
This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Deploying building simulation to enhance the experimental design of a full-scale empirical validation project

PLEASE CITE THE PUBLISHED VERSION

<http://www.ibpsa.org/building-simulation-2019>

PUBLISHER

IBPSA

VERSION

AM (Accepted Manuscript)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Mantesi, Eirini, Konstantinos Mourkos, Christina Hopfe, Rob McLeod, Paraskevi Vatougiou, Matthias Kersken, and Paul Strachan. 2020. "Deploying Building Simulation to Enhance the Experimental Design of a Full-scale Empirical Validation Project". figshare. <https://hdl.handle.net/2134/38180>.

Deploying Building Simulation to Enhance the Experimental Design of a Full-scale Empirical Validation Project

Eirini Mantesi¹, Konstantinos Mourkos¹, Christina J. Hopfe¹, Robert S. McLeod¹, Paraskevi Vatougiou¹, Matthias Kersken², Paul Strachan³

¹ School of Architecture, Building and Civil Engineering, Loughborough University, UK

² Fraunhofer Institute for Building Physics IBP, Holzkirchen, Germany

³ Energy Systems Research Unit, Department of Mechanical and Aerospace Engineering, University of Strathclyde, Glasgow, UK

Abstract

Empirical validation of building simulation results is a complex and time-consuming process. A well-structured and thorough experimental design is therefore a crucial step of the experimental procedure. A full-scale empirical validation study is planned to take place within IEA EBC Annex 71: “Building energy performance assessment based on in situ measurements”. The experimental data are currently being gathered in two experiments being conducted at the Fraunhofer IBP test site at Holzkirchen in Germany. This paper describes the methodology followed during the experimental design of the project. Particular focus is on how Building Performance Simulation (BPS) was used to assist the preparation of the actual experiment, to determine suitable test sequences, magnitudes of heat inputs and temperature variations. A combination of both deterministic and probabilistic simulation (using the method of Morris) is employed to replicate the actual experiment and to assess the sensitivity of the model to uncertain input parameters. A number of experimental errors are identified in the experiment, primarily concerning the magnitude of heat inputs. Moreover, the paper includes a discussion on lessons learned from the simulations and on the reliability, reproducibility and limitations of the suggested experimental design procedure.

Introduction

There are numerous international policies and frameworks currently in force, aiming to help reduce energy consumption in the building environment to tackle the challenges imposed by climate change. As a result, dynamic modelling becomes more widely used for building performance assessment (Clarke & Hensen, 2015) and to demonstrate compliance with building regulations (Raslan & Davies, 2010). Consequently, there is a need to confirm that building performance simulation (BPS) programs are able to provide accurate simulation predictions.

Considerations regarding the input uncertainties and modelling assumptions are two areas that require attention in BPS to enhance the physical correctness of the model and quality of simulation results (Coakley et al., 2014; Mantesi et al., 2018). A BPS model contains hundreds of input variables and parameters. Current state-of-the-art BPS tools have several limitations related to air flow, lighting, HVAC systems, and occupants, among others

(Clarke & Hensen, 2015). Therefore, a detailed validation methodology is necessary to be able to create a reliable comparison between software and reality. Judkoff and Neymark (1995), described the validation methodology adopted by NREL preceding the BESTEST project [also adopted in ASHRAE Standard 140 (ANSI/ASHRAE, 2014)], which incorporated three kinds of tests:

- **Analytical verification:** the output from a program algorithm is compared to the results provided by analytical solutions under simple boundary conditions.
- **Comparative testing:** it is used to compare a simulation program to other programs. This approach includes sensitivity testing and inter-model comparison.
- **Empirical validation:** this allows calculated results from a program to be compared with monitored, experimental data from a real building, test cell or laboratory experiment.

Empirical validation of building simulation results is a complex process. It can include a high level of uncertainty in the experiment, it is considered expensive and time consuming (Ryan & Sanquist, 2012), yet it can test the combined effect of all the internal errors in a program (Lomas et al., 1997). There are a number of previous studies focused on empirical validation of simulation results (Lomas et al, 1997; Loutzenhiser et al., 2007; Kalyanova et al., 2009; Strachan et al., 2015; Cattarin et al., 2018).

A large-scale empirical validation exercise was conducted as part of the IEA EBC Annex 21 project on “Environmental Performance” to assess the predictive ability of dynamic thermal simulation programs. The project compared the results of 25 BPS tools commonly used in EU, USA and Australia at the time (Lomas et al., 1997). Subsequently, several other successful international projects for empirical validation purposes followed including IEA EBC Annex 42 project on “The Simulation of Building-Integrated Fuel Cell and Other Cogeneration Systems” and IEA EBC Annex 43 on “Testing and Validation of Building Energy Simulation Tools” (IEA EBC projects, n.d.). As part of IEA EBC Annex 43, Loutzenhiser et al. (2007) empirically validated the ability of seven solar radiation models implemented in four BPS codes in calculating irradiated solar energy on buildings. A series of experiments was performed in an outdoor test cell. Similarly, as part of a

different project within IEA EBC Annex 43, Kalyanova et al. (2009) used a full-scale outdoor test facility to empirically validate the accuracy of five BPS tools in modelling double-skin facades.

