appropriate when it reproduces with sufficient accuracy the dynamics observed in the data. Such model cannot be known in advance. In agreement with the existing literature [22], the most appropriate model must be chosen among a set of models through a model selection procedure.

The set of RC models used in the Sereine data analysis is given in Appendix A. The set comprises first order models up to 6th order models. Up to 3 boundary conditions can be considered, like an adjacent unheated space, a crawlspace or unheated attics. Figure 5 shows 3 examples of different boundary conditions: a house with an adjacent unheated

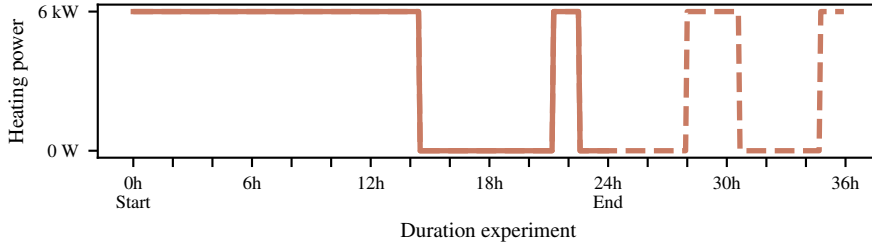


Figure 4: Sereine heating power control follows an optimized pseudo-random signal (here cut at 36 h) [18]

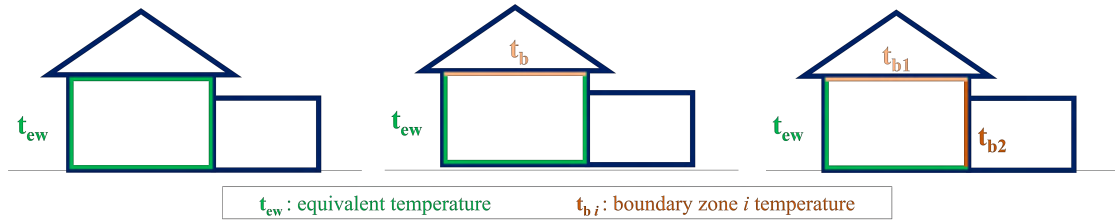


Figure 5: Accounting for unheated boundary zones in the RC model set: either (1) an equivalent temperature t_{ew} is inferred for the entire envelope or (2) boundary zones are included in the model (through temperatures $t_{b1,2}$), with or without thermal capacity

garage and unheated attics can be modeled with an overall single boundary condition, or the measurements can be split into 2 or 3 boundary conditions. When applicable, all resulting models are considered in the model selection procedure.

As for the model selection procedure, the literature [23, 24] suggests that a likelihood ratio-based selection procedure is appropriate for RC model selection. However, these papers primarily aim to find models suitable for accurately predicting building thermal dynamics. In this research project, with a particular focus on physical interpretability in this research project, previous work [18, 20] suggests that the Bayesian Information Criterion (BIC) achieves higher physical interpretability of the estimated parameters. Therefore, a BIC procedure is applied in the Sereine data analysis.

Part of model error influence is solved with the use of a stochastic grey-box color[25]. Although it is not the orthodox approach presented in Kennedy and O’Hagan [15], the use of a stochastic Gaussian process catches part of the model error and part of the random variability error. In addition, the model error is also minimised through the RC model selection procedure, which basically finds an optimum between number of parameters, henceforth higher interpretability, and low residuals. It is important to note that without model selection, an unfit RC

model, albeit stochastic, is unlikely to provide an accurate uncertainty estimation.

Finally, an automated parameter initialisation procedure is adopted from Faure et al. [26]. In summary, this procedure uses the highly practically identifiable first-order model ti (1 capacitance, 1 resistance as shown in Appendix Appendix A) to initialise the other models. Faure et al. [26] demonstrated that this initialisation algorithm significantly improved the convergence of the Sereine set of models.

2.3.2. Sereine test output: HTC and H_{tr}

The selected stochastic RC model takes as input the outdoor equivalent temperature and the heating power to predict the indoor air temperature. The Heat Transfer Coefficient, i.e., the overall heat transfer through the envelope, air infiltrations comprised, is directly inferred from the inverse problem resolution with the selected RC model.

However, heat transfer through air infiltration is variable and cannot be considered as an intrinsic property of the envelope. In the end, the HTC is itself variable and will depend on wind speeds and temperature differences. The HTC as a measurand is per se not repeatable.

The Sereine method separates therefore the whole-building HTC into an infiltration transfer co-

efficient H_{inf} (which depends on weather conditions during the test) and a transmission heat transfer coefficient H_{tr} (see Equation 1 from the ISO 13789 standard [4]). Coefficient H_{tr} is an intrinsic property and, therefore, a preferable measurand that can be more easily compared to theoretical calculations.

$$HTC = H_{tr} + H_{inf} \quad (1)$$

The Sereine data analysis can make in the end two consecutive estimations: first, a regular HTC estimation from the measured heating power P_h , and then an H_{tr} estimation based on a corrected heating power P'_h defined as follows: $P'_h = P_h + P_{inf}$.

The heat power gained or lost by infiltration, P_{inf} , is conventionally defined by Equation (2):

$$\begin{aligned} P_{inf} &= H_{inf}(T_{ext}-T_{int}) \\ &= Q_{inf}C_{p,v,air}(T_{ext}-T_{int}) \end{aligned} \quad (2)$$

where:

- H_{inf} heat transfer coefficient [W/K] by infiltration
- T_{ext} the outdoor air temperature [K]
- T_{int} the indoor air temperature [K]
- Q_{inf} the air flow rate by infiltration [m^3/h]
- $C_{p,v,air}$ the volumetric heat capacity of air [$Wh/m^3/K$]

The estimation of Q_{inf} is based on the standard EN 16798-7 (method 1) [27]. This simplified aeraulic model uses as inputs an ISO 9972-compliant [28] air permeability test results, time-varying wind speed measurements, main building dimensions, and outdoor pressure coefficients. A detailed explanation of this model can be found in Thébault and Millet [29].

Uncertainty of the infiltration model is accounted for, as detailed in the next section. In particular, the outdoor pressure coefficients (mainly depending on the nearby air barriers around the building) and the wind speed measurements (taken from the nearest weather station, can differ from the local site at 10 meters height). These combined uncertainties typically result in a 50 % error in the estimated infiltration flow rate P_{inf} , but its impact on the H_{tr} result remains generally acceptable when the building is highly airtight, and when the test conditions are not excessively windy.

This infiltration model is admittedly a simplification of the complex air flows by infiltration during the Sereine test. However, it represents a reasonable compromise between not accounting for air infiltration and using accurate but expensive methods like gas tracer measurements. It effectively captures most of the variability due to air infiltration, except when wind speeds are extremely high or air leakages substantial [29, 30].

