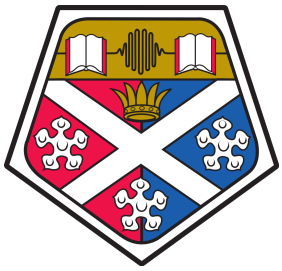
Sensitivity Analysis and Bayesian Calibration of Building Energy Models



Energy System Research Unit

University of Strathclyde, Glasgow, United Kingdom

Thesis submitted for the degree of Doctor of Philosopy

Filippo Monari
filippo.monari@strath.ac.uk

29th February 2016

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:
Date:

Acknowledgements

I thank the BRE Trust and the EPSRC Training Grant for providing me with the economical resources necessary to undertake this research.

I am mostly grateful to all the members of the Energy System Research Unit (ESRU) of the University of Strathclyde for their support and valuable advices. In particular I want to thank my supervisor Dr Paul Strachan, who guided me during the path of my PhD. He always gave me the trust and freedom necessary to pursue my own ideas and research directions.

During my time spent as PhD student at the University of Strathclyde, I met many friends who made my staying in Glasgow sunnier than the actual meteorological conditions. I thank all other ESRU PhD students, in particular John, Eric, Gavin, Graeme, Carmina, Katie, Tom and Valentina.

I will always hold dear my friendships with Stefania, Aldo, Licia, Sandra and Alex. My best wishes and good luck to Isabelle, and I hope her name will be Clementine. I thank my team mate, Luca, for playing as better as he can, although he has two left feet. I thank Roberta for sharing with me every once in a while a glass of good Scotch, during my first period in Glasgow. Hopefully we will do it again.

I am especially grateful to my colleagues architects, Alessandra, Anna, Clara, Maddalena, with whom I shared the office in the last two years of my PhD, and Giulia. Honestly, I have not always understood them, but the trips, the lunches, the dinners, and the laughs together must mean something. Although the friendships with them greatly decreased my productivity, the time spent together will always be among my dearest memories. I am not absolutely kidding.

Finally, I dedicate this thesis to my family. To my mum, Tiziana, and to my brother, Alessandro, who always stayed close to me and let me free to take my own decisions, also when it was not easy to do so. Especially to my father, Luciano, who though me the very science of life, that is to live with honesty, irony and goodness. In the end it is just a journey. To them goes all of my love.

Dedication

allora, come si sta al mondo a capire il giusto?
eh eh si stava bene, ma adesso...tu come stai?
si stava meglio quando si stava peggio?
eh sì, e ai tempi antici eran tutti felichi.
mai abbandonare la strada vecchia per una nuova.
non ti preoccupare la strada è sempre quella...
cosa hai fatto da quando sono andato via?
sono andato in Scozia a fare un PhD.
cos'è un PhD?
un dottorato di ricerca.
Dio bono, te sempre a studiare, ma com'è che hai una testa così! E Alessandro e la Mamma?
tu non capisci, sei indietro come i maroni dei cani. Ale e la Mamma stanno bene. Oh! Alessandro ha aperto una palestra. È in gamba.
se se, devi fare sempre diverso. Bravo Alessandro! Dai un bacio alla Mamma.
oh dove vai! Aspetta bene un attimo. Ho scrtto sta roba qui. Te la volevo fare leggere, ma tu l'inglese lo sai poco eh?
fammi un po' vedere. Eh povero me i miei sette figli e mia moglie incinta! Se è lunga! Mo che, mo che, mo che, tieni pure.
eh lo sapevo io. Non ti preoccupare poi te la spiego. Più che altro speriamo che vada bene, altrimenti...
vedrai che andrà bene. Anche perchè è l'unica che hai scritto no?
secondo te?!
va bhè, io vado. Ti voglio bene.
anch'io ti volglio bene Capo. Pero' potevi stare un po' di più...
ma cosa vuoi mai, noi in quelle faccende lì contiamo come il due di coppe quando briscola è bastoni.
allora ci vediamo più vecchi?
sicuramente.

Abstract

The current sate of the art of Building Energy Simulation (BES) lacks of a rigorous framework for the analysis, calibration and diagnosis of BES models. This research takes this deficiency as an opportunity for proposing a strongly mathematically based methodology serving such purposes, providing: a better consideration of the modelling uncertainties, means to reduce BES model complexity without oversimplification, and methods to test and select different modelling hypotheses depending on field observations. Global Sensitivity Analysis (GSA), Gaussian Process Regression (GPR) in a quasi-Bayesian set up and Markov Chain Monte Carlo (MCMC) methods are the foundations upon which the proposed framework is built. It couples deterministic BES models and stochastic black box models, thus having the physical and probabilistic representation of real phenomena complementing each other. It comprises four phases: *Uncertainty Analysis*, *Sensitivity Analysis*, *Calibration*, *Model Selection*.

The framework was tested on a series of trials having increasing difficulty. Relatively simple preliminary experiments were used to develop the methodology and investigate strengths and weaknesses. They showed its capabilities in treating measurement uncertainties and model deficiencies, but also that these aspects influence the estimation of model parameters. More detailed experiments were used to fully test the efficacy of the method in analysing complex BES models. Novel techniques, based on Bootstrap and Smoothing with Roughness Penalty, for the determination of the uncertainties of multidimensional model inputs, were introduced. The framework was proven effective in adequately simplifying BES models, in precisely identifying parameters, causes of discrepancies and improvements, and in providing clear information about which model was the most suitable in describing the observed processes.

This research delivers a powerful tool for the analysis, diagnosis and calibration of BES models, which substantially improves the current practice and that can be already applied to solve many practical problems, such as the investigation of energy conservation measures, model predictive control and fault detection.

Contents

List of Figures

List of Tables

Chapter 1

F1 on dirt roads

Detailed BES models have reached high level of detail in representing accurately the main phenomena determining the thermal and energetic performance of buildings. However this achievement in terms of fidelity in representing reality entails a high level of complexity. In particular, building energy simulation models have complicated structures with many in-built sub-models that attempt to represent the physical reality, and large numbers of input parameters. In order to build them accurately, a considerable amount of information and data of high quality is required, characterising the present and future situation (up to a reasonable time) of the building being modelled, which often is not available to the modeller. All the knowledge used to create a certain model is subject to a certain degree of uncertainty. Even more, despite the great detail of BES models, their representation of the phenomena behind the thermal and energetic behaviour of buildings is not exact; therefore every model will have deficiencies and inadequacies in its structure. This contributes to further increase uncertainties that, because of the intricate network of interactions linking the various model inputs, unpredictably propagates through the computer code, resulting in uncertainty in the model outputs which, when compared with actual measurements, show often significant levels of discrepancy. These reasons led to questions about the actual reliability of BES model predictions. The building simulation community is very aware of these problems, highlighted by several studies over the last few years, concluding that there is a gap between model predictions and actual building performances which has to be bridged, and a lot of effort is being put in researching methods to diagnose BES models and reconcile model outputs with field observations.

Just to give an idea of the level of uncertainty involved, assume that it is actually possible to build a BES model which has no inadequacies in it, and having the right inputs it returns the exact result. In the most common case, the

information to build such a model will be gathered through drawings, material properties and plant specifications. Further investigation might be done after the construction phase with measurements, surveys, and audit, in order to better specify HVAC and occupancy schedules, and to check data from the design stage. Blower door tests could be carried out to measure infiltration. It might be also possible to measure weather data from a nearby meteorological station, to have a better understanding of the micro-climate around the building. However, the achieved picture of the building will be inevitably related to particular operational conditions, and it will become increasingly unrepresentative as its life progresses. Something not observable during the construction phase might have partially invalid data based on design stage information, the properties of the building will change during time, due to its dependency on external conditions or the wear of its components, and the weather forecasts and occupancy schedules will become increasingly less accurate as the prediction horizon recedes. Heat transfer mechanisms due to convection, infiltration and ventilation are difficult to observe and the model parameters characterising them often can not be accurately defined. For example, according to literature ([4], [119], [36], [12] [83] and [16]), simple errors in the consideration of occupational patterns and building operation by the users, as well as wrong estimation of building services features, loads from appliances and their usage hours, may lead to optimistic predictions and to substantially underestimate the actual energy consumption. In a few cases, the actual figures were up to three times higher than the predicted ones. Even more, there are uncertainties involved in the monitoring, due to observation errors or stochastic variability of the observed processes. Technical limits and difficulties may produce measurement errors which are not negligible, and the measurement activity, may become not possible due to economic constraints. It is then clear that it is impossible to recreate an exact representation of a real process through a model, because of inaccuracies involved in the act of observation, or errors that inevitably will affect the modelling.

The scenario just depicted becomes even more complicated if the several approaches of modelling the same phenomena are considered. Building energy simulation tools are over-engineered programs offering several ways, involving different degrees of detail and physical accuracy, of modelling the different processes contributing to the determination of the thermal and energetic performance of buildings [25]. For example airflow can be represented through a schedule of constant flow rates, with an airflow network, or by coupling computational fluid dynamics with the thermal model.

The different uncertainties involved, model inadequacies and the variety of structures that a BES model can have, leads to *over-parametrisation* and to model *equifinality*. The former phenomenon is mainly due to interactions and correlation between model inputs, for different combinations of which the same model can return practically equal outputs. The latter is the capability of different models, based on different assumptions and combinations of model inputs, to provide similar predictions. Thus, especially when similarity with observed performances has to be met, it is often not easy to determine the most suitable model and the configuration of its inputs. Analogous problems are highlighted in [9] and [8] for hydrological models. Experience has demonstrated on several occasions that a more detailed model does not always yield better results. More detail often involves more uncertainties or even the impossibility of adequately defining the additional parameters and sub-models. For example in airflow network models, most parameters, like pressure coefficients, can be only reasonably guessed.

The picture of BES models emerging from this initial discussion is that of tools which are extremely capable but also extremely difficult to manage. Indeed they could be compared to F1 cars running on dirt roads. They are very powerful, very precise machines, which run on roads that do not allow them to express their complete potential, and which make driving them very difficult. In particular it is not clear what should be the aim of a modeller approaching a modelling problem. Should modellers try to represent real phenomena at the best of their physical understanding, or rather allow for approximations in order to achieve quickly a reasonable evaluation of the building energy performance? The former point of view appears more suitable in research, and the latter in practice; nonetheless the debate is open. On one hand, since it is known, to a certain level, how the physical processes being modelled work, the primary objective of the modeller should be to represent as faithfully as possible such processes rather improve their understanding through adequate research. On the other hand, in the light of the premises stated above, it is recognised that often the uncertainties involved in the modelling do not allow the adequate specification of all the requirements needed to represent in detail what is being observed. Therefore, it is believed that there should be a relation of direct proportionality between the amount and quality of information available and degree of detail of the BES model.

A recent contribution [25] sets out the state of the art of Building Energy Simulation, its objectives and future developments. The purpose given to BES models in this study is shared in this thesis. BES models are not oracles, giving precise predictions and answers, but they are support tools in understanding

complex systems. They provide the means to represent the thermal and energetic behaviour of buildings and to test hypothesis and assumptions. This function of BES models becomes particularly important and relevant when simulation is used in analysing existing buildings which, through adequate monitoring, can provide the data and information to disprove or support the principles and concepts underpinning a particular model. However, the capabilities of BES models in describing real world phenomena have to be enhanced in order to effectively serve this important purpose. It is necessary to augment the deterministic approach adopted by building energy models with stochastic methods, by considering modelling uncertainties, and the stochastic character of the phenomena that they aim to replicate.

The increasing monitoring activity due to the widespread deployment of smart meters, as well as the improvement of data handling systems, is making available a significant amount of data which can be used to better understand the thermal and energetic behaviour of buildings and improve modelling techniques. Recently, sensitivity analysis and calibration have been increasingly employed in characterising BES models against metered data, in analysing uncertainties, and in reconciling their predictions with field observations. In particular a great effort has been put in developing and defining sensitivity analysis and calibration procedures effective in treating building energy models. The research described in this thesis fits within this context, and aims to develop a comprehensive probabilistic approach to the analysis of detailed building energy models against real measurements. Global Sensitivity Analysis and Gaussian Process Regression in a quasi-Bayesian set up are the main tools on which the research in built. The proposed framework, involves four parts: *Uncertainty Analysis*, *Sensitivity Analysis*, *Calibration* and *Model Selection*.

The next section defines the objectives and gives an overview of the methodology and high level descriptions of its phases. A summary of the undertaken research path follows. Finally the structure of this manuscript is outlined.

1.1 Objective and Overview

The main objective of this research is to develop tools and propose a framework for analysing, diagnosing, calibrating and comparing detailed building energy models with the help of information coming from real measured data. The proposed framework aims to solve modelling problems involving the evaluation of BES model performances against measurements, which can be divided in two typologies: validation problems and calibration problems. The former are characterised by significant amount of provided information, allowing the almost complete definition the detailed building energy model, which should then be a quite accurate virtual replica of the observed process. The purpose is to demonstrate the capability of a simulation program and of the built model in reproducing the monitored data. Calibration, involves less prior information about the possible model able to characterise the observed phenomena, and the objective is to identify a suitable modelling structure and model parameters according to a given dataset, in order to built a predictor of the actual building performance. In practice, calibration can be used to aid validation, due to the uncertainties involved and the impossibility of providing adequate specifications for every modelling aspect, requiring then the choice of assumptions by the analyst. In particular, using calibration methods to reduce modelling uncertainties and verifying modelling assumptions is often helpful in demonstrating the validity of a simulation program or model.

When facing these problems, several issues need to be tackled:

- On the basis of the available information the modeller has to decide the initial assumptions and structure of the model.
- The several uncertainties involved need to be adequately considered.
- The capability of the model in representing the observations has to be assessed.
- Eventual causes of discrepancies have to be identified and provided for.
- Often several models need to be compared with each other in order to decide which is the most suitable.

The deterministic approach adopted by building energy models does not allow an exhaustive answer to many of these questions, especially those concerning uncertainties consideration. Through normally employed sensitivity analysis techniques it is possible to investigate the propagation of the model parameter uncertainties in the model and to assess the consequent uncertainties in the model

outputs. However most of the methods employed nowadays in analysing BES models neglect model deficiencies in representing the observed phenomena and observation errors. The full consideration of the modelling uncertainties would be of great benefit also in comparing and ranking BES models.

Therefore it is proposed to complement detailed building energy models with stochastic methods and probabilistic black-box models in order to provide for this deficiency. In particular, the methodology that will be explained is able to effectively consider uncertainties, to gain information from the measured data by identifying unknown model inputs, and to estimate parameter and prediction uncertainties conditional on the measurements. Most of all, it gives to the modellers the possibility to test their hypotheses and beliefs, consisting of different model structures, against the evidence coming from field observations and to choose one model among the many others. The steps of the methodology are:

- *Uncertainty Analysis*: the initial (prior) uncertainties related to model inputs and measurements are characterised. All the variables involved are treated as stochastic, by assigning to them prior probability density distributions describing possible errors in their definition. Such probability density distributions are determined according to the information from literature review, given specifications, and experience.

- *Sensitivity Analysis*: most of the time it is not feasible to consider all model inputs; moreover many of them are not clearly identifiable because of their weak effects on the model outputs. To overcome this problem a three step sensitivity analysis is performed with the objectives of screening model inputs and identifying the most important factors, measuring the adequacy of the screening, reducing prior uncertainties and obtaining indications about a suitable calibration set up.

- *Calibration*: this allows the identification of unknown inputs, the estimation of model parameter uncertainties conditional on the field observations, and the assessment of the robustness of the predictions. In this stage the uncertainty from measurement activity and model inadequacy are considered, and causes of discrepancy are investigated.

- *Model Selection*: according to likelihood based ranking criteria the different built models are compared in order to select the most suitable for representing the given dataset.

General descriptions of these four steps, with particular focus on the reasons that led to the adoption of the methods employed, is given in the following section.

1.1.1 Uncertainty Analysis

Ideally, especially in calibration and validation exercises, modelling uncertainties and their propagation through the model would be routinely considered in order to aid the comparison between model outputs and measurements, and to allow the robustness and reliability of the predictions to be assessed. However in practice, this is seldom the case because of the lack of support tools allowing users to take into account uncertainties in predictions. The only attempt known to the author to embed uncertainty analysis in building energy simulation programs is [70] based on the dynamic simulation tool ESP-r [23].

The possible sources of uncertainties in computer models include ([65]):

- *Parameter Uncertainty*: it is the uncertainty in model inputs which might be unknown, which vary from context to context or which are dependent on other variables.
- *Model Inadequacy*: the model will always contain some degree of error in representing real processes, thus even if parameter uncertainty is negligible, the observed process can exhibit variability which the computer model is not able to explain.
- *Residual Variability*: it is the property of a process to assume different values for different observations in the same conditions. It may be due to actual unpredictability of the phenomenon or to some unobserved conditions acting on the process.
- *Parametric Variability*: it is the uncertainty involved in using the model to make predictions in unknown, uncontrolled or unspecified conditions. In BES models this kind of uncertainty is related mainly to the variability of the weather factors determining the boundary conditions.
- *Observation Errors*: in calibration and validation studies measurement errors add further degrees of uncertainties that should be taken into account.
- *Representation Uncertainty*: complicated computer models, like BES models, involve complicated relations among the different part of the calculation. Thus even though the computer code, acting as a deterministic function, is actually known, its complexity as well as possible errors make its output, with respect to a particular input configuration, unknown until the simulation is actually run. Running the code for each possible input configuration is not feasible; thus there is additional uncertainty to consider.

In [65] this was referred to as *Code Uncertainty*. This term was deemed to be ambiguous in this context since it could have been interpreted as uncertainties in the coding activity itself. In order to avoid misunderstanding it was changed as here indicated.

All these types of uncertainty will find natural places in the the proposed framework. This phase mainly focuses on the characterisation of *Parameter Uncertainty*.

It is possible to distinguish two kind of model parameters in building energy models: vectorial (or multidimensional) inputs, and scalar (or unidimensional) inputs. The former consists of variables described by time series and therefore their values change during the experiment and simulations. To adequately characterize the uncertainties for such factors it is necessary to define multidimensional probability density distributions depicting time varying marginal probability density distributions and correlation patterns between observations at different time steps. Indeed, especially for weather factors such as wind speed, time varying monitoring conditions, as well as unobserved phenomena influencing the measurements, may cause time varying magnitudes of the measurement errors. The latter can be assumed constant during the simulation and the experiment. To suitably represent their uncertainties it is sufficient to define unidimensional probability density distributions.

The main objective of the *Uncertainty Analysis* phase is to define the prior probability density distributions describing the variabilities of the various factors entering in the calculations (prior probability density distributions). A reasonable assessment of model parameter uncertainties is a decisive step in order to perform a sensible analysis. Through the definition of prior probability density distributions, the space wherein solutions will be searched is determined. An underestimation of model parameter uncertainties, results in the analysis exploring a narrow solution space from the beginning thus producing misleading results. On the contrary an overestimation may lead to a difficult identification of reasonable solutions, especially if the model is *over-parametrised*. Furthermore, model sensitivity can be dependent on the defined variation ranges as well.

Uncertainty analysis is not an easy task, and should be based on reliable information, or on inference performed through adequate statistical tools. In particular, while Uncertainty Analysis for scalar inputs is routinely performed effectively, by considering information from literature, trial simulations, experience and good sense, the same is not true for vectorial parameters, like weather factors, which are at least as important. Vectorial inputs are rarely taken into account

and, usually, their uncertainties are simply represented with constant offsets from measured mean vectors. Such an approach neglects the time-varying conditions affecting the measurements of these entities, which may have different levels of uncertainty during the monitoring period. In this work, a procedure based upon Bootstrap and Smoothing with Roughness Penalty is used to infer, from the data, plausible multi-dimensional distributions representing the uncertainties related to the measurements of multidimensional model inputs. Smoothing with Roughness Penalty has been deemed to be a particular interesting option, due to its accurate control on the power of the smoothing and low computational load. In this way it is possible to investigate, rigorously, how the random variability of the observed processes changes over time. It will be shown, through a case study, that this indeed leads to different results and more sensible considerations.

1.1.2 Sensitivity Analysis

Sensitivity analysis techniques are methods able to characterise, qualitatively or quantitatively, the influences of model inputs on model outputs. In the study of detailed building energy models, sensitivity analysis has been increasingly employed in a stand-alone fashion or as a step in a more articulated procedure, mostly in order to investigate uncertainty propagation through the computer code and identify the most influential inputs. Especially the latter task is often undertaken according to qualitative screening, through the widely applied Morris Method, and no actual quantification of parameter effects is performed. Qualitative screening is often chosen since it is computationally cheap to perform; nonetheless it can lead to the selection of too few parameters, thereby excessively reducing the degrees of freedom of the model. The main consequence is to work with over-simplified models which poorly represent the original model.

This research proposes an approach to sensitivity analysis of BES models employing qualitative and quantitative methods whose aim is to aid calibration and validation studies. The main objectives are to measure the extent of the simplifications induced by considering only qualitative sensitivity results, and to gain information in order to reduce prior parameter uncertainties, thus facilitating parameter identification at a later stage. In particular, quantitative methods cannot be applied directly considering every model input on its own due to the prohibitive computational load. Thus, qualitative screening is performed to reduce model dimensionality and group parameters. Its efficacy is then assessed by quantifying to what extent the model resulting from considering only the retained factors is representative of the original BES model. The additional number

of simulations required is used to reduce prior input uncertainties and to acquire knowledge about which inputs are more powerful in driving the calibration of the model.

The Morris Method and the Sobol Method have been employed because they are model independent GSA techniques ([108]). It is believed that these two properties are indispensable for a sensitivity analysis methodology in order to be applied to BES models. Differently from local sensitivity analysis methods, which change one factor at a time, and thus are unable to capture higher order interactions, GSA measures the sensitivity of a model by varying all its inputs at the same time, hence it is able to give an adequate picture of first as well as higher order effects. Model independence underlines the capability of a sensitivity technique to perform well regardless of the mathematical structure of the model, so as to effectively treat linearity as well as non-linearity. Even more, GSA can have a probabilistic character by attributing to each model factor a probability distribution regulating its variations, and thus it perfectly integrates with the proposed treatment of uncertainties.

Commonly, the calculation of sensitivity indexes is done on scalar model responses. However BES models produce vectorial outputs, and in most approaches the computation of sensitivity measures has been empirically adjusted by carrying out results for each time step, or by considering integrals of output variables and distances from reference values. In the former case, while it is possible to see how the model sensitivity changes during time, a large and redundant amount of information is produced, which is hard to analyse and summarise. The latter approach produces more concise results, but at the same time does not consider the dynamics of the model outputs in an adequate way. To achieve concise information about model sensitivity, while considering output dynamics, new rigorous approaches based on Principal Component Analysis (PCA) are employed to summarise vectorial model responses. In particular, an expansion of the Morris Method is proposed.

1.1.3 Calibration

Calibration of BES models is often defined as the process of using an existing building simulation program and tuning its various inputs so that the observed energy use matches closely with that predicted by the simulation program ([96]). It has the potential to be applied in several context including, for example, the creation of base-line models, the evaluation of energy conservation measures, the development of model predictive controls, the breakdown of the energy consump-

tions, and the improvement of building simulation programs.

In the past it has been performed according to ad-hoc manual iterative procedures and mathematical analytical methods involving deterministic optimisation. The former approach has been the most used, and, because of its empirical character, its heterogeneity, its dependence on analyst skill and expertise, and its lack of reproducibility and documentation, doubts and scepticism about its rigorous foundation and correctness have been raised. Thus, while the advantages of calibration are well understood, it has not been largely adopted, especially in practice, and the development of a robust and strongly mathematically based framework, which can aid BES model calibration, is still an open issue. The topic is attracting increasing research efforts and it is perceived as an indispensable step to improve Building Energy Simulation practice.

An effective calibration framework for building energy models, should be able to treat detailed models, to consider prior information, to process high resolution datasets by extracting as much information as possible from them while ignoring noise and measurements errors, and to return probabilistic solutions averaging the gained information and the initial uncertainties. Recent researches have outlined calibration methodologies for complex computer models based on Bayesian techniques, which seem more apt than the previous approaches to answer the requirements just listed (*Bayesian Calibration*).

Bayesian Calibration is becoming an object of increasing interest from the building simulation community, for its natural capability to effectively treat modelling and measurement uncertainties, and detailed computer models, like BES models, through a rigorous approach based on Bayes' Probability Theory. Recent studies ([66] and [125]) proved its capability in calibrating BES modes, and demonstrated that its performance is generally better than those of commonly employed deterministic and stochastic optimisation routines. Also appealing are its capability to include prior modeller knowledge in the analysis, by setting up suitable prior probability density distributions for the calibration parameters; the possibility, granted by the Bayesian Paradigm, to update the calibration results as new data and information become available; and the opportunity to adopt likelihood based criteria to rank and select different models.

In contrast with the common perception, it is believed that modeller dependency is not necessarily a drawback. Calibration of detailed BES models is a complex problem whose solution cannot be blindly entrusted to an algorithm. It would be much better to provide the analysts with the tools to test their beliefs against field observations within a well defined mathematical framework, so that

information from the modeller and from the data could combine to find the solution to the calibration problem. The Bayesian paradigm has exactly this idea at its base, and it is intrinsically present in many manual iterative procedures, although with the lack of a rigorous mathematical framework. Indeed, as will be discussed during the Literature Review, in many manual iterative calibration studies it is possible to glimpse the Bayesian concept of upgrading models according to the knowledge acquired through the various iterations of the procedure. Therefore, *Bayesian Calibration* is considered the perfect bridge between a blind search for an optimum and the seeking of a solution based exclusively on personal expertise and judgement.

For these reasons, *Bayesian Calibration* constitutes the basis upon which the proposed calibration framework for complex BES models was built. This calibration method has been applied to detailed computer models directly or by employing supportive probabilistic black-box probabilistic emulators. Although the former approach is attractive, since it avoids additional uncertainties due to the creation of the probabilistic black-box emulator, the latter seems actually more suitable for the purpose of calibrating BES models, especially because of the consequently significant reduction in computational time. In particular, Bayesian inference methods usually require a large number of model runs and, as shown by their applications on models of complexity similar to BES models ([33] and [68]), when simulation time is not negligible, the time needed to carry out the needed calculations can become prohibitive. GPR ([93]) has been chosen as the framework for building the probabilistic emulator for its flexibility and generality. In particular, GPR models are non-linear non-parametric data driven models allowing the representation of different phenomena with the same modelling structure.

The calibration framework that will be described in this work aims to improve the current state of the art by building on and blending together the works in [51], [13] and [5]. The main novelties are the capability of the method to effectively consider highly dimensional datasets and model outputs, and to take into account multiple target variables and multiple datasets during the calibration process.

1.1.4 Model Selection

Models are convenient mathematical representations of observed real world phenomena. Because of the impossibility to completely understand all the causes and conditions generating real processes, observation errors and modelling uncertainties, each model will be subject to a certain degree of error and inadequacy,

resulting in uncertainties in its predictions. These aspects, especially in complicated *over-parametrised* models like BES models, cause *equifinality*, meaning that different models with different input combinations can provide similarly good fit of the considered target data, thus, making it very difficult to find a clear solution to the calibration problem. In particular, in order to avoid the introduction of unnecessary uncertainties, it is important to clearly identify the simplest model giving an adequate representation of the observed processes and additional details should be included only if leading to significant improvements.

In Statistics this concept is routinely adopted in performing model selection. Statistical models are easily upgraded and modified, and several criteria are available for their comparison and ranking according to the goodness of the provided predictions and their level of detail. For example Likelihood Ratios, Akaike Information Criterion, Bayes Information Criterion and Hannan-Quinn Criterion are commonly used for this purpose ([118]). Although BES models are significantly more complex and their upgrade and assessment are not so easy as for statistical models, the idea to use similar means for model comparison and selection is interesting. In order to achieve this goal, this research proposes to use Bayes Factors, in addition to the commonly employed Normalised Mean Bias Error (NMBE) and Coefficient of Variation Root Mean Squared Errors (CVRMSE) in order to undertake model selection.

NMBE and CVRMSE have been used in a number of studies to assess the capabilities of BES models to fit measured data and they can be considered a sort of standard. According to [1] they are defined as:

$$NMBE = \frac{\sum_{i=1}^{N}(y_i^* - \hat{y}_i)}{N\bar{y}^*} \tag{1.1}$$

$$CVRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{N}(y_i^* - \hat{y}_i)^2}{N}}}{\bar{y}^*} \tag{1.2}$$

where y_i^* are the observations, $\hat{y}_i$ are the predicted values, $\bar{y}^*$ is the empirical observation mean and N is the number of observations. The former represents the mean bias between model predictions and measurements as a percentage of the observation mean, and indicates a general overestimation or underestimation of the model output with respect to the observed values. The latter represents the variability in the residuals as a percentage of the observation mean and indicates how well the model captures the dynamic trends of the data. These two indexes can be coupled together as indicated in [22] in order to have an overall Goodness

Of Fit (GOF) criterion:

$$GOF = \frac{\sqrt{2}}{2}\sqrt{NMBE^2 + CVRMSE^2} \tag{1.3}$$

It is important to keep in mind, when interpreting these measures of fit, that they are expressed as a percentage of the mean of the observed values. Therefore NMBE, CVRMSE and GOF can reach values relatively high, when the average of the measurements is close to zero. In order to help the reader to correctly understand these GOF criteria, the empirical averages of the measured data are always indicated in the table listing their calculated values.

These GOF criteria can be easily calculated, and are apt for assessing models when it is required only to meet a certain similarity with the measured data. Nonetheless they can be easily subject to model *equifinality*, and it is not clear how several target variables should be considered. Thus, when the objective is to identify the best model besides achieving a good match with the measurements, likelihood based criteria like Bayes Factors are more suitable since they are less subject to model *equifinality* and able to easily consider multiple target measurements and datasets.

Bayes Factors are a Bayesian approach to rank two competing scientific hypotheses. In this case the competing hypotheses are two models, which have to be compared, and the one with highest probability to generate the observed data has to be selected. The calculation of Bayes Factors involves the estimation of the marginal likelihoods of the two competing models. Marginal likelihood estimation can be performed through the built probabilistic meta-model. This is a further strength for *Bayesian Calibration* which allows the adoption of such measures of model goodness with respect to other calibration approaches.

1.2 Research Methodology

The PhD research involved a review of the state of the art of sensitivity analysis and calibration of building energy models, the development of methods improving the current practice, and their testing against synthetic and real experiments. The research activity was not a linear path consisting in the orderly sequence of the above mentioned phases. Rather, they often overlapped and alternated depending on the emergence of new problems to solve and ideas to test. A brief summary which tries to describe as clearly as possible this research path follows.

The research started, as usual, by reviewing the state of the art about sensi-

tivity analysis and calibration of BES models. In this initial stage it was possible to understand the weaknesses and the strengths of the sensitivity analysis and calibration approaches commonly adopted in analysing BES models, and identify the necessary improvement to develop better methods. The trail opened by Heo and colleagues in a series of studies ([46], [49], [48], and [47]) was deemed to be the right research path to undertake. These works described the first applications of *Bayesian Calibration* to the solution of problems related to building energy modelling, and led to deeper investigation of the subject by reviewing contributions from other fields, mainly Machine Learning, Econometrics, Computer Science and Statistics. The previous works that had the most influence on the current research were [5], [6], [7], [13], [50], [52] and [65]. [65] is the first paper establishing a rigorous Bayesian approach to the calibration of computer models. It defines the mathematical framework and constitutes the basis upon which the works by Heo and colleagues were developed. Such mathematical framework was, then, modified in [50], [52], [7] in order to effectively consider vectorial model responses. Finally [13] provided a means to consider multiple calibration targets. The calibration framework that will be presented drew great inspiration from these studies, and tries to contribute to the literature by blending them together, so as to propose a general calibration framework for detailed computer models, particularly focused on BES models.

As the understanding of the problems that were necessary to solve in order to calibrate models grew, it became clear that it was necessary to deal effectively with the large degree of *over-parametrisation*, by reducing model dimensionality. On several occasions, sensitivity analysis had already been applied for this purpose. In particular, in many studies the Morris Method was used in performing a qualitative screening of the model parameters. However, as the complexity of the model analysed increased, such approach to dimensionality reduction was deemed to be insufficient, and particularly subject to over-simplification. The work in [94] was found particularly interesting in complementing the normal practice of screening the model inputs on the basis of qualitative information. The result was a three step sensitivity analysis procedure that allowed rigorous preparation for the calibration phase.

The last part of the literature review involved the investigation of criteria used to compare and rank models. The widely applied criteria based on the sum of squared errors between model predictions and measurements were found to give not decisive evidence supporting the adoption of one model with respect to another, and the focus was moved to likelihood based criteria that are rou-

tinely used for model selection in Statistics. Bayes Factors naturally integrated in the *Bayesian Calibration* framework, and due to their model independence and robustness were adopted as the model ranking criterion of reference.

Two main tasks were identified, in order to carry out the analysis:

- Running BES models according to a certain input design in order to perform Sensitivity Analyses and Monte Carlo Simulations.
- Perform numerical optimisation and integration of joint posterior probability density distribution functions.

and suitable computer code was written to undertake them.

ESP-r [23], was chosen as modelling environment of reference. A Python library which is able to automatically generate ESP-r scripts containing all the instructions to run models for a given set of input vectors and collect the results was developed. This library can consider variations in material properties, airflow network parameters, measurements, infiltration rates and thermal bridges. Hence it represents a general tool which can be used to run batch simulations in ESP-r according to a certain design of the considered variables.

Algorithms performing numerical optimization and integration were implemented in R and Fortran 95 programming languages. At first prototyping was undertaken in R, which, by being an interpreted language for statistical computation, provided most of the needed functions and allowed for easy and fast debugging of the code. Once enough confidence had been achieved, the most computational expensive routines were ported to Fortran 95, which offers significantly more computational power. The Open MP library was used in order to take advantage of modern multi-core processors, and R interface functions were linked to the resulting Fortran 95 routines, thus creating a flexible computational environment. Finally for the most complex, and therefore computationally demanding, case studies, the ARCHIE-West High Performance Computing facility was employed. In order to fully benefit from its parallel architecture, the Fortran 95 code was improved by using the Open MPI libraries. The correctness of the code resulting from the several iteration of the programming activity was ensured, through comparisons of the results achieved by performing calculations with different versions of the developed software.

In parallel with the software development activity, experiments involving the analysis of data synthetically generated and coming from the monitoring of real buildings were undertaken in order to tests the capability of the evolving analysis framework. Preliminary investigations were performed on experiments involving

a multilayer wall and a test-box. The latter was particularly significant in understanding the capability of the calibration method proposed, and constituted its first real test. Both these two exercises were at first undertaken considering synthetically generated observations. In such a way the true solution was known and it was possible to ensure that the right answer was returned. Subsequently, the data from the real experiments were used as well, in demonstrating the calibration procedure and testing its limits.

This was followed by the analysis of more complicated experiments involving small domestic buildings. These studies involved the analysis of very detailed BES models, and the presented procedure was fully applied in order to spot model improvements, and aid model validation. Especially, the large number of variables that it was necessary to consider represented a significant challenge. Nonetheless the methods performed well, and were able to give useful insight into which actions to take in order to improve the capability of the model in representing the observed data.

All case studies used to prove the developed methods were experiments performed in the context of the International Energy Agency (IEA) Energy in Buildings and Communities (EBC) Annex 58. The objective of this international project was to develop the necessary knowledge to achieve reliable on site dynamic testing and data analysis methods that could be used to characterise the actual energy performance of building components and whole buildings. This project called several teams of experts to work on issues like development of quality procedures for full scale testing, development of quality procedures for dynamic data analysis, guidelines for building performance characterisation and predictions, gathering well documented high quality dynamic data for validation purposes, and applications of dynamic whole building test data. The research in this thesis greatly benefited from the environment established and found an ideal context in which to develop. It was possible to make comparisons with other techniques, exchange knowledge with experts in Statistics and Data Analysis which had been very useful in the development of the method, and establish links and connections with colleagues from different institutions which continue.

1.3 Thesis structure

This work will unfold in the following chapters:

- Literature Review.
- Uncertainty and Sensitivity Analysis.

- Calibration of BES Models.
- Initial Experiments.
- Detailed Experiments.
- Conclusions.
- Appendix.

The Literature Review chapter discusses recent research contributions involving Sensitivity Analysis and Calibration of BES models. Particular attention is given to studies where Sensitivity Analysis has been used as a preparatory step to model calibration and to works applying *Bayesian Calibration*.

The Uncertainty and Sensitivity Analysis chapter opens with a discussion about the uncertainties involved in computer experiments that is necessary to consider in calibration and validation studies. Then some novel approaches for the consideration of the uncertainties relative to vectorial model inputs, are explained. The chapter closes with the description of the three step Sensitivity Analysis procedure suggested for performing dimensionality reduction of BES models.

The Calibration of BES Models chapter describes the probabilistic and mathematical framework adopted to perform *Bayesian Calibration*.

The Initial Experiments chapter describes the preliminary case studies used to develop the methodology. In particular virtual experiments and real experiments are discussed.

The Detailed Experiments chapter involves the explanation of more complex real experiments wherein the capability of the method was applied to assist the validation of detailed BES models.

In the Conclusions chapter the findings from the PhD research are discussed and particular emphasis is given to the discussion of the strengths and limitations of the current research, and to the identification of possible future development applications and expansion.

In the Appendix the reader can find additional material which can help in a better understanding of the methods employed.

The adopted notation and terminology are explained in glossaries and lists of symbols attached at the end of this thesis.

Chapter 2

Literature Review

In order to keep this chapter to a reasonable length, it was deemed appropriate to focus the literature review on sensitivity analysis and on calibration of BES models, which are the two central parts of this research. Uncertainty analysis for unidimensional model parameters is a well established practice, itself based on literature review. Therefore, references will be given from time to time during the discussion of the performed experiments, in order to adequately justify the adopted prior uncertainties. However, significant contributions are in [34] and [70]. Regarding multidimensional model inputs, the treatment of their uncertainties is one of the novelties of this research, and detailed discussions of the adopted techniques can be found in [35], [91] and [90]. Model selection is a wide subject, with different proposed criteria depending on the field of knowledge. The objective of this work is to extend the common practice of Building Energy Simulation, by supporting the usage of routinely employed measures with criteria coming from Statistics, especially Bayes Factors. Therefore, the interested reader is referred to [1] for a discussion of the usually adopted goodness of fit criteria for BES models, and to [118] for an overview on criteria coming from Statistics. A detailed discussion of Bayes Factors and the corresponding calculation methods is contained in [64] and [121].