A key observation, as highlighted in (Strachan et al., 2015) is that most monitoring projects have not been designed to provide the comprehensive coverage required for validation of an entire building in simulation programmes. A full-scale building empirical validation experiment was conducted within the IEA EBC Annex 58 project: “Reliable building energy performance characterization based on full-scale dynamic measurements” (IEA Annex 58, 2015). The BPS validation scenario in Annex 58 was designed to test the basic functionalities of BPS tools, mainly focussing on the thermal performance of the building envelope (i.e. transmission losses, thermal bridges, solar gains). In (Strachan et al., 2015) the authors described the empirical validation methodology followed during the Annex 58 project. Among others, the experimental design was characterised as an essential step of the methodology.

A full-scale empirical validation is highly demanding of time and cost. It is therefore essential that the experiments fully excite the building dynamics through a range of boundary conditions and that all important heat and mass flow paths are comprehensively monitored. Consequently, a well-structured, thorough experimental design is a crucial step of the process. The means for undertaking the experimental design in a full-scale empirical validation is to model the selected building, using BPS, in a representative local climate dataset. There are two objectives in undergoing such a step. First, to design the overall experiment by specifying test sequences and suitable experimental configurations that allow to cover the range scientific questions by the following validation. And second, to design the monitoring scheme and determine instrumentation requirements. The latter is achieved with the use of sensitivity studies to identify important simulation parameters that need to be measured, or measured with increased accuracy.

Research Aim

The aim of the paper is to describe the procedure followed to specify the requirements of the experiment for a full-scale empirical validation exercise, currently being conducted within IEA Annex 71 project: “Building energy performance assessment based on in situ measurements” (IEA Annex 71, n.d.). The particular focus is on how BPS was used to define and test the different stages of the experiment, in terms of test sequences, magnitudes of heat input and temperature variations, along with instrumentation requirements.

The research objectives were:

- To create a preliminary synthetic dataset of the experiment to evaluate the usability of the real measurement dataset, before the actual experiment is conducted.

- To conduct a Sensitivity Analysis (SA) to identify important parameters that need to be measured more accurately during the experiment.

Methodology

A whole building monitoring project is currently being conducted on the Fraunhofer IBP’s Twin Houses at Holzkirchen, Germany (Fig.1) (IBP Fraunhofer, 2012). The two houses are of residential layout, detached, identical, located side-by-side to allow direct comparison. They are equipped with extensive measurement and control equipment. The weather data are recorded in-situ. The houses are air-tight, insulated with U-values ranging between 0.29 W/m²K and 0.22 W/m²K for the external walls. The structure comprises externally-insulated brick walls for the external envelope and honeycomb brick walls for the internal partitions, insulated concrete for the ground floor and first floor slabs and lightweight timber-framed roof. All windows are double-glazed with a glazing U-value of 1.2 W/m²K and with electric external roller blinds. The empirical validation exercise is focused on the ground floor and the attic. The cellar space is treated as a boundary condition. A constant flow rate balanced mechanical ventilation system is in operation.



Figure 1: South view of one of the Twin Houses in Holzkirchen, Germany.

The outcome of this BPS validation exercise aims to build on the findings of the empirical validation experiment conducted within IEA Annex 58 (IEA Annex 58, 2015). The former BPS validation scenario in Annex 58, was designed to test the basic functionalities of BPS tools, mainly restricted to the thermal performance of the building envelope. The current project aims to investigate further important aspects, such as building services equipment and the impact of occupancy. Consequently, the two houses are now equipped with different heating systems (i.e. electric radiators compared to underfloor heating) and the synthetic users based on stochastic variations of occupancy profiles (Flett & Kelly, 2017). These synthetic user profiles include internal heat loads, humidity gains and automatically operated windows and doors.