2.3.3. Uncertainty propagation of sensor systematic errors

So far, the Sereine data analysis does not account for sensor and estimated variable-related bias. As identified earlier, sensor-related bias can occur when the sensors themselves exhibit systematic errors, or when the spatial variability of the measurand is significant. Similarly, coefficients such as solar absorbance and pressure coefficients (used to compute the equivalent outdoor temperatures and estimate the infiltration airflow) can also contain significant biases relative to their estimated values.

After the model calibration sub-process, the Sereine data analysis adds an uncertainty propagation for systematic errors. The principles and validation of this second step, introduced in Thébault and Bouchié [7] and applied in de Carvalho Araujo et al. [31], involve combining model calibrations with a quasi-Monte Carlo method:

1. Systematic uncertainties are defined, each with corresponding probability density functions;
2. a Sobol sampling produces a set of n samples, each added to the measurement data, resulting in n biased datasets;
3. Each biased dataset is processed using the Sereine analysis procedure to produce an HTC estimate;
4. All n estimates are summed to form a new probability density function;
5. A 95 % confidence interval is inferred from the new density function.

The latest novelty is to include a convergence criterion, based on the overall variance variation. With a Sobol sampling of the inputs, it allows stopping the propagation when the convergence criterion is met, thereby avoiding unnecessary computations. The resulting HTC uncertainty is henceforth named u_{step2} .

Bounds and probability densities for the considered systematic uncertainty sources are presented in Appendix B.

2.4. Conclusion: uncertainty quantification in the Sereine test

On-site thermal performance estimation of a building envelope is prone to a number of uncertainty sources that jeopardise the estimation process as well as its repeatability and reliability, in particular for tests of short duration. Section 2 has detailed the Sereine experimental setup and data analysis, and has identified the most influential sources of error.

Among those, two significant sources of error are to this day not accounted for in the Sereine analysis: weather conditions and building-type. Yet, literature and previous work suggest it is necessary to quantify the weather related and building type related uncertainty. This paper will now focus on quantifying these uncertainties and proposing a method to account for them in the final estimation of the Heat Transfer Coefficient.

3. Methodology: quantification of the building type and weather related bias

Weather conditions and building type significantly influence the accuracy of a Heat Transfer Coefficient estimation. A numerical experiment setup that has therefore been developed to assess and quantify this influence is described in this section.

3.1. Numerical experiment setup

The objective of the numerical experiments is to (1) assess the variability of a Sereine test result under weather conditions representative of French climates and (2) quantify the associated uncertainties. The numerical setup relies on a building energy model in which a Sereine test, with modeled Sereine sensors, is simulated.

For a simulated Sereine test to be analysed identically to an actual on-site test, the building energy simulation needs to provide 5-minute time-step simulated data. The modeling tool and its algorithms need therefore to be physically relevant at this frequency. The simulation program EnergyPlus has been chosen not only for its modeling abilities [32] but also because it is open source and widely used in the building energy community [33].



Figure 6: France representative weather data for the numerical Sereine tests

The simulated Sereine tests are performed year-round, once a week, except during the first week of the simulation to avoid the influence of the initialisation process on the first day. Between two tests, the indoor temperature set-point is 20 °C. For each simulated year, there are, therefore 51 Sereine tests.

A numerical setup allows for easy changes in the weather data to represent one or more climates. Since the funding project foresees deployment in the French metropolitan area, the IWEC weather data for 21 cities have been chosen to represent the French climatic variability. Figure 6 shows the 21 locations of the IWEC data, either in France or in border locations. As summarised in Figure 7, with 51 tests a year and 21 sets of yearly weather data, more than 1000 Sereine tests are simulated.

It should be noted that the weather data used in this study do not represent extreme weather conditions. This does not affect the validity of the methodology because it is highly unlikely that a Sereine test would be conducted during extreme events, such as a winter storm or a heatwave. Consequently, high wind speeds and elevated outdoor temperatures are de facto excluded, and the weather variability used in this methodology is based on typical data.

3.2. Numerical quantification of weather related bias for two building-archetypes

The ambition is to reconstruct a faithful estimation of the actual uncertainty for the HTC es-

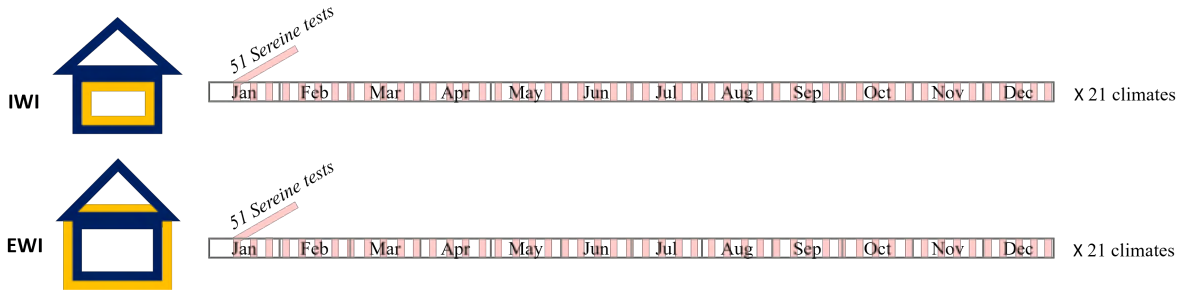


Figure 7: Numerical test bed: two building-types, 51 Sereine tests over the entire year repeated for each of the 21 weather data

climate. Therefore, the next section will quantify the weather-related uncertainty to establish a final coverage interval, as defined in the VIM [13]: an "interval containing the set of true quantity values of [the] measurand with a stated probability, based on the information available". It is important to note that this interval is not a "confidence interval" as it fundamentally differs from the statistical concept, as explained in the Guide to the expression of uncertainty in measurement [17].

Determining the weather conditions that allow for accurate HTC estimations and quantifying their associated uncertainty must be accomplished through a large number of experiments conducted under representative weather conditions. Conducting an extensive on-site experimental campaign is not feasible due to its time-consuming costly nature. Instead, numerical experiments have been set up.

As described in Section 3, the Sereine test is virtually implemented in two numerical energy models representing externally and internally insulated houses, and it is performed over 50 times across 21 climates. The building energy models are based on the geometry of a case study later used for the experimental validation (see section 4). However, the envelope composition have been modified to represent typical IWI and EWI house envelopes in France. Like any building energy model, a theoretical HTC can be obtained: $HTC^* = 89.3 W/K$ and $H_{tr}^* = 82.2 W/K$ for the EWI house and $HTC^* = 112.9 W/K$ and $H_{tr}^* = 106.7 W/K$ for the IWI house.