This chapter will outline recent contributions which had a significant impact on this research and which were deemed relevant in depicting the current state of the art of sensitivity analysis and calibration of building energy models. As the work of literature review proceeded, it became increasingly evident that sensitivity analysis and calibration procedures were often coupled together, in different ways, in order to meet the objectives of a particular study. Thus it is not easy, and in most cases incorrect, to classify a contribution as a sensitivity study or a calibration study. However, in order to give a clear structure to this chap-

ter, it was necessary to attempt to undertake such classification. The reviewed studies were divided in two categories, according to the personal judgement of the author. Works involving only sensitivity analysis, or in which calibration is performed in order to infer uncertainties conditional on measurements, afterwards used as sensitivity analysis inputs, were classified as sensitivity studies. Researches adopting sensitivity analysis as a means of reducing model dimensionality and modelling uncertainties, or to improve model parameter identifiability in a preparatory phase to model calibration were labelled as calibration studies.

2.1 Sensitivity Analysis

Sensitivity analysis has been increasingly applied in a stand-alone fashion or as a step in more structured procedures, in order to investigate how model parameter uncertainties influence model behaviour and to address the following issues:

- identifying the most influential variables,
- quantifying output uncertainty,
- understanding the relations between inputs, and inputs and outputs,
- supporting decision making,
- aiding model calibration and validation.

An extensive review of sensitivity analysis techniques applied to BES models can be found in [117] while a more general treatment focusing on the different problems that different sensitivity analysis methods are suitable for solving is in [108].

In [2] a validation method for building energy simulation codes is proposed, which aims to improve the common practice of comparison between simulation outputs and experimental results by also taking into account the uncertainties in the former. In particular, uncertainties in the numerical results are determined through a sensitivity analysis carried out by an adjoint-code method.

In [29] the authors apply the Morris Method [76] and Monte Carlo simulation to identify the most influential parameters and the overall output uncertainty for a monthly quasi-steady simplified regulatory model describing a residential building in Turin, Italy. A similar approach is adopted in [109] to investigate a complex dynamic ESP-r model representing an apartment building in Spain. In this study the Morris Method in its extended version [18] is used to evaluate the first and second order effects of the several model parameters. The authors also outline a framework to classify effect typologies.

In [114] a three step sensitivity analysis is performed on a detailed EnergyPlus model of the INCAS experimental platform of the French National Institute of Solar Energy in Le-Bourget-du-Lac, France. The objectives are to determine influential parameters, to identify the influence of parameter uncertainty on the building performance and to quantify the model output uncertainty. The first two steps of the procedure consist of applying local sensitivity and correlation analysis in order to single out the most important factors and then to group model

inputs. Finally global sensitivity analysis is undertaken to quantify model output uncertainty and apportion it among the selected most influential parameters.

In [37], the authors analyse a complex building model having a large number of model parameters (1000) through variance, L^1 norm and L^2 norm based sensitivity measures. Also a method to break down the sensitivity of the model according to its many part and sub-models is explained. In order to speed up the calculations a meta-model based on Support Vector Regression is used to approximate the detailed BES model.

[89] investigates the effect of uncertainties from primary and secondary sources of pressure coefficients, on summer energy consumptions of a night ventilated office building, for several European climates. These are assessed for an EnergyPlus model provided with an airflow network. In particular a simple sensitivity analysis is carried out on an EnergyPlus model provided with an airflow network, by assessing the effects of variations in pressure coefficient values for different model design and weather conditions, showing that these parameters are one of the main factors in determining air flow rates.

[14], [53], [15], [105] and [11] are examples wherein sensitivity analysis is used to support decision making at different levels. In the first four of these studies, different sensitivity and uncertainty analysis approaches are applied to drive the building design and retrofit according to different objectives. Particularly interesting are [14], [15], [105]. In [14] an approach is described for assessing the effect of parameter uncertainties on the BES model output based on sensitivity analysis. The objective of the study is to probabilistically evaluate the possibility that a certain design fails in meeting prescribed energy consumption requirements, in a similar way as done for structural performances. At first, the Morris Method is used to screen the several model inputs; secondly the Fourier Amplitude Sensitivity Test is used to rank the retained factors and quantify their influence on on the model responses. Finally, the model outcome uncertainties are investigated by Monte Carlo simulation. The probability of design failure is calculated from the empirical cumulative distribution of the model output, which was derived from the last step. The proposed methodology is applied to a typical residential Danish building, modelled with the simulation program Be06, developed by the Danish Building Research Institute.

In [15] a methodology for investigating the effect of model parameter variations in Near Zero Energy (NZE) building energy performances is developed. The procedure unfolds in three main steps. At first a detailed BES model is optimized several times through an evolutionary algorithm to find combination

of parameters resulting in a NZE building and populate empirical probability density distributions. Such distributions are then used as inputs to a Monte Carlo simulation in order to investigate the total model output variance. Finally a back-tracking search was used to identify the parameters most important in determining the building energy performances. The object of the study was a NZE house in Quebec, Canada.

[105] propose a non probabilistic treatment of the uncertainties aimed at providing a simple procedure for robust decision making. The described method seems particularly suitable in selecting one of different possible retrofit scenarios. In particular, in the paper sequential models are used to evaluate the implementation of different energy conservation measures; the most convenient solution is then selected according to the Wald, Hurwicz, and Savage decision rules ([75]). The decision making framework is demonstrated on a mid-sized office building in Cambridge, United Kingdom.

[11] present a methodology to assess the most convenient energy conservation measures to implement on building community scale. The proposed decision making framework is based on a multi-criteria decision making method and multi-attribute utility theory, and involves parameter screening with the Morris Method, calibration and probabilistic sensitivity analysis. The proposed method is demonstrated through a model of a housing stock of approximately 15000 properties, located in Salford, UK.

An interesting approach to the assessment of uncertainties and sensitivity of computer models, which is particularly apt in preparation for a subsequent calibration or validation study is the coupling of GSA with the Generalised Likelihood Uncertainty Estimation (GLUE). GLUE is a flexible and simplified Bayesian approach to uncertainty assessment of computer models which uses non-rigorous likelihood measures. Its relation with a more rigorous Bayesian approach is shown in [81] wherein the link between GLUE and Approximate Bayesian Computation is demonstrated.

GLUE was primarily developed as a calibration procedure for hydrological and environmental models ([8], [9]) and more recently has been coupled with GSA in order to assess and identify the factors more influential in calibrating a certain model ([102] and [94]). It is based upon Monte Carlo simulation and on the mapping between input and output according to a generalised likelihood function, which measures how well the model output generated from a certain combination of inputs matches the observed data. In particular the generalised likelihood function attributes to each combination of inputs a weight, approximating its

probability to produce the target data. Such weights can be used in sampling procedures to make inferences about the model parameters conditional on the observations.

As already mentioned, GLUE was at first applied in the calibration of hydrological and environmental models. In [8] a hydrological model of the Institute of Hydrology Gwy catchment in Mid-Wales is analysed. GLUE is used to evaluate the posterior uncertainties for model inputs and outputs as time evolves. The authors highlight that often model output distributions are different from normal. In this case the usually employed calculations of the uncertainties, based on the Normality assumption, are not applicable and it is necessary to use simulation methods, like GLUE. The paper also outlines how GLUE can account for new incoming data, thus updating estimates and reducing the relative uncertainties, and how it can be used to calculate model output confidence bands for prediction in new conditions, thus facilitating the comparison with measurements not used in previous analyses.

In [9] three case studies are discussed. The subject of the first example involves the modelling of the rainfall run-off of the Maimai catchment in New Zealand. The aim is to investigate the sensitivity of GLUE to the choice of the likelihood measure. The results suggested that the choice of a particular function influence the analysis outcomes, but it is not critical in order to obtain good performances from the method, and that it is possible to adopt ad-hoc approximated likelihood measures according to the different aspects to be assessed. In the second investigation, GLUE is applied to models representing latent and sensible heat fluxes from the land surface to the atmosphere mainly due to evaporation. The aim in this case is to investigate model performance depending on different levels of detail. The analysis concluded that, simple models are often able to return outcomes in terms of predictions, prediction uncertainties, and parameter identification similar to the more complicated models. The last example had similar objectives and outcomes as the second. In this case, predictions from dispersion models, based on random particle tracking methodology, were compared with detailed measurements of velocity distributions in an over-bank flow in the Flood Channel Facility at HR Walligford, in order to assess their goodness. Different model structures were considered, showing that simpler models were able to give a more adequate representation of the observed process. Nonetheless all the models adequately matching the measurements had substantially different structures and assumptions, thus highlighting *equifinality*.

In [102] and [94], GLUE was coupled with GSA, in order to gain additional information from the large amount of model runs that such kind of sensitivity analysis technique requires.

In [102], a model describing the short and medium range dispersion of radionuclides was analised. Firstly a Monte Carlo Simulation was carried out, in order to derive mean and confidence bounds for model predictions, and investigate model sensitivity through correlation analysis and scatter-plots between model inputs and model outputs. Secondly the outputs from this operation were analysed with the GLUE in order to identify the model parameters most effective in improving the similarity between model output and target measurements.

In [94] the Sobol Method and the GLUE approach are combined to analyse a model representing a simple chemical system. At first, model sensitivity is characterised without considering the observed data, by using the Sobol Method to apportion the model output variance among the model parameters, and identify the most influential ones. Afterwards GLUE was used to map model parameters according to the measured data. Particularly interesting is the analysis performed on the weights returned by the generalised likelihood measure adopted. A second sensitivity analysis was carried out on these variables, by means of scatter plots, correlation analysis, PCA and Sobol Method. In particular the calculated sensitivity indexes from the Sobol Method gave clear indications in ranking calibration parameters according to their capability in improving the match between model predictions and target observations.

Despite the fact that GLUE can be used for model calibration, there are concerns about its application to building energy simulation models, related to its slow convergence and the empirical character of the adopted measure of goodness of the fit. GLUE can show a slow convergence rate in estimating posterior statistics if the zones of high probability of the model parameter joint posterior distribution are distant from the zones of high probability of the model parameters joint prior distribution used as input of the initial Monte Carlo Simulation. This is often the case rather than an exception, because of deficiencies of BES models in depicting the observed processes. Even more, if not critical for the good performance of the method, the choice of a particular generalized likelihood function has influence, especially on the calculated posterior uncertainties. Therefore it is deemed more suitable for preliminary investigations, preceding more rigorous calibration procedures. In particular, the Sensitivity Analysis framework outlined in [102] and [94] is judge to be particularly suitable in analysing BES models, and in gaining information aiding their validation and calibration. However, the di-

rect application of this approach on BES models is not feasible due to the large number of inputs and the consequent unmanageable number of model simulations required. Thus, a previous dimensionality reduction of the computer model is required, by identifying most important model parameters and by grouping the various factors. Many of the reviewed contributions proved that the Morris Method is a suitable tool to perform this preliminary task.

The *Sensitivity Analysis* procedure presented in this research involves three steps, and it combines the Morris Method with GSA and GLUE. The main aim of this procedure is to adequately reduce model dimensionality and prior parameter uncertainties, in order to aid subsequent calibration or validation studies. In particular, to simplify a model only according to results from qualitative screening is insufficient and potentially it might lead to over-simplification. It should, instead, become good practice to quantify the amount of variance attributable to the retained inputs in order to adequately justify model simplifications. This is achieved by quantifying the efficacy of the qualitative screening through the Sobol Method. The additional number of simulations required is then processed with GLUE, which complements the previous result by identifying additional parameters useful in improving similarity between model predictions and measured data, and reducing the prior model parameter uncertainties. This procedure is explained and demonstrated in a case study in the following chapters.

2.2 Calibration

Comprehensive reviews of calibration techniques can be found in [96] and [28]. The former is part of the ASHRAE project RP-1051 which aimed to gather and investigate the best tools, techniques, approaches and procedures from the existing body of research and develop a coherent and systematic calibration methodology that included both parameter estimation and determination of the uncertainty in the calibrated model. [28] integrates and extend [96] with more recent contributions.

Both [96] and [28] use the classification given in [24]:

- manual iterative, based on pragmatic intervention
- manual iterative, based on a suite of comparative graphical displays
- manual iterative, based on the use of special tests and analytical procedures
- mathematical analytical for automatically adjusting the input parameters to reduce the discrepancy between measured and predicted data

For the purpose of this study it is deemed more suitable to modify it as follows:

- manual iterative
- mathematical analytical for automatically adjusting the input parameters
- mathematical analytical based on probabilistic/stochastic methods

The suggested classification groups together the three manual iterative classes and an additional class has been added for methods specifically based on probabilistic and stochastic techniques. These approaches are becoming the object of increasing interest for their natural capability to effectively treat uncertainties and deserve to be discussed separately.

2.2.1 Manual Iterative Calibration

Calibration of BES models according to manual iterative procedures has been the most applied approach in the past. It has been applied in several contexts, ranging from practice to research and it consists of an ad hoc iterative series of steps wherein the analyst gathers information and data about the building object of the study, and according to its judgement, experience and intuition tries to identify a suitable model by tuning its inputs in order to achieve a satisfying match with the measured target data. It has been the most used approach in the past and it has been employed in practice as well as research contexts.

The study described in [24] is an example of *Manual Iterative Calibration* employed for research purposes, namely the improvement and the empirical validation of the dynamic simulation program ESP-r. It was carried out within the PASSYS European project whose aim was to investigate the effectiveness of passive solar architectures, in particular conservatories. Being research oriented, it presented some major differences with respect to most other calibration works of the time. The subject is a very particular building, that is a test-cell, and the quality and resolution of the used data were very high due to the elevated monitoring standards adopted during the experiments. The validation method adopted in the study involved:

- The creation of a blind simulation model built upon the initially available information.
- The assessment of the goodness of the fit with the measurements by means of sensitivity analysis and Monte Carlo simulation in order to estimate parameters and the confidence bounds for the predictions.

- The identification of model upgrades, through statistical analysis of the residuals, and correlation analysis between residuals and model parameters.

This procedure was iteratively repeated until a satisfying agreement was obtained between measurements and simulations, making sure that each model modification was carefully selected and justified from experimental evidence. The workflow consisted of calibrating firstly a model without the conservatory and successively with the component installed, so that it was possible to investigate eventual benefits from the adoption of the passive solar architectural element. Once enough confidence was acquired in the modelling of the building component applied to the test-cell, the model was extrapolated to studying a real scale building. Particularly interesting was to notice how the modelling team was able to find improvements to the model according to the knowledge gained from new observations.

Two significant examples of *Manual Iterative Calibration*, employed in investigating retrofit solutions for commercial buildings, are [82] and [126]. Both the studies depict procedures to guide the iterative calibration process. The approach presented in [82] was developed in applications over a period of 7 years on 15 office buildings in Brazil and consists of three main steps:

- Simulation from building design plans: it consists of the construction and evaluation of a model according to building design plans and documentation about plant systems, occupancy and equipments, without visiting the site. This stage can be further refined by analysing envelope parameters, through parametric simulation, and by adjusting schedules according to hourly energy demand.

- Audit and walk through: using knowledge from the previous step the analysts visit the building in order to collect additional information and measurement to further improve the model.

- End-use phase: the measurements of energy consumption are divided according to energy-end uses in lights, equipments and air conditioning, in order to fine tune the schedule and internal power density.

Hourly and monthly data were used to calibrate the several models. From step to step the match between simulation results and measurements improved systematically.

A more articulate methodology, built upon the concept of base load analysis of the monthly energy consumption, is proposed in [126] and involved:

- Base case modelling: the analyst builds a model by using a reliable simulation tool, according to the design specifications as well as the available information at the time of the analysis.

- Base load analysis: the main energy consumptions like electricity and gas are decomposed in base loads and exterior temperature dependent loads. By comparing the results with on site measurement it is possible to further break-down the energy consumptions and accordingly refine the model.

- Mid-season calibration: during this period, the energy consumptions are mostly independent of heating and cooling. This makes possible the fine tuning of factors regarding lighting, HVAC and equipment.

- Site interview and confirmation: additional site interviews are used to confirm the design information used in model construction.

- Heating/cooling season calibration: when the agreement for mid-season calibration is satisfying, the analyst can proceed with this step in order to calibrated the parameters influencing heating and cooling loads.

- Validation of the calibrated model: finally, the model can be validated according to the chosen goodness of fit criteria and eventual graphical comparison between measurements and predictions.

- Investigation of promising energy conservation measures: energy conservation measures are implemented in the calibrated model and the most effective ones identified.

The method was successfully applied on a 26 storey office building in Seoul. The calibration was carried out considering monthly data of electricity and gas consumption.

In [123] energy audit and sensitivity analyses are combined to aid model calibration. The procedure involves the following phases:

- calibration of power and schedules of constant loads, such as lights and plug loads.

- simulation of particular design days for thermal load analysis wherein a simple local sensitivity analysis is performed, for the parameters governing the main heat flow path, in order to understand how the model can be further improved.

- Assessment over the whole year.
- Final refinements.

The procedure was demonstrated on a public office building in Brazil which was modelled in EnergyPlus.

Many other studies defining a more or less complex series of steps, through which the analyst improves the initial model according to new information acquired from further investigation, can be found in literature. A significant body of work has been produced by Haberl and colleagues ([111], [20], [21], [56], [55], [112] and [113]). These studies developed *Manual Iterative Calibration* procedures based on the availability of accurate measurements regarding specific aspects of building energy performances and on a detailed breakdown of the building energy consumption. Often the models are run over specific periods in order to isolate causes of discrepancy. Graphical analysis tools, like signatures, are widely employed and indexes representative of the main energy loads are used for comparison with similar buildings. All these works employ models built with the dynamic simulation program DOE-2 and their purpose is the evaluation of energy conservation measures and retrofit designs. Particularly interesting is the concept of signature, firstly developed by [122] and then expanded by [111]. These graphical tools can be defined as the description of trends in variations of a certain target variable as functions of changes in the model inputs. In the former study they are employed to help HVAC simulation engineers in identifying the impact of different parameters on air handling units (AHUs) performance. In this case the signature of each parameter was defined as the rate of change in the AHU heating or cooling energy consumption as functions of variation in that parameter. [111] refines the concept of signature by defining two types: calibration signatures and characteristic signatures. Calibration signatures describe graphical deviation between measured energy consumptions and simulated energy consumption as a function of the average dry bulb temperature. Characteristic signatures indicate the sensitivity of a building energy model relative to each parameter. They provide a predictable shape according to changing an input parameter by a certain amount based on the calibration signatures.

A potential step forward for this kind of calibration is illustrated in [88] and [87]. The authors describe and apply a methodology which aims to be evidence based, meaning that changes to the input parameters should only be made according to available evidence under well defined priorities. In particular a hierarchy is established among the sources of information depending on their reliability and model changes should not be made if the evidence comes from a more reliable

source. The use of the most capable modelling tools and data having higher resolution than monthly is strongly recommended, in order to avoid simplifications and cancellation errors, and ultimately to have a better description of the dynamics of the observed phenomenon. New proposals are the employment of a version control software to keep track of and document the calibration process, improving its reproducibility, and to adopt zoning strategies more representative of the real building set-up. In many contemporary calibration studies, the building floor was divided in a core and four perimeter zones. The authors argued that it would be better to identify the various zones according to space function, position relative to the exterior, measured data available and systems used to condition the internal environment. This zoning strategy was named zone-typing. The steps constituting the iterative procedure explained in these papers are:

- Preparation: it involves the set up of the version control, the construction of the initial model, the hierarchical organization of the source of information, and the selection of the model validation criteria.
- Obtain readily accessible information: consisting of collecting all the available information about the building.
- Update model inputs: the model is updated with the information from the previous step. At each update a new model version should be created.
- Zone-typing: the geometry and zoning of the initial model are revised according to information gathered by visiting the building.
- Construction: the constructions used for building the model are verified.
- HVAC and plant: factors relative to system, plant and equipments are updated according to previously collected information and where possible through direct observation and measurements.
- Internal loads: internal loads are updated according to data from monthly bills and where possible from real measured data.
- Error check: the model is run and reviewed to assure its good quality.

The resulting model is then tested against the validation criteria chosen. If it does not meet the validation requirements, causes of discrepancy are investigated by graphical analysis of the residuals and by employing new information coming from further measurements and direct observations and the procedure is iterated. The method is demonstrated on a large office building in Ireland.

In all the studies above cited, it is possible to notice the attempt to define a clear stepped procedure aiming to establish a homogeneous approach to *Manual Iterative Calibration*. However, the proposed procedures are somewhat different and there is no particular evidence supporting the adoption of one of them rather than another. Thus, probably, *Manual Iterative Calibration* is destined to keep its character of ad-hoc procedure dependent on the skills, expertise and personal judgement of the modeller, which constitutes at the same time its weakness and its strength. Both [96] and [28] agree that the main obstacles impeding a widespread adoption of *Manual Iterative Calibration* are the diversity of the procedures adopted and the lack of systematic documentation, resulting in low reproducibility and difficult assessment of results reliability. In particular in [96] the author refers to *Manual Iterative Calibration* as follows: "historically, the calibration process has been an art form that inevitably relies on user knowledge, past experience, statistical expertise, engineering judgement, and an abundance of trial and error". Nonetheless it is not constrained in a strict framework and has great flexibility only limited by the capability of the analyst and of the modelling tool. Therefore, if applied in an organized way, it has often led to satisfactory results. The majority of the cited studies rely on the criteria indicated in [1] to assess the predictive performance of the models and to label them as calibrated. Therefore, the agreements with the measured data achieved in this work is around 5% and 15% for monthly analysis and around 10% and 30% when hourly data were used, in terms of NMBE and CVRMSE respectively. Although it was not possible to infer precise information about the average time required to perform model calibration, it is possible to envisage the need of several days since different procedures advised multiple surveys and audits. Major concerns to the application of this typology of calibration are related to the total impossibility to consider the uncertainties involved in the modelling.

2.2.2 Mathematical Analytical Calibration

Mathematical Analytical Calibration employs sensitivity analysis and optimization techniques to determine possible solutions to the calibration problem. It often uses statistical tools like Monte Carlo Simulation and statistical tests, but differs sensibly from a fully probabilistic/stochastic approach since calibration is treated as a standard optimization problem.

The first strongly mathematically based approach to calibration was presented by Reddy and colleagues in a series of papers related to the ASHRAE research project RP-1051 ([97], [98] and [116]). The methodology is composed of four

parts:

- The definition of a set of influential parameters and schedules depending on building type and heuristics along with their estimates and variation ranges.
- Coarse grid search through Monte Carlo Filtering. The heuristically retained parameters are classified as strong or weak, according to a convenient sensitivity analysis method, and narrower ranges for the strong ones are defined. In particular, statistical tests are used to compare empirical density distributions of parameter samples giving acceptable and unacceptable model realisations, depending on the values of NMBE and CVRMSE, in order to assess the statistical significance of each factor in improving the goodness of the fit with the target data.
- The refinement of the coarse grid search by performing a guided grid search and identification of a set of solutions.
- Estimation of model predictions and corresponding uncertainties relative to the identified set of solutions.

For the first time, it is recognised that is insufficient to give a single set of parameter values as the solution to the calibration problem, but it is necessary to provide a set of possible solutions, describing the relative uncertainties. In [116] an extension to the methodology depicted in [97, 98] is presented. The authors focused especially on the guided search step, and on the problem of parameter identifiability. Regarding the former aspect, the employment of an optimization routine adopting a suitable objective function was suggested. The latter issue is related to the fact that calibration is considered a highly under-determined problem wherein the information provided by the measured data is not sufficient to infer all the model inputs. In order to overcome this problem, the authors proposed the use of approximations of the partial derivatives and Hessian matrix of the adopted objective function, to identify model parameters to which model calibration is most sensitive and that are least correlated with other inputs. This methodology was tested on synthetic and real case studies, with better performance in the former cases.

In [22] a similar calibration method is described and applied to an unoccupied building in Lleida (Spain). The methodology can be summarised as follows:

- To gather building description and the data for calibration and validation periods.

- To construct the initial model.
- To define the best estimates for the calibration parameters and the probability distributions describing the relative uncertainties. In this way the search space is defined.
- To perform a Monte Carlo filtering according to the chosen GOF criteria and based on Latin Hyper Cube Sampling (LHS).
- To perform regional sensitivity analysis in order to identify weak and strong parameters. Here the Pearson Chi Squared Test is used to compare the prior and the posterior (after Monte Carlo Filtering) distributions. For strong parameters the two distributions should be sensibly different and hence the p-value of the test should be lower than a chosen threshold.
- To assess new estimates and narrower variation ranges for the strong variables.
- To iterate the process until no further strong parameters are identified.

As previously mentioned, a certain number of solution vectors is used to make predictions and define uncertainties. The procedure seems to depend on the threshold chosen for the p-value of the Pearson Chi Squared Test, on the size of the Latin Hyper Cube sample and on the criteria chosen for performing the Monte Carlo Filtering. In particular the latter should be chosen accordingly to the purpose of the analysis. For the p-value, the authors suggested the use of a value of 0.01, while the suggested size of the Latin Hyper Cube sample was 15-19 times the number of parameters.

Monte Carlo Simulation was used in a different framework in [61], [27] and [26]. The former describes a simplified approach for building energy model calibration based on the coupling of normative energy calculation methods and LHS in order to perform Monte Carlo Filtering. In particular, a quasi-steady state model is created according to regulations; suitable probability distributions for the model parameters are then chosen in order to define an initial input space. Such parameter space is then explored by LHS, and model runs are carried out accordingly. Then, only the input vectors giving NMBE, CVRMSE and GOF lower than certain thresholds are retained and used to infer suitable values for the calibration parameters. The method is used to create a baseline model to investigate the effectiveness of energy conservation measures. In [27] and [26] an analytical procedure combining the evidence-based procedure depicted in [88]

and Monte Carlo Simulation is presented and applied respectively. The steps of the procedure are:

- Data gathering/building audit.
- Evidence-based BES model development.
- Bounded grid search.
- Refined grid search.
- Uncertainty analysis.

The first two phases involve the creation of a BES model as depicted in [88]. In particular the analyst collects building data and information and uses them to build a detailed BES model according to a hierarchy of information sources. Version control software is employed to keep track of the modelling process. Afterwards, best-guess estimates and probability density functions are applied to the model inputs and Monte Carlo simulation is used to identify promising solution vectors yielding values of the chosen GOF criterion below a certain threshold. The procedure can be applied iteratively refining the search grid, through regional sensitivity analysis, until a satisfying match with the measurements is obtained. In the last step, the uncertainties for the calibration parameters are assessed by considering the identified solution vectors. The method is used to calibrate an EnergyPlus model of the Nursing Library at the National University of Ireland.

Mathematical iterative calibration offers a more general and rigorous approach than manual iterative procedures to the calibration of BES models. Through Monte Carlo simulation and optimization techniques it is possible to effectively explore the input parameter space looking for reasonable solutions. However uncertainty consideration is not satisfactory. Observation errors, and model inadequacy seems to be neglected in the reviewed contributions, while model parameter uncertainties are only partially considered. In particular, even if at the beginning of the several described procedures probability density distributions are used in describing model parameter uncertainties, it is not clear how they propagate during the analysis and influence the final outcomes, and the adopted objective functions or goodness of fit criteria do not account for observation errors. Although the need to provide a set of solutions is recognised, calibration is still considered as a normal optimisation problem and only a few solution vectors are returned. These few estimated parameter values cannot give an adequate description of the uncertainties related to these estimates. Potentially, techniques

like Monte Carlo Filtering are able to provide for such issues, however for the same reasons as GLUE, their convergence may be very slow.

2.2.3 Probabilistic/Stochastic Calibration

Probabilistic/Stochastic Calibration involves using grey-box models or supportive black-box probabilistic models in order to perform parameter identification by likelihood maximization or Maximum A Posteriori (MAP) estimation, when a Bayesian framework is used. This class of methods differs from the previous in the fact that they approach the calibration problem probabilistically, inferring full posterior probability density distributions as solutions.

Grey-box models are hybrid models built upon physical knowledge of a phenomenon and statistical analysis of observed data. They are simple in the sense that they are built by few equations and parameters, and flexible meaning that they can be easily extended or reduced. A relevant body of work about the grey-box modelling of building energy performances was done by Madsen and colleagues. Some of these contributions deemed to be most significant for this research are reviewed in the following.

In [72] stochastic differential equations were used to model an RC network representing the heat dynamics of a single storey wood-build house, located in Denmark, subject to Pseudo Random Binary Sequance (PRBS) of heat pulses. Model parameter estimation was achieved by maximising the likelihood using the Kalman Filter. The estimated model was then validated through correlation analysis and spectral analysis of the residuals, in order to verify their similarity to white noise. The main building parameters were determined in a previous study so it was possible to assess the correctness of the estimates. The resulting thermal resistances and solar aperture were in agreement with the older results while the heat capacity was slightly overestimated.

Since many grey box models, having different levels of complexity, can fit the same data set, methods for model selection are necessary. A possible approach is explained in [3]. The objective was to identify the simplest model able to adequately represent the metered internal temperature of a single storey building in Denmark, subject to PRBS heat injections. The suggestion was to use a forward selection strategy in which models are assessed according to likelihood ratio tests and relative performances. The modeller starts with the simplest possible design and gradually increases its complexity. New variants are accepted only if they provide sensible improvements.

In [60] and [78] wall components installed on test cells and monitored under

real weather conditions were investigated with different techniques. In the former, ARMAX models were employed to estimate the physical parameters of a wall element tested in round robin experiments performed at several European PASLINK test sites, through Output Error Methods and Prediction Error Methods. The validation and selection of the achieved models was done according to the following principles:

- Fit to the data: model residuals should result similar to white noise.
- Internal validity: the model has to be able to predict well part of the data not used for calibration.
- External validity: model results should agree with previous experiment in different conditions.
- Dynamic stability
- Identifiability: the model parameters should be uniquely identified from the data.
- Simplicity: the model should be as simple as possible.

In [78] grey box models based on stochastic differential equations were used to describe a wall component installed on a test cell at the Building Component Test Laboratory at Plataforma Solar de Almeria, in Spain. The objective was to estimate the main parameters characterising the construction element, namely transmittance, solar absorptance and effective heat capacity, by using Maximum Likelihood Estimation through Kalman Filter. Similarly, as in [72], model validation was performed by checking the similarity of the residuals with white noise in the time and frequency domains. The best model was selected according to likelihood ratio tests such as described in [3].

The use of black box probabilistic models is particularly suitable for calibrating detailed building energy models and GPR in a quasi-Bayesian framework is extensively employed. A review of such methodologies can be found in [99]. The main steps involved are:

- Quantifying the uncertainties for the model inputs.
- Parameter screening (if needed)
- Generating a meta-model or emulator of the computer model

- Performing model calibration of the original BES model against the field observations, with the built emulator.

The main concept is to augment the deterministic approach normally adopted by computer model in describing physical phenomena through the employment of probabilistic emulators. Such probabilistic emulators are mainly consisting of meta-models built upon GPR, and are able to accurately mimic the behaviour of the original computer models within the input parameter space considered in their creation. They allow for the formulation of a likelihood equation, probabilistically describing the relation between model inputs, model outputs, and target measurements, and for the effective consideration of the uncertainties. If a Bayesian framework is adopted it is possible to include prior knowledge in the analysis by setting up prior probability distributions for the parameters. Parameter estimation is then performed by maximising the likelihood, in a similar way as for grey-box models, or by integrating the relative joint posterior distribution through MCMC methods in the case a Bayesian treatment is adopted. This approach has been applied at first for calibrating computer models for general applications and only recently it has been used for the solution of building energy model calibration problems.

The first work describing in detail a Bayesian approach to the calibration of computer model is [65]. In this study the authors explain a mathematical framework, which will constitute the basis for a series of successive researches, capable of exhaustively treating modelling uncertainties according to the Bayes' Probability Theory. Even model inadequacy is carefully considered by including a specific term in the probabilistic meta-model proposed. The method is demonstrated on a computer model predicting dispersion and deposition of radioactive material due to accidental release, and on a hydrological model.

In [52] a similar approach is depicted and applied to the calibration of computer models simulating a charged particle accelerator and the spot-weld process of metal sheets. The same spot-weld model is also treated in [6] along with the analysis of a computer model modelling the kinetics of simple chemical reactions, in developing a framework for model validation. In this case calibration is used as instrument to investigate model flaws and possible improvements. The procedure was outlined as follows:

- To specify model inputs and relative uncertainties, in order to build an input/uncertainty map.
- To select the validation criteria according to the objective.

- Data collection and design of the computer experiment.
- Approximation of the computer model by creating an emulator or meta-model.
- Analysis of the model output and comparison with field data. This step involves the the statistical/probabilistic modelling of the data and the real calibration.
- Feedback information into the current validation exercise and feed-forward information in future validation activities. The former is relative to the improvement of the model according to the achieved results. The latter consists of using the current validation outcomes to assess the validity of future models for which field data are missing.

All the works above mentioned analysed models returning scalar model responses. If vectorial outputs are produced instead it is however possible to perform the analysis by indexing each discrete value with additional parameters. This approach is adopted in [7], wherein a model to predict the effect of vehicle collisions is investigated, and in [44], wherein a stochastic model describing the random evolution of the state of a biochemical system over time is calibrated, by using time as additional an input to index the vectorial outputs. In particular, in [44] two datasets coming from different sources were used in the calibration of the model in object. These datasets were returning conflicting estimates for the calibration parameters, and a method to overcome the issue is explained. The author allowed each dataset to have its own calibration parameters by expressing them as the combination of common terms and normally distributed random variables, quantifying the variation due to the particular dataset in use.

However, when model output consists of very long time series, which is often the case for BES models, this may lead to infeasible computational loads. A more convenient approach involves the approximation of vectorial outputs with basis expansions, retaining only a certain number of basis vectors sufficient to reach an adequate accuracy. The coefficients of the expansion are then used as calibration targets instead of the real data. Examples are [43], [5] and [50]. In [43] a computer model for cosmic simulations was calibrated against synthetically generated target data. PCA was used to decompose the computer model output, achieving a substantial reduction of its dimensionality. In [5], wavelet functions were employed to approximate the responses of a model predicting the loads on vehicle suspensions caused by stressful events. PCA was also used in [50]

to decompose two dimensional grids produced by a computer model predicting deformations of steel cylinders due to implosions by surrounding explosive.

For applications to building energy models, procedures similar to the one outlined in [65] were used in [48], [47], [66] and [49]. In the studies by Heo and colleagues ([48], [47] and [49]) quasi-steady state normative models and transient models were calibrated against monthly gas consumption data with the purpose of creating baseline models to evaluate the effectiveness and convenience of different energy conservation measures. In [66] an office building in South Korea was calibrated through deterministic and stochastic techniques showing that the latter provided more reasonable results. A common feature of the three works above mentioned is an initial screening of the parameters by applying the Morris Method ([76]) to reduce the dimensionality of the parameter space and select the most influential factors which should also be more easily identifiable.

The reviewed works revealed a theoretical advantage of *Probabilistic/Stochastic Calibration* compared to *Manual Iterative Calibration* and *Mathematical Analytical Calibration*. Unlike *Manual Iterative Calibration* and *Mathematical Analytical Calibration*, *Probabilistic/Stochastic Calibration* allows for a satisfactory treatment of the uncertainties, which are consistently considered during all the calibration process, and the adoption of a Bayesian approach, which allows modellers to rigorously include their expertise in the mathematical framework of the calculations. In particular, prior uncertainties and analyst personal knowledge, represented by the chosen probability density distributions for the calibration parameters, are averaged with the information coming from the measured data, resulting in a set of full posterior probability density distributions, constituting the solution to the calibration problem. Both the approaches, by using grey-box models or through probabilistic black-box emulators, were shown to be able to perform adequately. However, it is believed that the modelling activity can greatly benefit from a well defined modelling environment established by the adoption of a detailed simulation program. Such simulation tools are the result of knowledge accumulated over years of research and development, and give the opportunity to model very complex phenomena, like airflow and plant systems. Grey-box models, even if based on the same theoretical bases, do not offer such a well defined modelling environment and at each analysis an ad-hoc model has to be created. Although dynamic simulation programs may have steep learning curves, it seems unwise not to take advantage of all the modelling power they have to offer. Supportive probabilistic black box models are particularly suitable for performing calibration of models built with such tools. For these reasons, it is

on *Probabilistic/Stochastic Calibration* through probabilistic black-box emulators in a Bayesian framework (*Bayesian Calibration*) that this research focused on.

Chapter 3

Uncertainty and Sensitivity Analysis

In validation and calibration studies involving detailed building energy models having a large number of inputs (greater than 100), analysts find it difficult to understand which parameters to focus on in order to improve the match with the given target measurements. In particular, many parameters will have negligible effects on the model outputs, therefore varying them will be ineffective for calibration and validation purposes. Such parameters will also be subject to identifiability problems, since it will not be possible to identify them clearly from the data. In addition, the calculation load of the calibration framework that will be presented in the next chapter is dependent on the number of variables considered. Thus it is important to be able to consider only the parameters necessary to lead to a sensible analysis.

In these cases, it is necessary to reduce *over-parametrisation* and model dimensionality in order to facilitate the calculations. The methods that will be outlined in this Chapter have these objectives. In particular it is important to characterise the uncertainties related to the model inputs and determine, consequently, the sensitivity of the model in order to reduce the degrees of freedom of the BES model.

Uncertainty analysis, while it is routinely effectively performed for scalar model factors, is often neglected or performed in an approximated way for vectorial model inputs. Thus, particular attention will be given to these kind of variables, describing a possible approach based on Bootstrap and Smoothing with Roughness Penalty to the analysis of their uncertainties. As the literature review showed, qualitative parameter screening is usually employed in reducing the dimensionality of the BES model, and it is undertaken often through the Morris

Method. Qualitative sensitivity analysis methods are computationally convenient but they do not quantify the model output variance considered by working with the retained parameters only. This may lead to over-simplification, neglecting variables important in improving the match with the given data. In this study a sensitivity analysis procedure based on three steps: Factor Screening (FS), Factor Prioritising (FP) and Factor Fixing (FF), and Factor Mapping (FM), is presented which aims to be a more comprehensive and rigorous approach to model reduction than the common practice.

3.1 Uncertainty Analysis (UA)

The *Uncertainty Analysis* phase can be defined as the study of the variability of all the free parameters in a model. Its objective is to characterise model input uncertainties according to modeller beliefs and data gained from available specifications and literature. Uncertainty characterisation is performed probabilistically by attributing, to each considered factor, probability density distributions summarising the initial information (prior probability density distributions). In this way it possible to establish the parameter space where the analysis will take place (search space), which will have boundaries, and areas of different densities, according to the defined prior probability density distributions. The search space, resulting from *Uncertainty Analysis*, will be the input of the *Sensitivity Analysis*, and will be further refined by considering its outcome, before being employed as the solution domain in the *Calibration* step. In particular, it is important to work in an adequate search space, therefore the *Uncertainty Analysis* is a particularly important step in order to perform a sensible analysis. As general rule of thumb, it is preferable to have a slight overestimation of the model parameter uncertainties, in order to not constrain the analysis in a small search space at the beginning.