The procedure for undertaking the experimental design for the Twin Houses full-scale empirical validation project comprised the following steps:

1. **Modelling the selected houses using BPS.** The Twin houses were modelled in EnergyPlus V8.8 (EnergyPlus, n.d.) using up-to-date, post-construction drawings of the building geometry and surrounding environment, construction details of the existing fabric, material properties from the manufacturers, surface properties and thermal bridges measured during the previous Annex 58 validation study, infiltration rates measured by a blower door test, information on HVAC systems based on manufacturers' specification, constant user-specified ventilation flow rates (representing the air flow rates that will be used in the actual experiment) and the Test Reference Year (TRY) weather file of Munich, as the average climate data for this location.
2. **Creating stochastic occupancy profiles.** The validation scenario of Annex 71 includes the impact of occupancy. Occupying the Twin Houses with real humans would bring some disadvantages for the experiment and introduce large uncertainties. To avoid this, it was decided to represent realistic deviations in internal heat gains and building operation using "synthetic users". A number of different occupancy profiles were developed for the simulation models; one for the experiment and 99 more for the sensitivity analysis. These represented typical room-wise usage profiles and corresponding internal heat gains based on the use of each space, assuming a typical four-members family of two adults and two children.
3. **Specifying the experimental sequence/experimental schedule.** The different phases of the experiment were specified based on the purpose of the validation exercise and the different parameters for investigation. Three main user profiles were created, each having a different hypothesis. User 1 aimed to investigate if simulation programs are able to handle user interactions like small room-wise occupancy differences and some basic building service equipment, operating in identical setpoint temperature in all rooms and night setback. User 2 was a more complex realistic situation, including operating internal doors and external windows and having different set temperature profiles in the individual rooms. User 3 intended to check if simulation programs are handling the thermal and energetic influences of moisture effects. The experiment included also an initialisation and reinitialization phase, where all the rooms were set to the same constant temperature, aiming to bring both buildings to identical initial conditions. A free-floating phase was included at the end of the experimental schedule to test if simulation programs are able to handle small heat inputs dominated by solar gains under summer conditions correctly. Finally, it was decided to include a Fault Detection and Diagnosis (FDD) phase and a free Pseudo-Random Binary Sequence (PRBS). These

two phases are not primarily intended for the validation experiment but serve the activities of other subtasks within Annex 71 by providing realistic high accuracy measurement data.

4. **Specifying uncertain input parameters.** All input parameters fed to the simulation models were subject to a certain level of uncertainty. Firstly, a list of all uncertain input parameters was made, along with their base values. The uncertain parameters were identified considering what will/can be measured as part of the experiment and what is the information that will be released to the modelling teams of the empirical validation exercise. The base values of these parameters were specified to the best of existing knowledge (e.g. material properties were based on manufacturers' data, surface properties were measured in the previous Annex 58 empirical validation project). Then a uniform distribution with a fixed relative range of 20% was assigned to each parameter to account for major unexpected differences, resulting from measurement uncertainty, system faults and others.
5. **Running the deterministic simulations.** The selected buildings were simulated using the specified experimental schedule and the TRY weather data aiming to replicate the actual experiment. The purpose of this step was to identify and correct possible errors in the experimental specification.
6. **Running the probabilistic simulations.** A Sensitivity Analysis (SA) was carried out to identify important simulation parameters, i.e. parameters that have the most significant influence over simulation predictions and that need to be measured carefully during the actual experiment. The details of this sensitivity analysis are described in the related section later in this paper.

Fig.2 illustrates a schematic representation of the experimental design methodology adopted as part of the Annex 71 full-scale empirical validation project.

The SA was performed using Python SALib (SALib, n.d.). 570 simulations were conducted using JEPlus¹ v1.7 (JEPlus, n.d.). The method of Morris (Morris, 1991; Campolongo et al., 2007) was adopted, as a screening method, to determine the sensitivity of the models to all uncertain input parameters. The Morris sampling method, allows the quantification of the importance of factors as well as those factors that have a direct or indirect effect on the examined output. Its experimental approach is based on 'one-factor-at-a-time' (OAT) experiments, in which the impact of changing the value of each of the chosen factors is evaluated in turn (Giglioli and Saltelli, 2008). This characteristic of Morris technique and its convergence of the sampling procedure (i.e. stability of the rank order) (Morris, 1991; McLeod et al., 2013; Hopfe & Hensen, 2011) qualifies it for this study's SA.

¹ JEPlus is an EnergyPlus simulation manager, used to execute and control multiple simulation.

Table 1 includes the uncertain input parameters investigated as part of the SA, along with their base, maximum and minimum values.

Deterministic Results

The actual full-scale empirical validation project was replicated using BPS. Deterministic simulation was used to create a preliminary synthetic dataset of the actual experiment. The predicted internal air temperatures and heating demand were analysed per zone for each of the different experimental phases and for both Twin Houses. The aim of this step was to assess if the heat input would be enough to excite the dynamics of the fabric, to check if the specified internal heat gains would result in severe overheating that might pose safety issues in the actual experiment and to test whether there were any unpredicted issues in the experimental specification that might compromise the validity of the experiment output. The internal air temperature of both houses, along with the dry-bulb temperature, are plotted in Fig. 3 for one of the main living areas, the ground-floor living room, for three consecutive days in the experimental phase User 2. The heating demand in the living room is plotted for both houses in Fig.4, for the same three-day period in User 2 phase.