Figure 8 shows how the HTC estimations from 48-hour data in the IWI house vary throughout the entire year and across all 21 climates. The theoretical HTC value of the building energy model is provided as a reference with a $\pm 20\%$ margin. The

HTC estimations are notably less accurate in summer than in winter. This is likely due to the smaller temperature difference between the indoor and outdoor environment, which affects both the accuracy and precision of the HTC estimation.

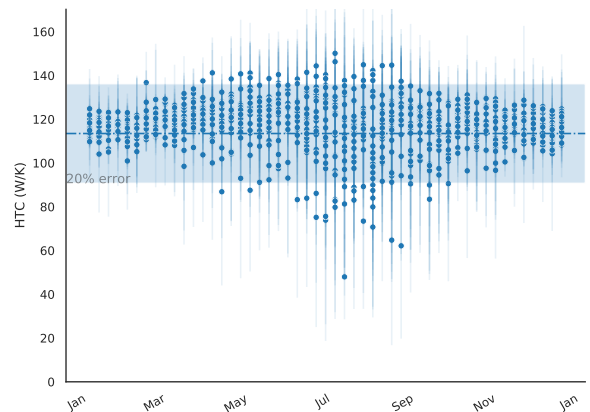


Figure 8: HTC variability across entire year: estimations over 48-hour data show highest variability in summer and lowest in winter

Figure 9 then focuses on the direct influence of the outdoor temperature on the 48-hour HTC estimations of the IWI house. To achieve this, an average outdoor temperature is calculated based on the observed outdoor temperature during the 48-hour analysed data.

Two observations can be made from Figure 9: (1) temperature is clearly influential, with a slight increase in HTC values as outdoor temperatures rise and (2) variability remains moderate up to an average of approximately $8^\circ C$ beyond which variability increases significantly.

The initial objective is to quantify a maximum bias observed in HTC estimations with the Sereine test, regardless of the season. Performing this quan-

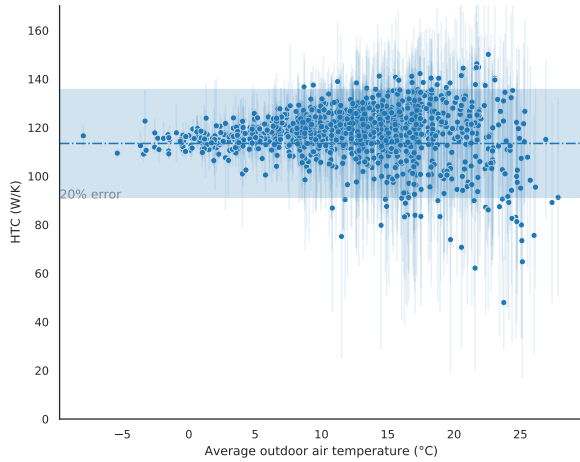


Figure 9: HTC variability dependent on outdoor air temperature

tification once for the entire year would be somewhat misleading, as it would not accurately reflect the evolution of variability across different seasons.

Instead, a maximum bias is calculated as a function of the average outdoor temperature during the test. The calculation principle, illustrated in Figure 10, is as follows: (1) calculate the average outdoor temperature during the test and designate it as T_{thres} , (2) define a moving window ranging from $[T_{thres} - 5 \text{ to } T_{thres}]$, and consider all results within this interval for step (3), where the 95th percentile provides the maximum bias for the T_{thres} moving window. For example in Figure 10, with $T_{thres} = 15 \text{ }^\circ\text{C}$, the maximum observed bias is approximately 17 %, denoted $MOB^{IWJ}|_{thres=\bar{T}}$.

The calculation of the maximum bias is performed for four test durations: 12 hours, 24 hours, 36 hours, and 48 hours. The results are displayed in Figure 11. The same calculations are carried out for the externally insulated version of the house across six different test durations, and the results are presented in Figure 12. Ideally, the maximum bias should be estimated for a broader range of durations. However, each estimation is computationally expensive, which is why we have focused on multiples of 12 hours.

From Figure 11, it is evident that a 12-hour Sereine test in a well-insulated house exhibits maximum biases exceeding 60 % even in winter. This indicates that when combining uncertainty propagation with the maximum bias for a 12-hour test, the resulting confidence interval is likely not reasonable. Consequently, estimating the HTC based

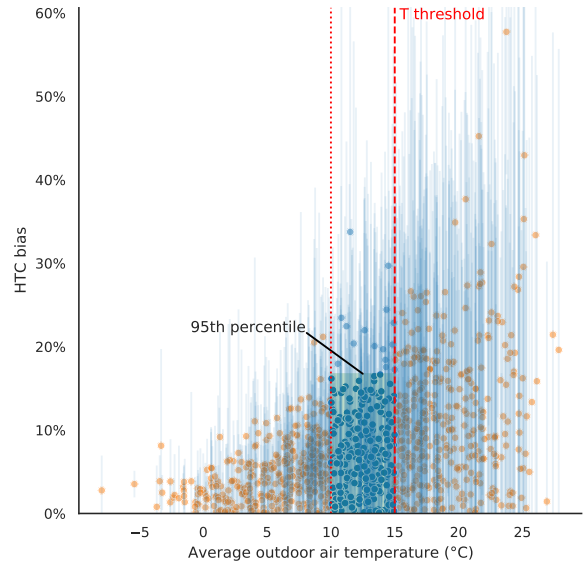


Figure 10: Maximum bias calculation: moving window principle. In this example, the average temperature is $T_{thres} = 15 \text{ }^\circ\text{C}$. The 95th bias percentile over the temperature range 10–15 $^\circ\text{C}$, i.e. the observed bias to add to the measurement uncertainty, is 17 %.

on a 12-hour test is not feasible for an insulated house. Lower maximum bias values are achieved with longer test durations. In general, a winter test is feasible within 24 hours, and an early mid-season test can be accomplished within 48 hours.

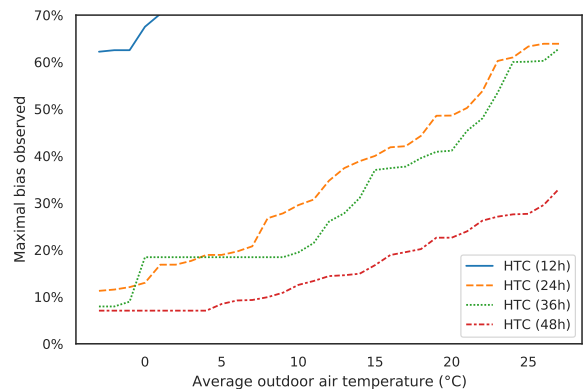


Figure 11: Maximum bias for an internally insulated house

As for the externally insulated house, the lowest maximum biases are obtained with test durations longer than 48 hours. In winter, when average outdoor temperatures are lower than 8–10 $^\circ\text{C}$, a 48-hour test should yield acceptable final uncertainties. In mid-season however, the maximum bias

grows and is likely to make the final HTC estimation uncertainty too high to be acceptable for stakeholders. Longer durations, 72 h or 96 h, may then be needed for HTC measurements in mid-season.