Uncertainty analysis is a well established and defined practice for scalar model inputs, and nowadays is routinely and effectively performed. Usually it consists of undertaking a literature review based on previous studies and given specifications, in order to define the probability density distributions characterising the initial beliefs about the variability of model parameters. Afterwards such distributions are sampled according to the analysis to be undertaken. For example, LHS is usually adopted in carry out Monte Carlo simulations, while factorial sampling is employed in performing parameter screening with the Morris Method.

In contrast, vectorial (or multidimensional) BES model inputs, like weather factors, are often neglected or their uncertainty characterisation is over-simplified by considering constant uncertainty magnitudes over time. These model inputs are at least as important as the scalar counterparts and deserve more rigorous analyses of their variabilities. In particular, the values of such variables are usually determined through measurements, which may be affected by errors of different magnitudes due to monitoring conditions changing over time. A useful example is usually given by the wind speed. *Systematic errors* of the sensors employed to measure this variable are usually dependent on its magnitude (Table 6.4), and it is reasonable to assume a similar behaviour for *random errors*. Figure 6.12 shows the wind speed measurements contained in the dataset relative

to EXPERIMENT2, and the analysis of their uncertainties with the proposed methods. It is possible to see that the variability of the observed values increases for higher velocities, which is likely to be due to larger *random errors*. This aspect is correctly reflected by the standard errors inferred, firstly through Bootstrap, secondly with Smoothing with Roughness Penalty. A reasonable approach to the investigation of the uncertainties relative to vectorial model inputs should consider such aspects by modelling the time series resulting from the measurements and investigating the local variability of the data. Therefore the following discussion will focus on the treatment of the uncertainties of multidimensional variables and a procedure for their analysis, quantification and characterisation based on Bootstrap and Smoothing with Roughness Penalty will be explained.

3.1.1 Uncertainty for multidimensional variables

Measurements are inevitably affected by errors, which can be divided into two kinds ([57]): *systematic errors* and *random errors*. The former are intrinsic properties of the sensors used in the monitoring. They can be assumed constant during time or as functions of the measured values and producing always the same bias in the data, meaning that a certain sensor always overestimates or underestimates the *true value* of the measured variable. The latter kind of errors are unpredictable and they are produced by the stochastic character of the monitored process or by the effects of unobserved phenomena on the measurements. Generally, they are assumed to be normally, independent and identically distributed (iid) variables. Under such assumptions, the model assumed for a certain measured variable ($x(t)$) including the error terms is the following:

$$\boldsymbol{x} = \tilde{\boldsymbol{x}} + \boldsymbol{s} + \boldsymbol{\varepsilon} \tag{3.1}$$

where $\boldsymbol{x}$ is the measurement, $\tilde{\boldsymbol{x}}$ is the *true value* of $x(t)$, $\boldsymbol{s}$ represents *systematic errors* and $\boldsymbol{\varepsilon}$ indicates *random errors*.

The properties of $\boldsymbol{s}$ would allow the data to be corrected accordingly if the exact magnitude and direction of the errors were known. However, often only information about their maximum bounds are provided. In these cases, their magnitudes and directions can be treated as random unidimensional variables, by setting up suitable probability distributions centred around zero from which to generate values. *Systematic errors* can then be simulated by drawing values from such probability density distributions and adding them to the data.

The consideration of *random errors* requires the modelling of the measured

time series. Usually their estimation is performed through smoothing techniques ([31],[54] and [62]), and it should be based on a priori probabilistic model describing the variability of $\boldsymbol{\varepsilon}$, in order to avoid generating spurious data by excessive smoothing. Establishing a prior error model for the *random errors* is not an easy task as the modelled entity is hidden in the data and unpredictable. However often useful information about the local accuracy of the measurements can be inferred by evaluating their local variance. For example, when data sampled at high frequencies are available, Bootstrap ([35]) can be used to calculate averages and relative standard errors over suitable time intervals. In this way it is possible to infer a reasonable prior probabilistic model for the random noise affecting the averaged time series. In Equation (3.2) the same symbols as in Equation (3.1) have been used but they refer to the Bootstrap outcomes.

$$\boldsymbol{x} = \tilde{\boldsymbol{x}} + \boldsymbol{\varepsilon} \tag{3.2}$$

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Lambda}^{-1}) \tag{3.3}$$

$$\boldsymbol{\Lambda}^{-1} = \text{diag}(se_i^2;\ i = 1, ..., N) \tag{3.4}$$

where $\text{diag}(\cdot)$ represents an operator that creates a diagonal matrix with elements comprising the given arguments, $\mathcal{N}(\cdot)$ indicates a multidimensional Normal probability density distribution, se_i are the estimated standard errors, $\boldsymbol{\Lambda}$ is the precision matrix of $\boldsymbol{\varepsilon}$ and N is the length of $\boldsymbol{x}$.

By considering the model in Equation (3.3) as prior probability density distribution for the *random errors*, Smoothing with Roughness Penalty can take place, with the purposes of refining the prior error model considered, investigating correlation patterns in the *random errors* and providing for missing values in the measurements. In this framework the measured time series is represented through a suitable basis expansion:

$$\boldsymbol{x} = \sum_{c=1}^{C} \boldsymbol{\psi}_c \hat{v}_c + \boldsymbol{\varepsilon} = \boldsymbol{\Psi}\hat{\boldsymbol{v}} + \boldsymbol{\varepsilon} \tag{3.5}$$

where $\boldsymbol{\psi}_c$ are realizations of the chosen basis function $\psi_c(t)$, having as coefficients $\hat{v}_c$.

It is important to notice that, in Equation (3.5), the unknown vector $\tilde{\boldsymbol{x}}$ has been represented by the function:

$$\hat{x(t)} = \sum_{c=1}^{C} \psi_c(t)\hat{v}_c$$

Its derivatives are used to quantify the roughness of the time series and control the power of the smoothing. In this study B-splines were used as $\psi_c(t)$.

The smoothing is performed by estimating the coefficients in $\hat{\boldsymbol{v}}$ according to a regularised least square criterion, penalising the roughness of the data, which is expressed as a function of the derivatives of $\hat{x(t)}$:

$$(\boldsymbol{x} - \boldsymbol{\Psi}\hat{\boldsymbol{v}})^T \boldsymbol{\Lambda} (\boldsymbol{x} - \boldsymbol{\Psi}\hat{\boldsymbol{v}}) + \eta \hat{\boldsymbol{v}}^T \boldsymbol{R} \hat{\boldsymbol{v}} \tag{3.6}$$

where η is a parameter controlling the power of the smoothing and $\boldsymbol{R}$ is the matrix quantifying the roughness of the data.

A suitable choice to represent this entity is the curvature of $\hat{x(t)}$, that is the square of its second time derivative. This measure of roughness is suggested in [91] and it is based on the rationale that an infinitely smooth function, like a straight line, has its second derivative always equal to zero, while a highly variable function will show, at least over some ranges, large values for it second derivative. Therefore $\boldsymbol{R}$ is defined as:

$$\boldsymbol{R} = \int \frac{d^2\boldsymbol{\Psi}}{dt} \left(\frac{d^2\boldsymbol{\Psi}}{dt} \right)^T dt$$

where $\boldsymbol{\Psi}$ represents the basis system defined by the functions $\psi_c(t)$.

By minimising Equation (3.6) with respect to $\boldsymbol{v}$, estimates for such variables ($\hat{\boldsymbol{v}}$) are given by:

$$\hat{\boldsymbol{v}} = \boldsymbol{S}\boldsymbol{x}$$

where $\boldsymbol{S} = (\eta \boldsymbol{R} + \boldsymbol{\Psi}^T \boldsymbol{\Lambda} \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \boldsymbol{\Lambda}$, and it projects the $\boldsymbol{x}$ in the space panned by the basis functions $\psi_c(t)$. An estimate of $\tilde{\boldsymbol{x}}$ can, then, be calculated as:

$$\hat{\boldsymbol{x}} = \boldsymbol{\Psi}\hat{\boldsymbol{v}}$$

Thus, in the smoothing, the standard errors calculated through Bootstrap, act mostly as weights relative to the accuracies of the values in $\boldsymbol{x}$, so that the smoothing model tries to obtain a closer fit for observations with low se_i.

$\hat{\boldsymbol{v}}$ depends on the smoothing parameter η since the matrix $\boldsymbol{S}$ is a function of it. For values of η close to zero the model in Equation (3.5) tries to fit exactly the observations even if this causes over-fitting. For values of η approaching ∞ the model will perform a standard linear regression which can be poorly representative of the main dynamical trends. Thus this parameter is particularly important and must be chosen carefully. [91] suggests to determine its value by minimising the Generalised Cross Validation criterion (GCV), which is a function of the Sum of

Squared Errors (SSE) and the degrees of freedom of the smoothing model ($\mathrm{df}(\eta)$):

$$GCV = \frac{N^{-1}SSE}{(N^{-1}\mathrm{tr}(\boldsymbol{I} - \boldsymbol{S}))^2} = \frac{N}{N - \mathrm{df}(\eta)} \frac{SSE}{N - \mathrm{df}(\eta)} \tag{3.7}$$

where $\mathrm{df}(\eta) = \mathrm{tr}(\boldsymbol{S})$.

GCV can be seen as a discounted mean squared error measure according to the degrees of freedom as a function of η. In particular since GCV is a function depending only on one variable, it is particularly easy to find the value of η minimising it by employing common optimisation routines. This approach was adopted in this work and the software package described in [90] was used to carry out the necessary calculations.

Because of the assumption of the noise being Gaussian (Equation (3.3)), $\boldsymbol{v}$ will be normally distributed with means $\hat{\boldsymbol{v}}$ and covariance matrix $\boldsymbol{S}\boldsymbol{\Lambda}^{-1}\boldsymbol{S}^T$. Hence for the property of Gaussian distributions the following can be assumed for $\boldsymbol{x}$ ([91]):

$$\boldsymbol{x} \sim N(\boldsymbol{\Psi}\hat{\boldsymbol{v}}, \boldsymbol{\Psi}\boldsymbol{S}\boldsymbol{\Lambda}^{-1}\boldsymbol{S}^T\boldsymbol{\Psi}^T) \tag{3.8}$$

The probability density distributions defined by Equation (3.8) can be used to draw random samples for the multi-dimensional variables. The *systematic errors* terms can then be added in order to rebuild the model depicted in Equation (3.1).

The differences produced by considering the uncertainties for multidimensional inputs through a simplified method and the outlined methodology is highlighted in Section 6.5. In particular, the analysed model was significantly sensitive to wind speed if the uncertainties for such weather factor were simply estimated as the average of the standard errors calculated through Bootstrap; while the same variable had a relatively low importance if its uncertainties were assessed with the method described above. It is believed that the approach described in this section is more rigorous than the common practice, thus providing more sensible results.

3.2 Sensitivity Analysis (SA)

A three-step sensitivity analysis is proposed involving different settings, objectives and methods according to the tasks to perform ([106]):

- Factor Screening (FS): the Morris Method is applied to the model in order to gain qualitative information about parameter effect magnitudes and understand which variables may have major influences on the model responses.

- Factor Prioritising (FP) and Factor Fixing (FF): the Sobol Method, a variance based sensitivity analysis technique, is used to quantify the amount of variance that can be attributed to individual parameters or group of parameters. This allows the identification of those inputs which should be tuned carefully in order to minimise the model output variance and those inputs which can be fixed to default values because they are responsible for negligible model output variations.

- Factor Mapping (FM): GLUE is employed to weight the simulations, and the relative model input vectors are mapped according to their probabilities to produce model realizations close to the target measurements. Furthermore the importance of the different model parameters in improving the match between model outputs and measurements is assessed.

FS could be seen as redundant since the same information can be gained from more detailed sensitivity results using FP and FF. However, variance based methods are particularly simulation intensive and because of the large number of parameters usually involved in detailed building energy models, the number of simulations needed would be hardly manageable. Thus the Morris Method, which has a substantially lower computational burden, is used to gather qualitative information and reduce the dimensionality of the problem by grouping together the parameters having small effects. Then the Sobol Method is used to assess the efficacy of the screening.

Usually these sensitivity analysis techniques are applied on scalar model responses. BES models produce vectorial outputs and it was necessary to extend them by following the principles outlined in [17] and [67]. In particular, PCA (A.2) is extensively applied to treat vectorial outputs. Descriptions of the methods used to perform FS, FP, FF and FM with particular emphasis on the modifications adopted in order to treat multidimensional outputs follows.

3.2.1 Factor Screening

In this step the Morris Method [76] is employed to identify the most important factors governing the model in order to reduce its dimensionality to a feasible extent for the next phases. The model inputs are divided into Most Important Factors (MIF) and Least Important Factors (LIF), and eventually grouped.

The Morris Method characterises the sensitivity of the model to the $p-th$ input through the concept of elementary effects ($ee_{p,l}$), which can be described

as partial derivative approximations:

$$ee_{p,l} = \frac{f(\boldsymbol{z}_l + \boldsymbol{e}_p \zeta_{p,l}) - f(\boldsymbol{z}_l)}{\zeta_{p,l}}$$

where $f(\cdot)$ represents the model evaluated at a certain input vector $\boldsymbol{z}_l$, $\boldsymbol{e}_p$ is a zero vector where only the $p-th$ position is equal to one and $\zeta_{p,l}$ is the applied $l-th$ variation to the $p-th$ input.

A chosen number L (usually within the range [20, 50]) of elementary effects are calculated, for each model input, according to a factorial design, representing the parameter space, defined as described in [19], which allow the required information to be obtained with a number of simulations (M) linearly proportional to the number of inputs (P): $M = L(P+1)$. The empirical absolute means ($\mu^\star{}_p$):

$$\mu^\star{}_p = \frac{1}{L}\sum_{l=1}^{L} |ee_{p,l}|$$

and the standard deviations (ς_p):

$$\varsigma_p = \sqrt{\frac{1}{L-1}\sum_{l=1}^{L}(ee_{p,l} - \bar{ee}_p)^2}$$

of the derived samples of elementary effects, where $\bar{ee}_p$ is the empirical mean of the elementary effects relative to the $p-th$ model parameter, characterise respectively the magnitude and typology of each input effect. In particular the magnitudes of first order effects are proportional to $\mu^\star{}_p$, while parameters having high ς_p have significant higher order effects.

To handle the high dimensionality of the ESP-r vectorial outputs it has been necessary to extend the method. For this purpose PCA was used to decompose the generated simulation data set, so that each simulation output $\boldsymbol{y}_m$ is represented as follows:

$$\boldsymbol{y}_m = \sum_{q=1}^{Q} \boldsymbol{k}_q w_{m,q} = \boldsymbol{K}\boldsymbol{w}_m + \boldsymbol{\varepsilon} \tag{3.9}$$

where Q is the number of retained orthonormal bases, $\boldsymbol{k}_q$, and $\boldsymbol{K}$ is the matrix having as columns such bases. $w_{m,q}$ and $\boldsymbol{w}_m$ are respectively the corresponding coefficients and vector of coefficients. In particular Q is determined in order to achieve a good approximation of each $\boldsymbol{y}_m$, so that $\boldsymbol{\varepsilon}$ represents a negligible amount of the total variability. In this way, the initial data set of dimensionality $N \times M$, is reduced to Q independent sets of dimensionality $M \times 1$ suitable to be

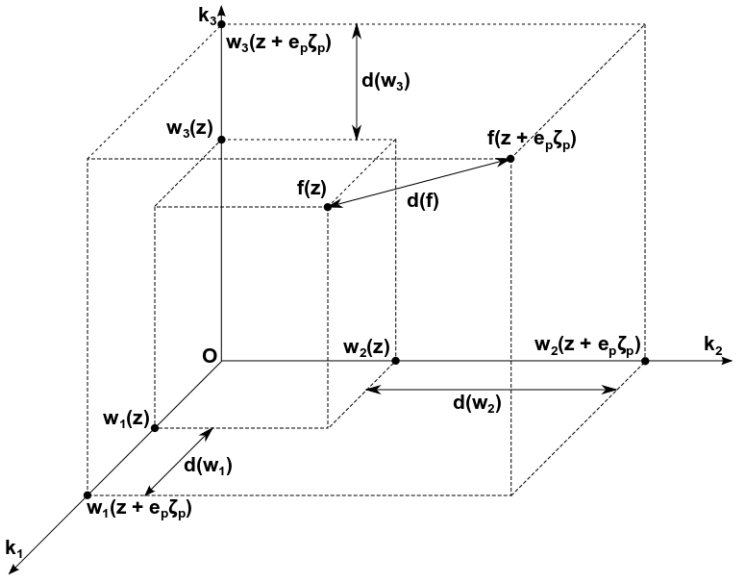


Figure 3.1: Elementary effect representation in the coordinate system defined by PCA relative to the $p-th$ model input and $m-th$ model output.

separately processed by the Morris Method.

In particular, each simulation output is represented in the space defined by $\boldsymbol{k}_q$ as depicted in Figure 3.1 where Q has been assumed equal to 3 to aid visualisation. Therefore, the elementary effects can be generalized as follows:

$$EE_{p,l} = \frac{\sqrt{\sum_{q=1}^{Q} \mathrm{d}(w_{p,q,l})^2}}{\zeta_{p,l}} \tag{3.10}$$

and the sensitivity indexes adopted by the Morris Method as:

$$M^{\star}{}_p = \frac{1}{L}\sum_{l=1}^{L} |EE_{p,l}| \tag{3.11}$$

$$SD_p = \sqrt{\frac{1}{L-1}\sum_{l=1}^{L}(EE_{p,l} - \bar{EE}_p)^2} \tag{3.12}$$

where $\bar{EE}_p$ is the empirical mean of $EE_{p,l}$.

Due to the orthonormal properties of $\boldsymbol{k}_q$, the $EE_{p,l}$ can be seen as an approximations of the directional derivative with respect to the $p-th$ input in the reference system defined by PCA. As for the indexes $\mu^{\star}{}_p$ and ς_p, $M^{\star}{}_p$ indicating the magnitude of the $p-th$ parameter main effect on the model output, and SD_p, measuring the dispersion of the elementary effects, can be used to assess the power and the typology of relation between the $p-th$ model inputs and the

model output. The same means used in analysing the original sensitivity indexes, like $\mu^\star - \varsigma$ plots, can be used for interpreting $M^\star{}_p$ and SD_p.

The Morris Method requires the direct link between one combination of inputs and the associated outputs in order to correctly calculate the elementary effects. Therefore it is proposed to create variations for the multi-dimensional variables by taking the iso-probability lines of the inferred distributions (Equation (3.8)) and considering the defined quantiles as model parameters by including them in the factorial design. The same approach can be used to add *systematic errors*. This procedure does not produce completely random samples and slightly overestimates uncertainties, since random and *systematic errors* always add, but it provides adequate variations while keeping the parameters to a reasonable number. Sensitivity indexes can be dependent on the variation ranges applied. To limit this aspect all input samples were scaled and centred so that each of them has mean equal to zero and standard deviation equal to one.

3.2.2 Factor Prioritising and Factor Fixing

The purpose of this stage is to extend the qualitative outcomes from the previous analysis by quantifying the amount of model output variance attributable to each model parameter or group of parameters. Through Factor Prioritising the MIF parameters are ranked according to the fraction of model output variance contributed by the parameters independently. This provides a priority scale for identifying which variables it is necessary to know accurately in order to reduce most of the model output variance. Factor Fixing provides complementary objectives. In this case, MIF factors are ranked depending on the fraction of model output variance for which they are responsible for, including interactions between parameters. This gives information about which factors it is possible to fix to default values because their associated uncertainties have negligible influence on model outputs. In this phase, the effectiveness of the previous parameter screening is assessed by calculating the portion of model output variance attributable to the LIF group. The Sobol Method ([110]) has been used to undertake these tasks.

It is based on the decomposition of total (or unconditional) model output variance into its conditional components. By defining with $\boldsymbol{z}_i$ a set of inputs and with $\boldsymbol{z}_{-i}$ its complement, the variance of the model output, $\boldsymbol{y}$, can be decomposed

as follows ([71]):

$$\mathrm{V}(\boldsymbol{y}) = \mathrm{V}(\mathrm{E}(\boldsymbol{y}|\boldsymbol{z}_i)) + \mathrm{E}(\mathrm{V}(\boldsymbol{y}|\boldsymbol{z}_i)) \tag{3.13}$$

$$\mathrm{V}(\boldsymbol{y}) = \mathrm{V}(\mathrm{E}(\boldsymbol{y}|\boldsymbol{z}_{-i})) + \mathrm{E}(\mathrm{V}(\boldsymbol{y}|\boldsymbol{z}_{-i})) \tag{3.14}$$

where $\boldsymbol{y}|\boldsymbol{z}_i$ and $\boldsymbol{y}|\boldsymbol{z}_{-i}$ indicate the conditionality of variances ($\mathrm{V}(\cdot)$), and estimates ($\mathrm{E}(\cdot)$), on knowing $\boldsymbol{z}_i$ and $\boldsymbol{z}_{-i}$ respectively.

Equations (3.13) and (3.14) allow the definition of two sensitivity measures of major importance. In particular, by normalising these equations with $\mathrm{V}(\boldsymbol{y})$ it is possible to derive the two following indexes as fractions of the total output variance ([108]):

$$S_i = \frac{\mathrm{V}(\mathrm{E}(\boldsymbol{y}|\boldsymbol{z}_i))}{\mathrm{V}(\boldsymbol{y})} \tag{3.15}$$

$$ST_i = \frac{\mathrm{E}(\mathrm{V}(\boldsymbol{y}|\boldsymbol{z}_{-i}))}{\mathrm{V}(\boldsymbol{y})} \tag{3.16}$$

S_i indicates the portion of $\mathrm{V}(\boldsymbol{y})$ which can be attributed to the first order effect of $\boldsymbol{z}_i$ and it is named the first order effect. Parameters with high values for S_i are responsible for most of the output variance and by knowing their *true values* it is possible to reduce output uncertainty at least proportionally to the sum of the S_i indexes, since higher order effects might actually contribute as well. This can be seen directly from Equation (3.13). Since $\mathrm{V}(\boldsymbol{y})$ is a constant, factors with high $\mathrm{V}(\mathrm{E}(\boldsymbol{y}|\boldsymbol{z}_i))$ have low expected output variance ($\mathrm{E}(\mathrm{V}(\boldsymbol{y}|\boldsymbol{z}_i))$).

ST_i represents the portion of $\mathrm{V}(\boldsymbol{y})$ left by leaving only $\boldsymbol{z}_i$ unknown, i.e. the portion of $\mathrm{V}(\boldsymbol{y})$ attributable to all the effects (including first and high order effects) of $\boldsymbol{z}_i$ and it is called the total effect. In particular, setting parameters with negligible ST_i, to default values should leave a negligible output uncertainty. Similarly if $\boldsymbol{z}_i$ has negligible influence, $\mathrm{V}(\mathrm{E}(\boldsymbol{y}|\boldsymbol{z}_{-i}))$ will be high since for different $\boldsymbol{z}_{-i}$ the estimates of the output are sensibly different and thus $\mathrm{E}(\mathrm{V}(Y|\boldsymbol{z}_{-i}))$ assumes small values.

The S_i index is used in performing FP, while the ST_i index is employed in carrying out FF. Also, differences in the values of the two indexes indicates how $\boldsymbol{z}_i$ act on the model outputs. For similar S_i and ST_i the relative factors have linear and additive effects while for high ST_i and low S_i, they exert their influences by interacting with other inputs or through non-linearities. For example for linear additive models $S_i = ST_i$ and $\sum_i S_i = 1$; while for non-linear models $S_i < ST_i$ and $\sum_i ST_i = 1 > 1$, since different ST_i may account for the same higher order effects.

The multi-dimensional integrals involved in the evaluation of the estimates and variances in equations Equation (3.15) and Equation (3.16) are calculated through Monte Carlo estimation. Several estimators have been proposed to perform this task and a comparison study can be found in [107]. In this work the estimator proposed in [106] was chosen.

As in the previous case the sensitivity indexes described above are defined for scalar model outputs. Their calculation can be extended according to [67] in order to account for vectorial model responses. The outlined methodology is based on PCA decomposition of the covariance matrix of the model outputs. In particular let $\boldsymbol{Y}$ be a $N \times M$ matrix having as columns the simulation outputs, with $\boldsymbol{Y}$ centred so that each row has mean 0. Then the total variability of the data set represented by $\boldsymbol{Y}$ can be defined as the trace ($\mathrm{tr}(\cdot)$) of its empirical covariance matrix ($\boldsymbol{\Sigma}$):

$$\boldsymbol{\Sigma} = \frac{1}{M}\boldsymbol{Y}\boldsymbol{Y}^T$$
$$\mathrm{V}(\boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{\Sigma})$$

It is shown in [67] that the data set consisting of the sum of the principal components has the same variance as the original data set and, since it is composed by unidimensional variables it can be used to replace $\boldsymbol{Y}$ in Equations (3.15) and (3.16). It is worth mentioning that the determinant of $\boldsymbol{\Sigma}$ can also be used as a measure of $\mathrm{V}(\boldsymbol{Y})$.

In performing the Sobol Method, it is proposed to use random seeds as parameters to which link the variations of multidimensional inputs.

3.2.3 Factor Mapping

Factor Mapping is an extension to normal sensitivity analysis which can provide useful information about which parameters it is necessary to focus on, in calibration and validation studies. During FS, FP and FF the measures of sensitivity were relative to model output. This may lead to neglecting some variables having a relatively low influence in the model, but important for achieving a good fit with the given monitored data. FM, by considering the target measurements as well, provides for this, and integrates the results from the previous phases. It aims to identify input vectors more likely to produce model realisations close to the target observations and thus to determine which model parameters are more powerful in improving the similarity between model predictions and measurements. To perform this task the GLUE framework ([8]) was chosen, in particular

the methodology described in [94].

GLUE is a simplified Bayesian method allowing inference about posterior estimates of model parameters and model outputs, which is conditional on the measured data. In the usual Bayesian approach the joint posterior distribution of the model inputs ($\boldsymbol{z}$) given the observations ($\boldsymbol{y}^*$) is defined as:

$$p(\boldsymbol{z}|\boldsymbol{y}^*) \propto p(\boldsymbol{y}^*|\boldsymbol{z})p(\boldsymbol{z}) \tag{3.17}$$

where $p(\boldsymbol{y}^*|\boldsymbol{z})$ is the likelihood of $\boldsymbol{y}^*$ and $p(\boldsymbol{z})$ is the joint prior probability distribution of $\boldsymbol{z}$. It is then possible to infer the posterior estimates and uncertainties for the model inputs and outputs by evaluating the following integrals:

$$\mathrm{E}(\boldsymbol{y}|\boldsymbol{y}^*) = \int f(\boldsymbol{z})p(\boldsymbol{z}|\boldsymbol{y}^*)d\boldsymbol{z} \tag{3.18}$$

$$\mathrm{cov}(\boldsymbol{y}|\boldsymbol{y}^*) = \int (f(\boldsymbol{z}) - \mathrm{E}(\boldsymbol{y}|\boldsymbol{y}^*))(f(\boldsymbol{z}) - \mathrm{E}(\boldsymbol{y}|\boldsymbol{y}^*))^T p(\boldsymbol{z}|\boldsymbol{y}^*)d\boldsymbol{z} \tag{3.19}$$

$$\mathrm{E}(\boldsymbol{z}|\boldsymbol{y}^*) = \int \boldsymbol{z}p(\boldsymbol{z}|\boldsymbol{y}^*)d\boldsymbol{z} \tag{3.20}$$

$$\mathrm{cov}(\boldsymbol{z}|\boldsymbol{y}^*) = \int (\boldsymbol{z} - \mathrm{E}(\boldsymbol{z}|\boldsymbol{y}^*))(\boldsymbol{z} - \mathrm{E}(\boldsymbol{z}|\boldsymbol{y}^*))^T p(\boldsymbol{z}|\boldsymbol{y}^*)d\boldsymbol{z} \tag{3.21}$$

where $f(\cdot)$ indicates the computer model. Often the evaluations of the integrals in Equations (3.18), (3.19), (3.20) and (3.21) are difficult for detailed computer models requiring the employment of Markov Chain Monte Carlo or Importance Sampling methods and the creation of meta-models to speed up the calculations. Additionally the definition of a proper likelihood equation is not always possible due to the lack of information.

The GLUE approach assumes that the model parameters are generated directly from their prior distributions and an approximate likelihood measure instead of an accurate one. In particular a suitable function depending on the SSE between model realisations and measured data is assumed as approximation of a proper likelihood measure. Such function assigns higher weights to model simulations ($\boldsymbol{y}_m$), and thus to the relative input vectors, having low SSE and vice versa and it is called weighting function ($\omega(\cdot)$). The definition of $\omega(\cdot)$ is problem dependent and examples can be found in ([9]). In this study it has been defined as follows ([94]):

$$\omega(\boldsymbol{y}_m|\boldsymbol{z}_m, \alpha) \propto (\frac{1}{2N}\sum_{i=1}^{N}(y_{m,i} - y_i^*)^2)^{-\alpha} \tag{3.22}$$

α can be used to regulate the power of the weighting by more or less concentrating higher values of $\omega(\cdot)$ around $\boldsymbol{y}^*$ and is empirically chosen in order to achieve a reasonable distribution of the weights over the simulation sample.

It is then possible to weight each $\boldsymbol{y}_m$ as follows:

$$\omega_m = \frac{\omega(\boldsymbol{y}_m|\boldsymbol{z}_m, \alpha)}{\sum_{m=1}^{M} \omega(\boldsymbol{y}_m|\boldsymbol{z}_m, \alpha)} \tag{3.23}$$

As Equation (3.22) is an approximation of a proper likelihood measure, ω_m are approximations of posterior probabilities having drawn the model inputs from their prior probability distributions, and can be used to simplify Equations (3.18) to (3.21):

$$\mathrm{E}(\boldsymbol{y}|\boldsymbol{y}^*) \approx \hat{\boldsymbol{y}}' = \sum_{m=1}^{M} \boldsymbol{y}_m \omega_m \tag{3.24}$$

$$\mathrm{cov}(\boldsymbol{y}|\boldsymbol{y}^*) \approx \mathrm{cov}(\boldsymbol{y})' = \sum_{m=1}^{M} (\boldsymbol{y}_m - \hat{\boldsymbol{y}}')(\boldsymbol{y}_m - \hat{\boldsymbol{y}}')^T \omega_m \tag{3.25}$$

$$\mathrm{E}(\boldsymbol{z}|\boldsymbol{y}^*) \approx \hat{\boldsymbol{z}}' = \sum_{m=1}^{M} \boldsymbol{z}_m \omega_m \tag{3.26}$$

$$\mathrm{cov}(\boldsymbol{z}|\boldsymbol{y}^*) \approx \mathrm{cov}(\boldsymbol{z})' = \sum_{m=1}^{M} (\boldsymbol{z}_m - \hat{\boldsymbol{z}}')(\boldsymbol{z}_m - \hat{\boldsymbol{z}}')^T \omega_m \tag{3.27}$$

These equations can be estimated through Bootstrap procedures using ω_m as sampling weights. Similarly by sampling with replacement the input vectors assuming as sampling probabilities ω_m it is possible to infer posterior samples and empirical probability distributions for the model parameters. By analysing differences between the inferred posterior probability distributions and the assumed prior probability distributions it is possible to assess the importance of each parameter in driving model outputs towards good matches with the target observations since variables most influencing the goodness of the fit will show larger variations. Thus the value of α is particularly important. Too high a value of this parameter may produce a weight distribution dominated by few ω_m, leading to underestimation of the posterior parameter uncertainties. On the other hand, too low a value of α may generate a practically uniform distribution of ω_m over the different input vectors, precluding useful information being obtained from subsequent sampling.

Similar information can be inferred by processing the values of the weighting

function with sensitivity analysis techniques, such as the Sobol Method, thus quantifying the importance of the considered model variables in calibrating the computer model. In this case, first order and total effects represent fractions of V(ω) (where $\boldsymbol{\omega} = [\omega_1, ..., \omega_m, ..., \omega_M]$). Parameters with high S_i are most responsible for changing the goodness of fit between model outputs and field measurements (i.e. are the most important factor for model calibration). ST_i for LIF can be used to roughly assess the variability of the model output which could contribute to improve model calibration, but that is neglected by fixing such inputs.

Through first order and total parameter effects it is also possible to assess the degree of *over-parametrisation* of a model. In particular big differences between S_i and ST_i mean that the goodness of the match between model outputs and measurements is governed by higher order effects and interactions leading to several optimal input vectors. It is important to notice that even if all the parameters are set to their optimal values there still may be discrepancies between simulation results and monitored data, mainly because of model inadequacy. In this case it may be necessary to improve the model (e.g. through higher resolution modelling), or it may indicate a deficiency in the simulation program.

The main problem with the GLUE method is its slow convergence rate in the estimation of Equations (3.24) to (3.27) if the zones of high probability of the joint posterior distribution are distant from the zones of high probability of the joint prior distribution. In this case a large number of model simulations is required to achieve a good estimate. This issue is partially mitigated by coupling GSA and GLUE since the former makes available a large set of model outputs which can then be processed by the latter.

If applied in an iterative fashion GLUE can be used to perform model calibration ([8]), but due to its slow convergence and the empirical character of the adopted measure of goodness of the fit, it is deemed more suitable for preliminary investigation, preceding more rigorous calibration analyses, as described in the next chapter.

Chapter 4

Calibration of BES models

This chapter describes the core of this research, that is the development of a rigorous framework for probabilistic calibration of detailed building energy models. The main principles upon which such framework was developed was an effective consideration of the modelling uncertainties, and of the prior knowledge coming from the analyst expertise and from the given data, as well as the capability of identifying upgrades effective in creating models more representative of the observes processes. To achieve this goal and to contain the calculation time to a reasonable extent, GPR in a quasi-Bayesian set up is used to build fast running emulators of the original BES models, in order to support their calibration.

The previous works in [13], [5] and [50] constitute the basis upon which the mathematical framework described in this Chapter was developed. Concepts and methods outlined in these papers were blended together and modified in order to treat the particular problem of BES model calibration. The main differences and novelties with respect to these previous studies are the following. Vectorial variable boundary conditions are considered, by representing them with adequate basis expansions derived through PCA. Particular care was taken in building the *Difference Model* in such a way to not interfere with the inference of the calibration parameters. In many real applications it will be necessary or convenient to consider more than one observed variable as calibration targets. An example may be the calibration of a multi-zone model, wherein the temperatures of adjacent zones are correlated and dependent on each other. The identification of model parameters can benefit from these dependency relations between calibration target variables. Therefore the methodology to build dependent Gaussian Processs (GPs) depicted in [13], is used for expanding the framework in order to take into account these aspects. Finally, instead of normally employed MCMC methods, SAN and AIS were used respectively for training the BES model emulator and

infer the calibration parameters. In particular the latter, allowing the calculation of marginal likelihoods, made possible the adoption of Bayes Factors as model selection criteria.

The Chapter starts with an overview about the main concepts behind GPs, upon which GPR is based. Then, the mathematics upon which the calibration methodology is built, is explained in detail, and it will be explained how to expand the proposed probabilistic models in order to consider multiple target variables. The employed quasi-Bayesian approach, requires the specification of prior probability distributions for all the variables involved in the calculations. In particular, the adoption of a quasi-Bayesian approach, instead of a fully Bayesian approach, allows the reduction of the main problem in more easy to solve sub-problems, simplifying the calculations, at the cost of a less then full consideration of some kind of modelling uncertainties. Thus, the differences between a quasi-Bayesian and a Bayesian treatment of the problem, as well as the rationale behind the choices of the prior distribution for the various variables involved in the calculations, will be discussed. The Chapter ends explaining the use of Bayes Factor in performing model selection and their calculation.

4.1 Gaussian Processes

GPs are collections of random variables, any finite number of which are jointly normally distributed. In this context they are employed as stochastic models in performing non-linear regression, in order to build the probabilistic emulator of the BES model to calibrate. Such regression framework is named Gaussian Process Regression (GPR) and is deeply discussed in [93].

GPs are defined by a mean function ($\mu(\cdot)$) and a covariance function ($cf(\cdot)$). They determine probability density distributions over the functions apt to represent observations of a stochastic variable ($\boldsymbol{y}$). Thus, it is possible to indicate that $\boldsymbol{y}$ is represented through a certain GP by writing:

$$\boldsymbol{y} \sim GP(\mu(\cdot), cf(\cdot))$$

$\mu(\cdot)$ is responsible for the main trends, while $cf(\cdot)$ determines how different observations are related to each other, as well as the stochastic character of the process itself. In this study, GPs are used mainly for interpolation, and it is not necessary to specify a proper mean function in order to achieve good performances. It is, instead, convenient to assume such function equal to zero, since the data can always be centred accordingly. Therefore the choice of the covariance function will completely characterize the GPR model.

GPs can be seen as continuous versions of discrete multidimensional Gaussian distributions. $cf(\cdot)$ determines the entries of the covariance matrix of the observed variable ($\boldsymbol{\Sigma}$) as functions of the model parameters ($\boldsymbol{z}$), and links different observations and relative inputs according to similarity criteria. Thus, the $i, j-th$ entry of $\boldsymbol{\Sigma}$ is determined according to the $cf(\cdot)$, the $i-th$ input vector ($\boldsymbol{z}_i$), and the $j-th$ input vector ($\boldsymbol{z}_j$):

$$\Sigma_{i,j} = cf(\boldsymbol{z}_i, \boldsymbol{z}_j)$$

and the entire covariance matrix is defined by:

$$\boldsymbol{\Sigma} = cf(\boldsymbol{Z}, \boldsymbol{Z})$$

where $\boldsymbol{Z}$ is the matrix having as rows the vectors $\boldsymbol{z}$.

An arbitrary function will not be in general a valid covariance function, but fortunately several covariance functions, apt to represent different kind of observed processes, are available from literature. The interested reader is referred again to [93], while the covariance functions adopted in this work will be explained

in detail in Section 4.2.2. $cf(\cdot)$ will have its own parameters, controlling different aspects of the functions being distributed according to the relative GP, like amplitude, smoothness, and frequency. They are usually called hyper parameters (hp), in order to not be confounded with the actual model parameters.