The simulation predictions showed insignificant differences in the amplitude of the internal air temperatures of the two houses, although the heat up times after night setback were found to be different between the two heating systems, as expected (Fig.3). However, a noticeable difference was obvious in the predicted heating demand of the two different heating systems (i.e. electric heaters in House N2 and underfloor heating in House O5), as shown in Fig.4.

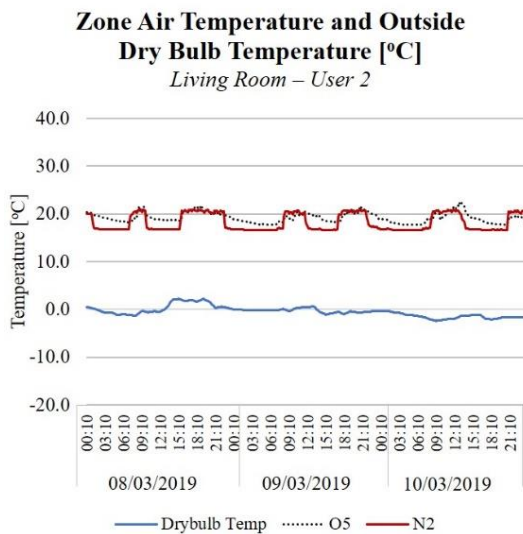


Figure 3: Zone air temperature in the ground floor living room. Simulation results for both Twin Houses (with O5 using underfloor heating and N2 using electric heaters), plotted against outdoor dry bulb temperature for three consecutive days in User2.

Zone Total Heating Rate [W]
Living Room - User 2

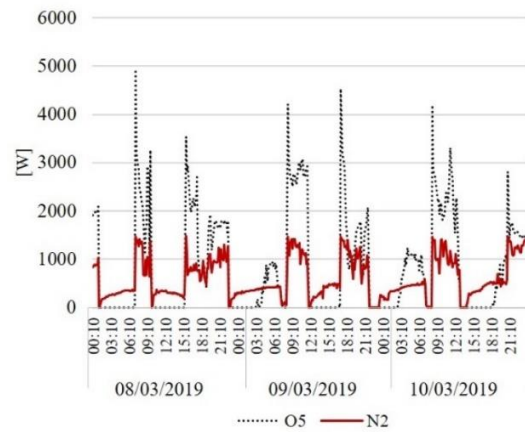


Figure 4: Zone heating demand in the ground floor living. Simulation results for both Twin Houses (see Fig 3), for three consecutive days in User 2.

During the preparation of the experimental specification, BPS highlighted two important issues with regards to internal air temperatures. The first was that the initial high magnitude of internal heat gains in the kitchen (representing cooking processes), resulted in severe overheating, with temperature reaching above 50°C. This was partly due to the lack of mechanical ventilation in the room and partly to the fact that in a real cooking process a large proportion of the energy used does not go to the room air but to the food. There were two options to resolve this issue; either to introduce a supply air point in the kitchen, or to decrease the amount of internal heat gains in the zone. The second option was decided as more suitable for the purposes of the experiment because the natural air exchange through the operable kitchen door was one of the validation goals.

Another finding of the deterministic simulation was with regards to the PRBS experimental phase. The initial 2000W PRBS signal proved to be excessive for the rooms, resulting in very high indoor air temperatures. To solve this problem, the PRBS signal was reduced to 700W to avoid exceeding 35°C but to maintain a signal high enough to significantly excite the building's thermal mass to allow for parameters identification for low order models (investigated in other subtasks of Annex 71).

Finally, a third finding of the deterministic simulations was that the initial experimental specification, which had the heating setpoint directly connected to the occupancy during User 2, produced very fragmented heating inputs. Although this might have served well the purposes of the validation exercise, it is not a realistic scenario. The room-wise setpoints were adjusted to continue heating during periods of absence shorter than 2 hours.

Many other assumptions in the experimental design were confirmed by the simulation results, giving assurance to the involved teams regarding the usability of the data. As an example the heating capacities, ventilation flow rates

and the duration of the individual experimental phases can be mentioned.

Sensitivity Analysis

The results of the SA are shown in Fig.5 and Fig.6 for House O5 (underfloor heating). Similar findings are derived for the other house (N2 – electrical heaters). For reasons of brevity, only the results of experimental phase User 2 for House O5 heating demand are included in this paper.