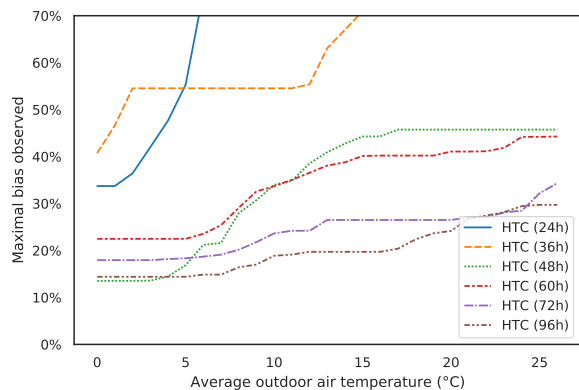


Figure 12: Maximum bias for an externally insulated house

The final step is to account for the maximum bias in the overall uncertainty of the HTC estimate. The maximum bias is assumed to follow a uniform distribution. Thus, the uncertainty on the \widehat{HTC} estimate is calculated as follows:

$$u_{MOB} = \frac{MOB|_{thres=\bar{T}}}{\sqrt{3}}$$

The final HTC uncertainty estimation, denoted u_{HTC} , is calculated using the classical formula:

$$u_{HTC} = \sqrt{(u_{HTC}^{step2})^2 + u_{MOB}^2}$$

Here, u_{HTC}^{step2} represents the HTC uncertainty resulting from the systematic error propagation at step 2 (as described in Section 2.3.3).

It is important to note that this is not the place to establish whether there exists a specific reference value for the maximum bias. While a 50–60 % bias was already considered too high, as it could render HTC estimations non-significant when combined with other sources of uncertainty, the acceptability of a 20 % or a 30 % maximum bias is not yet relevant. What truly matters is the final confidence interval, which combines the maximum weather-related bias with the other sources of uncertainty. We believe it is the responsibility of the test operator and, ultimately, the stakeholders to determine whether or not the final HTC uncertainty is acceptable. We further discuss this perspective in Section 5.



Figure 13: INCAS platform at the French National Solar Energy Institute (source picture: CEA)

4. Experimental assessment of the Sereine test

To validate the concept of incorporating a maximum observed bias and, consequently, assess the reproducibility of the Sereine method, an experimental campaign was conducted on the I-BB house at the INCAS platform. Nine Sereine tests were carried out, and this section presents the HTC estimations obtained from these tests. Additionally, an analysis of the uncertainty, both with and without the maximum bias, and relatively to the weather conditions.

4.1. The case study : INCAS platform house I-BB

To validate the feasibility and reproducibility of the Sereine method, a comprehensive case study is conducted on house I-BB. This house is a full-scale building located within the experimental INCAS platform, as depicted in Figure 13, at the French National Solar Energy Institute (INES). The platform is situated in Le Bourget-du-Lac (N: 45°650, E: 5°867). House I-BB has a concrete cast structure and is externally highly insulated.

A co-heating test has been conducted to establish a reference value for the HTC, resulting in $HTC_{coheating}^* = 80.7 \text{ W/K}$ ($\pm 7.2 \%$ calculated in accordance with Gori et al. [6]). Additionally, a pressurization test has been carried out and repeated to assess the air permeability level, yielding $n_{50} = 3.5 \text{ h}^{-1}$ ($\pm 5.8 \%$).

4.2. Overview of the experimental campaign

Table 1 provides an initial summary of key details about the tests, including the test dates, the

Table 1: Summary of practical information for the experimental campaign conducted on house I-BB. The tests are arranged in order of average outdoor temperature for ease of analysis.
 * long datasets have been truncated for the analysis.

| | Started on | Average 48-hour temperature | Average 48-hour wind speed | Duration dataset |
|---------|----------------------|-----------------------------|----------------------------|------------------|
| Test #1 | Nov 3 rd | 6.4 °C | 1.55 m/s | 96 h |
| Test #2 | Oct 23 rd | 6.7 °C | 1.25 m/s | 48 h |
| Test #3 | Apr 14 th | 7.4 °C | 5.48 m/s | 94 h |
| Test #4 | Oct 27 th | 7.8 °C | 1.08 m/s | 48 h |
| Test #5 | Oct 15 th | 8.6 °C | 1.32 m/s | 48 h |
| Test #6 | Oct 11 th | 10.2 °C | 3.03 m/s | 48 h |
| Test #7 | Apr 23 rd | 13.5 °C | 1.50 m/s | 90 h * |
| Test #8 | Oct 8 th | 13.5 °C | 3.35 m/s | 48 h |
| Test #9 | Oct 19 th | 14.8 °C | 3.89 m/s | 48 h |

recorded average outdoor temperature, and the amount of available data.

Tests #1 and #2 were conducted under the coldest conditions, with average temperatures below 7 °C, while tests #7, #8 and #9 were performed under milder weather conditions with average temperatures exceeding 13 °C. Based on the numerical findings, it is expected that tests #1 and #2 will exhibit better accuracy than tests #7, #8 and #9. To be more precise, it is not excluded that tests #7, #8 and #9 perform well in isolation, but it is likely that, when repeated, their overall accuracy is lower.

4.3. Including the maximum bias: an illustration with test #7

Figure 14 provides an illustrative representation of the novel uncertainty quantification of the \widehat{HTC} estimate. The initial uncertainty, shown in blue, predicts an error in the range of 7–30 %. It is unexpectedly low with a 12-hour dataset, and the confidence interval is far from the reference value. After 24 hours, the HTC estimate converges toward the reference value and its confidence interval contains the reference value.

When the HTC confidence interval includes the maximum bias, shown in brick-orange in Figure 14, the overall uncertainty significantly increases for short test durations. It takes more than 48 hours for the overall uncertainty to decrease below a 50 % error and over 72 hours to decrease below 33 %. For a 12-hour duration, the overall uncertainty is much larger than without the maximum bias, providing a more realistic representation of the actual uncertainty of the HTC estimate. In other words, it means that with a 12-hour dataset and under other

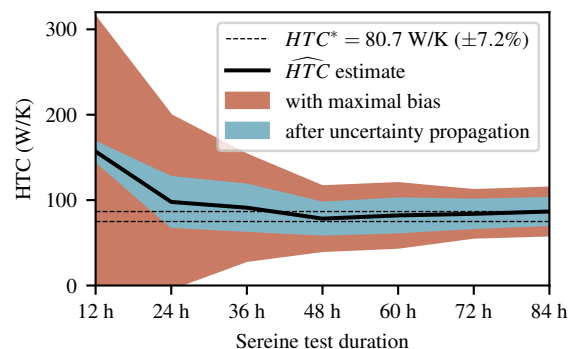


Figure 14: Convergence of the HTC estimation and maximum bias - illustration on test #7 on the I-BB EWI house

measurement conditions, a significant variability of \widehat{HTC} estimates is very likely to be observed.