Having a dataset, consisting of a set of observations and relative inputs vectors, a GPR model can be trained by inferring the hp maximising the likelihood function of the corresponding multidimensional Gaussian probability density distribution, or, in case a Bayesian approach is adopted, their joint posterior probability density distribution function. Once the hp have been inferred, it is possible to define the multidimensional Normal distribution representing the GP for the training input vectors, and to make predictions for new input vectors ($\boldsymbol{z}^*$). In particular GPR models, instead of point predictions, return a full predictive distribution, consisting of the probability density distribution of the model output ($\boldsymbol{y}^*$) at the vectors $\boldsymbol{z}^*$, conditional on the training dataset. Such distribution can be derived from the joint probabilistic model for the training and prediction sets ([10]):

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}^* \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} cf(\boldsymbol{Z}, \boldsymbol{Z}) & cf(\boldsymbol{Z}, \boldsymbol{Z}^*) \\ cf(\boldsymbol{Z}^*, \boldsymbol{Z}) & cf(\boldsymbol{Z}^*, \boldsymbol{Z}^*) \end{bmatrix}\right) \tag{4.1}$$

as:

$$\boldsymbol{y}^* | \boldsymbol{Z}^*, \boldsymbol{Z}, \boldsymbol{y} \sim$$
$$N(cf(\boldsymbol{Z}^*, \boldsymbol{Z}) cf(\boldsymbol{Z}, \boldsymbol{Z})^{-1} \boldsymbol{y},\ cf(\boldsymbol{Z}^*, \boldsymbol{Z}^*) - cf(\boldsymbol{Z}^*, \boldsymbol{Z}) cf(\boldsymbol{Z}, \boldsymbol{Z})^{-1} cf(\boldsymbol{Z}, \boldsymbol{Z}^*)$$

where $\boldsymbol{y}^*$ indicates the values of the stochastic process evaluated at the vectors $\boldsymbol{z}^*$ and $\boldsymbol{Z}^*$ is the matrix having as rows $\boldsymbol{z}^*$. $cf(\boldsymbol{Z}, \boldsymbol{Z}^*) = cf(\boldsymbol{Z}, \boldsymbol{Z}^*)^T$ are the cross-covariance matrices, and they describe how the observations used during the training of the model and the model predictions covary. $cf(\boldsymbol{Z}^*, \boldsymbol{Z}^*)$ is the covariance matrix of the predictions only.

Recent works in [51] and [13], outlined a framework expressing GPs as a convolutions of normally distributed variables, and smoothing kernels. In particular in the latter case, white noise and stationary Gaussian kernels are used in order to build GPs of dependent variables. The concept described in [13] was particularly useful in expanding the method in order to consider multiple target variables during calibration.

4.2 Calibration

The proposed calibration method employs a probabilistic emulator, built upon GPR in a quasi-Bayesian framework, of the original BES, in order to effectively consider the stochastic character of the calibration problem, and perform inference of the unknown model parameters. Such an emulator, will be able to accurately mimic the behaviour of the computer model within the input parameter space of interest, and to give a probabilistic representation of the BES model through the definition of a proper likelihood equation, linking probabilistically model inputs, model outputs and observations. Furthermore the Bayesian paradigm allows for an effective treatment of the uncertainties, and to consider prior knowledge, by setting up suitable prior probability density distributions.

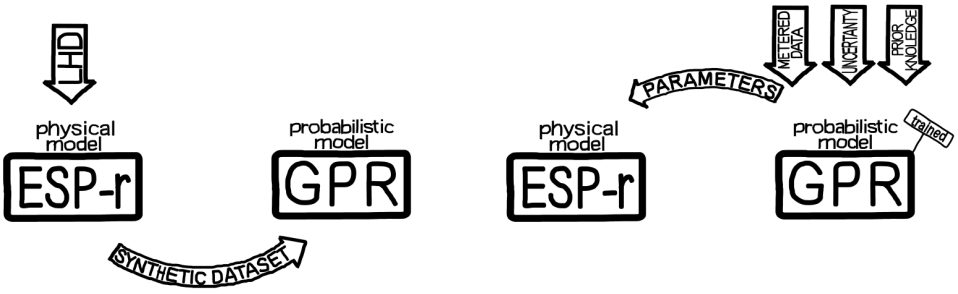
There are two steps involved in the calibration procedure:

- *Training*: in this phase (Figure 4.1a) the probabilistic emulator (*Training Model*) is built by fitting a GPR model to a synthetic dataset consisting of a significant sample of BES outputs. Such a sample is generated by running the model in a Monte Carlo fashion according to a LHS design, effectively exploring the input parameter space defined by the selected prior probability density distributions. The experience accumulated during this research indicated that a number of model simulations equal to ten times the number of calibration parameters, gives generally good performances. The fit is performed by estimating the values of the hp of the GPR model maximising the relative joint posterior probability density distribution using Simulate Annealing (SAN).

- *Identification*: in this second stage (Figure 4.1b) a similar probabilistic model representing the observed target variables, is formulated. This model will be composed of two independent terms, one representing the variability of the measured data that the BES model can represent (*Calibration Model*), and one depicting the variability of the measurements that, mainly because of model inadequacy, the computer model cannot describe (*Difference Model*). The *Calibration Model* will be linked to the *Training Model*, by deriving the conditional probability distribution of the observations on the training data set, and used to infer marginal posterior probability density distributions for the calibration parameters. The *Difference Model* will be used to investigate causes of discrepancies between model predictions and measurements. In this phase, inference is performed by integrating the joint posterior probability density distribution functions for the unknown

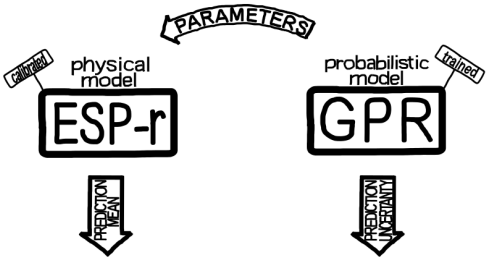
parameters through Annealed Importance Sampling (AIS).

The main rationale behind this two step procedure is the following. Firstly, during the *Training*, the deterministic BES is transformed in a stochastic process. Secondly, in the *Identification*, it is assumed that the measured data can be generated in most part by the same process with the help of a complement providing for possible model deficiencies and other major causes of discrepancy. This hypothesis is then verified against the observed data.



(a) *Training* phase.

(b) *Identification* Phase.



(c) Prediction Phase.

Figure 4.1: Graphical representation of the calibration process.

The GPR model can be used to make predictions (Figure 4.1c). However GPR models of structures similar to those adopted here are effective in interpolation but have poor performance in extrapolation. In particular, their capabilities of emulating the BES become increasingly worse as the condition and inputs, for which predictions are needed, become increasingly different from those used in *Training*. In this cases the predictive distributions will have means close to zero and large variances. Thus, when it is necessary to provide predictions, it is advised to use the BES model output as the mean vector and to use the GPR model only for calculating confidence bounds. Such an approach should provide

conservative results since the predictive uncertainties will be overestimated to a certain extent. Even more it is reasonable to have relatively large uncertainties for predictions subject to boundary conditions not observed before. Nonetheless the extrapolation capability of the framework must be improved, and this will be object of future research. In the case studies presented in the next two chapters confidence bands will be provided only for conditions similar to those used in training the model, and the model predictions are the outputs of the actual BES models.

4.2.1 Premises of the methodology

Calibration of computer models involves the observation of real phenomena, the building of a model representing such phenomena, and the inference of unknown model parameters according to the information acquired through monitoring activity. Here some preliminary concepts and mathematical notation are established in order to facilitate the reading of the following sections explaining the mathematics underpinning the method.

A real process is observed subject to S known variable boundary conditions, like weather factors or particular experimental solicitations, and it is characterized by T unknown calibration parameters. According to the objective of the study, during the monitoring, R target variables are accurately measured. A computer model representing the observed phenomenon will be built by imposing on it the S variable boundary conditions, and it will have parameters corresponding to the unknown calibration parameters. Its main objective will be to adequately predict the R monitored target variables, and so it will have to be able to produce R corresponding outputs.

To reduce the computational burden to a feasible extent, the dataset dimensionality is decreased through PCA. PCA is used to determine optimal empirical basis expansions for variable boundary conditions, simulation outcomes and observed target variables. In order to effectively apply PCA, it is necessary to have multiple observation of the real process subject to different boundary conditions. This is achieved by dividing in periods the time series resulting from the measurements of the involved variables. To clarify this point it may be useful make an example. Let assume that the experimental dataset, contains 1000 measurements for each boundary condition and target variable, and that it is deemed approriate to split this dataset in 4 different periods of 250 measurements. Similarly, let assume that, by running the model suitably, 200 simulations are obtained. Eachone of such simulations is then splitted in 4, for a total of 800 periods of 250 time

steps. The same results could be achieved by running independently the model initialised adequately, therefore these latter periods are just called simulations. In the following M^* indicates the number of periods used to split the observed data (4 in the exmple jut made), M refers to the total number of simulations after their division in periods (800 in the exmaple just made), and N is the period length (250 in the example just made).

Thus, let be:

- $\boldsymbol{x}_m^{(s)}$ indicates the $s-th$ variable environmental condition, in the $m-th$ simulation, imposed on the BES model.
- $\boldsymbol{y}_m^{(r)}$ indicates the $m-th$ simulation for the $r-th$ model output.
- $\boldsymbol{\theta}_m$ indicates the $m-th$ model input vector, form LHS design, having elements corresponding to the unknown calibration parameters.
- $\boldsymbol{X}^{(s)}$ denote the $N \times M$ matrix, having as columns the vectors $\boldsymbol{x}_m^{(s)}$.
- $\boldsymbol{Y}^{(r)}$ denote the $N \times M$ matrix, having as columns $\boldsymbol{y}_m^{(r)}$.
- $\boldsymbol{Y} = [\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(r)}, \ldots, \boldsymbol{Y}^{(R)}]$.

Let $\boldsymbol{X}^{(s)}$ and $\boldsymbol{Y}$ be centred so that their columns have mean equal to zero. Then, by following the procedure explained in Section A.2 such variables can be represented as:

$$\boldsymbol{y}_m^{(r)} = \sum_{q=1}^{Q} \boldsymbol{k}_q w_{m,q}^{(r)} + \boldsymbol{\varepsilon}^{(r)} = \boldsymbol{K}\boldsymbol{w}_m^{(r)} + \boldsymbol{\varepsilon}^{(r)} \tag{4.2}$$

$$\boldsymbol{x}_m^{(s)} = \sum_{c=1}^{C^{(s)}} \boldsymbol{\psi}_c^{(s)} v_{m,c}^{(s)} + \boldsymbol{\varepsilon}^{(s)} = \boldsymbol{\Psi}^{(s)}\boldsymbol{v}_m^{(s)} + \boldsymbol{\varepsilon}^{(s)} \tag{4.3}$$

or

$$\boldsymbol{Y} = \boldsymbol{K}\boldsymbol{W}^T + \boldsymbol{\varepsilon}^{(r)} \tag{4.4}$$

$$\boldsymbol{X}^{(s)} = \boldsymbol{\Psi}^{(s)}(\boldsymbol{V}^{(s)})^T + \boldsymbol{\varepsilon}^{(s)} \tag{4.5}$$

where:

- Q and $C^{(s)}$ are the numbers of basis used for approximating the model simulations and the $s-th$ boundary condition respectively.
- $\boldsymbol{k}_q$ and $\boldsymbol{\psi}_c^{(s)}$ are the empirical basis defined by applying PCA for the model outputs, and $s-th$ boundary condition data sets respectively.

- $\boldsymbol{K}$, $\boldsymbol{\Psi}^{(s)}$ are the matrices having as columns $\boldsymbol{k}_q$ and $\boldsymbol{\psi}_c^{(s)}$ respectively.
- $\boldsymbol{\varepsilon}^{(r)}$ and $\boldsymbol{\varepsilon}^{(s)}$ indicates the errors due to the approximations.
- $\boldsymbol{W}$ and $\boldsymbol{V}^{(s)}$ respectively are the $MR \times Q$ and $M \times S$ matrices, defined as:

$$\boldsymbol{W} = \begin{bmatrix} \hat{\boldsymbol{W}}^{(1)} \\ \vdots \\ \hat{\boldsymbol{W}}^{(r)} \\ \vdots \\ \hat{\boldsymbol{W}}^{(r)} \end{bmatrix}$$

$$\boldsymbol{V}^{(s)} = [\boldsymbol{v}_1^{(s)}, ..., \boldsymbol{v}_m^{(s)}, ..., \boldsymbol{v}_M^{(s)}]^T = [\boldsymbol{v}_1^{(s)}, ..., \boldsymbol{v}_c^{(s)}, ..., \boldsymbol{v}_C^{(s)}]$$

were $\boldsymbol{W}^{(r)} = [\boldsymbol{w}_1^{(r)}, ..., \boldsymbol{w}_m^{(r)}, ..., \boldsymbol{w}_M^{(r)}]^T = [\boldsymbol{w}_1^{(r)}, ..., \boldsymbol{w}_q^{(r)}, ..., \boldsymbol{w}_Q^{(r)}]$,

Estimates for the coefficients $w_{m,q}^{(r)}$ and $v_{m,c}^{(s)}$, grouped in the vectors $\boldsymbol{w}_m^{(r)}$ and $\boldsymbol{v}_m^{(s)}$, are given by:

$$\hat{\boldsymbol{w}}_m^{(r)} = (\boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{K}^T\boldsymbol{y}_m^{(r)}$$
$$\hat{\boldsymbol{v}}_m^{(s)} = (\boldsymbol{\Psi}^{(s)})^T\boldsymbol{\Psi}^{(s)})^{-1}(\boldsymbol{\Psi}^{(s)})^T\boldsymbol{x}_m^{(s)}$$

Then such estimates can then be substituted in Equations 4.2 and 4.3, and the matrices $\hat{\boldsymbol{W}}$, $\hat{\boldsymbol{W}}^{(r)}$ and $\hat{\boldsymbol{V}}^{(s)}$ can be defined similarly as $\boldsymbol{W}$, $\boldsymbol{W}^{(r)}$ and $\boldsymbol{V}^{(s)}$.

Afterwards, the target variables are represented as sum of the basis expansion derived by projecting them in the space spanned by $\boldsymbol{k}_q$, and a term accounting for eventual residual variance that cannot be represented through $\boldsymbol{k}_q$. Thus, let:

- $\boldsymbol{y}_{m^*}^{(r)*}$ indicate the $m^* - th$ observation period.
- $\boldsymbol{Y}^{(r)*}$ denote the $N \times M$ matrix, having as columns $\boldsymbol{y}_{m^*}^{(r)*}$.
- $\boldsymbol{Y}^* = [\boldsymbol{Y}^{*(1)}, ..., \boldsymbol{Y}^{(r)*}, ..., \boldsymbol{Y}^{*(r)}]$.

Let $\boldsymbol{Y}^*$ be centred as $\boldsymbol{Y}$. Then such variables can be approximated as:

$$\begin{aligned} &\boldsymbol{y}_{m^*}^{(r)*} = \sum_{q=1}^{Q} \boldsymbol{k}_q w_{m,q}^{(r)*} + \sum_{d=1}^{D} \boldsymbol{h}_d u_{m^*d}^{(r)} + \boldsymbol{\varepsilon}_{m^*}^{(r)} = \\ &\boldsymbol{K}\boldsymbol{w}_{m^*}^{(r)*} + \boldsymbol{H}\boldsymbol{u}_{m^*}^{(r)} + \boldsymbol{\varepsilon}_{m^*}^{(r)*} = \\ &\boldsymbol{\Phi}\boldsymbol{b}_{m^*}^{(r)} + \boldsymbol{\varepsilon}_{m^*}^{(r)*} \end{aligned} \tag{4.6}$$

or

$$\boldsymbol{Y}^* = \boldsymbol{K}(\boldsymbol{W}^*)^T + \boldsymbol{H}\boldsymbol{U}^T + \boldsymbol{\varepsilon}^* \tag{4.7}$$

where:

- $\boldsymbol{w}_{m^*}^{(r)*}$ are defined analogously as $\boldsymbol{w}_m^{(r)}$, and estimates for them are provided by:
$$\hat{\boldsymbol{w}}_m^{r*} = (\boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{K}^T\boldsymbol{y}_{m^*}^{(r)*}$$
- $\boldsymbol{W}^*$ and $\boldsymbol{U}$ are respectively $M^*R \times Q$ and $M^*R \times D$ matrices defined as:
$$\boldsymbol{W}^* = \begin{bmatrix} \boldsymbol{W}^{(1)*} \\ \vdots \\ \boldsymbol{W}^{(r)*} \\ \vdots \\ \boldsymbol{W}^{(r)*} \end{bmatrix}; \ \boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}^{(1)} \\ \vdots \\ \boldsymbol{U}^{(r)} \\ \vdots \\ \boldsymbol{U}^{(r)} \end{bmatrix}$$
where:
  - $\boldsymbol{W}^* = [\boldsymbol{w}_1^{(r)*}, ..., \boldsymbol{w}_{m^*}^{(r)*}, ..., \boldsymbol{w}_{M^*}^{(r)*}]^T = [\boldsymbol{w}_1^{*(r)}, ..., \boldsymbol{w}_q^{(r)*}, ..., \boldsymbol{w}_Q^{*(r)}]$;
  - $\boldsymbol{U}^{(r)} = [\boldsymbol{u}_1^{(r)}, ..., \boldsymbol{u}_{m^*}^{(r)}, ..., \boldsymbol{u}_{M^*}^{(r)}]^T = [\boldsymbol{u}_1^{(r)}, ..., \boldsymbol{u}_d^{(r)}, ..., \boldsymbol{u}_D^{(r)}]$.
- $\boldsymbol{H}$ is the matrix having as columns $\boldsymbol{h}_d$ and $\boldsymbol{\Phi} = [\boldsymbol{K}, \boldsymbol{H}]$.
- $\boldsymbol{b}^{(r)} = [(\boldsymbol{w}_{m^*}^{(r)*})^T, (\boldsymbol{u}_{m^*}^{(r)})^T]^T$.

Similar results can be achieved by projecting the $s-th$ boundary condition influencing the real experiment in the m^*-th period ($\boldsymbol{x}_{m^*}^{(s)*}$) on the basis $\boldsymbol{\psi}_c^{(s)}$. It is important to notice that, the resulting coefficients ($\hat{\boldsymbol{v}}_{m^*}^{(s)*}$) will be equal to $\hat{\boldsymbol{v}}_m^{(s)}$, since the actual time series were imposed on the model, and used in calculating the latter. However, for clarity reasons, and since $\hat{\boldsymbol{v}}_{m^*}^{(s)*}$ can appear in different orders, two different notations are adopted. In particular let be $\hat{\boldsymbol{V}}^*$ the following matrix:

$$\boldsymbol{V}^* = \begin{bmatrix} \hat{\boldsymbol{V}}^{(1)*} \\ \vdots \\ \hat{\boldsymbol{V}}^{(s)*} \\ \vdots \\ \hat{\boldsymbol{V}}^{(S)*} \end{bmatrix}$$

where $\hat{\boldsymbol{V}}^{(s)*}$ is defined analogously to $\hat{\boldsymbol{V}}^{(s)}$, but it contains the coefficients $\hat{\boldsymbol{v}}_{m^*}^{(s)*}$.

Thus, the $\boldsymbol{y}_{m^*}^{(r)*}$ are expressed as sum of two terms represented by different basis expansions. In most of the cases, it will not be possible to represent $\boldsymbol{y}_{m^*}^{(r)*}$ with enough accuracy by only considering the variation modes represented by the basis $\boldsymbol{k}_q$, so it is necessary to consider a supplementary set of basis ($\boldsymbol{h}_d$) to achieve a sufficient level of precision. These two contributions represents different aspects of the measured data. The former, represents the variability of the measurements which can be represented by the BES and its emulator, and the *Calibration Model* will be built using it. The latter describes the variability of the data which can not be explained by the BES and its emulator, probably because of model inadequacies, and it will underpin the *Difference Model*.

In [5] the same basis used to built the *Calibration Model* were also employed in building the *Difference Model*. In [50] $\boldsymbol{h}_d$ were chosen according to available information about model deficiencies and the consequent differences between its outputs and the actual measurements. The former approach may cause confounding between the parameters of the *Calibration Model* and *Difference Model*, as discussed in [5]. The latter approach requires knowledge that, because of the variety of the processes that BES models aim to represent and their variability, is usually unavailable. For these reasons a different approach was used in defining $\boldsymbol{h}_d$.

In particular, $\boldsymbol{h}_d$ and the relative coefficients were defined by difference as follows. In order to simplify the calculations, the *Calibration Model* and the *Difference Model* were built so as to be independent. This was achieved by defining $\boldsymbol{h}_d$ to be complementary to $\boldsymbol{k}_q$. Let the *Difference Vectors* ($\boldsymbol{\delta}_{m^*}^{(r)}$) be defined defined as:

$$\boldsymbol{\delta}_{m^*}^{(r)} = \boldsymbol{y}_{m^*}^{(r)*} - \boldsymbol{K}(\boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{K}^T\boldsymbol{y}_{m^*}^{(r)*} \tag{4.8}$$

so that they represents the residuals of the measured data after a least square fit based upon $\boldsymbol{k}_q$. Let be $\boldsymbol{\Delta}$ the matrix collecting all the $\boldsymbol{\delta}_{m^*}^{(r)}$:

$$\boldsymbol{\Delta} = [\boldsymbol{\Delta}^{(1)}, ..., \boldsymbol{\Delta}^{(r)}, ..., \boldsymbol{\Delta}^{(R)}]$$

where $\boldsymbol{\Delta}^{(r)}$ denotes the $N \times M$ matrix, having as columns $\boldsymbol{\delta}_{m^*}^{(r)}$. $\boldsymbol{h}_d$ are derived by applying PCA on $\boldsymbol{\Delta}$. In particular, only a certain number (D) of basis is retained and used in approximating the *Difference Vectors*. Estimates for the $\boldsymbol{u}_{m^*}^{(r)}$ ($\hat{\boldsymbol{u}}_{m^*}^{(r)}$) are then calculated as:

$$\hat{\boldsymbol{u}}_{m^*}^{(r)} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{\delta}_{m^*}^{(r)} \tag{4.9}$$

In this way the basis set $\boldsymbol{K}$ and $\boldsymbol{H}$ will be orthogonal and they will describe vari-

ation modes that are complementary to each other, allowing to treat separately the two corresponding models.

Having the estimates $\hat{\boldsymbol{w}}_m^{r*}$ and $\hat{\boldsymbol{u}}_{m^*}^{(r)}$ it is possible to define the matrices $\hat{\boldsymbol{W}}^*$, $\hat{\boldsymbol{W}}^{(r)*}$, $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{U}}^{(r)}$ analogously to $\boldsymbol{W}^*$, $\boldsymbol{W}^{(r)*}$, $\boldsymbol{U}$ and $\boldsymbol{U}^{(r)}$.

By looking at Equations (4.2 – 4.6), it is possible to notice that from period to period the only variables changing are the coefficients in $\boldsymbol{w}_m^{(r)}$, $\boldsymbol{w}_{m^*}^{(r)*}$, $\boldsymbol{u}_{m^*}^{(r)}$ and $\boldsymbol{v}_m^{(s)}$ (and their estimates $\hat{\boldsymbol{w}}_m^{(r)}$, $\hat{\boldsymbol{w}}_m^{r*}$, $\hat{\boldsymbol{u}}_{m^*}^{(r)}$ and $\hat{\boldsymbol{v}}_m^{(s)}$), since the relative bases are constants. Therefore, instead of using the actual measured and synthetic data as parameters and target variables of the GPR model, it is possible to use the relative coefficients, so reducing significantly the dataset dimensionality. In particular, the dimensionality of the problem is reduced from $(M + M^*) \times N$ to $(M + M^*) \times Q + M^* \times D$. In this sense it is important to notice that the same set of basis vectors ($\boldsymbol{k}_q$) was assumed in representing all the R outputs and target variables. Therefore the problem dimensionality reduction will be more effective for outputs and target variables showing similar shapes and trends.

Finally, let the $M \times P$ and $M^* \times P$ (where $P = S + T$ is the total number of inputs for the GPR model) matrices $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$ be the training input matrix and the observation input matrix having as rows the vectors:

$$\boldsymbol{z}_m = [\boldsymbol{\theta}_m, ..., \hat{\boldsymbol{v}}_m^{(1)}, ..., \hat{\boldsymbol{v}}_m^{(s)}, ..., \hat{\boldsymbol{v}}_m^{(S)}]$$
$$\boldsymbol{z}_{m^*}^* = [\boldsymbol{\theta}^*, ..., \hat{\boldsymbol{v}}_{m^*}^{(1)*}, ..., \hat{\boldsymbol{v}}_{m^*}^{(s)*}, ..., \hat{\boldsymbol{v}}_{m^*}^{(S)*}]$$

where $\boldsymbol{z}_m$ are known and $\boldsymbol{z}_{m^*}^*$ are the input vectors characterizing the observations, composed by known $\hat{\boldsymbol{v}}_{m^*}^{(s)*}$ and unknown calibration parameters contained in $\boldsymbol{\theta}^*$.

4.2.2 The mathematics of the method

In the following the mathematical formulation behind the probabilistic models representing the BES model and the observed data as well their coupling will be explained in details. Firstly the mathematical models behind the *Training* and the *Identification* phases will be described considering only one target variable, therefore the superscript "(r)" will be dropped. Following this, it will be shown how to consider multiple calibration targets, with particular focus on the necessary modification to the previously depicted models. At the end of the section covariance functions and prior probability density distributions will be discussed, as well as the employment of the *Difference Model* for identifying model improvements.

Training

The objective of the training phase is to create a probabilistic meta-model, emulating the BES model within the space defined by the prior probability density distributions chosen for the model parameters. This is achieved by fitting a GPR model to a previously generated sample of simulation outputs, that is by finding the values for the hp, of the used covariance functions, maximising the relative joint posterior probability density distribution. This probabilistic model will then be used to condition the estimation of the calibration parameters.

The model chosen to represent the $m-th$ simulation outcome ($\boldsymbol{y}_m$) is the following:

$$\boldsymbol{y}_m = f(\boldsymbol{z}_m) + \boldsymbol{\varepsilon} \tag{4.10}$$

where $f(\cdot)$ is the BES emulator and $\boldsymbol{\varepsilon}$ is assumed to be white noise with precision λ. $\boldsymbol{\varepsilon}$ represents *Representation Uncertainty* and the inaccuracy due to the approximation of the actual BES with a black-box meta-model.

A Normal-Gamma model is used to probabilistically represent the model depicted by equation 4.10, as in [50]:

$$p(\boldsymbol{y}_m, \lambda|\boldsymbol{z}_m) \propto p(\boldsymbol{y}_m|\lambda, \boldsymbol{z}_m) \times \mathcal{G}(\lambda|shp, rt) \tag{4.11}$$

where:

$$p(\boldsymbol{y}_m|\lambda, \boldsymbol{z}_m) \propto \frac{\lambda^{\frac{N}{2}}}{2\pi^{\frac{N}{2}}} exp\{-\frac{\lambda}{2}(\boldsymbol{y}_m - \hat{\boldsymbol{y}}_m)^T(\boldsymbol{y}_m - \hat{\boldsymbol{y}}_m)\} \tag{4.12}$$

$$\mathcal{G}(\lambda|shp, rt) \propto \frac{rt^{shp}}{\Gamma(shp)} \lambda^{shp-1} exp\{-rt\lambda\} \tag{4.13}$$

Equation (4.12) is the likelihood of the $m-th$ simulation ($\boldsymbol{y}_m$) depending on λ and the input vector $\boldsymbol{z}_m$, for which $f(\cdot)$ returns the predictions $\hat{\boldsymbol{y}}_m$. Equation (4.13) is the Gamma prior probability density distribution used to represent λ, where shp and rt are respectively its shape and rate parameters.

By representing $f(\cdot)$ with the basis expansion defined in Equation (4.4), Equation (4.12) becomes:

$$p(\boldsymbol{y}_m|\lambda, \boldsymbol{z}_m, \boldsymbol{w}_m) \propto \lambda^{\frac{N}{2}} exp\{-\frac{\lambda}{2}(\boldsymbol{y}_m - \boldsymbol{K}\boldsymbol{w}_m)^T(\boldsymbol{y}_m - \boldsymbol{K}\boldsymbol{w}_m)^T\} \tag{4.14}$$

It is important to notice that $\boldsymbol{w}_m$ are not least square estimates, but they are the corresponding unknown regression variables. $\hat{\boldsymbol{w}}_m$ arise naturally by completing the square (A.1), and rearranging Equation (4.14) as shown in Equation 4.15.

They will be the new target variables for the regression model.

$$\begin{aligned} &p(\boldsymbol{y}_m|\lambda, \boldsymbol{z}_m, \boldsymbol{w}_m) \propto \\ &\lambda^{\frac{N}{2}} exp\{-\frac{\lambda}{2}(\hat{\boldsymbol{w}}_m - \boldsymbol{w}_m)^T \boldsymbol{K}^T \boldsymbol{K}(\hat{\boldsymbol{w}}_m - \boldsymbol{w}_m) - \\ &\frac{\lambda}{2}(\boldsymbol{y}_m{}^T(\boldsymbol{I} - \boldsymbol{K}(\boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{K}^T)\boldsymbol{y}_m\} \end{aligned} \tag{4.15}$$

Equation (4.15) factorises as follows:

$$\begin{aligned} &p(\boldsymbol{y}_m|\lambda, \boldsymbol{z}_m, \boldsymbol{w}_m) \propto \\ &\lambda^{\frac{Q}{2}} exp\{-\frac{\lambda}{2}(\hat{\boldsymbol{w}}_m - \boldsymbol{w}_m)^T \boldsymbol{K}^T \boldsymbol{K}(\hat{\boldsymbol{w}}_m - \boldsymbol{w}_m)\} \times \\ &\lambda^{\frac{N-Q}{2}} exp\{-\frac{\lambda}{2}\boldsymbol{y}_m{}^T(\boldsymbol{I} - \boldsymbol{K}(\boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{K}^T)\boldsymbol{y}_m\} \end{aligned} \tag{4.16}$$

The likelihood of the complete simulation set ($\boldsymbol{Y}$) is then defined as product of M contributions equal to Equation (4.16):

$$\begin{aligned} &p(\boldsymbol{Y}|\lambda, \boldsymbol{Z}, \boldsymbol{W}) \propto \\ &\prod_{m=1}^{M} \lambda^{\frac{Q}{2}} exp\{-\frac{\lambda}{2}(\hat{\boldsymbol{w}}_m - \boldsymbol{w}_m)^T \boldsymbol{K}^T \boldsymbol{K}(\hat{\boldsymbol{w}}_m - \boldsymbol{w}_m)\} \times \\ &\prod_{m=1}^{M} \lambda^{\frac{N-Q}{2}} exp\{-\frac{\lambda}{2}(\boldsymbol{y}_m)^T(\boldsymbol{I} - \boldsymbol{K}(\boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{K}^T)\boldsymbol{y}_m\} \end{aligned} \tag{4.17}$$

The last term on the right hand side of Equation (4.17) can be included in Equation (4.13), which is updated with the information coming from the data:

$$\mathcal{G}(\lambda|shp', rt') \propto \frac{(rt')^{shp'}}{\Gamma(shp')}\lambda^{shp'-1} exp\{-rt'\lambda'\} \tag{4.18}$$

where, shp' and rt' are the posterior values of the shape and rate parameters respectively:

$$shp' = shp + \frac{M(N-Q)}{2} \tag{4.19}$$

$$rt' = rt + \frac{1}{2}\sum_{m=1}^{M} \boldsymbol{y}_m{}^T(\boldsymbol{I} - \boldsymbol{K}(\boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{K}^T)\boldsymbol{y}_m \tag{4.20}$$

Therefore it is possible to focus the rest of the mathematical formulation on the first term in the right hand side of Equation (4.17). In order to correctly perform GPR it is necessary to rearrange such term basis-wise. This can be done

easily since the matrix $\boldsymbol{K}^T\boldsymbol{K}$ is diagonal. The result is:

$$p(\boldsymbol{Y}|\lambda,\boldsymbol{Z},\boldsymbol{W}) \propto p(\hat{\boldsymbol{W}}|\lambda,\boldsymbol{Z},\boldsymbol{W}) \propto \prod_{q=1}^{Q} \lambda^{\frac{M}{2}} exp\{-\frac{1}{2}(\hat{\boldsymbol{w}}_q - \boldsymbol{w}_q)^T \boldsymbol{C}_q (\hat{\boldsymbol{w}}_q - \boldsymbol{w}_q)\} \tag{4.21}$$

where $\boldsymbol{C}_q = \text{diag}(\lambda \boldsymbol{k}^T \boldsymbol{k};\ m = 1,\ldots,M)$. GPR is performed by introducing zero mean GP prior distributions on the variables $\boldsymbol{w}_q$:

$$\boldsymbol{w}_q \propto GP(0, \varrho_q(\cdot))$$

where $\varrho_q(\cdot)$ indicates the adopted covariance function, with hp collected in the vector $\boldsymbol{\vartheta}_q$. The set of hp of all the Q covariance functions, $\varrho_q(\cdot)$, is indicated by the vector $\boldsymbol{\vartheta}$.

$\boldsymbol{w}_q$ can be integrated out ([10]), so achieving the expression of the probability of $\hat{\boldsymbol{W}}$ dependent only on λ, $\boldsymbol{Z}$ and $\boldsymbol{\vartheta}$:

$$p(\hat{\boldsymbol{W}}|\lambda,\boldsymbol{Z},\boldsymbol{\vartheta}) \propto \prod_{q=1}^{Q} (2\pi)^{-\frac{M}{2}} |(\lambda \boldsymbol{C}_q)^{-1} + \varrho(\boldsymbol{Z},\boldsymbol{Z})|^{-\frac{1}{2}} exp\{-\frac{1}{2}\hat{\boldsymbol{w}}_q [(\lambda \boldsymbol{C}_q)^{-1} + \varrho(\boldsymbol{Z},\boldsymbol{Z})]^{-1} \hat{\boldsymbol{w}}_q\} \tag{4.22}$$

By applying the Bayes' Theorem it is possible to derive the expression for the joint posterior probability density distribution for the unknown parameters of the GPR model (λ and $\boldsymbol{\vartheta}$):

$$p(\lambda,\boldsymbol{\vartheta}|\hat{\boldsymbol{W}},\boldsymbol{Z}) \propto p(\hat{\boldsymbol{W}}|\lambda,\boldsymbol{Z},\boldsymbol{\vartheta}) \times \mathcal{G}(\lambda|shp', rt') \times p(\boldsymbol{\vartheta}) \tag{4.23}$$

where $p(\boldsymbol{\vartheta})$ indicates the prior probability density distributions chosen for the hp of the covariance functions.

In the proposed quasi-Bayesian framework it is not needed to derive full posterior marginal probability density distributions for λ and $\boldsymbol{\vartheta}$, but it is sufficient to infer only the relative MAP values, since these parameters are then fixed in *Identification*. Thus, solutions to the fitting problem are calculated by maximising Equation (4.23) through Simulate Annealing (SAN) (B.2). The inferred precision and hp will constitute the hypothesis that probabilistically define the BES model according to the built GPR model. In particular the $k-th$ GPR model having λ and $\boldsymbol{\vartheta}$ set to the estimated values will be indicated by $\mathcal{M}_k$.

Identification

In this phase a probabilistic model depicting the observed target variable is built and linked to $\mathcal{M}_k$, in order to allow to perform inference of the unknown calibration parameters, consistent with the BES behaviour. The model representing the observations from the m^*-th experimental period ($\boldsymbol{y}^*_{m^*}$) is composed of the physical model emulator, $f(\cdot)$, evaluated at the unknown inputs, $\boldsymbol{z}^*_{m^*}$, characterising the experiment (*Calibration Model*), a stochastic term, $\Delta(\cdot)$, dependent only on the known variable boundary conditions, accounting for model inadequacy and correlated noise (*Difference Model*), and a white noise term, $\boldsymbol{\varepsilon}^*_{m^*}$, with precision $\lambda^*_{m^*}$, representing *Residual Variability* and *Observation Errors*:

$$\boldsymbol{y}^*_{m^*} = f(\boldsymbol{z}^*_{m^*}) + \Delta(\boldsymbol{x}^{(s)*}_{m^*}) + \boldsymbol{\varepsilon}^*_{m^*} \tag{4.24}$$

The steps leading to the derivation of joint posterior probability density distribution for the unknown parameter, are similar to those explained in the previous section. Therefore only the main passages will be repeated, and particular focus will be given in explaining the linking of the two models.

As in the *Training* phase a Normal-Gamma model is used to probabilistically characterize the model depicted in Equation (4.24):

$$p(\boldsymbol{y}^*_{m^*}, \lambda^*_{m^*} | \boldsymbol{z}^*_{m^*}) \propto p(\boldsymbol{y}^*_{m^*} | \lambda^*_{m^*}, \boldsymbol{z}^*_{m^*}) \times \mathcal{G}(\lambda^*_{m^*} | shp^*_{m^*}, rt^*_{m^*}) \tag{4.25}$$

where:

$$\begin{aligned} &p(\boldsymbol{y}^*_{m^*} | \lambda^*_{m^*}, \boldsymbol{z}^*_{m^*}) \propto \\ &\frac{(\lambda^*_{m^*})^{\frac{N}{2}}}{2\pi^{\frac{N}{2}}} exp\{-\frac{\lambda^*_{m^*}}{2}(\boldsymbol{y}^*_{m^*} - \hat{\boldsymbol{y}}^*_{m^*} - \hat{\boldsymbol{\delta}}_{m^*})^T(\boldsymbol{y}^*_{m^*} - \hat{\boldsymbol{y}}^*_{m^*} - \hat{\boldsymbol{\delta}}_{m^*})\} \end{aligned} \tag{4.26}$$

$$\mathcal{G}(\lambda^*_{m^*} | shp^*_{m^*}, rt^*_{m^*}) \propto \frac{(rt^*_{m^*})^{shp^*_{m^*}}}{\Gamma(shp^*_{m^*})} (\lambda^*_{m^*})^{shp^*_{m^*}-1} exp\{-rt^*_{m^*}\lambda^*_{m^*}\} \tag{4.27}$$

Equation (4.26) is the likelihood of the m^*-th observation ($\boldsymbol{y}^*_{m^*}$) depending on $\lambda^*_{m^*}$ and $\boldsymbol{z}^*_{m^*}$, for which $f(\cdot)$ and $\Delta(\cdot)$ return respectively $\hat{\boldsymbol{y}}^*_{m^*}$ and $\hat{\boldsymbol{\delta}}_{m^*}$. Equation (4.27) is the prior probability density distribution assumed for $\lambda^*_{m^*}$, where $shp^*_{m^*}$ and $rt^*_{m^*}$ are the shape and rate parameters. It is important to notice that, unlikely the training model, in this case each observation period has its own precision parameter ($\lambda^*_{m^*}$).