The absolute value of μ^* shows the ranking order of the input parameters, in other words the overall influence of each input factor on the simulation output (Fig.5). The higher the μ^* , the more influential the parameter (Campolongo et al., 2007; McLeod et al., 2013). A graphical representation of σ vs μ^* (given in the scatter graphs of Fig.6) is given to evaluate the monotonicity of the input parameters. If the input factors are positioned below the $\sigma/\mu^* = 0.1$ line then their behaviour is considered linear. If the input factors are positioned between the lines $\sigma/\mu^* = 0.1$ and $\sigma/\mu^* = 0.5$ then they are monotonic. If the input factors are between the lines $\sigma/\mu^* = 0.5$ and $\sigma/\mu^* = 1$ they are almost-monotonic. Finally, if they are above the $\sigma/\mu^* = 1$ line they are considered highly non-linear and non-monotonic (McLeod et al., 2013).

The SA indicated that the most significant input factor was found to be the effect of thermal bridges, having a linear effect on the heating demand. Moreover, the mechanical ventilation supply flow rate of the living room and the attic space were found among the most influential input parameters, also showing a linear effect on the heating demand. Finally, the temperature of the cellar boundary condition and the hot water flow rate of the underfloor heating system were found to be two more influential parameters.

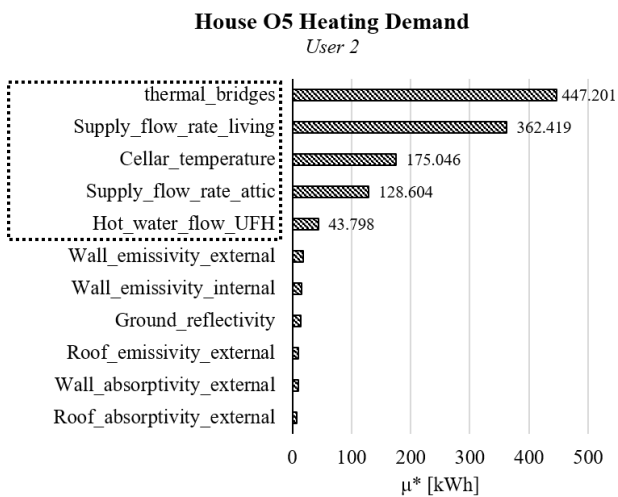


Figure 5: Morris analysis on zone heating demand of House O5 (underfloor heating): sensitivity ranking. The black, dotted rectangle encloses the parameters with a significance factor > 5%.

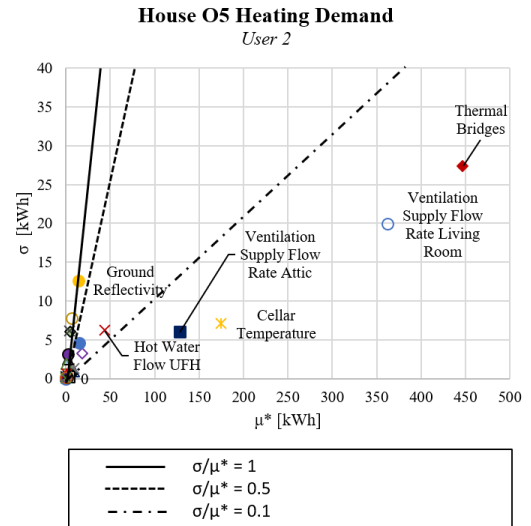


Figure 6: Morris plot of absolute mean (μ^*) and standard deviation (σ) on zone heating demand of House O5 (underfloor heating), showing the 6 out of 27 most sensitive parameters.

To distinguish between significant and insignificant input parameters, a significance factor of 5% on the actual heating demand was considered as the threshold. All factors that had an influence on the output above 5% were considered significant, all other factors were taken as non-influential.

Looking at the results of the SA at the zone level², the airtightness of the exterior and interior (operable) door, along with the opening area of the operable window were found to have a relatively noticeable impact on the simulation predictions of the room-wise heating demand. Moreover, the heating demand in several zones was also sensitive to the specification of the exterior wall and roof surface emissivity and absorptivity.

As a result of these findings parts of the instrumentation, (e.g. the underfloor heating flowmeter), have undergone a second calibration process. The entire measurement chain from the sensor to the data acquisition system was calibrated. Regarding the thermal bridges a higher number of junctions were analysed in detail than was intended before. Regarding the operable door and window the design of the tracer gas measurement was changed in a way to provide better data on the air flows through both openings.

Discussion

BPS is often associated with the concept of a *virtual laboratory* used to conduct *virtual experiments* to assess the performance of hypothetical, alternative design and operation scenarios and to find quantifiable answers to “what-if” design questions (Loonen et al., 2014; Clarke & Hensen, 2015). This paper has presented a methodology describing how BPS was used to assist the experimental design of a full-scale empirical validation project. The different phases of the experiment were

² The graphs were not included in this paper due to the limited length of the manuscript.

replicated using EnergyPlus. Based on these virtual experiments a synthetic dataset was created to investigate whether the anticipated outcome of the monitoring project would be adequate to meet the objectives of the empirical validation study and to identify possible faults in the experimental specification, that might pose a risk to the project and affect the usability of the results. Moreover, these tests were also used to determine the requisite degree of scaling up of internal heat gains and/or the prolongation of different experimental phases, before the empirical experimental design is finalised.