4.4. Reproducibility of the HTC measurement on house I-BB

Let us now compare the results of all 9 Sereine tests conducted on the I-BB house.

Figure 15 displays the 48-hour \widehat{HTC} estimates of all 9 Sereine tests on house I-BB and the 72-hour \widehat{HTC} estimates for the three longer tests. Both the initial uncertainty (in blue) and the novel uncertainty including the maximum bias (in brick-orange) are shown. The co-heating reference value $HTC_{coheating}^*$ is provided with its associated 7.2 % uncertainty.

As expected, Figure 15 illustrates that confidence intervals without maximum bias do not always intersect with the HTC^* reference value: four out of nine tests fail to do so. With the expanded uncertainty calculation, the confidence intervals intersect

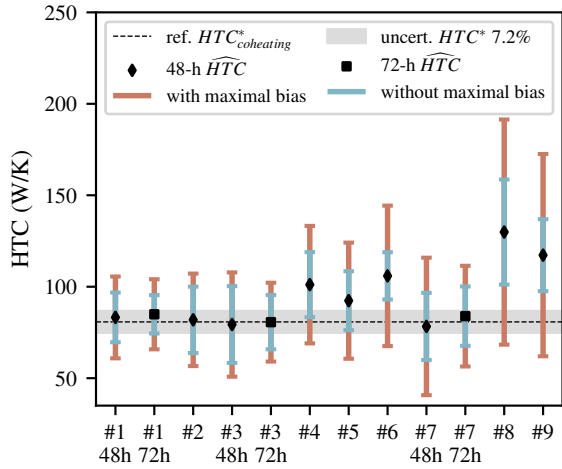


Figure 15: All 48-hour HTC estimations and three 72-hour tests – EWI house I-BB

the reference for all 9 tests. Figure 15 also demonstrates how a longer test duration results in a lower overall uncertainty for the \widehat{HTC} estimate.

Figure 16 illustrates the influence of average outdoor temperatures on the 48-hour \widehat{HTC} estimates and its associated uncertainty calculation. As expected, higher outdoor temperatures result in greater variability in \widehat{HTC} estimates, as seen in tests #7, #8 and #9. For example, Test #7 performs well, but test #8, under almost identical conditions yields a very different estimate. In contrast, lower temperatures generally lead to both good accuracy and precision, except perhaps for Test #4, where the poorer precision is compensated for by a confidence interval that intersects with the reference value.

It is evident that variability in estimates caused by outdoor conditions is not adequately accounted for in the uncertainty calculation without the inclusion of the maximum bias, shown in blue in Figure 16. In this case, the confidence intervals remain similar, regardless of the outdoor temperature and, consequently, they do not reflect the true uncertainty associated with on-site HTC measurements.

5. Discussion

In this paper, we have presented the Sereine test procedure in detail, analysed its sources of uncertainty, and numerically quantified the weather- and building type-related maximum bias, which has

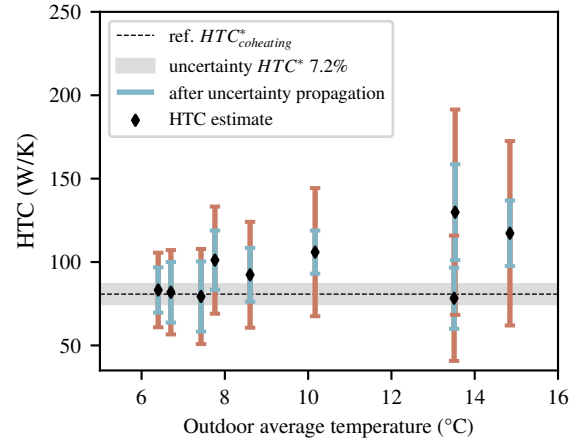


Figure 16: HTC estimation variability as a function of outdoor temperature – EWI house I-BB

been added to the uncertainty quantification of the HTC estimation. Through repeated tests on an experimental house, we have demonstrated that the novel expanded coverage interval includes the reference value in all measurements.

On the influence of outdoor temperature

It is not surprising that outdoor temperature was found to have a significant impact on the HTC uncertainty. Earlier work [18] revealed that higher temperatures were associated with lower physical interpretability of the HTC estimation, i.e., the HTC estimate was either far from the reference value and/or that its uncertainty interval was wider. The literature on other methods also supports the influence of weather conditions on accuracy [9, 8, 11]. The novelty of the Sereine method lies in quantifying the maximum bias and providing a coverage interval that almost always contains the true value.

On the inter- and extrapolation of the maximum bias

The question of whether the numerical bias quantification can be inter- and extrapolated is not trivial. It is reasonable to expect that the maximum bias would be much lower in a non-insulated house than in a well-insulated one, all other things being equal. However, without further numerical tests, inter- and extrapolation becomes a delicate task.

From a physical perspective, we can expect a continuous behaviour, without any form of tipping point. Linearity, though, is unlikely, and a logistic behaviour might be more plausible. This aspect remains to be explored in future research.

In addition, it is worth noting that house I-BB has an HTC^* reference value of 80.7 W/K ($\pm 7.2\%$) whereas the externally insulated numerical model has a reference of 89.3 W/K. This means that the numerical bias was extrapolated towards lower values of HTC. It can be expected that the expanded coverage interval has been somewhat underestimated.

In any case, it would be prudent to quantify the bias for moderately insulated building models to gain insights into interpolation, and for highly insulated models to explore the validity of extrapolation. Additional numerical tests should be complemented by more field experiments. A larger experimental campaign was launched in 2022 and will continue until 2024, with up to 200 tests expected.

Another open question regarding interpolation is that the maximum biases have only been calculated for multiples of 12 hours. As a consequence, the only test durations admissible are also multiples of 12 hours. Whether interpolation of the maximum bias is feasible has not been assessed but could be interesting. It would avoid making the calculation for all biases at minute time-steps. In the end, having 1-hour time-step maximum bias would be advantageous, as it would enable convergence checks and potentially reduce the overall test duration in real-time.

Another point worth discussing is the outdoor boundary condition. Section 2.2 showed how the Sereine kit measures an overall equivalent temperature. This makes the Sereine method different from other on-site data based methods that use an outdoor air temperature but not less compliant to the ISO 13789:2017 standard [4]. The ISO 13789 standard refers indeed to the ISO 6946:2017 standard [34] which opens the possibility to use other measurand than the air temperature to represent the external environment "when appropriate to the purpose of the calculation". Our standpoint is that measuring an outdoor equivalent temperature is an advantage for the Sereine test to be used as a conventional reference.