By representing $f(\cdot)$ and $\Delta(\cdot)$ according to the basis expansion defined in

Equation (4.6), Equation (4.26) becomes:

$$p(\boldsymbol{y}^*_{m^*}|\lambda^*_{m^*}, \boldsymbol{z}^*_{m^*}) \propto \\ \frac{(\lambda^*_{m^*})^{\frac{N}{2}}}{2\pi^{\frac{N}{2}}} exp\{-\frac{\lambda^*_{m^*}}{2}(\boldsymbol{y}^*_{m^*} - \boldsymbol{\Phi}\boldsymbol{b}_{m^*})^T(\boldsymbol{y}^*_{m^*} - \boldsymbol{\Phi}\boldsymbol{b}_{m^*})\} \tag{4.28}$$

and the likelihood for the whole observation set ($\boldsymbol{Y}^*$), given the input and precision sets ($\boldsymbol{Z}^*$ and $\boldsymbol{\lambda}^*$) is:

$$p(\boldsymbol{Y}^*|\boldsymbol{\lambda}^*, \boldsymbol{Z}^*) \propto \\ \prod_{m^*=1}^{M^*} (\lambda^*_{m^*})^{\frac{Q^*}{2}} exp\{-\frac{\lambda^*_{m^*}}{2}(\hat{\boldsymbol{b}}_{m^*} - \boldsymbol{b}_{m^*})^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}(\hat{\boldsymbol{b}}_{m^*} - \boldsymbol{b}_{m^*})\}\times \\ \prod_{m^*=1}^{M^*} (\lambda^*_{m^*})^{\frac{N-Q^*}{2}} exp\{-\frac{1}{2}\lambda^*_{m^*}(\boldsymbol{y}^*_{m^*})^T(\boldsymbol{I} - \boldsymbol{\Phi}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T)\boldsymbol{y}^*_{m^*}\} \tag{4.29}$$

where $Q^* = Q + D$.

Similarly as before, the second terms on the right hand side of Equation (4.29) can be used to update the Gamma probability density distribution in Equation (4.27). In particular the $m^* - th$ updated shape ($shp^{*\prime}_{m^*}$) and rate ($rt^{*\prime}_{m^*}$) parameters result:

$$shp^{*\prime}_{m^*} = shp^*_{m^*} + \frac{N - Q^*}{2} \tag{4.30}$$

$$rt^{*\prime}_{m^*} = rt^*_{m^*} + \frac{1}{2}(\boldsymbol{y}^*_{m^*})^T(\boldsymbol{I} - \boldsymbol{\Phi}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T)\boldsymbol{y}^*_{m^*} \tag{4.31}$$

so that the prior probability density distribution for the whole set of precision parameters ($\boldsymbol{\lambda}^*$) can be defined as:

$$p(\boldsymbol{\lambda}^*|\boldsymbol{shp}^*, \boldsymbol{rt}^*) \propto \prod_{m^*=1}^{M^*} \mathcal{G}(\lambda^*_{m^*}|shp^*_{m^*}, rt^*_{m^*}) \tag{4.32}$$

where the vectors $\boldsymbol{shp}^*$ and $\boldsymbol{rt}^*$ respectively contain the parameters $shp^*_{m^*}$ and $rt^*_{m^*}$ relative to all the $\boldsymbol{y}^{(r)*}_{m^*}$. In the case that equal precision parameters are assumed for all the observed experimental periods, Equation (4.32) becomes analogous to Equation (4.18).

The matrix $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ is diagonal allowing to rearrange the first term on the right hand side of Equation (4.29) basis wise, and to perform GPR. Even more it is convenient to represent the likelihood equation as factorisation of two independent terms, one depicting $f(\cdot)$ ($p(\hat{\boldsymbol{W}}^*|\boldsymbol{\lambda}^*, \boldsymbol{Z}^*)$) and one representing $\Delta(\cdot)$

$(p(\hat{\boldsymbol{U}}|\hat{\boldsymbol{V}}^*, \boldsymbol{\lambda}^*, \boldsymbol{Z}^*))$:

$$p(\boldsymbol{Y}^*|\boldsymbol{\lambda}^*, \boldsymbol{Z}^*) \propto p(\hat{\boldsymbol{W}}^*, \hat{\boldsymbol{U}}|\boldsymbol{\lambda}^*, \boldsymbol{Z}^*) \propto p(\hat{\boldsymbol{W}}^*|\boldsymbol{\lambda}^*, \boldsymbol{Z}^*) \times p(\hat{\boldsymbol{U}}|\hat{\boldsymbol{V}}^*, \boldsymbol{\lambda}^*, \boldsymbol{Z}^*) \quad (4.33)$$

where:

$$p(\hat{\boldsymbol{W}}^*|\boldsymbol{\lambda}^*, \boldsymbol{W}^*, \boldsymbol{Z}^*) \propto \prod_{q=1}^{Q} |\boldsymbol{C}_q^*|^{\frac{1}{2}} exp\{-\frac{1}{2}(\hat{\boldsymbol{w}}_q^* - \boldsymbol{w}_q^*)^T \boldsymbol{C}_q^* (\hat{\boldsymbol{w}}_q^* - \boldsymbol{w}_q^*)\} \quad (4.34)$$

$$p(\hat{\boldsymbol{U}}|\boldsymbol{\lambda}^*, \hat{\boldsymbol{V}}^*, \boldsymbol{U}) \propto \prod_{d=1}^{D} |\boldsymbol{G}_d|^{\frac{1}{2}} exp\{-\frac{1}{2}(\hat{\boldsymbol{u}}_d - \boldsymbol{u}_d)^T \boldsymbol{G}_d (\hat{\boldsymbol{u}}_d - \boldsymbol{u}_d)\} \quad (4.35)$$

$\boldsymbol{C}_q^* = \text{diag}(\lambda_{m^*}^* \boldsymbol{k}_q{}^T \boldsymbol{k}_q;\ m^* = 1, ..., M^*)$ and $\boldsymbol{G}_d = \text{diag}(\lambda_{m^*}^* \boldsymbol{h}_d{}^T \boldsymbol{h}_d;\ m^* = 1, ..., M^*)$.

The formulation continues by introducing GP prior probability density distributions on the variables $\boldsymbol{w}_q^*$ and $\boldsymbol{u}_d$, and by deriving the relative marginal probability density distributions for $\hat{\boldsymbol{w}}_q^*$ and $\hat{\boldsymbol{u}}_d$ in analogous ways as previously explained in the training phase. Since $p(\hat{\boldsymbol{W}}^*|\boldsymbol{\lambda}^*, \boldsymbol{Z}^*, \boldsymbol{W}^*)$ and $p(\hat{\boldsymbol{U}}|\boldsymbol{\lambda}^*, \hat{\boldsymbol{V}}^*, \boldsymbol{U})$ are independent, and for clarity reasons, they are explained separately. The former is used to infer the unknown parameters in the vector $\boldsymbol{\theta}^*$ while the latter can help in identifying correlation patterns between variable boundary conditions and *Difference Vectors*, thus providing information about possible upgrades to the BES model.

The GP prior probability density distributions assumed for $\boldsymbol{w}_q^*$ are the same adopted in the *Training* phase, and the previously inferred values of $\boldsymbol{\vartheta}$ are used as hp. Hence, the marginal probability distribution of $\hat{\boldsymbol{W}}^*$, with respect to $\boldsymbol{W}^*$:

$$p(\hat{\boldsymbol{W}}^*|\boldsymbol{Z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\vartheta}) \propto \prod_{q=1}^{Q} |(\boldsymbol{C}_q^*)^{-1} + \varrho_q(\boldsymbol{Z}^*, \boldsymbol{Z}^*)|^{-\frac{1}{2}} exp\{-\frac{1}{2}(\hat{\boldsymbol{w}}_q^*)^T [(\boldsymbol{C}_q^*)^{-1} + \varrho_q(\boldsymbol{Z}^*, \boldsymbol{Z}^*)]^{-1} \hat{\boldsymbol{w}}_q^*\} \quad (4.36)$$

In order to estimate the calibration parameters according to the original BES model behaviour, it is necessary to condition the probability density distribution depicted by Equation (4.36) respect to the model $\mathcal{M}_k$, which probabilistically links BES model inputs to BES model outputs. In particular $p(\hat{\boldsymbol{W}}^*|\boldsymbol{Z}^*, \boldsymbol{\lambda}^*, \mathcal{M}_k)$ is the predictive distribution of $\mathcal{M}_k$ at the input vectors $\boldsymbol{z}_{m^*}^*$, which can be expressed as factorisation of Q terms relatively to the Q target vectors $\hat{\boldsymbol{w}}_q^*$. The

joint distribution for the vectors $\hat{\boldsymbol{w}}_q$ and $\hat{\boldsymbol{w}}_q^*$ results:

$$\begin{bmatrix} \hat{\boldsymbol{w}}_q \\ \hat{\boldsymbol{w}}_q^* \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} \varrho(\boldsymbol{Z},\boldsymbol{Z}) + \boldsymbol{C}_q^{-1} & \varrho(\boldsymbol{Z},\boldsymbol{Z}^*) \\ \varrho(\boldsymbol{Z}^*,\boldsymbol{Z}) & \varrho(\boldsymbol{Z}^*,\boldsymbol{Z}^*) + (\boldsymbol{C}_q^*)^{-1} \end{bmatrix}\right) \tag{4.37}$$

Then, the conditional distribution of $\hat{\boldsymbol{w}}_q^*$ with respect to $\hat{\boldsymbol{w}}_q$ results:

$$\hat{\boldsymbol{w}}_q^* | \hat{\boldsymbol{w}}_q, \boldsymbol{\lambda}^*, \boldsymbol{Z}, \boldsymbol{Z}^*, \boldsymbol{\vartheta}_q, \lambda \sim N(\boldsymbol{w}_q', \boldsymbol{\Sigma}_q')$$

where the conditional mean ($\boldsymbol{w}_q'$) and conditional covariance matrix ($\boldsymbol{\Sigma}_q'$) are defiend as follows:

$$\begin{aligned} \boldsymbol{w}_q' &= \varrho_q(\boldsymbol{Z},\boldsymbol{Z}^*)^T(\varrho_q(\boldsymbol{Z},\boldsymbol{Z}) + \boldsymbol{C}_q^{-1})^{-1}\hat{\boldsymbol{w}}_q \\ \boldsymbol{\Sigma}_q' &= \varrho_q(\boldsymbol{Z}^*,\boldsymbol{Z}^*) - \varrho_q(\boldsymbol{Z}^*,\boldsymbol{Z})[(\varrho_q(\boldsymbol{Z},\boldsymbol{Z}) + \boldsymbol{C}_q^{-1}]^{-1}\varrho_q(\boldsymbol{Z},\boldsymbol{Z}^*) \end{aligned}$$

and, consequently, it is possible to define $p(\hat{\boldsymbol{W}}^*|\boldsymbol{Z}^*, \boldsymbol{\lambda}^*, \mathcal{M}_k)$ as:

$$\begin{aligned} &p(\hat{\boldsymbol{W}}^*|\boldsymbol{Z}^*, \boldsymbol{\lambda}^*, \mathcal{M}_k) \propto \\ &\prod_{q=1}^{Q}(2\pi)^{-\frac{M^*}{2}}|\boldsymbol{\Sigma}_q'|^{-\frac{1}{2}}exp\{-\frac{1}{2}(\hat{\boldsymbol{w}}_q^* - \boldsymbol{w}_q')^T(\boldsymbol{\Sigma}_q')^{-1}(\hat{\boldsymbol{w}}_q^* - \boldsymbol{w}_q')\} \end{aligned} \tag{4.38}$$

Equation (4.38) is then multiplied by the prior probability density distributions assumed for the calibration parameters ($p(\boldsymbol{\theta}^*)$), and for the precision parameters (Equation (4.32)), in order to find their joint posterior probability density distribution according to the Bayes' Theorem:

$$\begin{aligned} &p(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*|\hat{\boldsymbol{V}}^*, \hat{\boldsymbol{W}}^*, \mathcal{M}_k) \propto \\ &p(\hat{\boldsymbol{W}}^*|\boldsymbol{Z}^*, \boldsymbol{\lambda}^*, \mathcal{M}_k) \times p(\boldsymbol{\theta}^*) \times p(\boldsymbol{\lambda}^*|\boldsymbol{shp}^*, \boldsymbol{rt}^*) \end{aligned} \tag{4.39}$$

By sampling calibration parameter vectors from the probability density distribution in Equation (4.39) it is possible to estimate the MAP and confidence intervals for such variables.

Different GP prior probability density distributions are, instead, assumed for the variables $\boldsymbol{u}_d$:

$$\boldsymbol{u}_d \sim GP(0, \varphi_d(\cdot)) \tag{4.40}$$

the covariance functions $\varphi_d(\cdot)$ have their own hp represented by the vectors $\boldsymbol{\nu}_d$. Their complete set is indicated by the vector $\boldsymbol{\nu}$. The probability of $\hat{\boldsymbol{U}}$ conditional

only on $\hat{\boldsymbol{V}}^*$, $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}$, results:

$$\begin{aligned} &p(\hat{\boldsymbol{U}}|\boldsymbol{\lambda}^*, \boldsymbol{\nu}, \hat{\boldsymbol{V}}^*) \propto \\ &\prod_{d=1}^{D}(2\pi)^{-\frac{M^*}{2}}|(\boldsymbol{G}_d^*)^{-1} + \varphi_d(\hat{\boldsymbol{V}}^*, \hat{\boldsymbol{V}}^*)|^{-\frac{1}{2}} \\ &exp\{-\frac{1}{2}(\hat{\boldsymbol{u}}_d)^T[(\boldsymbol{G}_d^*)^{-1} + \varphi_d(\hat{\boldsymbol{V}}^*, \hat{\boldsymbol{V}}^*)]^{-1}\hat{\boldsymbol{u}}_d\} \end{aligned} \tag{4.41}$$

Equation (4.41) multiplied by the prior probability density distributions of $\boldsymbol{\nu}$ ($p(\boldsymbol{\nu})$) and of $\boldsymbol{\lambda}^*$ (Equation 4.32) gives their joint posterior density distribution which can be used to make inference about these parameters:

$$p(\boldsymbol{\nu}, \boldsymbol{\lambda}^*|\hat{\boldsymbol{U}}, \boldsymbol{V}^*) \propto p(\hat{\boldsymbol{U}}|\boldsymbol{\lambda}^*, \boldsymbol{\nu}, \hat{\boldsymbol{V}}^*) \times p(\boldsymbol{\nu}) \times p(\boldsymbol{\lambda}^*|\boldsymbol{shp}^*, \boldsymbol{rt}^*) \tag{4.42}$$

Both Equations (4.39) and (4.42) are integrated with Annealed Importance Sampling (AIS) (B.3) in order to infer marginal posterior probability density distributions for $\boldsymbol{\theta}^*$, $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}$. AIS also gives the possibility to calcualate the marginal likelihood of the *Calibration Model* and *Difference Model*, thus allowing to use Bayes Factors for *Model Selection* as will be explained in Section 4.3. For completeness, the full joint posterior distribution is given in Equation (4.43).

$$\begin{aligned} &p(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}|\hat{\boldsymbol{W}}^*, \hat{\boldsymbol{U}}, \mathcal{M}_k, \hat{\boldsymbol{V}}^*) \propto \\ &\prod_{q=1}^{Q}(2\pi)^{-\frac{M^*}{2}}|\boldsymbol{\Sigma}_q'|^{-\frac{1}{2}}exp\{(\hat{\boldsymbol{w}}_q^* - \boldsymbol{w}_q')^T(\boldsymbol{\Sigma}_q')^{-1}(\hat{\boldsymbol{w}}_q^* - \boldsymbol{w}_q')\}\times \\ &\prod_{d=1}^{D}(2\pi)^{-\frac{M^*}{2}}|\boldsymbol{G}_d^{-1} + \varphi_d(\hat{\boldsymbol{V}}^*, \hat{\boldsymbol{V}}^*)|^{-\frac{1}{2}}exp\{-\frac{1}{2}(\hat{\boldsymbol{u}}_d)^T[\boldsymbol{G}_d^{-1} + \varphi_d(\hat{\boldsymbol{V}}^*, \hat{\boldsymbol{V}}^*)]^{-1}\hat{\boldsymbol{u}}_d\}\times \\ &p(\boldsymbol{\lambda}^*|\boldsymbol{shp}^*, \boldsymbol{rt}^*) \times p(\boldsymbol{\theta}^*) \times p(\boldsymbol{\nu}) \end{aligned} \tag{4.43}$$

Considering multiple target variables and datasets

In case R measured variables are available as calibration targets, it is possible to include them in the analysis without changing the probabilistic modelling framework by adopting the following approach. R probabilistic models are built for all the R targets according to the mathematical procedures depicted in the previous *Training* and *Identification* sections. Then, an overall probabilistic model is derived by modelling the covariance between the different target variables through adequate covariance functions. In this way, Equations (4.22), (4.36) and (4.41) keep the same forms, and only the vectors and matrices involved change their

structures.

In Equation (4.22), $\hat{\boldsymbol{w}}_q$, $\boldsymbol{C}_q$ and $\varrho_q(\boldsymbol{Z},\boldsymbol{Z})$ become respectively a MR vector and $MR \times MR$ matrices so defined:

$$\hat{\boldsymbol{w}}_q = [(\hat{\boldsymbol{w}}_q{}^{(1)})^T, ..., (\hat{\boldsymbol{w}}_q^{(r)})^T, ..., (\hat{\boldsymbol{w}}_q{}^{(r)})^T]^T$$

$$\boldsymbol{C}_q = \text{diag}(\boldsymbol{C}_q^{(r)};\ r = 1, ..., R)$$

where $\boldsymbol{C}_q^{(r)} = \text{diag}(\lambda^{(r)} \boldsymbol{k}_q{}^T \boldsymbol{k}_q;\ m = 1, ..., M)$;

$$\varrho_q(\boldsymbol{Z},\boldsymbol{Z}) = \begin{bmatrix} \varrho_q^{(1,1)}(\boldsymbol{Z},\boldsymbol{Z}) & \dots & \varrho_q^{(1,r)}(\boldsymbol{Z},\boldsymbol{Z}) & \dots & \varrho_q^{(1,R)}(\boldsymbol{Z},\boldsymbol{Z}) \\ \vdots & \ddots & \vdots & & \vdots \\ \varrho_q^{(r,1)}(\boldsymbol{Z},\boldsymbol{Z}) & \dots & \varrho_q^{(r,r)}(\boldsymbol{Z},\boldsymbol{Z}) & \dots & \varrho_q^{(r,R)}(\boldsymbol{Z},\boldsymbol{Z}) \\ \vdots & & \vdots & \ddots & \vdots \\ \varrho_q^{(R,1)}(\boldsymbol{Z},\boldsymbol{Z}) & \dots & \varrho_q^{(R,r)}(\boldsymbol{Z},\boldsymbol{Z}) & \dots & \varrho_q^{(R,R)}(\boldsymbol{Z},\boldsymbol{Z}) \end{bmatrix}$$

where $\varrho_q^{(l,k)}(\cdot)$ indicate the covariance functions between the $l-th$ and $k-th$ model outputs.

In Equations (4.36), $\hat{\boldsymbol{w}}_q^*$, $\hat{\boldsymbol{u}}_d$, $\boldsymbol{C}_q^*$, $\boldsymbol{G}_d$ become respectively M^*R vectors and $M^*R \times M^*R$ matrices with the following structures:

$$\hat{\boldsymbol{w}}_q^* = [(\hat{\boldsymbol{w}}_q^{(1)*})^T, ..., (\hat{\boldsymbol{w}}_q^{r*})^T, ..., (\hat{\boldsymbol{w}}_q{}^{(r)*})^T]^T$$

$$\hat{\boldsymbol{u}}_d = [(\hat{\boldsymbol{u}}_d^{(1)*})^T, ..., (\hat{\boldsymbol{u}}_d^{(r)})^T, ..., (\hat{\boldsymbol{u}}_d{}^{(r)*})^T]^T$$

$$\boldsymbol{C}_q^* = \text{diag}(\boldsymbol{C}_q^{(r)*};\ r = 1, ..., R)$$

$$\boldsymbol{G}_d = \text{diag}(\boldsymbol{G}_d^{(r)*};\ r = 1, ..., R)$$

where $\boldsymbol{C}_q^{(r)*} = \text{diag}(\lambda_{m^*}^{(r)*} \boldsymbol{k}_q{}^T \boldsymbol{k}_q;\ m^* = 1, ..., M^*)$ and $\boldsymbol{G}_q^{(r)} = \text{diag}(\lambda_{m^*}^{(r)*} \boldsymbol{h}_d{}^T \boldsymbol{h}_d;\ m^* = 1, ..., M^*)$.

The covariance matrices $\varrho_q(\boldsymbol{Z}^*,\boldsymbol{Z}^*)$ and cross-covariance matrices $\varrho_q(\boldsymbol{Z}^*,\boldsymbol{Z})$, become respectively $M^*R \times M^*R$ and $M^*R \times MR$ matrices defined as:

$$\varrho_q(\boldsymbol{Z}^*,\boldsymbol{Z}^*) = \begin{bmatrix} \varrho_q^{(1,1)}(\boldsymbol{Z}^*,\boldsymbol{Z}^*) & \dots & \varrho_q^{(1,r)}(\boldsymbol{Z}^*,\boldsymbol{Z}^*) & \dots & \varrho_q^{(1,R)}(\boldsymbol{Z}^*,\boldsymbol{Z}^*) \\ \vdots & \ddots & \vdots & & \vdots \\ \varrho_q^{(r,1)}(\boldsymbol{Z}^*,\boldsymbol{Z}^*) & \dots & \varrho_q^{(r,r)}(\boldsymbol{Z}^*,\boldsymbol{Z}^*) & \dots & \varrho_q^{(r,R)}(\boldsymbol{Z}^*,\boldsymbol{Z}^*) \\ \vdots & & \vdots & \ddots & \vdots \\ \varrho_q^{(R,1)}(\boldsymbol{Z}^*,\boldsymbol{Z}^*) & \dots & \varrho_q^{(r,R)}(\boldsymbol{Z}^*,\boldsymbol{Z}^*) & \dots & \varrho_q^{(R,R)}(\boldsymbol{Z}^*,\boldsymbol{Z}^*) \end{bmatrix}$$

$$\varrho_q(\boldsymbol{Z}^*, \boldsymbol{Z}) = \varrho_q(\boldsymbol{Z}, \boldsymbol{Z}^*)^T = \begin{bmatrix} \varrho_q^{(1,1)}(\boldsymbol{Z}^*, \boldsymbol{Z}) & \dots & \varrho_q^{(1,r)}(\boldsymbol{Z}^*, \boldsymbol{Z}) & \dots & \varrho_q^{(1,R)}(\boldsymbol{Z}^*, \boldsymbol{Z}) \\ \vdots & \ddots & \vdots & & \vdots \\ \varrho_q^{(r,1)}(\boldsymbol{Z}^*, \boldsymbol{Z}) & \dots & \varrho_q^{(r,r)}(\boldsymbol{Z}^*, \boldsymbol{Z}) & \dots & \varrho_q^{(r,R)}(\boldsymbol{Z}^*, \boldsymbol{Z}) \\ \vdots & & \vdots & \ddots & \vdots \\ \varrho_q^{(R,1)}(\boldsymbol{Z}^*, \boldsymbol{Z}) & \dots & \varrho_q^{(r,R)}(\boldsymbol{Z}^*, \boldsymbol{Z}) & \dots & \varrho_q^{(R,R)}(\boldsymbol{Z}^*, \boldsymbol{Z}) \end{bmatrix}$$

By following the work described in [13] is is possible to define rigorously the functions $\varrho_q^{(l,k)}(\cdot)$. In this study it is shown how it is possible to achieve the following closed form for $\varrho_q^{(l,k)}(\cdot)$, by building GPs as convolutions of white noise and Gaussian smoothing kernels:

$$\begin{aligned} &\varrho_q^{(l,k)}(\boldsymbol{z}_i, \boldsymbol{z}_j) = \\ &\tau_q^{(l)}\tau_q^{(k)}(2\pi)^{\frac{P}{2}}|\boldsymbol{A}_q^{(l)} + \boldsymbol{A}_q^{(k)}|^{-\frac{1}{2}} exp\{-\frac{1}{2}(|\boldsymbol{z}_i - \boldsymbol{z}_j| - \boldsymbol{a}_q^{(l)} - \boldsymbol{a}_q^{(k)})^T \boldsymbol{B}^{(l,k)} \\ &(|\boldsymbol{z}_i - \boldsymbol{z}_j| - \boldsymbol{a}_q^{(l)} - \boldsymbol{a}_q^{(k)})\} \end{aligned} \tag{4.44}$$

where: $\boldsymbol{B}^{(l,k)} = \boldsymbol{A}_q^{(l)}(\boldsymbol{A}_q^{(l)} + \boldsymbol{A}_q^{(k)})^{-1}\boldsymbol{A}_q^{(k)}$. The assumption made in defining the matrices $\boldsymbol{A}_q^{(l)}$ and $\boldsymbol{A}_q^{(k)}$, the vectors $\boldsymbol{a}_q^{(l)}$ and $\boldsymbol{a}_q^{(k)}$ and the parameters $\tau_q^{(l)}$, are explained in the next section.

In Equation (4.41) $\varphi_d(\hat{\boldsymbol{V}}^*, \hat{\boldsymbol{V}}^*)$ becomes a $M^*R \times M^*R$ matrix built as follows:

$$\varphi_d(\hat{\boldsymbol{V}}^*, \hat{\boldsymbol{V}}^*) = \begin{bmatrix} \varphi_d^{(1,1)}(\hat{\boldsymbol{V}}^*, \hat{\boldsymbol{V}}^*) & \dots & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & & \vdots \\ \boldsymbol{0} & \dots & \varphi_d^{(r,r)}(\hat{\boldsymbol{V}}^*, \hat{\boldsymbol{V}}^*) & \dots & \boldsymbol{0} \\ \vdots & & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \dots & \boldsymbol{0} & \dots & \varphi_d^{(R,R)}(\hat{\boldsymbol{V}}^*, \hat{\boldsymbol{V}}^*) \end{bmatrix}$$

where $\varphi_d^{(r,r)}(\cdot)$ are covariance functions used in building the *Difference Model* of the $r-th$ model output.

Therefore the *Difference Vectors* relative to different model outputs have been considered independent. This particular choice is due to the need to employ covariance functions for which a closed form for $\varrho_q^{(l,k)}(\cdot)$ is not available. Covariance functions having Gaussian form as the function depicted by Equation (4.44) are suitable for representing smooth process, (indeed the functions drawn from a GP defined by this kind of covariance function will be infinitely differentiable), while they have poor performances otherwise. In order to adequately analyse the *Difference Vectors*, very flexible covariance functions able to represent a wide range of variation modes must be employed since such time series often present discon-

tinuities like spikes and sharp corners. More details about the used covariance functions are given in the next section.

Multiple independent datasets can be consider by factorising the corresponding likelihood equations.

Covariance functions and hyper parameters

Depending on the purpose of the GPR model, and number of target variables to consider, three different covariance functions have been employed.

For *Training Model* and *Calibration Model* considering only one target variable, Square Exponential (SE) functions ([93]) were used as $\varrho_q(\cdot)$:

$$\varrho_q(\boldsymbol{z}_i, \boldsymbol{z}_j) = \tau_q \; exp\{-\frac{1}{2}\sum_{p=1}^{P} \beta_{q,p}|z_{i,p} - z_{j,p}|^2\} \tag{4.45}$$

The functions depicted in Equation 4.44 were used when it was necessary to account for multiple target variables in *Training* and *Identification*. In particular the matrices $\boldsymbol{A}_q^{(j)}$ (where $j = r, l$ or k) have been defined as:

$$\boldsymbol{A}_q^{(j)} = \text{diag}(\beta_{p,q}^{(j)};\; p = 1, ..., P)$$

and the elements of each $\boldsymbol{a}_q^{(j)}$ have been considered all equal. Thus, the covariance function of the $r-th$ model output results:

$$\varrho_q^{(r,r)}(\boldsymbol{z}_i, \boldsymbol{z}_j) = \frac{\pi^{\frac{p}{2}}}{\prod_{p=1}^{P} \beta_{p,q}^{(r)}}(\tau_q^{(r)})^2 exp\{-\frac{1}{4}\sum_{p=1}^{P} \beta_{p,q}^{(r)}|z_{i,p} - z_{j,p}|^2\} \tag{4.46}$$

and the cross-covariance function between the $l-th$ and $k-th$ outputs results:

$$\varrho_q^{(l,k)}(\boldsymbol{z}_i, \boldsymbol{z}_j) = \frac{\pi^{\frac{p}{2}}}{\prod_{p=1}^{P}(\beta_{p,q}^{(l)} + \beta_{p,q}^{(k)})}\tau_q^{(l)}\tau_q^{(k)} exp\{-\frac{1}{2}\sum_{p=1}^{P} \gamma_{p,q}^{(l,k)}(|z_{i,p} - z_{j,p}| - a_q^{(l,k)})^2\} \tag{4.47}$$

where $\gamma_{p,q}^{(l,k)} = \frac{\beta_{p,q}^{(l)}\beta_{p,q}^{(k)}}{\beta_{p,q}^{(l)}+\beta_{p,q}^{(k)}}$, and $a_q^{(l,k)} = a_q^{(l)} - a_q^{(k)}$.

$\beta_{p,q}$ and $\beta_{p,q}^{(j)}$ parameters are the rate of change of the GP functions relative to a certain parameter. For high values of these hp the output of the GPR model will show relevant variations also for small changes of the corresponding parameter. τ_q and $\tau_q^{(j)}$ parameters are the variances or amplitudes of the GP functions. For high values of these hp the GP functions will show high variation around its mean. The variables $a_q^{(l,k)}$ underline constant offsets between the trends of the

$l-th$ and $k-th$ model outputs or target variables.

In building the *Difference Model*, Spectral Mixture (SM) kernels [124] were adopted for their high capability to discover patterns in data:

$$\varphi_d^{(r,r)}(\boldsymbol{v}_i, \boldsymbol{v}_j) = \sum_{i=1}^{I} \tau_{i,d}^{(r)} \prod_{s=1}^{S} \prod_{c=1}^{C^{(s)}} exp\{-2\pi\beta_{i,c,d}^{(r,s)}(v_{i,c}^{(s)} - v_{j,c}^{(s)})^2\} cos(2\pi\phi_{i,c,d}^{(r,s)}(v_{i,c}^{(s)} - v_{j,c}^{(s)})^2) + \sigma_d^{(r)} \tag{4.48}$$

These kinds of covariance functions are flexible being able to represent periodic and non periodic variations of the process being modelled. In particular their flexibility can be increased by augmenting the number of functions being summarised (I), so that by superimposing more functions it is also possible to approximate, sharp edges and spikes present in the target data.

Similarly as before, $\beta_{i,c,d}^{(r,s)}$, indicate the rate of change of the functions being distributed according to the relative GP, respect to certain parameter. $\tau_{i,d}^{(r)}$, besides amplitudes and variances, can be also seen as the weights of the $i-th$ covariance function in the summation. $\phi_{i,c,d}^{(r,s)}$ parameters highlight periodic variation in the data linked to certain model inputs. They can be interpreted as frequencies, and for values of these hp tending to zero there will be no periodic variation relative to a certain input since the corresponding period will tend to infinity and vice versa.

It also important to notice that Equation (4.48) implies that different hp have been assumed for each $\varphi^{(r,r)}(\cdot)$. This make particularly reasonable the form chosen for $\varphi_d(\hat{\boldsymbol{V}}^*, \hat{\boldsymbol{V}}^*)$ since there are no common hp between two different $\varphi_d^{(r,r)}(\cdot)$, and thus the benefit deriving from considering the correlation between different target variables in their estimation is negligible.

Prior distributions

According to the Bayesian paradigm all the parameters involved in the calculations need prior probability density distributions reflecting the prior uncertainties and beliefs of the analyst.

In particular, the prior probability density distributions for the calibration parameters should be set up according to given specifications, information from literature and outcomes of a previously performed sensitivity analysis. Therefore, they are problem dependent and it is a modeller task to choose them suitably, in order to simplify the calculations without producing misleading results, for

example by assuming excessively informative probability density distributions. In the case of a lack of adequate information the most general choice would be to assume Uniform probability density distributions with suitable boundaries. The following discussion will focus on the prior probability density distributions for the hp and on the rationales behind them.

The prior probability density distributions for hp (Table 4.1) and precision parameters were selected in order to facilitate the calculations and to perform Automatic Relevance Determination (ARD) ([79]). ARD is especially useful in treating a large number of hp and consists in encouraging not influential hp to assume values which does not affect the entries of the GP covariance matrix. For example, in this study, Exponential probability density distributions have been used as prior probability density distributions for rate (β) and frequency (ϕ) parameters, so that β and ϕ parameters relative to weak model inputs will assume values close to zero.

Table 4.1: Hyper parameters prior probability density distributions.

COVARIANCE FUNCTION	PARAMETER	PRIOR
$\varrho_q(\cdot)$ and	$\beta_{p,q}^{(j)}$ and $\beta_{q,p}$	$\mathcal{E}xp(1)$
	τ_q	$\mathcal{G}(shp, \beta)$
$\varrho^{(r,r)}(\cdot)$	$\beta_{p,q}^{(r)}$	$\mathcal{E}xp(1)$
	$\tau_q^{(r)}$	$\mathcal{G}(shp, \beta)$
$\varrho^{(l,k)}(\cdot)$	$\beta_{p,q}^{(l)}$ or $\beta_{p,q}^{(k)}$	$\mathcal{E}xp(1)$
	$\tau_q^{(l)}$ or $\tau_q^{(k)}$	$\mathcal{G}(shp, \beta)$
$\varphi_d^{(r,r)}$	$\tau_{i,d}^{(r)}$	$\mathcal{E}xp(5)$
	$\beta_{i,c,d}^{(r,s)}$	$\mathcal{E}xp(5)$
	$\phi_{i,c,d}^{(r,s)}$	$\mathcal{E}xp(5)$
	$\sigma_d^{(r)}$	$\mathcal{E}xp(5)$
	$a^{(l,k)}$	$\mathcal{N}(\mathrm{E}(\boldsymbol{y}^{(l)} - \boldsymbol{y}^{(k)});\ \mathrm{V}(\boldsymbol{y}^{(l)} - \boldsymbol{y}^{(k)}))$

For the τ_q, $\tau_q^{(r)}$, $\tau_q^{(l)}$ and $\tau_q^{(k)}$ were used Gamma probability density distributions. The shape and rate parameters of these probability density distributions were set in order to have their means close to the empirical variances of the target variables of the GPR model.

By observing the prior distributions for the precisions of both simulation and observations it is possible to notice that their parameters are functions of the quality of the approximation achieved through the basis expansion. In particular, through some algebraic manipulations, it is possible to express the rate parameters as functions of the SSE due to the approximation of simulations or

field observations. For example considering Equation 4.18:

$$rt' = rt + \frac{1}{2}\sum_{m=1}^{M}(\boldsymbol{y}_m - \boldsymbol{K}\hat{\boldsymbol{w}}_m)^T(\boldsymbol{y}_m - \boldsymbol{K}\hat{\boldsymbol{w}}_m) \tag{4.49}$$

rt' is the sum of a constant (rt) and half the SSE due the least square fitting of the defined basis vectors. shp' (Equation (4.19)) contains the degree of freedom of the errors ($M(N-Q)$). The mean of $\mathcal{G}(\lambda|shp', shp')$ is defined as:

$$\frac{shp'}{rt'} = \frac{shp + \frac{M(N-Q)}{2}}{\beta + \frac{SSE}{2}} \tag{4.50}$$

which, for $shp = 0$ and $rt = 0$ is equal to the inverse of the variance of the residual from least square fit, hence their precision. Similar results can be derived for the precision parameters of the observations.

Therefore, it has been decided as prior setting to assume shp equal to zero and to use rt to include in the analysis the uncertainties due to *Observation Errors*, *Residual Variability* and *Representation Uncertainty*. Thus it is sensible to set rt close to zero during the *Training* phase since the simulation outputs are free from *Observation Errors* and *Residual Variability*, and to about half the SSE due to possible measurement errors in the *Identification* phase. Usually experimental specifications give enough information to estimate these uncertainties. For example by having observations of the internal temperature of a room at different locations it is possible to roughly estimate the magnitude of *Observation Errors*. Otherwise, if it is not possible to infer a reasonable estimate for rt from the given information, it can be treated as a free parameter, by attributing to it a prior probability density distribution reflecting the beliefs about the measurement accuracy, and inferring it during the analysis.

The prior probability density distributions for the hp of $\varphi_d(\cdot)$ are discussed in the next section.

Difference Analysis

Eventual model upgrades are identified according to the values assumed by the hp of $\varphi_d(\cdot)$ after training the *Difference Model*. The interpretation of such results is called *Difference Analysis*.

The training of the *Difference Model*, involve a large number of hp. In particular, their number can be very large depending on the number of components of the summation appearing in Equation (4.48) and on the number of considered

variable boundary conditions. For the $a^{(l,k)}$ were chosen prior normal probability density distributions reflecting the mean and variance of the offsets between two simulation outputs. For the other hp, it was decided to implement a strong ARD, in order push towards zero as more hp as possible and highlight only the boundary conditions that showed significant correlations with the *Difference Vectors*. Therefore for $\tau_{i,d}^{(r)}$, $\beta_{i,c,d}^{(r,s)}$, $\phi_{i,c,d}^{(r,s)}$ and $\sigma_d^{(r)}$ were adopted Exponential probability density distributions with rate parameter equal to 5.

According to the values assumed by the hp of $\varphi_d(\cdot)$, different causes of discrepancy between model outputs and target data can be identified. Thus, $\beta_{i,c,d}^{(r,s)}$ and $\phi_{i,c,d}^{(r,s)}$ having estimates sensibly different from zero, indicate respectively dynamic and periodic trends in the field observations correlated to certain boundary conditions. In this case the $\tau_{i,d}^{(r)}$ corresponding to the $\beta_{i,c,d}^{(r,s)}$ and $\phi_{i,c,d}^{(r,s)}$ higher than zero, will be higher that zero as well. All $\tau_{i,d}^{(r)}$ parameters close to zero and the additive variances $\sigma_d^{(r)}$ different from zero, highlight the possibility that correlated noise is present in the measurement. Such correlated noise will be normally distributed, with zero mean and covariance matrix $\sum_{d=1}^{D} \sigma_d \boldsymbol{h}_d \boldsymbol{h}_d{}^T$. It is important to say that correlation does not imply dependency, hence it is not possible to certainly identify discrepancy causes, but determining them according to the information derived from *Difference Analysis* increases the possibilities to chose effective model upgrades.

The analysis of these results is performed graphically, through scree plots, depending on the typology of the hp (i.e. β, ϕ, τ or σ parameters). For example in Figure 4.2 are shown the result from the *Difference Analysis* performed for model $\mathcal{M}_1$ relative to Twin House N2 (Chapter 6). In this case relative humidity (***Rh***), wind direction (***Wd***), convective heat gains in *kitchen* ($kitchen_{\boldsymbol{C/R}}$) and ambient temperature (***Te***) were highlighted. However, ***Rh***, which is used in ESP-r to calculate the longwave radiation towards the sky vault, was neglected since the experiment was performed in the ground floor of the building, which had an attic, making this heat flux negligible. This information led to the implementation of an airflow network in the model.