During this process a number of experimental errors were identified and corrected, primarily concerning the magnitude of heat inputs and potential risks posed by the resultant internal temperatures. The reliability of the experimental design relies on the awareness that a model is never an actual representation of reality. Hence the findings of the *virtual* experiment were used to assist decision-making for the actual experiment, knowing that there may be a discrepancy between the simulation results and the actual measurements. For example, the flexibility offered within the BPS environment sometimes contradicts what is feasible from an experimental point of view. Providing 10 kW internal gains for cooking into the kitchen is easy in BPS. In a real experiment on the other hand that requires a very substantial electrical installation. And even if the required electrical power can be provided to the room, a fire hazard would be introduced in the experiment, as a result of high air and surface temperatures.

In the deterministic simulation results potential experimental errors were not taken into account (except the intended “errors” for the FDD-phase), neither was the inevitable uncertainty in the values of a number of input parameters. To overcome this issue, a SA involving probabilistic simulation was employed to rank the impact of all uncertain input factors on the sensitivity of the output. This process revealed those factors that needed to be measured more carefully during the actual monitoring project. As an initial estimate a standardised assumption of a $\pm 20\%$ uniform distribution was used to explore the input ranges of the uncertain parameters (with respect to their normative or assumed base value). The rationale for this standardised approach, in the absence of more certain information, was to avoid introducing bias to the SA results, due to assigning variable ranges of uncertainty. However, it should be acknowledged that representing the actual range of uncertainty more realistically could be preferable to also account for the likely measurement uncertainty of different input factors. Doing so might reveal further influential parameters, that in a fixed uncertainty range would be disregarded as unimportant. For example, the supply flow rate of the ventilation system could be measured with an accuracy far better than $\pm 20\%$. However, the opening factor of an operable window might be associated with a higher degree of uncertainty in a house occupied with real users.

In the present SA all thermal bridges were treated globally, by assuming a simultaneous increase in all linear thermal bridges of the fabric. This decision was based on

the assumption that linear thermal transmittance (ψ -value) calculations would be performed for all the junctions between the different construction elements anyway. However, analysing the impact of each linear thermal bridge separately, would help prioritise which junctions need to be calculated more accurately and which could be ignored by relying on benchmark ψ -values.

Conclusion

This paper has presented a novel methodological procedure for incorporating *virtual BPS experiments* as a diagnostic precursor in support of a full-scale empirical validation exercise. The validation experiment is currently being conducted as part of the IEA Annex 71 project: “Building energy performance assessment based on in situ measurements”. The new methodology for the experimental design was introduced, by sequentially describing the different steps followed in the process and explaining how BPS assisted in creating and correcting the initial experimental schedule. The methodology comprised a combination of both deterministic and probabilistic simulations, each serving a different purpose.

Using deterministic simulation, the actual experiment was replicated in EnergyPlus in order to create a synthetic dataset. This process ensured that the expected outcome of the final experiment would be suitable to meet the objectives of the project, before the actual experiment took place. Using probabilistic simulation, the various uncertain parameters were ranked according to their impact on the output variable (zonal heating demand). As a result, the most influential factors, which need to be measured precisely during the actual experiment, were identified. The reliability, reproducibility and limitations of this methodology were discussed, along with suggestions for how it may be adapted to accommodate more complex situations, such as where parameter uncertainty can be better estimated or is non-uniform. This methodology demonstrates a universally applicable method for including uncertainty and sensitivity analysis into BPS experiments which could be used as an adjunct to the empirical validation process in a multitude of building performance research contexts. In addition, the techniques used here provide the basis for further virtual BPS experiments which could have wider applications in architectural and building services design development.

Acknowledgement

The work described in the paper was conducted as part of the International Energy Agency Energy in Buildings and Communities (IEA EBC) Annex 71 project. The English authors gratefully acknowledge support from Loughborough University and the University of Strathclyde. The German author would like to acknowledge support from the German Federal Ministry Economic Affairs and Energy for funding a part of the work described in this publication under the reference number 03ET1509A.