That the HTC is time-varying is not surprising, since it includes heat transfer through air leak-

ages, which are dependent on temperature difference and wind. The H_{tr} when defined as a function of indoor and outdoor air temperatures is however also time varying, since solar irradiation, radiative transfers with the environment as well as wind speed play a significant role in the actual thermal transfers through the envelope. This means that there is a significant model error in an outdoor air temperature defined H_{tr} definition as well. So considering the actual on-site measurement conditions of an HTC or an H_{tr} measurement, the outdoor equivalent temperature can be considered representative and "appropriate" in the sense of the ISO 6946:2017.

In a recent work, Li [35] studied the discrepancies between ISO-calculated U-values (and henceforth whole-dwelling HTCs) and values calculated from on-site measurements. Li [35] pinpoints the discrepancies to the fact that outdoor air temperature is used as a proxy for environmental temperature, hence introducing a model error in the calculation. His discussion as to whether a "true" HTC exists actually questions the HTC (and counterpart H_{tr}) as an adequate measurand. The HTC is indeed widely used as the reference indicator of the envelope performance [5] but can never be considered as an intrinsic and constant performance indicator. To tackle this, Li [35] proposes either to characterise the surface resistance relatively to environmental conditions or to define a scenario-specific HTC reference.

The Sereine method described in this paper rather proposes to define a conventional H_{tr}^* based on the equivalent temperature. The use of the SENS probes reduces the variability due to convective transfers, accounts for radiative transfers and direct solar radiation. Admittedly, the data analysis relies on an expert-guessed wall and roof tile emissivity. The resulting uncertainty is however propagated in the uncertainty analysis. As a result, the measured $\widehat{H_{tr}}$ presents a lower variability and is one step closer to an intrinsic property of the envelope and henceforth better suited for certification or performance contracting.

Last but not least, we would like to discuss the concept of expanded coverage interval. As suggested by Deb et al. [3], from a stakeholder point of view, larger intervals may be seen as a worse measurement that cannot be trusted. From where we stand, the expanded uncertainty is on the contrary

an opportunity.

When sources of uncertainty are quantified and accounted for, a large coverage interval is full of information: it means that effort needs to be put on data quality, measurement duration, number of sensors, etc. On this matter, the Sereine method does not so far include operator related uncertainty. Human related errors could however be influential. For the Sereine method to be reproducible in the strict sense of the GUM [17], this question should be addressed in the future. Understanding and identifying the sources of uncertainty can lead to directed efforts to reduce that uncertainty.

Consequently, we can consider that it is the task of the test operator to venture an opinion on the acceptability of the final HTC uncertainty, with in sight the expectations of the client. In any case, an expanded coverage interval is the actual assurance of reproducibility and is therefore a step closer to a reliable HTC estimation for certification.

Given the results of Deb et al. [3], stakeholders should probably be clearly made aware of the influence of all sources of uncertainty of an on-site HTC measurement, regardless of the method used. When uncertainty of the HTC measurement is transparently quantified, a compromise between cost, duration and accuracy will inevitably rise. For an HTC measurement to be both cheap and swift, it cannot be highly accurate, and vice-versa. This might come as a surprise for stakeholders and be an obstacle to measurement based verification of the envelope performance. Transparently and thoroughly quantifying the HTC uncertainty is therefore an important step towards acceptance of on-site methods by all parties involved.

6. Conclusion

In a global effort to reducing the carbon emissions from the building sector, reliable on-site thermal performance measurements could become a valuable asset to reducing the building intrinsic performance gap in newly built or retrofitted buildings. By a thorough uncertainty quantification approach, this paper aimed at identifying, quantifying and accounting for the sources of uncertainty in a Sereine test for Heat Transfer Coefficient measurement.

This paper first detailed the steps of a Sereine test and examined the sources of uncertainty at each step. Uncertainties related to weather and building type have been identified as significant factors affecting accuracy and precision. To the extent

of our knowledge, neither the Sereine method nor any other method accounts for these factors in their uncertainty estimation.

A numerical test bed has been set up to quantify the uncertainty related to the weather conditions. This quantification has been performed first on an internally insulated building model and then on an externally insulated model, showing significant differences in variability. These numerical tests have shown that short tests, typically those under 24 hours, and tests on externally insulated houses come with larger residual uncertainty. An exhaustive uncertainty estimation of the Sereine HTC measurement is then proposed as a coverage interval.

With this novel systemic approach, the Sereine method is put to the test in an experimental campaign. The coverage interval of all 9 tests contained the expected Heat Transfer Coefficient value suggesting a satisfactory reproducibility of the Sereine method in its new uncertainty analysis approach.

Finally, the paper argues that the acceptability of the final coverage interval is a shared task between the test operator and the client: the client may express a target value with an associated uncertainty target based on which the test operator may propose appropriate test conditions and duration. It is our belief that such agreement on a target uncertainty can only be achieved on the basis of a transparent and thorough uncertainty analysis. Uncertainty quantification holds indeed valuable information and should therefore be considered as an opportunity. In the end, only a consensual agreement on target performance values and associated uncertainty can advantageously nourish the grounds for effective energy performance verification and contracting.

Acknowledgments

This work has been supported by the PRO-FEEL program which funded the Sereine project. The authors also address their warmest thanks to Stéphanie Derouineau (CSTB) for her guidance on the project. Let also the participants to the Sereine uncertainty seminar, under the animation of Jade Deltour (CSTC), be thanked.

CRedit author statement

Sarah Juricic: Conceptualisation, Methodology, Writing - Original draft, Visualisation

Mickaël Rabouille: Conceptualisation, Methodology, Data curation, Software, Formal analysis **Arnaud Challansonnex:** Conceptualisation, Methodology, Formal analysis, Data curation, Software, Visualisation **Arnaud Jay:** Conceptualisation, Data curation **Simon Thébault:** Conceptualisation, Software, Formal analysis **Simon Rouchier:** Conceptualisation, Validation **Rémi Bouchié:** Conceptualisation, Methodology

Appendix A. RC models used in the Sereine data analysis

As to make the nomenclature self-explanatory, each model is named according to the nodes associated to a thermal capacity. The name given to the node depends on the position and the element it represents:

- a node representing a thermal capacity of an external wall is named **tw**. If an external wall is modeled with two capacities, the nodes are named **twtw2**;
- the node associated to the indoor air is named **ti**. All models have the **ti** node, but not all have a capacitance at this node, henceforth no **ti** in their name;
- a node associated to an internal mass is named **tm**;
- some models have a parallel transmittance H_b , which has consequently no capacitance and represents heat flux with very low thermal inertia. A letter **b** is then added to the model name;
- Additional boundary conditions nodes from adjacent instrumented zones are named **tb**. Up to two additional boundary conditions are modeled, named **tb1tb2**

With these naming rules, the structure of all models can be determined from its name. Table A.2 sums up all models used in the data analysis process of the Sereine test, with all model names, model orders, the number of boundary conditions they take into account and the degrees of freedom. The latter is used in the model selection process described in Section 2.3.1.