4.2.3 Bayesian vs. quasi-Bayesian

As stated at the beginning of the section the adopted calibration approach is not fully Bayesian. In particular the problem is decomposed in more easy to solve computationally cheaper sub-problems involving separate fits of three different models (*Training Model*, *Calibration Model* and *Difference Model*). A full Bayesian treatment would, instead, involve performing inference of the parame-

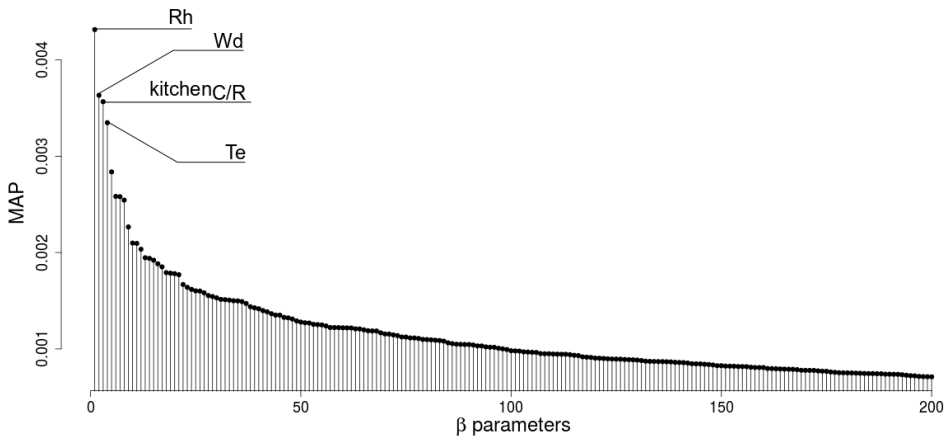


Figure 4.2: Example of analysis of β parameters.

ters λ, $\boldsymbol{\lambda}^*$, $\boldsymbol{\vartheta}$, $\boldsymbol{\theta}^*$ and $\boldsymbol{\nu}$ all at once, by considering simulation and observation data at the same time. Such an approach would often result in an hardly manageable computational load. The adoption of a quasi-Bayesian approach provides for such issues and it is advised in a series of studies ([65] and [6]) where it has been shown to provide results very similar to a fully Bayesian treatment of the problem.

The main consequence is the not full consideration of some type of uncertainties, as discussed in [65]. Regarding the particular proposed probabilistic modelling approach, it is believed that the simplifications resulting from the employment of a quasi-Bayes framework are reasonable. The λ and $\lambda^*_{m^*}$ parameters depend respectively on the simulation output and on the measured data so that their estimation can be carried out separately. The number of simulations (M) will be significantly greater than the number of observed periods (M^*), so that the loss of information caused by performing the inference of $\boldsymbol{\vartheta}$ considering only the former should be negligible. By construction, the *Calibration Model* and the *Difference Model* are independent, making it possible to treat them separately. Nonetheless they share the same precision parameters. However, in all the tackled case studies the prior probability density distribution for such parameters resulted to be very informative and practically equal estimates were returned by the two models.

Finally, fixing the $\boldsymbol{\vartheta}$ and λ during the *Identification* phase causes the incomplete consideration of the *Representation Uncertainty* in the inference of the calibration parameters, because different values of these parameters will produce slightly different outputs and relative uncertainties. Similarly fixing the param-

eters in $\boldsymbol{\nu}$ and in $\boldsymbol{\lambda}^*$ in making predictions produce a slight underestimation of the *Model Inadequacy*, *Observation Errors* and *Residual Variability*.

4.3 Model Selection

Bayes Factors are adopted as reference method in performing model comparison and selection. Bayes factors are a Bayesian approach to rank two competing scientific hypotheses deemed to explain observed real processes. In this context, the objective is to rank computer models according to their capabilities in describing the behaviour of real world systems. Thus, models $\mathcal{M}_i$ and $\mathcal{M}_j$ represent the two competing hypotheses to assess, and the one having the highest likelihood of generating a certain real measured dataset ($\boldsymbol{Y}^*$), has to be selected.

As with all Bayesian methods, Bayes Factors are based upon Bayes' Theorem and allow for consideration of prior beliefs about the possibility for a certain model to be a good representation of the observed real processes. Thus, if $p(\mathcal{M}_i)$ represents the prior probability for $\mathcal{M}_j$ generating $\boldsymbol{Y}^*$, $p(\mathcal{M}_j) = 1 - p(\mathcal{M}_i)$ will be the probability for $\mathcal{M}_j$ generating $\boldsymbol{Y}^*$. Consequently the posterior probability of a certain model $\mathcal{M}_k$ ($k = j$ *or* j) given the data ($\boldsymbol{Y}^*$) results:

$$p(\mathcal{M}_j|\boldsymbol{Y}^*) = \frac{p(\boldsymbol{Y}^*|\mathcal{M}_j)p(\mathcal{M}_j)}{p(\boldsymbol{Y}^*|\mathcal{M}_j)p(\mathcal{M}_j) + p(\boldsymbol{Y}^*|\mathcal{M}_i)p(\mathcal{M}_i)} \tag{4.51}$$

By considering the ratio between $p(\mathcal{M}_j|\boldsymbol{Y}^*)$ and $p(\mathcal{M}_i\boldsymbol{Y}^*)$:

$$\frac{p(\mathcal{M}_j|\boldsymbol{Y}^*)}{p(\mathcal{M}_i|\boldsymbol{Y}^*)} = \frac{p(\boldsymbol{Y}^*|\mathcal{M}_j)}{p(\boldsymbol{Y}^*|\mathcal{M}_i)} \frac{p(\mathcal{M}_j)}{p(\mathcal{M}_i)} \tag{4.52}$$

it is possible to define the Bayes Factor of $\mathcal{M}_j$ with respect to $\mathcal{M}_i$ ($\mathcal{B}_{j,i}$) as the ratio between posterior odds ($\frac{p(\mathcal{M}_j|\boldsymbol{Y}^*)}{p(\mathcal{M}_i|\boldsymbol{Y}^*)}$) and prior odds ($\frac{p(\mathcal{M}_j)}{p(\mathcal{M}_i)}$) of $\mathcal{M}_j$, and it corresponds to the ratio of the marginal likelihoods of the two models:

$$\mathcal{B}_{j,i} = \frac{p(\boldsymbol{Y}^*|\mathcal{M}_j)}{p(\boldsymbol{Y}^*|\mathcal{M}_i)} \tag{4.53}$$

In particular $p(\boldsymbol{Y}^*|\mathcal{M}_k)$ represents the probability that the data are given by the $k-th$ model. Thus $\mathcal{B}_{j,i}$ represents the evidence, coming from the field observations, that $\mathcal{M}_j$ is a better abstraction of the observed phenomena than $\mathcal{M}_i$, and it is independent from the prior probabilities of the two models.

For models with no free parameters $\mathcal{B}_{j,i}$ is equivalent to the Likelihood Ratios criterion. However, in the quasi-Bayesian framework adopted in this study,

$p(\boldsymbol{Y}^*|\mathcal{M}_k)$ are integrals in the parameter space defined by the unknown variables ($\boldsymbol{\theta}^*{}_k$, $\boldsymbol{\lambda}^*{}_k$ and $\boldsymbol{\nu}_k$):

$$p(\boldsymbol{Y}^*|\mathcal{M}_k) = \int p(\hat{\boldsymbol{W}}^*|\boldsymbol{\theta}^*{}_k, \boldsymbol{\lambda}^*{}_k, \mathcal{M}_k) \times p(\hat{\boldsymbol{U}}|\boldsymbol{\nu}_k, \boldsymbol{\lambda}^*{}_k, \mathcal{M}_k) \times p(\boldsymbol{\theta}^*{}_k) \times p(\boldsymbol{\lambda}^*) \times p(\boldsymbol{\nu}_k) d\boldsymbol{\theta}^*{}_k d\boldsymbol{\lambda}^*{}_k d\boldsymbol{\nu}_k \tag{4.54}$$

where the subscript k indicates the association with $\mathcal{M}_k$. The integral in Equation (4.54) is usually numerically estimated through Monte Carlo methods or approximations. An overview of such techniques is given in [64] and [121]. In this study Annealed Importance Sampling has been used to sample the marginal posterior distributions of the unknown parameters and at the same time estimate $p(\boldsymbol{Y}^*|\mathcal{M}_k)$.

In Table 4.2, the power of the evidence provided by $\mathcal{B}_{j,i}$ is interpreted according to the log_{10} scale suggested in [64] and in [121], which was adopted also in this research:

Table 4.2: Reference values for Bayes Factor interpretation

$log_{10}(\mathcal{B}_{j,i})$	$\mathcal{B}_{j,i}$	Evidence against $\mathcal{M}_i$
0 to 1/2	1 to 3.2	Not worth more than a mere mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

Finally, in validation exercises, the goodness of a model is qualitatively assessed by comparing the variations between MAP and specified values. The reason of this additional qualitative assessment is that being a validation experiment the specifications should be very accurate in order to put the analysts in the conditions to build accurate models. Thus it is expected that good models should be capable to provide a good fit of the observations with little changes in their parameters. This criterion can also be used to spot model inadequacy. In particular models having significant deficiencies tend to excessively alter their inputs in order to be able to adequately match the field observations.

Chapter 5

Initial Experiments

This chapter describes a series of experiments that were employed in developing the calibration method and in investigating its capabilities and limits. These relatively simple experiments required little computation and allowed the analysis of several tests involving increasing degree of complexity, which helped in understanding the limitations and the strengths of the developed calibration method.

The investigations that are described in the following, regard experiments carried out in the contexts of the DYNASTEE network (www.dynastee.org) and of the IEA EBC Annex 58 common exercises on a multilayer wall and a test box. In particular the experiments undertaken on the test box, which could be considered a scaled representation of a real building, represented the first important test, revealing that BES calibration is not a trivial problem, but one that requires a significant amount of prior information in order to be effectively solved.

Each experiment is divided in two parts. Firstly the *Virtual* subsection will describe the analyses carried out on synthetically generated observations. Since in these cases the real solutions were known, these experiments allowed to test the correctness and the capabilities identifying calibration parameters of the method. Secondly the *Real* subsection will explain the application of the calibration methodology on the actual performed experiments. GOF measures indicating the level of agreement achieved between the actual BES models and their emulators are listed in Appendix C.

5.1 The first brick in the wall

The first experiments involved the investigation of the thermal properties of a multilayer wall of a laboratory in an insulation factory in the south of Sweden, used in identification exercises in the context of the DYNASTEE network. The experiment was conducted by the EC Joint Research Centre, Institute for Energy and Transport in ISPRA, Italy, and consisted of monitoring the heat flux through the multilayer wall, as well as the external and internal temperatures conditioning the construction, for a period of one month. The construction element had three layers: a central core of gas concrete blocks (GC) of thickness 150 mm and insulation glass fibre boards (FB) of thickness 27 mm at both sides. The room at the inside face of the wall was heated with an electric heater. A fan was used to avoid air temperature stratification, and to achieve a good homogeneity of the heat flux distribution on the whole surface of the test component. Thermocouples and heat-flow meters were placed on both sides of the test wall in order to measure the surface temperatures and the heat flux at the inside surface of the wall (Figures 5.1, 5.2 and 5.3). The resulting dataset was composed of three time series of 1500 values with time step 0.5 hours.

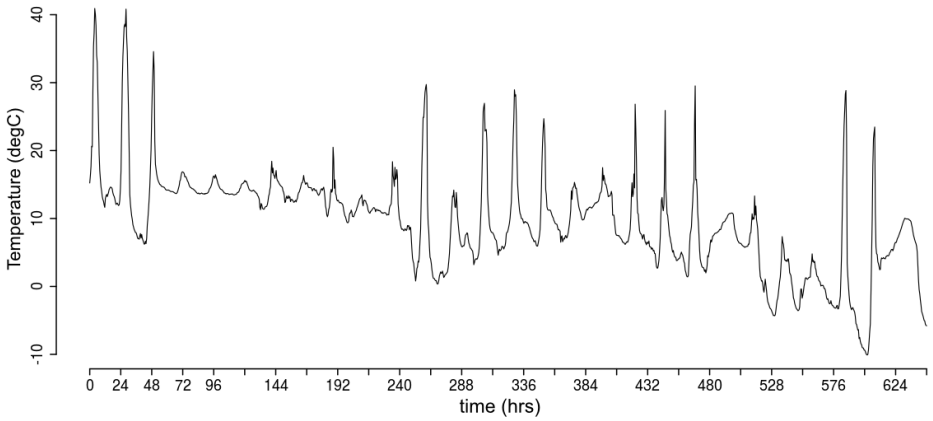


Figure 5.1: The Wall: measured external temperature.

At the end of the experiment, samples were taken from the test wall in order to determine the properties of each material. For the insulation boards, the data indicated in Table 5.1 were provided, and for the the concrete block only the density, equal to 552 $\pm$ 6 kg/m^3. The problem involved the inference of the thermal resistance and the thermal mass of the wall by identifying suitable values for glass fibre board specific heat ($FB_{\boldsymbol{c}}$), gas concrete block conductivity ($GC_{\boldsymbol{k}}$)

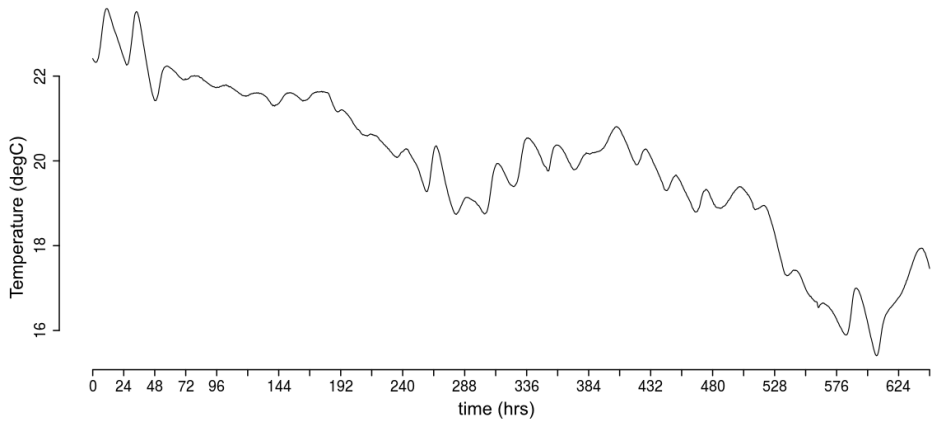


Figure 5.2: The Wall: measured internal temperature.

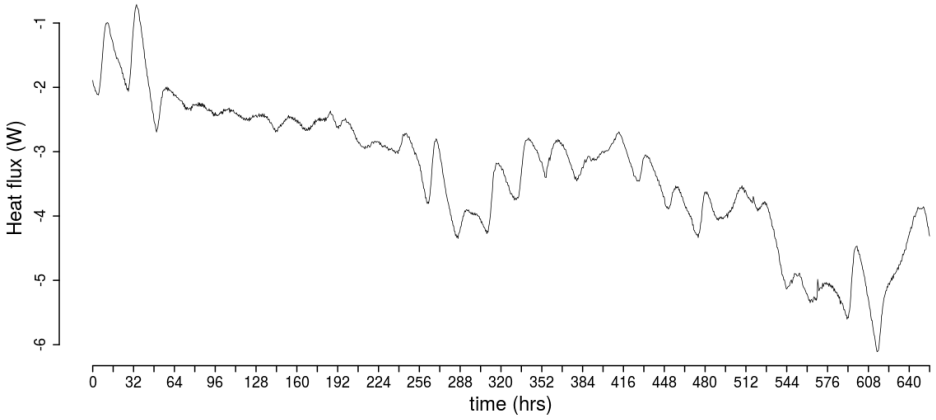


Figure 5.3: The Wall: measured heat flux.

and gas concrete block specific heat (GC_c).

5.1.1 Model

The analysis started by building a virtual replica of the real experiment within ESP-r. An ESP-r model comprising two thermal zones, one representing the inside of the laboratory, and the other depicting the exterior environment, divided by the test wall, was used to recreate the performed experiment (Figure 5.4).

In order to impose at the internal and external surfaces of the test wall the measured temperatures as boundary conditions, the following approach was adopted. Apposite control laws were set up in the two zones, so that their dry

Table 5.1: The Wall: glass fibre board material properties.

FEATURE	INTERNAL BOARD	EXTERNAL BOARD	AVERAGE
Dry density ($\boldsymbol{\rho}$) (kg/m^3)	114.8 ± 5.1	118.3 ± 4.4	116.6 ± 4.7
Conductivity ($\boldsymbol{k}$) ($\frac{W}{(mK)}$)	31.23 ± 0.04	31.31 ± 0.26	31.27 ± 0.17

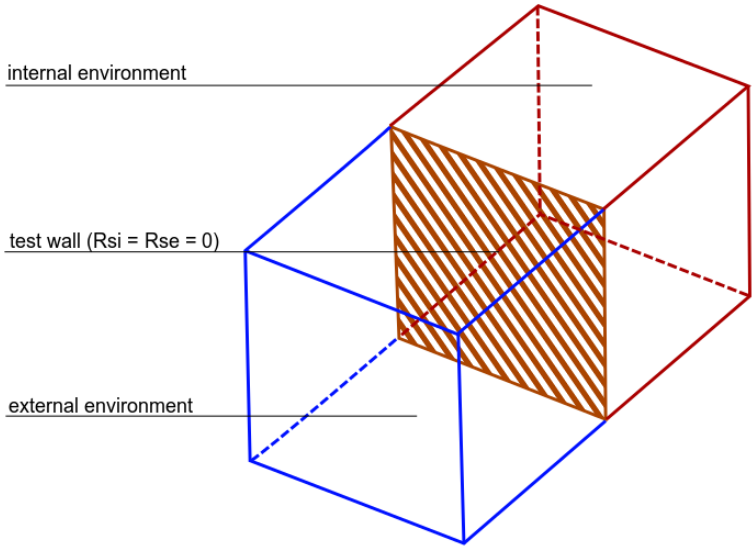


Figure 5.4: The Wall: Wire frame representation of the model.

bulb air temperatures were closely following the measured profiles. Very high convection heat transfer coefficients (1000 $W/(m^2K)$) were imposed on the internal and external surfaces of the test component in order to have their temperatures equal to the temperature of the respective air nodes. Emissivities and absorptivities at the internal surfaces bounding the two model zones were set to zero, thus avoiding radiative heat exchanges, and making conduction the only mechanism driving the heat flux through the test component.

A three layer construction, reproducing accurately the given specifications, was used to represent the test wall layer structure. However due to the negligible differences and uncertainties, the two fibre boards were considered having the same properties, equal to the average values indicated in Table 5.1. For the unknown model parameters the initial values and prior probability density distributions indicated in Table 5.2 were assumed. In particular, Uniform probability density distributions were chosen since no information was provided about possible likely values. The variation ranges were defined according to information

coming from literature, especially from [70] and [34].

Table 5.2: The Wall: calibration parameter prior probability density distributions and initial values.

PARAMETER	INITIAL VALUES	PRIOR DISTRIBUTION
$GC_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.12	$\mathcal{U}(0.05, 0.15)$
$GC_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	800.00	$\mathcal{U}(600, 1000)$
$FB_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	800.00	$\mathcal{U}(600, 1000)$

5.1.2 Virtual

A simple preliminary investigation about the sensitivity of the ESP-r model, revealed that $FB_{\boldsymbol{c}}$ had almost negligible effects on the model behaviour. To emphasise this aspect, Figure 5.5 shows the results from 30 model simulations wherein only the value of $FB_{\boldsymbol{c}}$ was varied in a Monte Carlo fashion within the range indicated in Table 5.2. It is clearly possible to notice that the 30 model responses are almost identical.

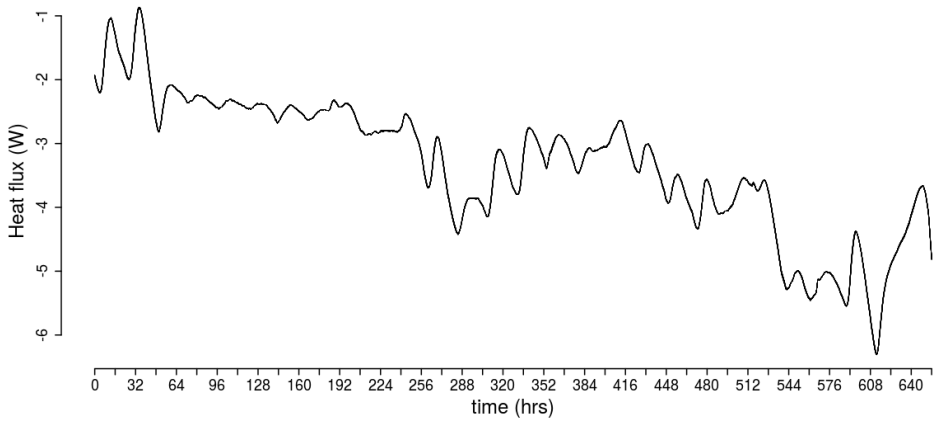


Figure 5.5: The Wall: 30 model simulation varying only $FB_{\boldsymbol{c}}$ in a Monte Carlo fashion.

This was taken as an opportunity to investigate the performances of the calibration framework in identifying calibration parameters having very weak effects on the model behaviour. Synthetic observations were generated for the initial model parameter values indicated in Table 5.2, and then contaminated with correlated noise and processes simulating model inadequacy, in order to investigate the effects of these aspects on the estimation of weak model parameters. In particular, it will be shown that the identification of model parameters depends on

the model sensitivity and that the characterisation of non-influential parameters is affected by measurement uncertainties and model deficiencies.

Virtual 1

In *Virtual 1* the ESP-r model was calibrated considering the synthetic observations as produced by the dynamic simulation program, that is considering the target data free from noise, and the model perfectly able to represent the observed process. The main objectives were to prove that the probabilistic emulator is able to adequately represent the behaviour of the computer model and that the calibration process is effectively capable of precisely estimating unknown model parameters. This example, also, provided a baseline to assess the results for the following virtual experiments, wherein the data were perturbed with different processes.

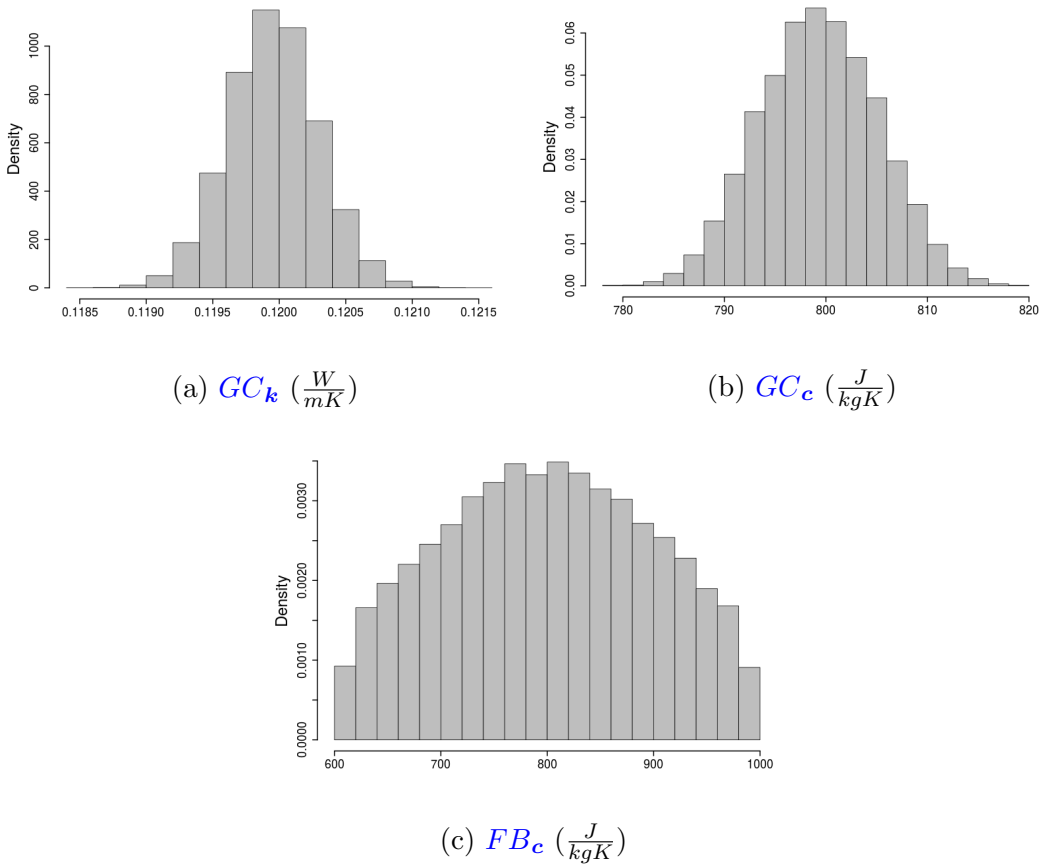
The outcomes of the calibration process, consisting of posterior probability density distributions, MAP estimates and 95% confidence intervals for the calibration parameters, are shown in Table 5.3 and Figure 5.6. It is possible to see that all the three calibration variables had been accurately identified and correctly estimated. $GC_{\boldsymbol{k}}$ and $GC_{\boldsymbol{c}}$ presented very low posterior variances and narrow confidence intervals. $FB_{\boldsymbol{c}}$, on the other hand, had a larger confidence interval and an empirical posterior density distribution which was spread over the initial variation range, but, nonetheless clearly peaked around its *true value*. In particular despite its negligible influence on the model, it had an MAP estimate particularly accurate.

It is also interesting to notice that the results reflected correctly the sensitivity of the model. Indeed, it is particularly sensible that calibration parameters are estimated with uncertainties inversely proportional to their identifiability, which is mainly determined by their capability of influencing the behaviour of the model.

Table 5.3: The Wall–*Virtual 1*: MAP estimates, 2.5%, 50% and 97.5% quantiles for the calibration parameters

PARAMETER	MAP	$Q2.5\%$	$Q50\%$	$Q97.5\%$
$GC_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.120	0.119	0.120	0.121
$GC_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	799.225	788.210	799.287	810.623
$FB_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	802.312	624.522	800.141	975.363

Furthermore, by investigating the correlation between the considered calibration parameters, (Figure 5.7) it was possible to observe a relevant negative



(a) $GC_{\boldsymbol{k}}$ ($\frac{W}{mK}$)

(b) $GC_{\boldsymbol{c}}$ ($\frac{J}{kgK}$)

(c) $FB_{\boldsymbol{c}}$ ($\frac{J}{kgK}$)

Figure 5.6: The Wall–*Virtual 1*: empirical posterior density distributions for the wall calibration parameters.

correlation between $GC_{\boldsymbol{c}}$ and $FB_{\boldsymbol{c}}$.

Such result was in agreement with the relations established by the physical model between the two variables. In particular, both contributed to the thermal mass of the wall, but the former in much larger measure than the latter, since the gas concrete block central layer is much thicker that the two fibre boards on the sides. Thus, big variations in $FB_{\boldsymbol{c}}$ were easily compensated by little variations in $GC_{\boldsymbol{c}}$ in the opposite direction and vice versa. However, in this case the identifiability of the two parameters was not compromised, because of the large difference in the model sensitivity to these two different parameters. In particular, since $GC_{\boldsymbol{c}}$ had the greater effect, this parameter was identified first and then $FB_{\boldsymbol{c}}$ was inferred subsequently. Therefore the identifiability of the weakest parameter was strengthened by its relation with the stronger one.

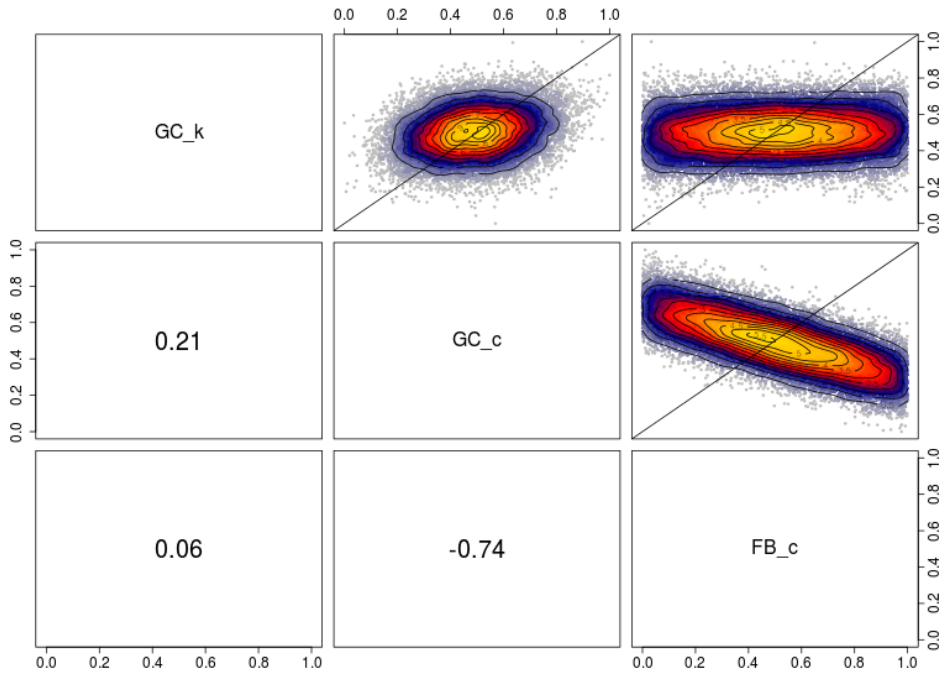


Figure 5.7: The Wall–*Virtual 1*: correlation between calibration parameters.

Virtual 2

Virtual 2 investigated the effects of noise in the data on the performance of the calibration method in identifying calibration parameters. Even more the capability of the *Difference Analysis* in revealing the presence of correlated noise in the data was tested.

An ARMA(1,1)/GARCH(1,1) ([104]) process was used to simulate the presence of correlated and heteroskedastic noise in the target observations (Figure 5.8).

Such a noise process, besides having correlation between its realizations at different time steps, has also correlation between the relative variances, thus simulating time varying uncertainties, which can, for example, generate unexpected spikes in the measurements. This situation is quite realistic and general. Indeed, despite the common assumption that the measurement noise is i.i.d., having processes similar to the considered one contaminating the data is often the case. The resulting synthetic observations are displayed in Figure 5.9.

The estimated MAP values and relative confidence intervals inferred for the calibration parameters are listed in Table 5.4, and the relative empirical posterior density distributions are depicted in Figure 5.11.

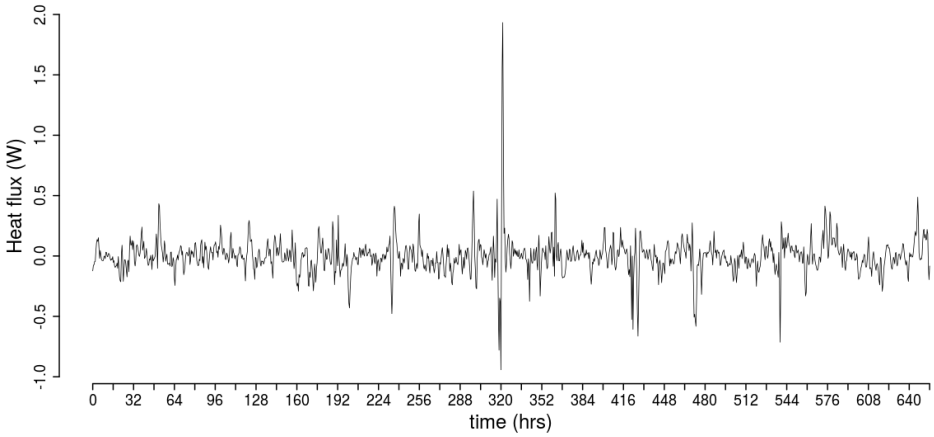


Figure 5.8: The Wall–*Virtual 2*: noise process.

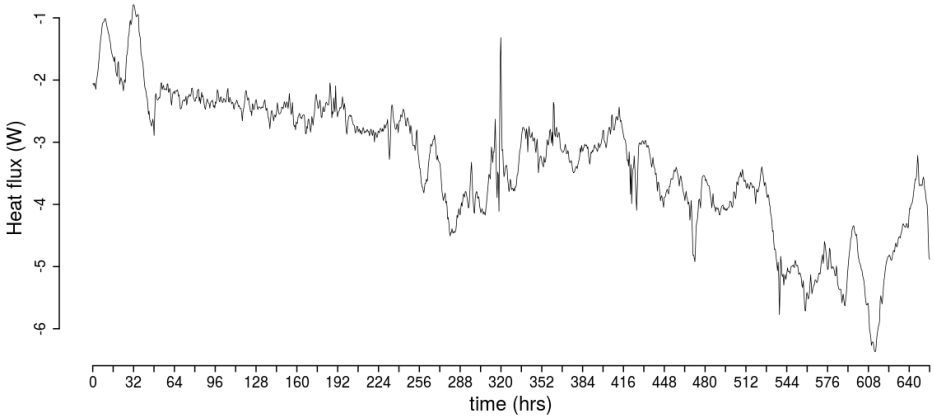


Figure 5.9: The Wall–*Virtual 2*: synthetic observations

$GC_{\boldsymbol{k}}$ was perfectly estimated and its confidence interval is unchanged with respect to the results of *Virtual 1*. Differently, the MAP values for $GC_{\boldsymbol{c}}$ and $FB_{\boldsymbol{c}}$ were not so accurate as before. While the former, although slightly underestimated, had a reasonable MAP and its *true value* is well within the calculated confidence interval, the latter has an empirical posterior probability density distribution squashed against the upper bound of its variation range, and it has been largely overestimated. Furthermore the directions and magnitudes of these inaccuracies are in agreement with the relation of negative correlation between the two variables.

It is interesting to notice that the parameters having inaccurate estimates were those with the major influences on the dynamic of the model. In particular

Table 5.4: The Wall– *Virtual 2*: MAP estimates, 2.5%, 50% and 97.5% quantiles for precision and calibration parameters

PARAMETERS	MAP		$Q50\%$	$Q97.5\%$
$GC_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.120	0.119	0.120	0.121
$GC_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	791.179	778.981	791.016	805.443
$FB_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	944.813	654.373	913.591	989.582

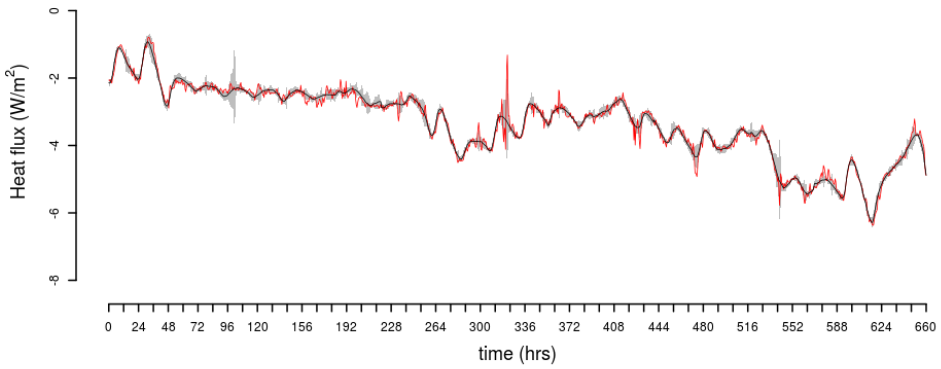
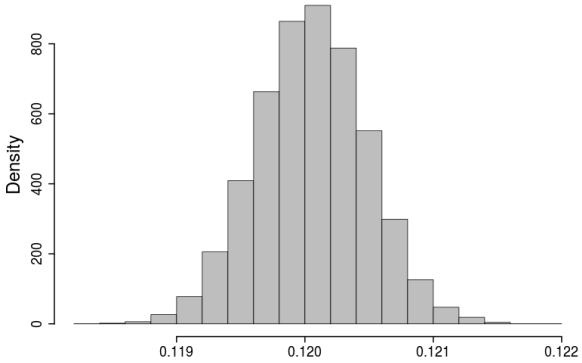


Figure 5.10: The Wall– *Virtual 2*: match between model prediction (black) and synthetic observations (red). In grey are indicated the 95% confidence intervals.

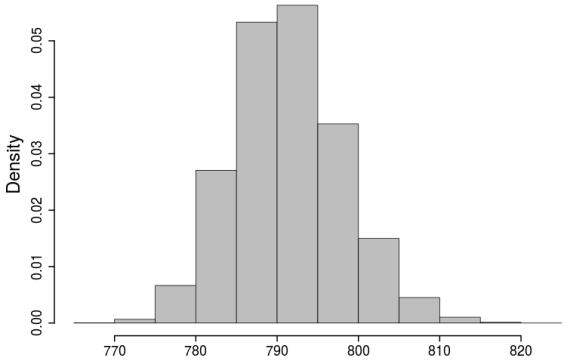
when the true process blends with another correlated process it changes its correlation and therefore its dynamics. When this occurs, it is impossible to neatly separate the two contributions, and some of the variability due to the perturbation process influences the estimation of the calibration parameters, which are tuned with respect to their *true values*, in order to better match the measurements. Indeed by calculating the CVRMSE, relative to the model output for the calibration parameter *true values* (-4.385) and the model output for the MAP values inferred in this analysis (-4.381), it is possible to notice that the latter was slightly lower. Therefore the model represented better the dynamics of the observed process, by having its inputs set to the MAP estimates, but the little improvement was due to a small over-fitting of the noise process and not to an actual better representation of the true process. A comparison between model predictions and synthetic observation is shown in Figure 5.10.

The results from the *Difference Analysis* revealed that the only two hyper parameters of the *Difference Model* having MAP estimates sensibly different from zero, were the additive variances ($\sigma_d^{(r)}$). In the light of these results and for the consideration explained in section 4.2.2, the method correctly identified correlated

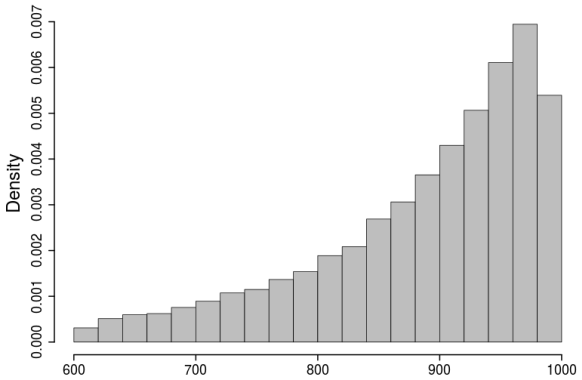
noise as the main reason of discrepancies between model output and observations.