References

- ANSI/ASHRAE. 2014. Standard 140-2014, Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs. Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.
- Berkeley, P., Haves, P. & Kolderup, E., 2014. Impact of Modeler Decisions on Simulation Results. In *2014 ASHRAE/IBPSA-USA Building Simulation Conference Atlanta, GA September 10-12, 2014, Conference Proceedings*.
- Campolongo, F., Cariboni, J. and Saltelli, A. 2007. An effective screening design for sensitivity analysis of large models, *Environmental Modelling and Software*, 22(10), pp. 1509–1518. doi: 10.1016/j.envsoft.2006.10.004.
- Cattarin, G., Pagliano, L., Causone, F. and Kindinis, A. 2018. Empirical and comparative validation of an original model to simulate the thermal behaviour of outdoor test cells, *Energy & Buildings*. 158, pp. 1711–1723. doi: 10.1016/j.enbuild.2017.11.058.
- Clarke, J. A. and Hensen, J. L. M. 2015. Integrated building performance simulation: Progress, prospects and requirements, *Building & Environment*, 91, pp. 294–306. doi: 10.1016/j.buildenv.2015.04.002.
- Coakley, D., Raftery, P. and Keane, M. 2014. A review of methods to match building energy simulation models to measured data, *Renewable and Sustainable Energy Reviews*. 37, pp. 123–141. doi: 10.1016/j.rser.2014.05.007.
- EnergyPlus (n. d.) EnergyPlus. Available at: <https://energyplus.net/> (Accessed: 28 May 2018).
- Flett, G. and Kelly, N. 2017. A disaggregated, probabilistic, high resolution method for assessment of domestic occupancy and electrical demand, *Energy & Buildings*. 140, pp. 171–187. doi: 10.1016/j.enbuild.2017.01.069.
- Giglioli N., Saltelli A. 2008. *Simlab 2.2. Reference Manual*. Ispra, Italy: Institute for Systems Informatics and Safety (Joint Research Centre, European Commission).
- Hopfe, C. and Hensen, J. L. M. 2011. Uncertainty analysis in building performance simulation for design support, *Energy & Buildings*. 43(10), pp. 2798–2805. doi: 10.1016/j.enbuild.2011.06.034.
- IBP Fraunhofer. 2012. Fraunhofer Institute for Building Physics IBP. Available at: <https://www.ibp.fraunhofer.de/en.html> (Accessed: 15 July 2018).
- IEA EBC Projects (n.d.) Completed Projects, Available at: <http://www.iea-ebc.org/projects/completed> (Accessed: 15/04/2019).
- IEA Annex 58, 2015. Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements, Available at: <http://www.ecbcs.org/annexes/annex58.htm>, 2011-15 (Accessed: 15/07/2018).
- IEA Annex 71 (n.d.) Building Energy Performance Assessment Based on In-situ Measurements. Available at: <http://www.iea-ebc.org/projects/project?AnnexID=71> (Accessed: 25 June 2018).
- JEPlus (n. d.) JEPlus User's Manual, Version 1.4. Available at: http://www.jeplus.org/wiki/doku.php?id=docs:manual_1_4 (Accessed: 12 March 2018).
- Judkoff, R. and Neymark, J. 1995. *International Energy Agency Building Energy Simulation Test (BESTEST) and Diagnostic Method*. Colorado: National Renewable Energy Laboratory (NREL).
- Kalyanova, O., Heiselberg, P., Felsmann, C., Poirazis, H., Strachan, P. and Wijsman, A. 2009. An Empirical Validation of Building Performance Software for Modelling of Double Skin Facades. Proceedings of the 11th International Building Performance Simulation Association Conference, Glasgow, July 27-30.
- Lomas, K.J., Eppel, H., Martin, C.J. & Bloomfield, D.P. 1997. Empirical validation of building energy simulation programs. *Energy & Buildings*, 26(3), pp.253–275.
- Loonen, R. C. G. M. et al. 2014. Simulation-based support for product development of innovative building envelope components, *Automation in Construction*, 45, pp. 86–95. doi: 10.1016/j.autcon.2014.05.008.
- Loutzenhiser, P. G., Manz, H., Felsmann, C., Strachan, P. and Maxwell, G. M. 2007. An Empirical Validation of Modeling Solar Gain Through a Glazing Unit with External and Internal Shading Screens. *Applied Thermal Engineering* 27 (2-3), pp.528-538.
- Mantese, E., Hopfe, C., Cook, M., Glass, J. & Strachan, P. 2018. The modelling gap: Quantifying the discrepancy in the representation of thermal mass in building simulation, *Building & Environment*, (131), pp. 74–98. doi: 10.1016/j.buildenv.2017.12.017.
- McLeod, R. S., Hopfe, C. J. and Kwan, A. 2013. An investigation into future performance and overheating risks in Passivhaus dwellings, *Building & Environment*. 70, pp. 189–209. doi: 10.1016/j.buildenv.2013.08.024.
- Morris, M.D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2), pp.161–174.
- Raslan, R. & Davies, M., 2010. Results variability in accredited building energy performance compliance demonstration software in the UK: an inter-model comparative study. *Journal of Building Performance Simulation*, 3(1), pp.63–85.
- Ryan, E.M. & Sanquist, T.F., 2012. Validation of building energy modeling tools under idealized and realistic conditions. *Energy & Buildings*, 47, pp.375–382. doi.org/10.1016/j.enbuild.2011.12.020.
- SALib (n. d.) SALib - Sensitivity Analysis Library in Python. Available at: <https://salib.readthedocs.io/en/latest/> (Accessed: 19 January 2019).
- Strachan, P. et al. 2015. Whole model empirical validation on a full-scale building, *Journal of Building Performance Simulation*, 1493(January 2016), pp. 1–20. doi: 10.1080/19401493.2015.1064480.