As an illustration, we give here the representation in electrical analogy of models **tw**, **titw** and **tibtw**, respectively in Figures A.17, A.18 and A.19.

Table A.2: Sereine analysis RC model set

| Name | Model order | Number of boundary conditions | Degrees of freedom |
|-----------------|-------------|-------------------------------|--------------------|
| ti | 1 | 1 | 6 |
| tw | 1 | 1 | 7 |
| tib | 1 | 2 | 7 |
| tib2 | 1 | 3 | 8 |
| twti | 2 | 1 | 11 |
| twbti | 2 | 1 | 12 |
| twtm | 2 | 1 | 12 |
| twtw2 | 2 | 1 | 12 |
| twtb | 2 | 2 | 13 |
| twtitm | 3 | 1 | 16 |
| twtw2ti | 3 | 1 | 16 |
| twtw2bti | 3 | 1 | 17 |
| twtw2tm | 3 | 1 | 17 |
| twtbti | 3 | 2 | 17 |
| twtbtm | 3 | 2 | 18 |
| twtw2tb | 3 | 2 | 18 |
| twtb1tb2 | 3 | 3 | 19 |
| twtb1tb2ti | 4 | 2 | 22 |
| twtw2tbtm | 4 | 2 | 22 |
| twtw2tbtm | 4 | 2 | 23 |
| twtb1tb2ti | 4 | 3 | 23 |
| twtb1tb2tm | 4 | 3 | 24 |
| twtw2tb1tb2 | 4 | 3 | 24 |
| twtw2tbtitm | 5 | 2 | 27 |
| twtb1tb2titm | 5 | 3 | 28 |
| twtw2tb1tb2ti | 5 | 3 | 28 |
| twtw2tb1tb2tm | 5 | 3 | 29 |
| twtw2tb1tb2titm | 6 | 3 | 33 |

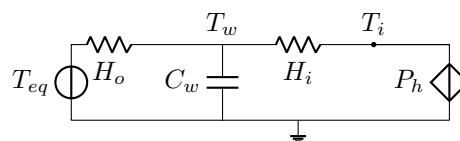


Figure A.17: Model *tw*

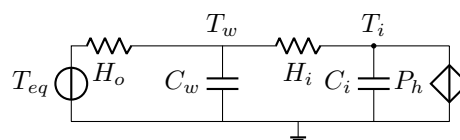


Figure A.18: Model *twti*

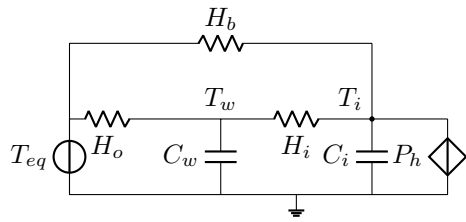


Figure A.19: Model *twbti*

Appendix B. Systematic error sources and quantification

The following table presents the bounds and probability densities for the considered systematic uncertainty sources.

| | Measurand | Potential causes of systematic error | Distribution | Systematic uncertainty range |
|------------------------|--|---|--------------|--|
| Sensor related bias | Indoor temperature | Fixed height of measurement, possibility of insufficient air mixing | Normal | ± 0.5 K |
| | Outdoor temperature | Sensor outdoor position, insufficient ventilation of the module | Normal | ± 1 K |
| | SENS plates temperatures | Local position of SENS sensor, heat transfer on the edges | Normal | ± 0.5 K |
| | Simple temperature sensor | Local position of sensor, local convective and radiative effects | Normal | ± 1 K |
| | Heat power | Meter error | Normal | ± 2 % |
| | Wind speed | Local position of weather station anemometer, uncharted obstructions, measurement error | Normal | ± 30 % |
| Measured coefficients | SENS plates absorbtivities | Measurement uncertainty, dust level | Normal | ± 5 % |
| | Blowerdoor leakage coefficient and exponent | Uncertainties related to the blowerdoor measurement documented in the ISO 9972 [28] | Normal | Based on blowerdoor test report |
| Estimated coefficients | Pressure coefficients CPwindw CPleew CProof | Wind direction, uncharted obstructions | Uniform | $[-0.7,0.3]^*$ $[-0.6,0.4]^*$ $[0.05,0.5]^*$ |
| | Walls and shutters absorptivities | Rough estimation from color chart: judgment error, plurality of wall absorptivities | Normal | ± 0.2 |

Table B.3: Systematic error cause and quantification (* min and max values tabulated according to EN 16798-7 (method 1) [27])

Appendix C. IWEC weather data for the numerical experiments

Table C.4 summarises the location of each IWEC weather file used for the numerical bias assessment in Section 3.

Table C.4: IWEC weather data

| Country | Location | French thermal regulation climatic class |
|---------|------------------|--|
| FRA | Strasbourg | H1b |
| FRA | Dijon | H1c |
| FRA | Bordeaux | H2c |
| FRA | Paris | H1a |
| FRA | Marseille | H3 |
| FRA | Brest | H2a |
| FRA | Clermont-Ferrand | H1c |
| FRA | Montpellier | H3 |
| FRA | Nantes | H2b |
| FRA | Nancy | H1b |
| FRA | Nice | H3 |
| FRA | Lyon | H1c |
| ESP | San Sebastian | H2c |
| ESP | Gerona | H3 |
| ITA | Olbia-Costa | H3 |
| ITA | Torino-Caselle | H1c |
| ITA | Pianosa-IGDG | H3 |
| BEL | Saint Hubert | H1b |
| GBR | London | H1a |
| CHE | Geneva | H1c |