(a) $GC_{\boldsymbol{k}}$ ($\frac{W}{mK}$)



(b) $GC_{\boldsymbol{c}}$ ($\frac{J}{kgK}$)



(c) $FB_{\boldsymbol{c}}$ ($\frac{J}{kgK}$)

Figure 5.11: The Wall– *Virtual 2*: empirical posterior density distributions for the wall calibration parameters.

Virtual3

Virtual3 studied how the presence of deficiencies in the model may affect the calibration outcomes and it tested the capabilities of the *Difference Model* is spotting correlation between *Difference Vectors* and boundary conditions. In order to simulate model inadequacies, a periodic function of the external temperature, $\boldsymbol{Te}$, (Equation 5.1) was added to the original synthetic observations, so as to provide the data with variability lying outside the domain of the trained model.

$$\Delta(\boldsymbol{Te}) = 0.3cos(\boldsymbol{Te}\frac{\pi}{50}) \tag{5.1}$$

The resulting observations are displayed in Figure 5.12.

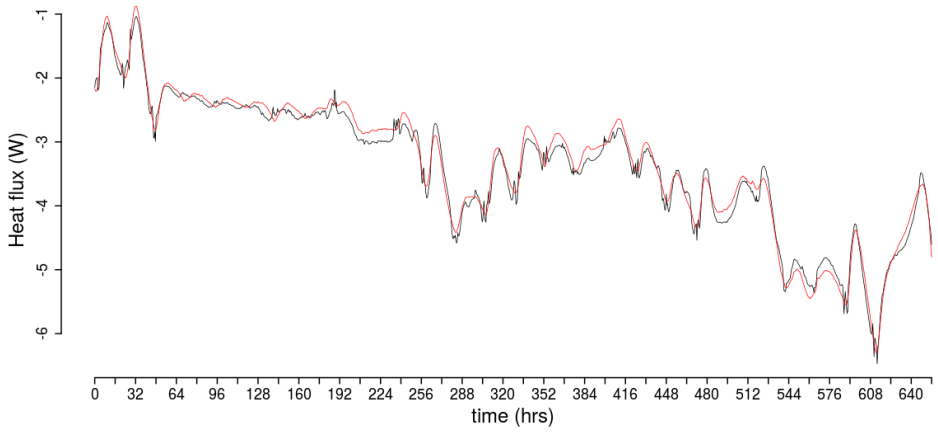
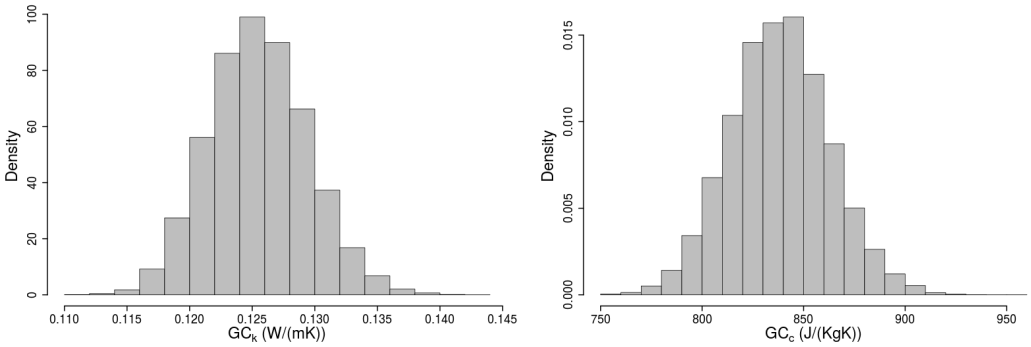


Figure 5.12: The Wall–Virtual3: original synthetic observations (red line) and synthetic observations with model inadequacy (black line).

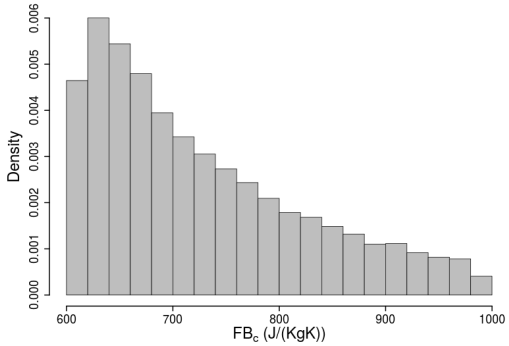
The calibration was undertaken by leaving out the last third of the data in order to test the predictive capabilities of the calibrated model over conditions not used in its training. Table 5.5 and Figure 5.13 show the outcome of the analysis.

Table 5.5: The Wall–Virtual3: MAP estimates, 2.5%, 50% and 97.5% quantiles for precision and calibration parameters.

PARAMETERS	MAP	$Q2.5\%$	$Q50\%$	$Q97.5\%$
$GC_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.125	0.118	0.125	0.134
$GC_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	837.601	791.630	838.195	887.172
$FB_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	659.549	610.101	700.874	958.090



(a) $GC_{\boldsymbol{k}}$ ($\frac{W}{mK}$) (b) $GC_{\boldsymbol{c}}$ ($\frac{W}{mK}$)



(c) $FB_{\boldsymbol{c}}$ ($\frac{W}{mK}$)

Figure 5.13: The Wall–Virtual3: empirical posterior density distributions for the wall calibration parameters.

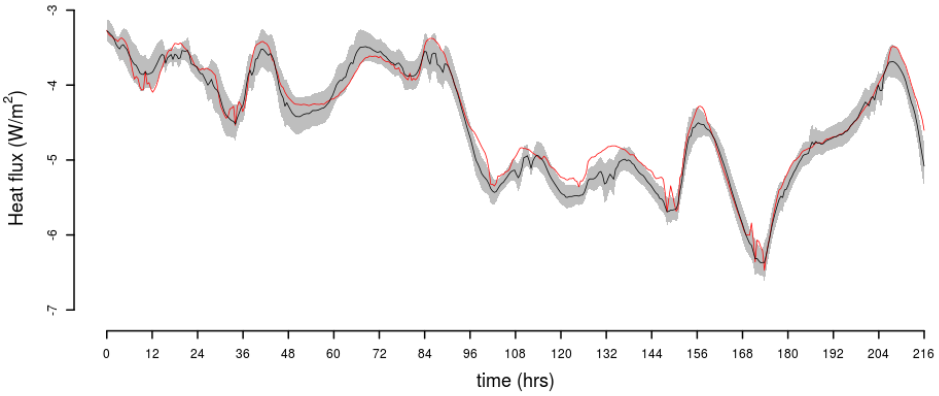


Figure 5.14: The Wall–Virtual3: comparison between model predictions (black) and observations (red) in the test period. In grey are indicated the 95% c.i..

The unknown model parameters were estimated with similar performances as *Virtual 2*. $GC_{\boldsymbol{k}}$ and $GC_{\boldsymbol{c}}$ were slightly overestimated but well identified and with their *true values* well within their confidence intervals. On the contrary $FB_{\boldsymbol{c}}$ had an empirical posterior density distribution concentrated against its lower bound and a MAP estimate abundantly lower than its *true value*. Once again it was possible to notice the influence of the model sensitivity and correlation on the results relative to $GC_{\boldsymbol{c}}$ and $FB_{\boldsymbol{c}}$. Furthermore, by calculating the CVRMSEs, over the training period, similarly as in *Virtual 2* it was possible to draw analogous conclusions. In particular, the CVRMSE relative to the model predictions for the calibration parameter *true values* (-3.949), was higher than the CVRMSE relative to the model output for the estimated calibration parameter values (-3.823). Thus, also in cases wherein the model presents inadequacies, calibration parameters are tuned with respect to their *true values*, in order to provide for deficiencies in the model.

The *Difference Analysis* successfully identified a periodic function of $\boldsymbol{Te}$ as the main cause of discrepancy between model output and observed values. The only two hyper parameters significantly different from zero were a frequency parameter relative to $\boldsymbol{Te}$ and the weight of the relative component of the SM covariance function.

Figure 5.10 shows a comparison between the synthetic observations and the model prediction over the test period. The model predictions followed adequately the target values and also their confidence bounds were reasonable.

5.1.3 Real

In this experiment the ESP-r model has been calibrated against the provided measured data. The true properties of the wall layers were not disclosed at the time of writing, in order to preserve the blind character of the identification exercise for future occasions. Thus, to prove the goodness of the calibrated model, the provided data set was divided in a training set consisting of the first two thirds of the data and in a test set comprising the last third of the data. The former was used to calibrate the ESP-r model, while the latter was used to perform external validation, by assessing the performance of the model in providing predictions according to data not used for its calibration. The results are numerically and graphically summarised in Table 5.6 and Figure 5.15.

The two main parameters, $GC_{\boldsymbol{k}}$ and $GC_{\boldsymbol{c}}$, were precisely identified. Both these variables have nicely bell shaped empirical posterior distribution, with low variances, as also indicated by their narrow confidence intervals. The glass fibre

Table 5.6: The Wall–*Real*: MAP estimates, 2.5%, 50% and 97.5% quantiles for precision and calibration parameters.

PARAMETERS	MAP	$Q2.5\%$	$Q50\%$	$Q97.5\%$
$GC_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.123	0.122	0.123	0.124
$GC_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	934.189	915.389	934.509	955.633
$FB_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	957.736	660.021	929.523	990.518

board specific heat, showed results similar to the outcomes of *Virtual 2* and *Virtual3*. Thus it is probable that the estimates of this parameter was influenced by the presence of noise in the data or deficiencies in the model. Its empirical posterior density distribution was mostly located near the upper bound of its variation range, and its confidence interval was significantly larger than those of the stronger parameters. However, considering the sensitivity of the model to this input (Figure 5.5), inaccuracies in its estimation should not cause relevant errors in the calibrated model predictions.

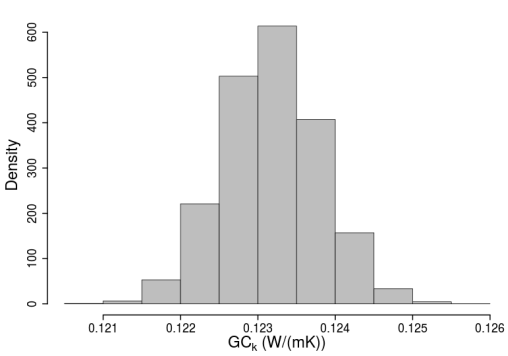
The results from the *Difference Analysis* revealed, besides the presence of correlated noise in the measurements, a periodic trend depending on the internal temperature as causes of discrepancies between model output and measurements. Unfortunately, due to the lack of data and information, it was not possible to

Table 5.7: The Wall–*Real*: goodness of fit criteria.

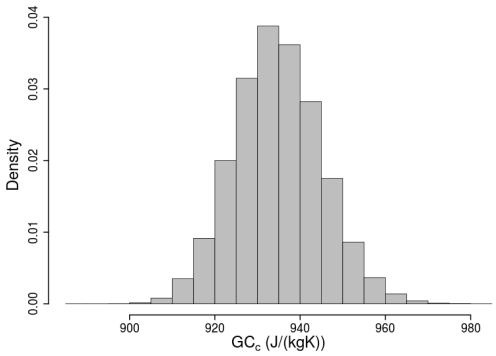
SET	NMBE (%)	CVRMSE (%)	GOF(%)
training$^{(1)}$	0.074	-3.546	2.508
test$^{(2)}$	-0.59	-2.317	1.690

(1) $\bar{\boldsymbol{y}}^* = -2.753\ kW$; (2) $\bar{\boldsymbol{y}}^* = -4.448\ kW$

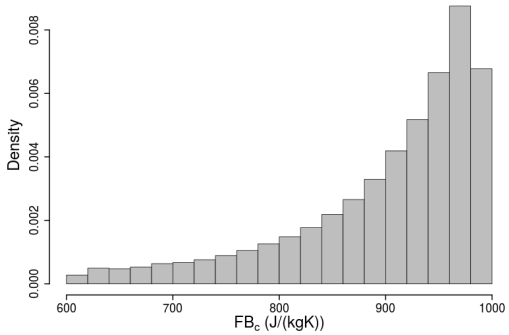
investigate further the causes of the mismatch and to provide upgrades for the model. Nonetheless the calibrated model showed good performances in predicting the measured data. In particular, the considered goodness of fit criteria (Table 5.7) had low values and the fit improved in the test period. The achieved fit for the training and test observations are displayed in Figures 5.16 and 5.17. It is also important to notice that the calculated confidence bands are reasonable.



(a) $GC_{\boldsymbol{k}}$



(b) $GC_{\boldsymbol{c}}$



(c) $FB_{\boldsymbol{c}}$

Figure 5.15: The Wall–*Real*: empirical posterior density distributions for the wall calibration parameters.

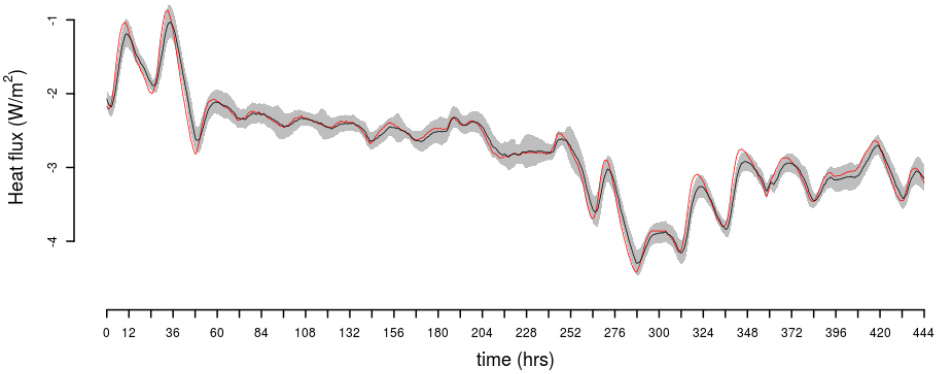


Figure 5.16: The Wall–*Real*: comparison between model predictions (black) and observations (red) during the training period. In grey are indicated the 95% confidence intervals.

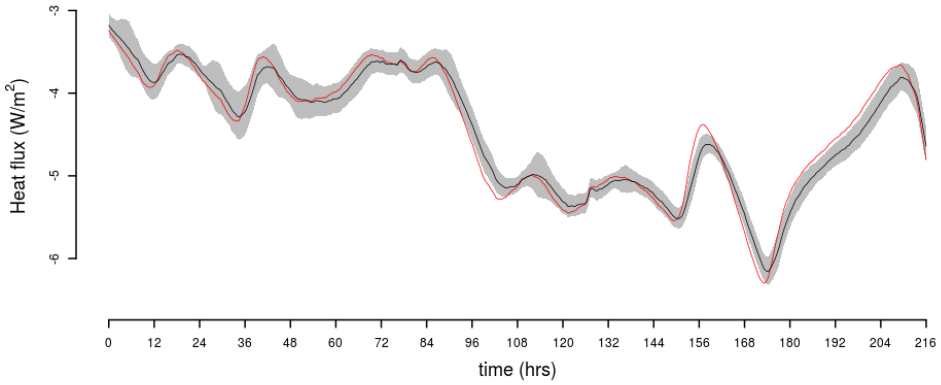


Figure 5.17: The Wall–*Real*: comparison between model predictions (black) and observations (red) during the test period. In grey are indicated the 95% confidence intervals.

5.1.4 Discussion

The virtual experiments undertaken on this simple case study, demonstrated that the presence of observation errors and inadequacies in the model, influences the estimation of the calibration parameters, especially those having weak effects on the model outputs. In particular, it is not possible to neatly separate the realization of the real process from noise affecting the observations and the model will always have deficiencies in representing real processes. These two aspects affect the estimation of the unknown model inputs, which will be tuned in order to match the calibration target values contaminated by noise as well as possible, and to provide for deficiencies in the modelling structure. Therefore, besides virtual experiments, where the actual solutions are known a priori, and the calibration targets can be considered noise free, the calibration process will yield different estimates for the calibration parameters according to the field data considered. Such estimates will be more close to each other for parameters having strong influences on the model behaviour, but they can be sensibly different for the weak model inputs. In real calibration experiments the term *true value* should be avoided in referring to parameter estimates and the inferred values should be considered as the most probable values, according to the considered field data.

For the real experiment, it is not possible to assess certainly the goodness of the achieved results, since at the time of writing the real specification of the wall was not disclosed. However due to the simplicity of the case study, the obtained reasonable estimates, and, most of all, the particular good match obtained be-

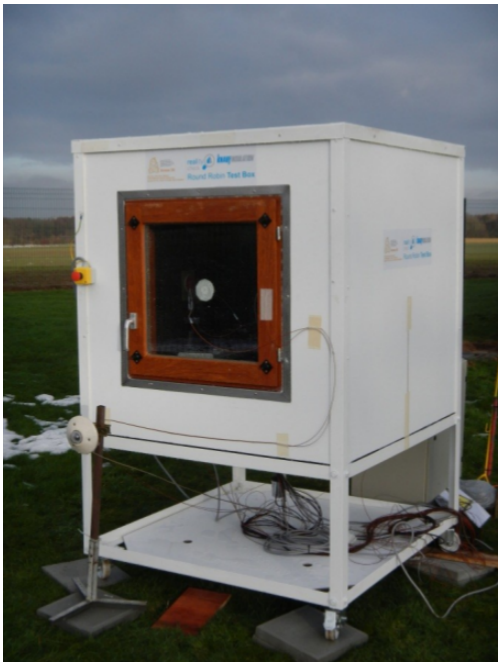
tween measured heat fluxes and model predictions, the author is confident that the inferred parameters are close to the actual properties of the materials making the three layer wall.

This first simple example gave the possibility to develop the basis of the method, to gain knowledge about its limitations and capabilities, and to build up the confidence necessary to undertake more complicated analysis.

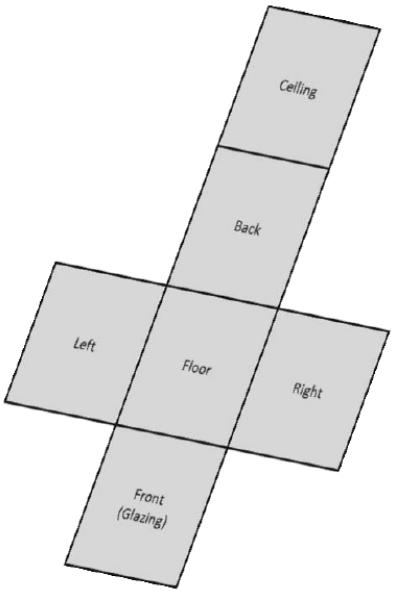
5.2 The little box(es)

The second series of experiments focuses on a test box used in a Round Robin Experiment, which aimed to develop identification and characterization methods for building energy models, in the context of the IEA EBC Annex 58.

The test box was built by a research team of the University of KU Leuven (KUL), participating in the Annex, which was the only party knowing its real composition. It had cubic form with internal dimensions $96 \times 96 \times 96$ cm. The roof, floor and walls had all identical composition and thickness of 12 cm. The front wall had a window of dimensions 60×60 cm, wherein the glazed part had an area of 52×52 cm. The whole structure was provided with a support which allowed the influence of the ground to be neglected (Figure 5.18a).



(a) Test box at the BBRI test site.



(b) Test box wall naming.

Figure 5.18: Test box.

The datasets used during the analyses were the results of measurements done during experiments conduced in Belgium at the Belgian Building Research Institute (BBRI) and in Spain at Building Component Test Laboratory (LECE) at Plataforma Solar de Almeria.

This section proceeds by firstly explaining the BBRI and LECE experiments, and the structures of the models used in the analyses. Secondly the carried out calibration experiments are depicted.

5.2.1 BBRI experiment

The experiment, conducted at the BBRI in Limelette, involved the monitoring of the internal conditions and the influencing weather factors for a period of 4 weeks, starting the 25th of January 2013 and ending the 28th of February 2013. The experiment developed in two phases: two initial weeks where a constant temperature of 25 ^oC was kept inside the box (Constant Temperature (CT)) and a following period of two weeks wherein the behaviour of the experimental facility was observed under free float conditions (Free Float (FFL)). During both of these phases the window was facing south. All the data including the heat inputs during the co-heating phase were recorded every 5 minutes. Two comprehensive data sets were gathered comprising 31 measured variables, of 3826 and 4124 observations, for the CT and FFL phases respectively. For more details about the experimental procedure and the collected data the reader is referred to [59].

During the analysis, the provided data was divided into training and test datasets. The former, which was used to undertake the calibration of the different models, was composed of the entire constant temperature period and the first four fifths of the free float phase, while the latter, consisting of the last fifth of the free-float phase, was employed to provide external validation for the calibrated models.

5.2.2 LECE experiment

This experiment was conducted at LECE Laboratory at Plataforma Solar de Almeria in the south of Spain and extended over a period of 44 days during the hot season, starting the 28th of May 2013 and ending the 10th of July 2013. Besides initial tests, performed in order to identify the most suitable heating device to employ, the experiment involved a constant temperature phase for a duration of 10 days, wherein the inside temperature was kept constant at a set point of 40 ^oC, followed by a period wherein the test box was subject to Randomly

Ordered Logarithmic Binary Sequence (ROLBS) of heat injections ([120]), which lasted 4 days and finally the experimental facility was monitored under free-float (FFL) conditions for a period of 9 days. As in the BBRI experiment, during all the tests the window was oriented towards south. For more details the reader is referred to [58].

The analysis focused only on the ROLBS experiment since it was deemed the most significant, and also representative of different operative conditions with respect to the BBRI experiment. The FFL data, which were not provided in a first place, were used to externally validate the calibrated models.

5.2.3 Models

Three models, of increasing levels of detail, were developed according to the outcomes of the analyses undertaken on the data coming from the BBRI experiment. An initial simple model ($\mathcal{M}_0$), was built according to the modellers experience and the provided information about the geometry test box. Since no specifications were given about material properties, which were initially set according to a simple sensitivity analysis, this virtual representation cannot be considered as a base line built on the provided data, and the assessment of its performance before calibration was deemed not meaningful. Therefore all the calculated GOF criteria and comparison with measured data, regarding this model are relative to its input parameter fixed to the calibrated values. Successively, the analysis highlighted the importance to consider the heat flux due to longwave radiation toward the sky and wind driven infiltration. The former aspect was considered in the model $\mathcal{M}_1$, while both of them were included in the final model ($\mathcal{M}_2$), thus gradually increasing the complexity accordingly to the result of the performed calibrations.

The two best models among the three analysed during the BBRI experiment were, then, used in trying to characterize the test box subjected to the ROLBS heating sequence in the LECE experiment. The assumptions and the structures upon which these three models were built are described in the following. A summary of the considered calibration parameters for the three models, comprising the relative prior probability density distributions, can be found in Table 5.8.

Initial model ($\mathcal{M}_0$)

The experiments were reproduced by building an ESP-r model of the box, and by imposing on it the measured variables conditioning the different phases. The

Table 5.8: Test Box: calibration parameter prior probability density distributions and initial values.

PARAMETER	INITIAL VALUE	PRIOR DISTRIBUTIONS $\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$
$wall_{\boldsymbol{k}}$ ($\frac{W}{mK}$)	0.1	$\mathcal{U}(0.01, 0.2)$	$\mathcal{U}(0.01, 0.2)$	$\mathcal{U}(0.01, 0.2)$
$wall_{\boldsymbol{\rho}}$ ($\frac{kg}{m^3}$)	1500	$\mathcal{U}(500, 2500)$	$\mathcal{U}(500, 2500)$	$\mathcal{U}(500, 2500)$
$wall_{\boldsymbol{\alpha}}$ (-)	0.25	$\mathcal{U}(0.1, 0.4)$	$\mathcal{U}(0.1, 0.4)$	$\mathcal{U}(0.1, 0.4)$
$window_{\boldsymbol{R}}$ ($\frac{m^2K}{W}$)	0.77	$\mathcal{U}(0.5, 0.95)$	$\mathcal{U}(0.5, 0.95)$	$\mathcal{U}(0.5, 0.95)$
$glass_{\boldsymbol{tr}}$ (-)	0.611	$\mathcal{U}(0.5, 0.72)$	$\mathcal{U}(0.5, 0.72)$	$\mathcal{U}(0.5, 0.72)$
$wall_{\boldsymbol{e}}$ (-)	0.85	fixed	$\mathcal{U}(0.7, 0.95)$	$\mathcal{U}(0.7, 0.95)$
$ROLBS_{\boldsymbol{C/R}}$[1] (-)	0.85	$\mathcal{U}(0.5, 1)$	$\mathcal{U}(0.5, 1)$	$\mathcal{U}(0.5, 1)$
$floor\_crack_{\boldsymbol{len}}$ (m)	0.1	absent	absent	$\mathcal{U}(0.0001, 0.5)$
$window\_crack_{\boldsymbol{len}}$ (m)	0.1	absent	absent	$\mathcal{U}(0.0001, 0.5)$
wall specific heat ($\frac{J}{kgK}$)	1600	fixed	fixed	fixed
glass conductivity ($\frac{W}{mK}$)	0.76	fixed	fixed	fixed
glass density ($\frac{kg}{m^3}$)	2710.00	fixed	fixed	fixed
glass specific heat ($\frac{J}{kgK}$)	837.00	fixed	fixed	fixed
glass emissivity (-)	0.83	fixed	fixed	fixed
glass absorptivity (-)	0.50	fixed	fixed	fixed

(1) considered only in LECE experiment.

ESP-r model was an exact geometric representation of the test box, but, since the construction specifications were not known, it was necessary to make assumptions about the envelope properties. The walls were considered having all the same properties, as suggested by the specifications, and approximated with construction components made of only one layer. The window frame was considered having the same composition of the walls, and the window consisting of a double glazed construction. The infiltration was neglected, according to the results from blower door tests, which recorded insignificant rates. Thermal bridges were neglected as well.

The variable conditions imposed on to the model consisted of external temperature (***Te***), vertical solar radiation (***Gv***) on the window plane, horizontal diffuse (***Ds***) and global horizontal (***Gh***) solar radiation, wind velocity (***Ws***) and direction (***Wd***), and relative humidity (***Rh***). In the constant temperature and ROLBS phases, the measured internal temperature profiles and heat injection sequences were respectively imposed as well through adequate control laws.

A simple preliminary sensitivity analysis suggested to consider as calibra-

tion parameters the wall conductivity ($wall_{\boldsymbol{k}}$), the wall density ($wall_{\boldsymbol{\rho}}$), the wall absorptivity ($wall_{\boldsymbol{\alpha}}$), the window thermal resistance ($window_{\boldsymbol{R}}$), and the glass optical transmittance ($glass_{\boldsymbol{tr}}$). The thermal mass of the glass and its conductivity proved to be negligible for the considered model outputs, and they were fixed to default values. The convective/radiative split coefficient ($ROLBS_{\boldsymbol{C/R}}$) relative to the heat gains from the ROLBS pulses was added as well to the set of free parameters in the calibration experiment involving the LECE dataset. The other parameters were fixed to default values.

The parameters $wall_{\boldsymbol{k}}$ and $wall_{\boldsymbol{\rho}}$ deserve some further comments. Due to the modelling approximations and assumptions, these two variables did not have a clear physical meaning. The former governed the heat transfer by conduction in the model, and, besides the normal heat fluxes through the walls of the box, it took into account also contributions due to thermal bridges at the corners, edges and window frame. It could be seen as the conductivity of a homogeneous building component showing the same conduction heat loss coefficient as the whole test box. Similarly the latter determines the heat capacity, thus it could be seen as the density of a homogeneous construction having effective thermal mass equal to the effective thermal mass of the real test box.

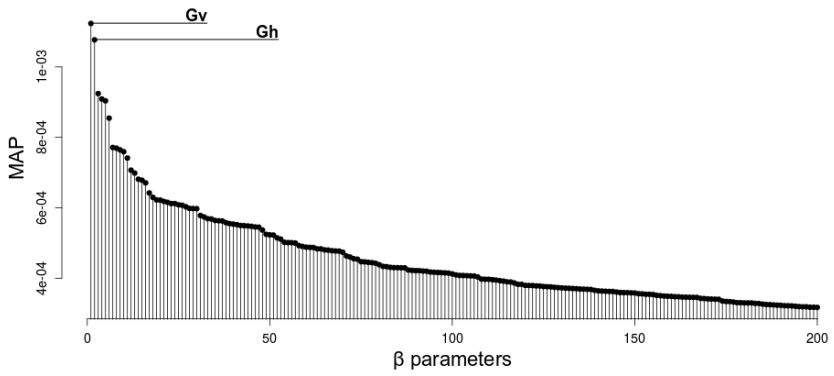
These assumptions and hypothesis resulted in the base model ($\mathcal{M}_0$), which was, then, upgraded according to the results of the undertaken analyses.

Including sky longwave radiation ($\mathcal{M}_1$)

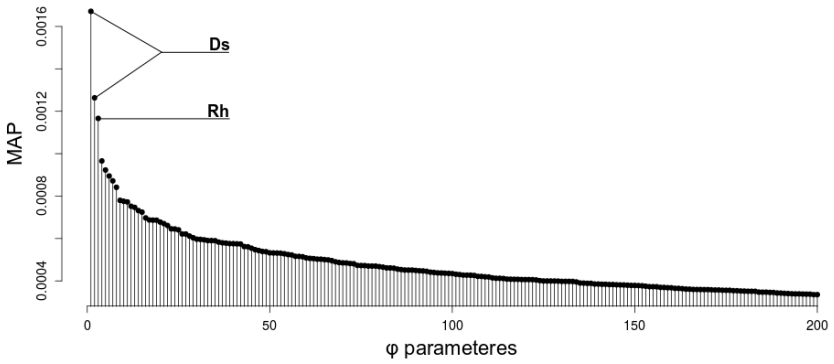
Following the results from the *Difference Analysis* relative to $\mathcal{M}_0$, it was observed that there were a number of boundary conditions weakly correlated to the *Difference Vectors*. In particular, Figures 5.19 and 5.20 show scree plots for the rate (β) and frequency (ϕ) parameters of the *Difference Model*, highlighting that wind speed ($\boldsymbol{Ws}$), wind direction ($\boldsymbol{Wd}$), global horizontal solar radiation ($\boldsymbol{Gh}$), direct solar radiation ($\boldsymbol{Ds}$), vertical solar radiation on the window plane ($\boldsymbol{Gv}$) and relative humidity ($\boldsymbol{Rh}$) are the model boundary conditions most correlated with the *Difference Vectors*.

Wind speed and wind direction, may affect several phenomena, like infiltration and convection. The former phenomenon was neglected, since the test box was assured to be extremely air tight. The latter was deemed to be not important in determining discrepancies since the low values of external surface resistances, especially at the outside surfaces, which had a negligible influence on the overall thermal resistance of the walls.

All the measured data relative to variables influencing solar processes, were



(a) β parameters.

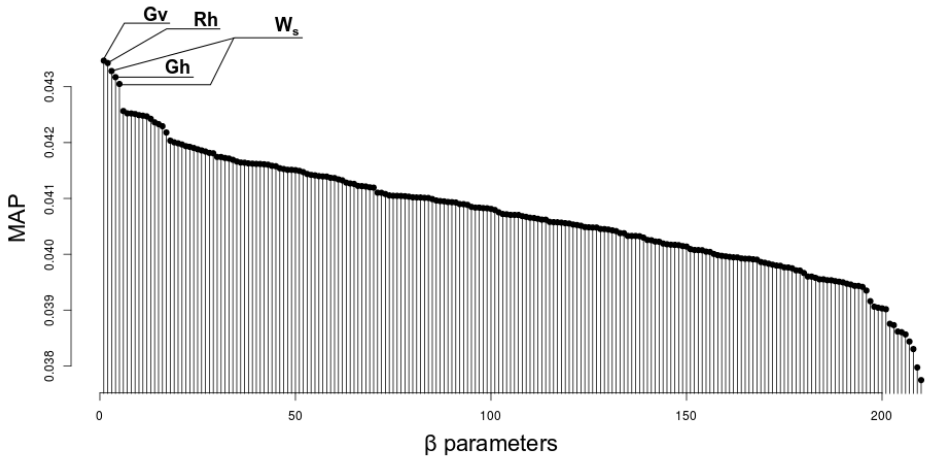


(b) ϕ parameters.

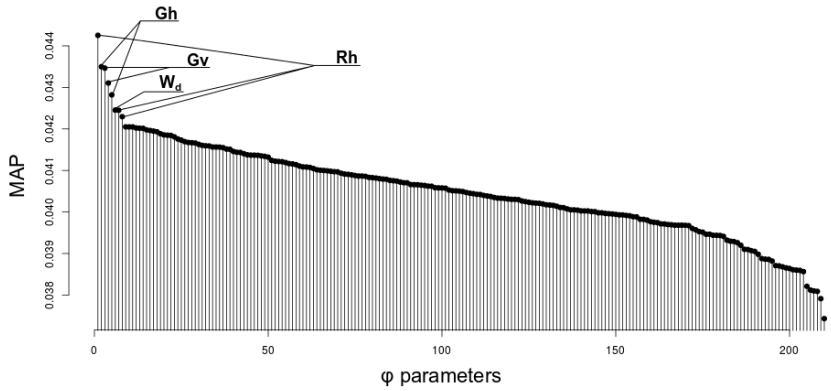
Figure 5.19: Test box–*Real* BBRI: $\mathcal{M}_0$ *Difference Analysis* for the constant temperature period.

already imposed on the model as well as the vertical solar radiation on the south wall of the box, and it was difficult to identify model improvements.

The remaining variable was the relative humidity, which is used in ESP-r in the calculation of the sky temperature, used to predict the longwave radiation heat flux towards the sky. This aspect was neglected at first in $\mathcal{M}_0$. The base model did not considered the wall emissivity, and while the global horizontal long wave radiation was provided, it was at first decided to not use it. Therefore, taking into account this aspect was deemed to be the most immediate and easy to implement upgrade. In particular, the wall emissivity ($wall_{\boldsymbol{\epsilon}}$) was added to the calibration parameter set, and the sky temperature, calculated from the measured global horizontal long wave radiation, was imposed on the model.



(a) β parameters.



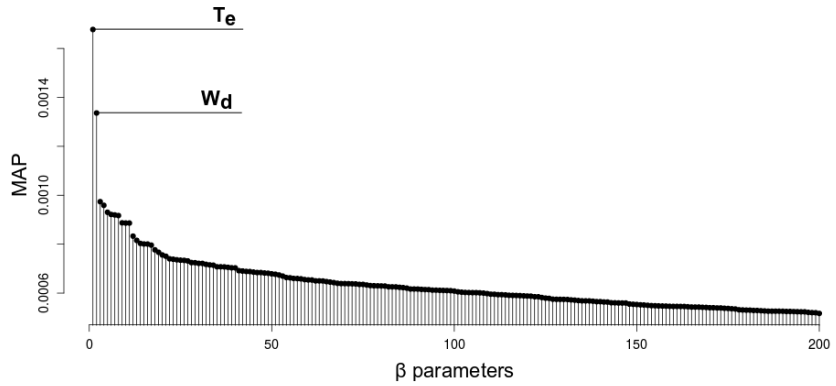
(b) ϕ parameters.

Figure 5.20: Test box–*Real* BBRI: $\mathcal{M}_0$ *Difference Analysis* for the free float period.

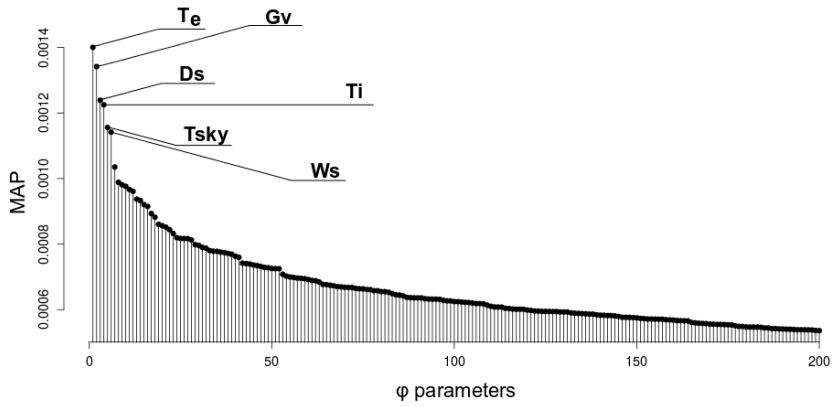
Final model ($\mathcal{M}_2$)

After calibrating $\mathcal{M}_1$, the *Difference Analysis* was still highlighting the possibility that wind speed ($\boldsymbol{Ws}$), wind direction ($\boldsymbol{Wd}$) and external temperature ($\boldsymbol{Te}$) could play a role in improving the match with the measurements (Figure 5.21 and 5.22).

Convection was neglected again for the same reason explained previously, and it was instead decided to test the air tightness hypothesis, by providing the model with an airflow network. It was known that, in order to install the equipment inside the box, a hole at the centre of its floor was made ready for the necessary wires. The researchers carrying out the experiment should then take care to adequately seal such an opening. This envelope feature and the window



(a) β parameters.



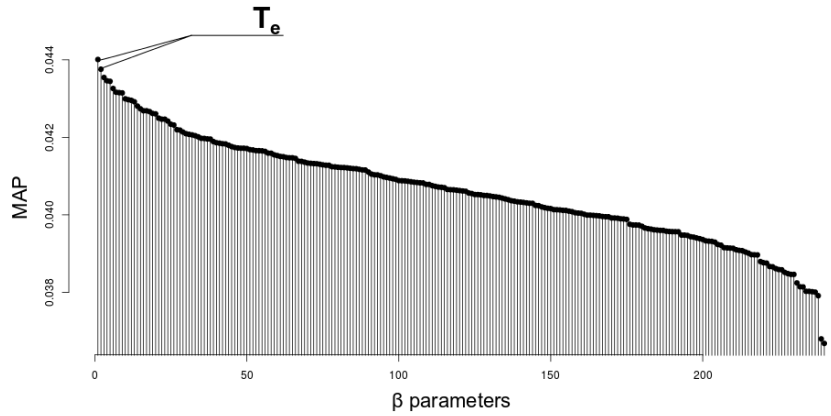
(b) ϕ parameters.

Figure 5.21: Test box–*Real* BBRI: $\mathcal{M}_1$ *Difference Analysis* for the constant temperature period.