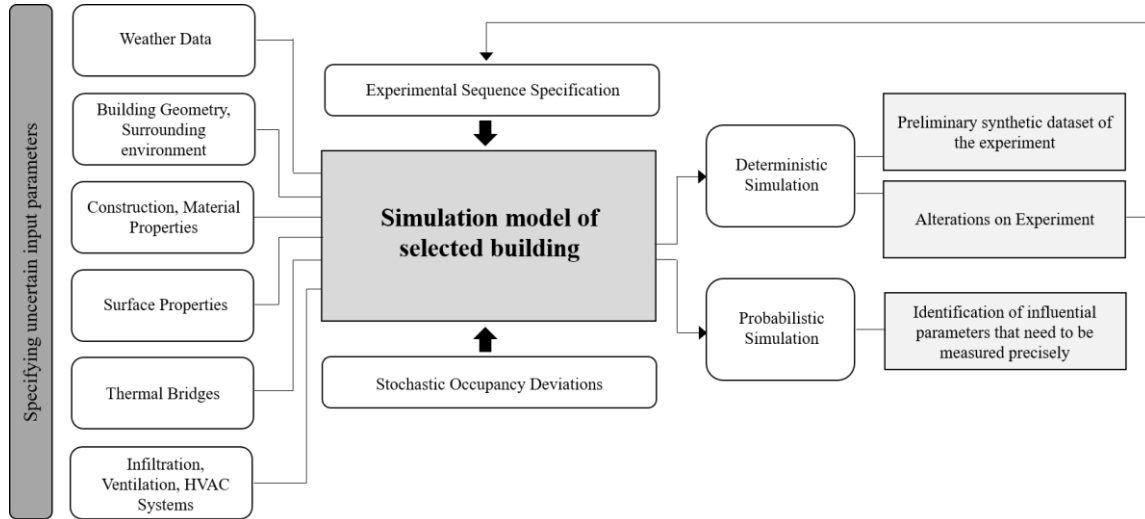


Figure 2: Graphical representation of the experimental design procedure conducted as part of the Annex 71 full-scale empirical validation project.

Table 1: Parameter uncertainties investigated in the SA: base, minimum and maximum values.

Category	Parameter	Base value	Min value	Max value	Units
Global parameters	Ground reflectivity	0.23	0.184	0.276	Dimensionless
	Snow modifier	4	3.75	4.75	Dimensionless
Surface properties	Cellar temperature	18	14.4	21.6	°C
	Roof emissivity (external)	0.9	0.72	0.99	Dimensionless
	Roof emissivity (internal)	0.9	0.72	0.99	Dimensionless
	Roof absorptivity (external)	0.63	0.504	0.756	Dimensionless
	Roof absorptivity (internal)	0.25	0.2	0.3	Dimensionless
	Wall emissivity (external)	0.9	0.72	0.99	Dimensionless
	Wall emissivity (internal)	0.9	0.72	0.99	Dimensionless
	Wall absorptivity (external)	0.23	0.184	0.276	Dimensionless
	Wall absorptivity (internal)	0.17	0.136	0.204	Dimensionless
Shading material properties	Blind conductivity	0.023	0.0184	0.0276	W/m·K
	Blind reflectivity	0.68	0.544	0.816	Dimensionless
	Blind emissivity	0.9	0.72	0.99	Dimensionless
	Blind to glass distance	0.06	0.048	0.072	m
Airflow parameters	Air mass flow coefficient for window	0.00001	0.000008	0.000012	kg/s·m
	Air mass flow exponent for window	0.7	0.56	0.84	Dimensionless
	Air mass flow coefficient for interior doors	0.02	0.016	0.024	kg/s·m
	Air mass flow exponent for interior doors	0.7	0.56	0.84	Dimensionless
	Air mass flow coefficient for exterior door	0.0002	0.00016	0.00024	kg/s·m
	Window opening area (in attic)	0.1	0.08	0.12	Dimensionless
Envelope	Thermal bridges	0	0	20	%
HVAC	Supply flow rate in living-room	100	80	120	m ³ /h
	Supply flow rate in attic	35	28	42	m ³ /h
	Electric heaters capacity	2000	1600	2400	W
	Electric heaters radiant fraction	0.3	0.24	0.36	Dimensionless
	Hot water flow in UFH	q ₀ *	0.8*q ₀	1.2*q ₀	m ³ /s

* q₀ is the hot water flow rate in each zone