References

- [1] P. de Wilde, D. Coley, The implications of a changing climate for buildings, *Building and Environment* 55 (2012) 1–7. doi:10.1016/j.buildenv.2012.03.014.
- [2] M. A. Mozzo, Setting the Energy Baseline For Performance Contracts, *Strategic Planning for Energy and the Environment* 21 (2001) 12–19. doi:10.1080/10485230109509567.
- [3] C. Deb, L. Gelder, M. Spiekman, G. Pandraud, R. Jack, R. Fitton, Measuring the heat transfer coefficient (HTC) in buildings: A stakeholder’s survey, *Renewable and Sustainable Energy Reviews* 144 (2021) 111008. doi:10.1016/j.rser.2021.111008.
- [4] Thermal Performance of Buildings — Transmission and Ventilation Heat Transfer Coefficients — Calculation Method, Standard ISO 13789:2017, International Organization for Standardization, 2017.
- [5] R. Fitton, R. Bouchié, M. Spiekman, R. Jack, U. Spindler, P. Groesdonk, M. J. Jiménez, A. Erkoreka, A. Marshall, C. Gorse, C. Deb, D. Alan, D. Farmer, G. Masy, V. Gori, G. Pandraud, J. Deltour, L. Gelder, S. Roels, T. Hughes, Building energy performance assessment based on in-situ measurements Challenges and general framework., 2021.
- [6] V. Gori, D. Johnston, R. Bouchie, S. Stamp, Characterisation of Uncertainties in Building Heat Transfer Estimates from Co-Heating Tests, 2023. doi:10.2139/ssrn.4361629.
- [7] S. Thébault, R. Bouchié, Refinement of the ISABELE method regarding uncertainty quantification and thermal dynamics modelling, *Energy and Buildings* 178 (2018) 182–205. doi:10.1016/j.enbuild.2018.08.047.
- [8] S. F. Stamp, Assessing Uncertainty in Co-Heating Tests: Calibrating a Whole Building Steady State Heat Loss Measurement Method, Ph.D. thesis, University College London, 2015.
- [9] R. Jack, D. Loveday, D. Allinson, K. Lomas, First evidence for the reliability of building co-heating tests, *Building Research & Information* 46 (2018) 383–401. doi:10.1080/09613218.2017.1299523.
- [10] J. Deltour, N. Heijmans, K. De Sloover, Assessing the building envelope performance during occupancy, *E3S Web of Conferences* 172 (2020) 22004. doi:10.1051/e3sconf/202017222004.
- [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. doi:10.1016/j.enbuild.2018.06.002.
- [12] Accuracy (Trueness and Precision) of Measurement Methods and Results — Part 1: General Principles and Definitions, Standard ISO 5725-1:1994, International Organization for Standardization, 1994.
- [13] The International Vocabulary of Metrology—Basic and General Concepts and Associated Terms (VIM), Technical Report 3rd edn., JCGM, 2012.
- [14] A. Q. Construction, EPILOG - Evaluation de la Performance intrinsèque de logements — Programme PACTE, <https://www.programmepacte.fr/epilog-evaluation-de-la-performance-intrinsèque-de-logements>, 2017.
- [15] M. C. Kennedy, A. O’Hagan, Bayesian calibration of computer models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2001) 425–464. doi:10.1111/1467-9868.00294.
- [16] D. Maillet, Y. Jarny, D. Petit, Problèmes inverses en diffusion thermique - Outils spécifiques de conduction inverse et de régularisation, *Techniques de l’Ingénieur* (2001) 29.
- [17] Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement, Technical Report First edition, JCGM, 2008.
- [18] S. Juricic, J. Goffart, S. Rouchier, A. Jay, P. Oberlé, Design of a short perturbation method for on-site estimation of a building envelope thermal performance, *Energy and Buildings* 269 (2022) 112211. doi:10.1016/j.enbuild.2022.112211.
- [19] L. Raillon, S. Rouchier, S. Juricic, pySIP: An open-source tool for Bayesian inference and prediction of heat transfer in buildings, in: *Congrès de La Société Française de Thermique*, Nantes, France, 2019, p. 8.
- [20] A. Challansonnex, G. Faure, M. Rabouille, S. Juricic, Méthodologie scientifique de la mesure de la performance de l’isolation de l’enveloppe des maisons neuves et rénovées, Technical Report, Unpublished, 2021.
- [21] R. Bouchié, S. Derouineau, C. Abele, J.-R. Millet, Con-

- ception et validation d'un capteur de mesurage de la température extérieure équivalente d'une paroi opaque d'un bâtiment, in: Conférence IBPSA France, Arras, France, 2014, p. 8.
- [22] H. Madsen, P. Bacher, G. Bauwens, A.-H. Deconinck, G. Reynders, S. Roels, E. Himpe, G. Lethé, Thermal Performance Characterisation Using Time Series Data – Statistical Guidelines, Technical Report Subtask 3 part 2, KU Leuven, 2016.
- [23] J. Leprince, H. Madsen, C. Miller, J. P. Real, R. van der Vlist, K. Basu, W. Zeiler, Fifty shades of grey: Automated stochastic model identification of building heat dynamics, *Energy and Buildings* 266 (2022) 112095. doi:10.1016/j.enbuild.2022.112095.
- [24] P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings, *Energy and Buildings* 43 (2011) 1511–1522. doi:10.1016/j.enbuild.2011.02.005.
- [25] 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. doi:10.1016/j.buildenv.2018.02.043.
- [26] G. Faure, S. Juricic, M. Rabouille, S. Rouchier, A. Challengonex, J. Arnaud, Towards automated stochastic rc model calibration for heat transfer coefficient inference, in: *under review at 13th Nordic Symposium on Building Physics*, 2023.
- [27] Energy performance of buildings - Ventilation for buildings - Part 7 : calculation methods for the determination of air flow rates in buildings including infiltration (Modules M5-5), Standard NF EN 16798-7:2017, AFNOR, 2017.
- [28] Thermal performance of buildings - Determination of airpermeability of buildings - Fan pressurization method, Standard ISO 9972:2015, International Organization for Standardization, 2015.
- [29] S. Thébault, J.-R. Millet, Cost-effective air flow rate estimations using blowerdoor and wind speed measurements to assess building envelope thermal performances, *Journal of Building Physics* 40 (2017) 504–529. doi:10.1177/1744259116659652.
- [30] A. Jay, A. Brun, S. Thebault, A. Fouquier, Dynamic infiltration airflow rate measurement thanks to tracer gas method: A case study at a dwelling scale, in: 15th Roomvent Conference, online, 2020.
- [31] L. de Carvalho Araujo, S. Thébault, L. Mora, T. Recht, A. Sempey, P. Schalbart, M. Cohen, Development of a short duration method to assess the envelope thermal performance of multi-family housings, *Building Simulation* (2023). doi:10.1007/s12273-022-0969-x.
- [32] NREL, Testing and validation, <https://energyplus.net/testing>, Online, accessed 24/01/2023.
- [33] T. Augspurger, E. Malliaraki, J. Hopkins, Open Source in Environmental Sustainability, 2023. URL: <https://report.opensustain.tech/chapters/index.html>.
- [34] Building components and building elements - Thermal resistance and thermal transmittance - Calculation methods, Standard ISO 6946:2017, International Organization for Standardization, 2017.
- [35] M. Li, Thermal performance of UK dwellings: assessment of methods for quantifying whole-dwelling heat loss in occupied homes, thesis, Loughborough University, 2022. doi:10.26174/thesis.lboro.19735699.v1.