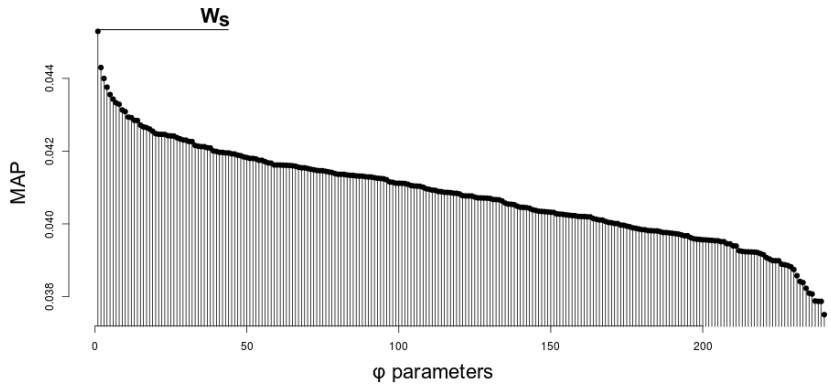
on the front wall, constitute possible places were small cracks may have occurred, thus causing small infiltration rates. Therefore an airflow network model was added to $\mathcal{M}_1$ in order to reflect these beliefs (Figure 5.23), and the length of its crack components ($window\_crack_{\boldsymbol{len}}$, and $floor\_crack_{\boldsymbol{len}}$) were included among the calibration parameters. Exponential prior probability density distributions with rate parameters equal to one were assumed for these variables, in order to reflect the information coming from the performed blower door tests, indicating infiltration rates close to zero.

5.2.4 Virtual

The *Virtual* experiment treated the problem of *equifinality*, which is a phenomenon common in over-parametrised model. In these cases the likelihood



(a) β parameters.



(b) ϕ parameters.

Figure 5.22: Test box–*Real* BBRI: $\mathcal{M}_1$ *Difference Analysis* for the free float period.

function will show multiple modes corresponding to the different solution vectors. Often only one solution is reasonable for the calibration problem, therefore it necessary to be able to handle multiple modes.

The phenomenon of *equifinality* was occurring while trying to calibrate $\mathcal{M}_0$ against synthetic observations, generated considering the calibration parameters set to the initial values listed in table 5.8, and as boundary conditions those determined by the FFL BBRI experiment. By sampling the joint posterior distribution with AIS, two modes were clearly identified for the parameters $wall_{\boldsymbol{k}}$ and $wall_{\boldsymbol{\rho}}$. These modes are depicted in Figure 5.24, where the marginal posterior probability density distributions are represented, and in Figure 5.25a, where the joint posterior probability density distribution for these two variables is displayed.

The parameter values associated with these two extrema are listed in Table 5.9. MODE1 is the sought solution, consisting in MAP estimates practically equal

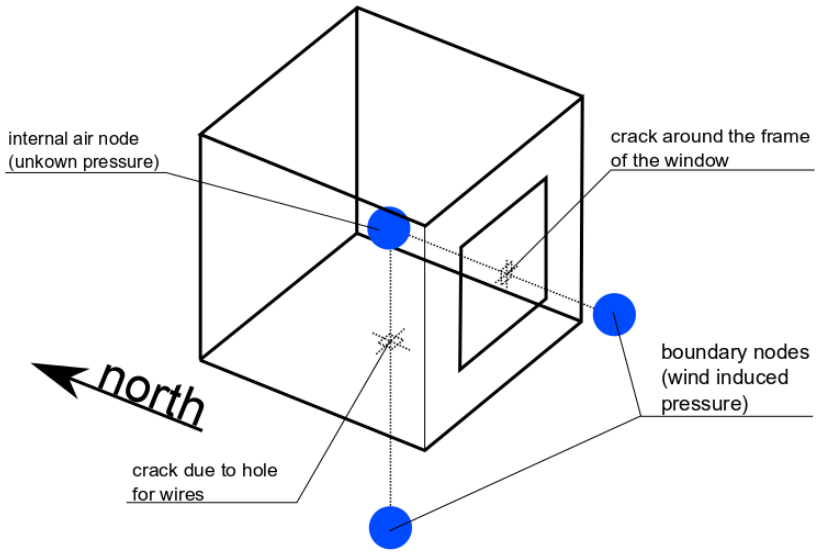


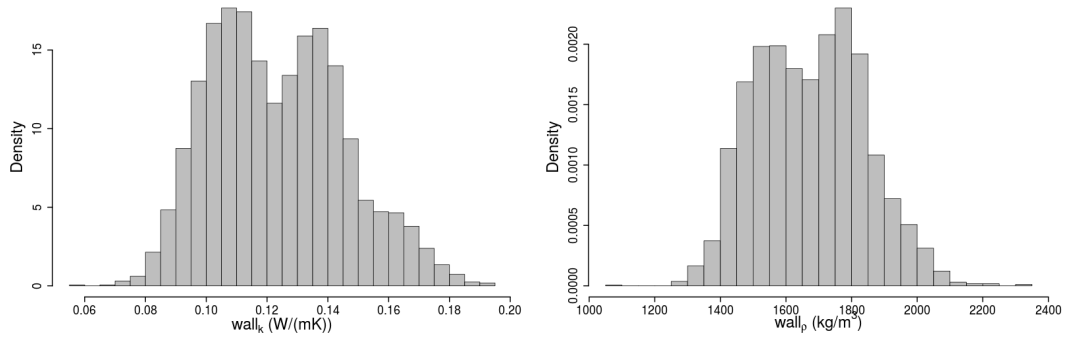
Figure 5.23: Test box at the BBRI test site.

Table 5.9: Test box–*Virtual*: MAP values for the two identified modes in the BBRI FFL experiment

PARAMETERS	MODE1	MODE2
$wall_{\boldsymbol{k}}$ $(W/(mK))$	0.109	0.135
$wall_{\boldsymbol{\rho}}$ (kg/m^3)	1545.000	1785.000
$wall_{\boldsymbol{\alpha}}$ (-)	0.230	0.220
$window_{\boldsymbol{R}}$ (m^2K/W)	0.289	0.299
$glass_{\boldsymbol{tr}}$ (-)	0.617	0.621

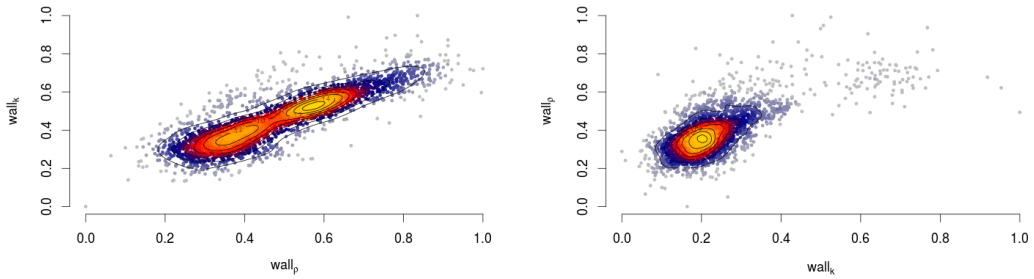
to the initial parameter values in Table 5.8, while MODE2 proposed overestimated values of wall conductivity and wall density. As Figure 5.26 shows, the two input vectors give model outputs fitting almost equally well the synthetic observations.

The parameters $wall_{\boldsymbol{k}}$ and $wall_{\boldsymbol{\rho}}$ were strongly correlated (correlation = 0.87), and having effects of similar magnitudes on the model output, so that by moving on the line, in the relative plane, defined by the two modes in Figure 5.25a (i.e. keeping constant ratio between $wall_{\boldsymbol{k}}$ and $wall_{\boldsymbol{\rho}}$), it is possible to periodically find good solutions to the calibration problem. This represent an identifiability problem, which can be avoided by using more information in performing the calibration step. By augmenting the amount of information, the searching of a solution is constrained in a better defined parameter space, therefore less subject to present multiple modes. There are two ways to add information: by setting up



(a) $wall_{\boldsymbol{k}}$ (b) $wall_{\boldsymbol{\rho}}$

Figure 5.24: Test box–*Virtual*: marginal posterior distributions considering only the FFL BBRI dataset.



(a) considering only FFL dataset. (b) considering CT and FFL dataset.

Figure 5.25: Test box–*Virtual*: joint posterior distributions for $wall_{\boldsymbol{k}}$ and $wall_{\boldsymbol{\rho}}$. samples are normalized in [0,1] to aid visualisation.

more informative prior probability distributions, or by using more data. In this case, the latter approach was adopted. In particular, two probabilistic models were built, relatively to the CT and FFL BBRI experiments, and the calibration was carried out by considering as likelihood equation the factorisation of the likelihoods of the these two models.

As Figures 5.25b, 5.27 highlight, the consideration of the data coming from the two experiments was successful in greatly improving the identifiability of the $wall_{\boldsymbol{k}}$ and $wall_{\boldsymbol{\rho}}$. Their joint and marginal posterior distributions are now bell shaped presenting only one mode, which identifies the correct values (Table 5.10). Even more, coupling the two data sets reduced the correlation between the two variables (correlation = 0.61). Thus the same calibration set up was used in processing the real measured data.

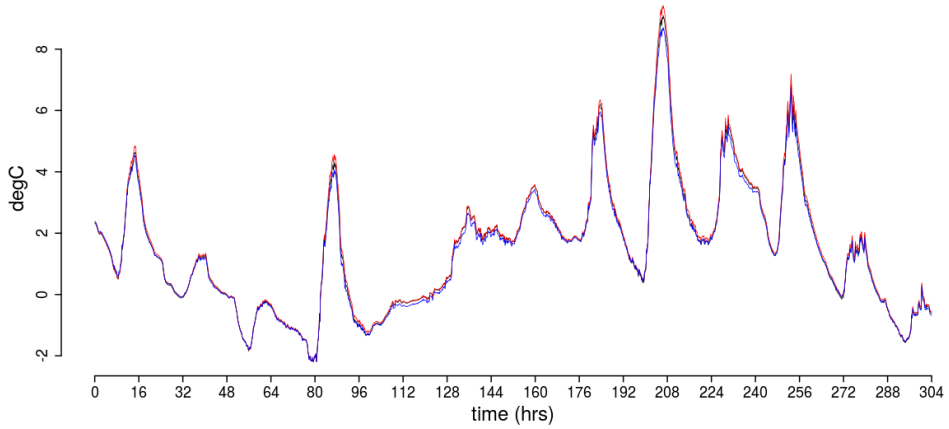
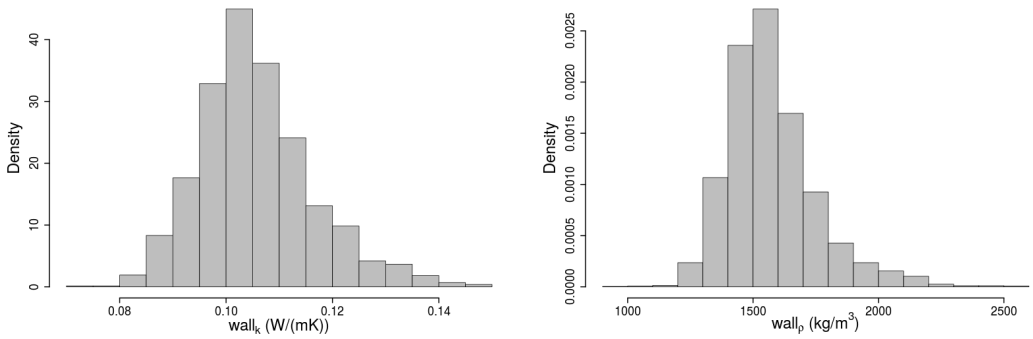


Figure 5.26: Test box–*Virtual*: comparison between synthetic observations (black line), model output for MODE1 (red line) and model output for MODE2 (blue line).

Table 5.10: Test box–*Virtual*: MAP estimates, 2.5%, 50% and 97.5% quantiles for calibration parameters, considering the CT and FFL datasets.

PARAMETERS	MAP	$Q2.5\%$	$Q50\%$	$Q97.5\%$
$wall_{\boldsymbol{k}}$	0.104	0.088	0.105	0.164
$wall_{\boldsymbol{\rho}}$	1528.587	1297.588	1544.244	2029.192
$wall_{\boldsymbol{\alpha}}$	0.230	0.153	0.219	0.330
$window_{\boldsymbol{R}}$	0.289	0.220	0.287	0.369
$glass_{\boldsymbol{tr}}$	0.631	0.574	0.632	0.693



(a) $wall_{\boldsymbol{k}}$ (b) $wall_{\boldsymbol{\rho}}$

Figure 5.27: Test box–*Virtual*: marginal posterior distributions considering the CT and FFL BBRI datasets.

5.2.5 Real

Firstly the three above described models were calibrated against the BBRI dataset and ranked according to Bayes Factors. Secondly the two best models were employed to characterise the test box according to the data from the LECE experiment. The sensible heat load needed to keep the test box internal temperature to the prescribed set point, and the internal dry bulb temperature resulting from the ROLBS solicitations were used as calibration targets respectively in the former and latter calibration experiments.

BBRI

Table 5.11, lists the MAP estimates and the confidence intervals for the calibration parameters relative to the three models used in representing the real test box.

$\mathcal{M}_1$ and $\mathcal{M}_2$ returned estimates for their common free parameters, substantially in agreement with each other. $\mathcal{M}_0$, while having similar $wall_{\boldsymbol{k}}$ and $window_{\boldsymbol{R}}$, presented values of the other variables slightly higher. Comparisons between the observed sensible heat load and internal dry bulb air temperature, and the relative model predictions are shown in Figure 5.28. As expected, since the similar calibration parameters estimates, $\mathcal{M}_1$ and $\mathcal{M}_2$ were able to provide similar fits of the observed data. This was further highlighted by the calculated goodness of fit criteria (Table 5.14), which for these two models had values quite close and significantly lower than those for $\mathcal{M}_0$. $\mathcal{M}_1$ had better performances during the free-float training period, while $\mathcal{M}_2$ was more accurate in the constant temperature training and in the free-float test periods. A comparison between $\mathcal{M}_2$ predictions and observations during the test period is shown in Figure 5.29.

The calculated values of marginal likelihood, indicated that the two upgrades implemented, namely a more accurate consideration of the longwave heat flux towards the sky, and the implementation of an airflow network, were decisive in achieving a better representation of the observed processes. In particular $log_{10}(\mathcal{B}_{2,1})$ was equal to 5.28, suggesting that despite the similar predictive capabilities of $\mathcal{M}_1$ and $\mathcal{M}_2$ the latter must be preferred. The posterior probability density distributions for the calibration parameters inferred through $\mathcal{M}_2$ are shown in Figure 5.30. Particularly interesting is the posterior distribution of $window\_crack_{\boldsymbol{len}}$, which seems to support the hypothesis that a small crack was present around the window frame causing small infiltrations.

Table 5.11: Test box–*Real* BBRI: MAP estimates, 2.5%, 50% and 97.5% quantiles for calibration parameters.

(a) $\mathcal{M}_0$

PARAMETERS	MAP	$Q2.5\%$	$Q50\%$	$Q97.5\%$
$wall_{\boldsymbol{k}}$	0.139	0.124	0.141	0.194
$wall_{\boldsymbol{\rho}}$	1062.799	638.35	1044.391	1243.059
$wall_{\boldsymbol{\alpha}}$	0.380	0.107	0.366	0.394
$window_{\boldsymbol{R}}$	0.862	0.517	0.846	0.929
$glass_{\boldsymbol{tr}}$	0.699	0.506	0.683	0.715

(b) $\mathcal{M}_1$

PARAMETERS	MAP	$Q2.5\%$	$Q50\%$	$Q97.5\%$
$wall_{\boldsymbol{k}}$	0.135	0.126	0.135	0.146
$wall_{\boldsymbol{\rho}}$	973.748	876.515	979.471	1093.676
$wall_{\boldsymbol{\alpha}}$	0.263	0.198	0.265	0.344
$window_{\boldsymbol{R}}$	0.904	0.593	0.870	0.943
$glass_{\boldsymbol{tr}}$	0.564	0.505	0.578	0.688
$wall_{\boldsymbol{e}}$	0.723	0.705	0.739	0.891

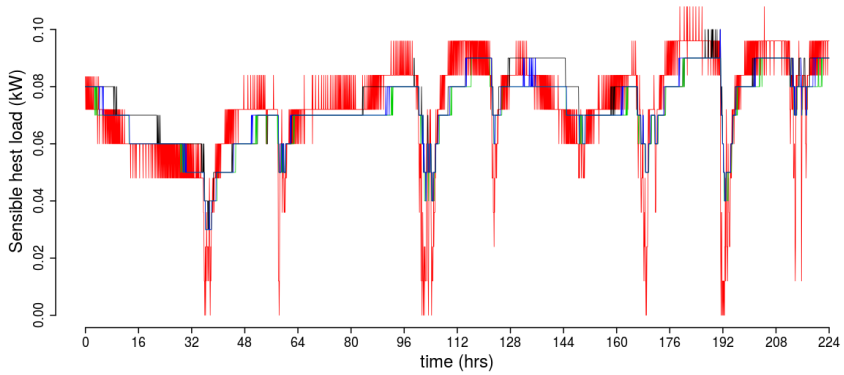
(c) $\mathcal{M}_2$

PARAMETERS	MAP	$Q2.5\%$	$Q50\%$	$Q97.5\%$
$wall_{\boldsymbol{k}}$	0.138	0.127	0.140	0.165
$wall_{\boldsymbol{\rho}}$	943.006	878.631	946.300	1037.310
$wall_{\boldsymbol{\alpha}}$	0.264	0.213	0.263	0.306
$window_{\boldsymbol{R}}$	0.872	0.537	0.826	0.941
$glass_{\boldsymbol{tr}}$	0.535	0.504	0.559	0.709
$wall_{\boldsymbol{e}}$	0.719	0.703	0.730	0.877
$floor\_crack_{\boldsymbol{len}}$	0.008	0.003	0.011	0.046
$window\_crack_{\boldsymbol{len}}$	0.024	0.004	0.036	0.167

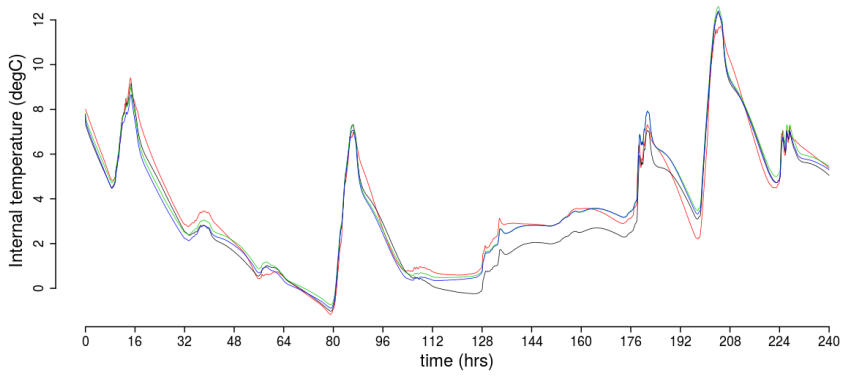
Table 5.12: Test box–*Real* BBRI: goodness of fit criteria.

MODEL	PERIOD	NMBE (%)	CVRMSE (%)	GOF (%)	$log_{10}(L)$
$\mathcal{M}_0$	CT–training[1]	2.30	16.92	12.07	32.45
	FFL–training[2]	9.26	16.86	13.60	
	FFL-test[3]	-18.54	23.93	21.41	
$\mathcal{M}_1$	CT–training[1]	2.54	15.73	11.26	34.58
	FFL–training[2]	-0.39	12.42	8.78	
	FFL–test[3]	-3.12	9.16	6.84	
$\mathcal{M}_2$	CT–training[1]	1.01	15.41	10.92	40.86
	FFL–training[2]	3.41	13.29	9.71	
	FFL–test[3]	3.48	9.06	6.86	

(1) $\bar{\boldsymbol{y}}^* = 0.07\ kW$; (2) $\bar{\boldsymbol{y}}^* = 3.70\ ^oC$; (3) $\bar{\boldsymbol{y}}^* = 3.78\ ^oC$.



(a) constant temperature



(b) free float

Figure 5.28: Test box–*Real* BBRI: comparison between observations (red) and model predictions ($\mathcal{M}_0$: black, $\mathcal{M}_1$: green, $\mathcal{M}_2$: blue), during the training period.

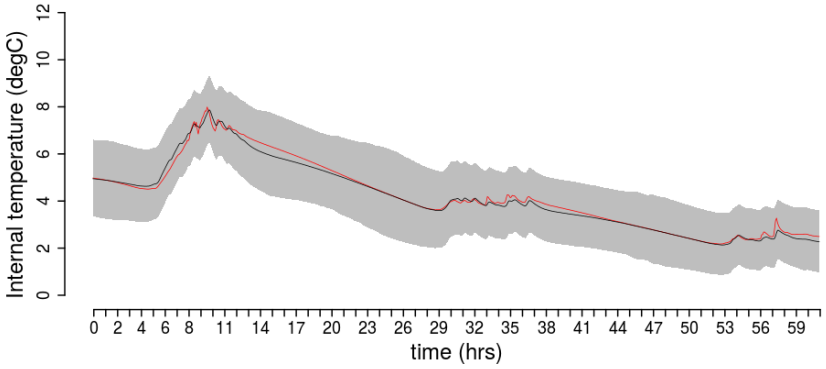
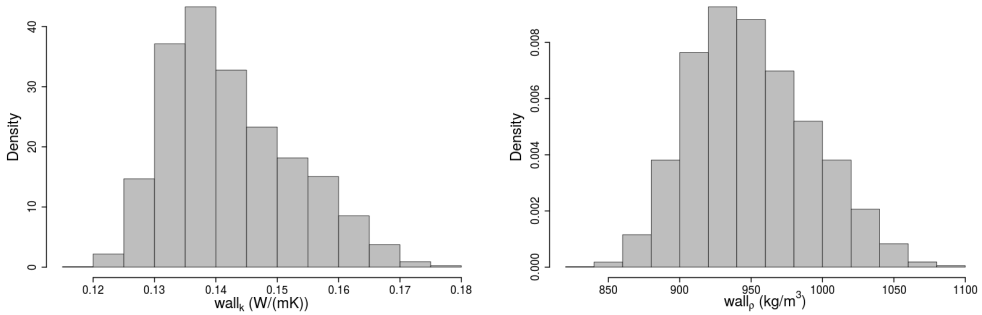
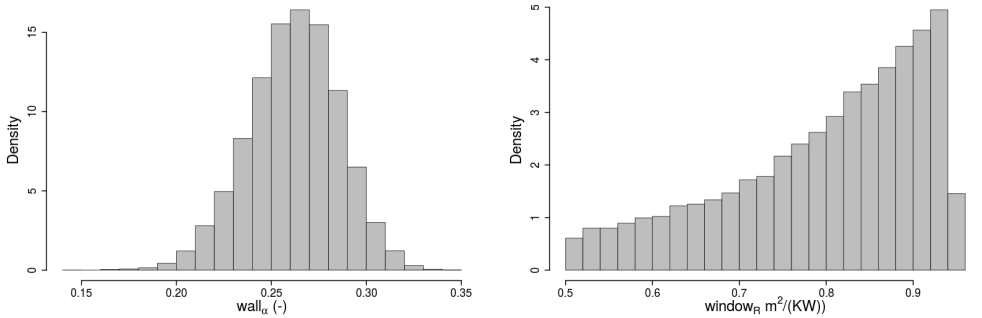


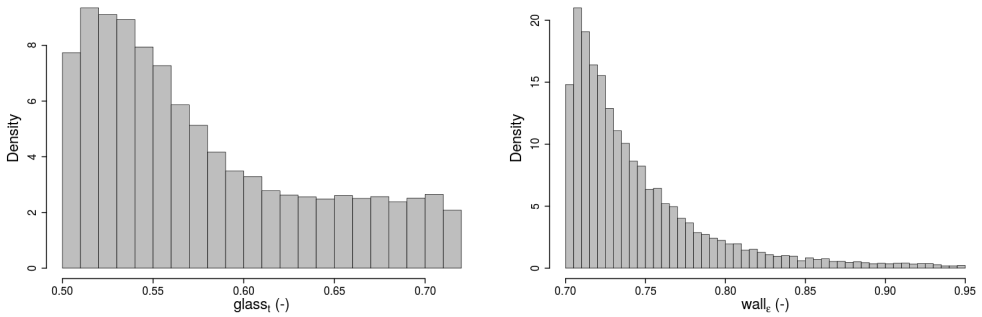
Figure 5.29: Test box–*Real* BBRI: comparison between $\mathcal{M}_2$ predictions (black) and observation (red) during the test period. In grey the 95% c.i. for the model predictions are indicated.



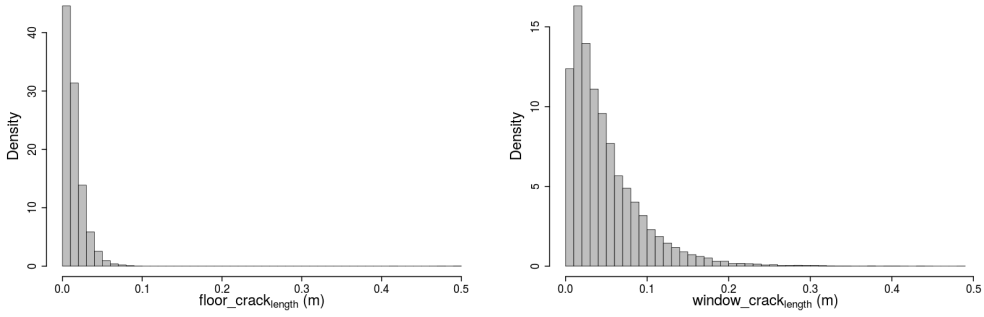
(a) $wall_{\boldsymbol{k}}$ (b) $wall_{\boldsymbol{\rho}}$



(c) $wall_{\boldsymbol{\alpha}}$ (d) $window_{\boldsymbol{R}}$



(e) $glass_{\boldsymbol{tr}}$ (f) $wall_{\boldsymbol{\epsilon}}$



(g) $floor\_crack_{\boldsymbol{len}}$ (h) $window\_crack_{\boldsymbol{len}}$

Figure 5.30: Test box–*Real* BBRI: $\mathcal{M}_2$ - empirical posterior density distributions for calibration parameters.

LECE

From the previous analysis it emerged that $\mathcal{M}_2$ provided the best characterisation of the test box in the BBRI experiment. However its predictive performance, as assessed by NMBE, CVRMSE and GOF, were very similar to those of $\mathcal{M}_1$. Therefore it was decided to use both these models in undertaking the investigation of the LECE dataset. The numerical results, comprising MAP estimates and confidence intervals for the calibration parameters, are listed in Table 5.13.

Table 5.13: Test box–*Real* LECE: MAP estimates, 2.5%, 50% and 97.5% quantiles for calibration parameters

(a) $\mathcal{M}_1$

PARAMETERS	MAP			
$wall_{\boldsymbol{k}}$	0.107	0.103	0.107	0.112
$wall_{\boldsymbol{\rho}}$	687.981	656.946	687.681	719.048
$wall_{\boldsymbol{\alpha}}$	0.347	0.291	0.345	0.391
$window_{\boldsymbol{R}}$	0.625	0.515	0.690	0.925
$glass_{\boldsymbol{tr}}$	0.543	0.507	0.545	0.593
$wall_{\boldsymbol{e}}$	0.753	0.705	0.781	0.935
$ROLBS_{\boldsymbol{C/R}}$	0.99	0.968	0.988	0.993

(b) $\mathcal{M}_2$

PARAMETERS	MAP			
$wall_{\boldsymbol{k}}$	0.114	0.107	0.114	0.120
$wall_{\boldsymbol{\rho}}$	692.618	619.512	692.643	761.072
$wall_{\boldsymbol{\alpha}}$	0.316	0.260	0.314	0.366
$window_{\boldsymbol{R}}$	0.644	0.524	0.659	0.914
$glass_{\boldsymbol{tr}}$	0.695	0.570	0.681	0.716
$wall_{\boldsymbol{e}}$	0.745	0.705	0.802	0.942
$ROLBS_{\boldsymbol{C/R}}$	0.987	0.954	0.983	0.997
$floor\_crack_{\boldsymbol{len}}$	0.008	0.003	0.011	0.076
$window\_crack_{\boldsymbol{len}}$	0.027	0.002	0.047	0.326

As before, these two models present estimates and posterior uncertainties for their free parameters particularly similar, as well as similar values of goodness of fit criteria (Table 5.14). $\mathcal{M}_1$ depicted slightly better the dynamics of the real test box, since a lower $CVRMSE$ while $\mathcal{M}_2$ performed averagely better since its lower $NMBE$. $\mathcal{M}_1$, presents also a lower GOF, but the differences of these three indexes were in the order of the 0.3%, making it it difficult to select one model over the other. $\mathcal{M}_1$ and $\mathcal{M}_2$ could be considered equally adequate in describing the measured internal dry air bulb temperature, according to these three goodness of fit criteria. Indeed, as shown by Figure 5.31, they provide similar fit of the

Table 5.14: Test box–*Real* LECE: goodness of fit criteria.

MODEL	PERIOD	NMBE (%)	CVRMSE (%)	GOF	$log_{10}(p(\boldsymbol{Y}^*\|\mathcal{M}_i))$
$\mathcal{M}_1$	ROLBS–training	-0.42	1.47	1.08	3.12
$\mathcal{M}_2$	ROLBS–training	0.39	1.72	1.24	6.97

$\bar{\boldsymbol{y}}^* = 34.784\ ^oC$

observed field data.

Their marginal likelihoods provided clearer information about their ranking. $log_{10}(\mathcal{B}_{2,1})$ was calculated to be 3.85, thus supporting $\mathcal{M}_2$ as the best model. The posterior probability density distributions for its calibration parameters are depicted in Figures 5.32 and 5.33.

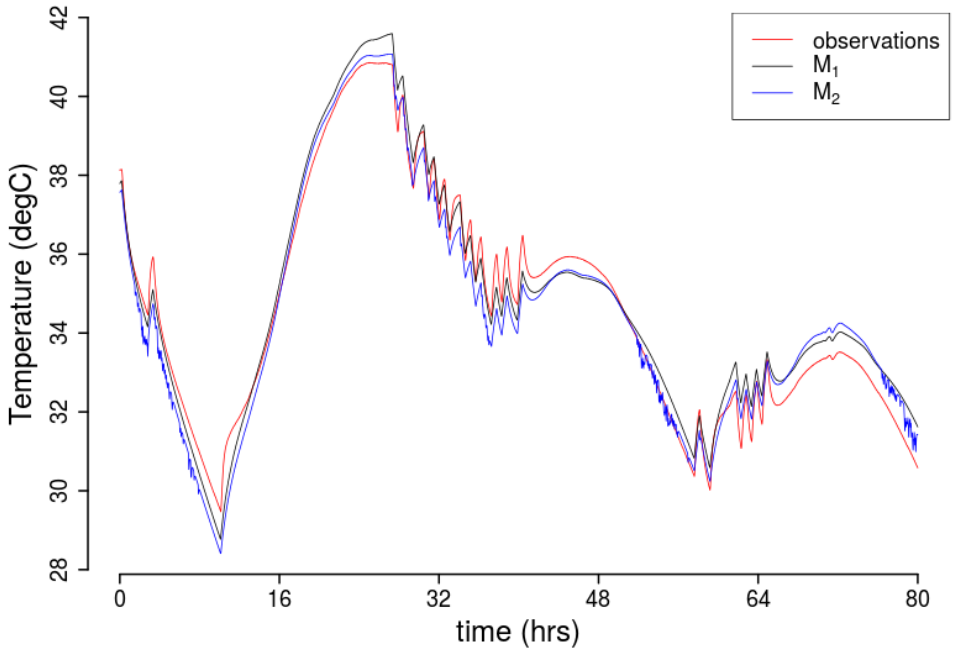
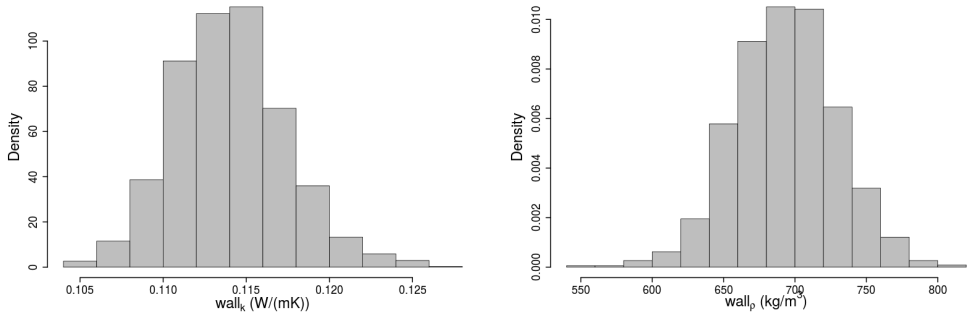
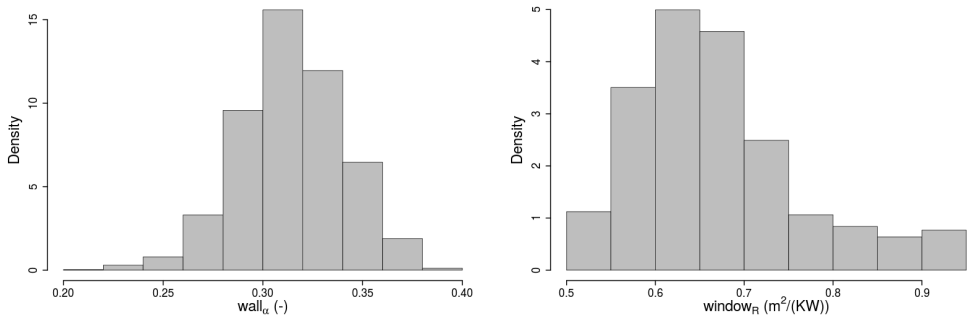


Figure 5.31: Test box–*Real* LECE: comparison between training observations and model predictions.

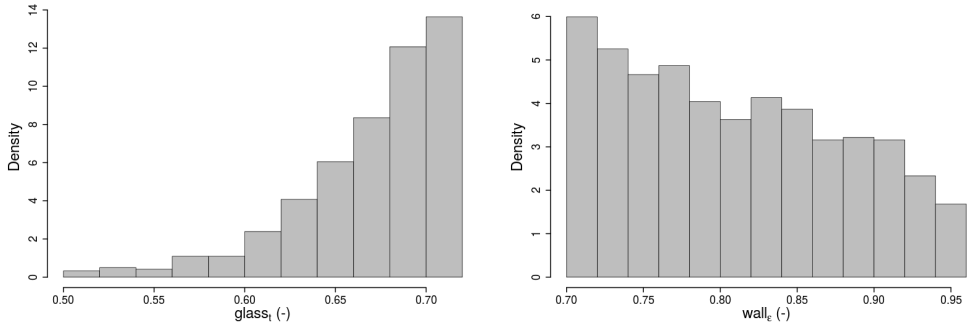
In this case the posterior probability density distribution for $window\_crack_{\boldsymbol{len}}$ did not show a clear peak different from 0. Nonetheless the inferred values for the two crack parameters, were strangely in agreement with those calculated from the BBRI calibration. This might just have been a coincidence, but it seemed to support the hypothesis that the infiltration in the test box was not negligible as assured by the specifications.



(a) $wall_{\boldsymbol{k}}$ (b) $wall_{\boldsymbol{\rho}}$



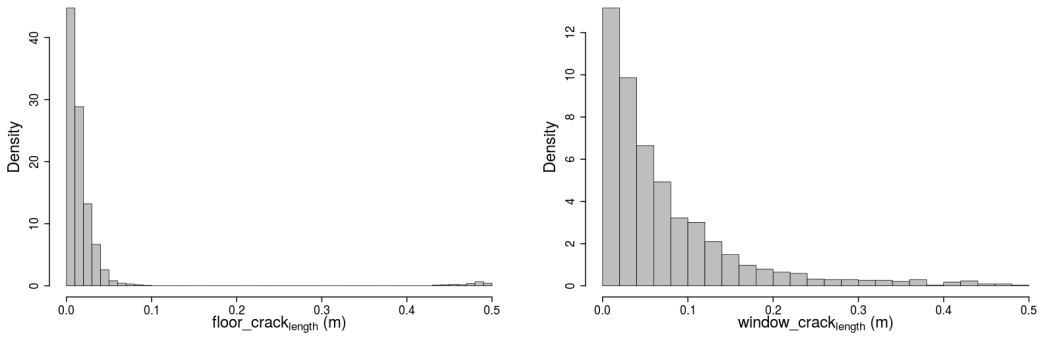
(c) $wall_{\boldsymbol{\alpha}}$ (d) $window_{\boldsymbol{R}}$



(e) $glass_{\boldsymbol{tr}}$ (f) $wall_{\boldsymbol{e}}$

Figure 5.32: Test box–*Real* LECE: $\mathcal{M}_2$ - empirical posterior density distributions for calibration parameters.

$\mathcal{M}_2$ was then used to predict the test box internal temperature during the free float period not provided with the calibration data and not used during the analysis, without achieving a good agreement. In particular, the model always overestimated the measurements.



(a) $floor\_crack_{len}$ (b) $window\_crack_{len}$

Figure 5.33: Test box–*Real* LECE: $\mathcal{M}_2$ - empirical posterior density distributions for calibration parameters.

5.2.6 Discussion

The *Virtual* experiment showed that even a relative simple model, as that considered, can be subject to *equifinality*, and, consequently, to identifiability problems. In most situations it is preferable to have only one solution to the calibration problem, especially considering that when multiple modes are present in the likelihood function some of them may just be due to an over-fit of the considered data, and the relative combination of input parameters may perform poorly when used to make predictions in different conditions from those considered in training the model.

Through this example, it was shown how to effectively circumvent these problems, by considering multiple datasets from different experiments. In this way, the same combination of model parameters has to provide good explanations for data observed under different boundary conditions, decreasing the possibilities of more than one combination of inputs providing an adequate match with the observed data.

Concerning the calibration exercises performed considering data coming for the real experiments, it was possible to achieve a good match with the target measurements coming from the BBRI experiment, while the analysed models showed difficulties in providing good representations of the test box subject to the ROLBS heat pulses. In particular the predictions of $\mathcal{M}_1$ and $\mathcal{M}_2$ were particularly inaccurate in predicting the internal temperature of the test box observed during the free float period of the LECE experiment.

A year after undertaking of the above described analyses, the real composition of the test box was disclosed, as well as estimates of its Heat Loss Coefficient

(HLC), which were derived through experiments in an environmental chamber. The former information were made available by the research team from the University of KU Leuven (KUL), participating in the Annex, and in the following this is referred to as the KUL dataset. The experiments in the environmental chamber were performed by a team of the Czech Technical University (CTU) in Prague, also participating at the Annex, and the resulting data will be referred to as the CTU dataset.

According to the KUL data the window frame had conductivity, density and specific heat, respectively equal to 0.17 $W/(mK)$, 700 kg/m^3 and 2070 $J/(kgK)$. For the glazing a U-value of 1.1 $W/(m^2K)$ was given, and the composition of the walls is described in Table 5.15.

Table 5.15: KUL data: composition of the walls (from inside to outside).

LAYER	THICKNESS (mm)	k $(\frac{W}{mK})$	ρ $(\frac{kg}{m^3})$	cp $(\frac{J}{kgK})$
$fibre\_cement\_board1$	36	0.35	1250	1470
XPS insulation	60	0.034	25	1450
$fibre\_cement\_board2$	16	0.35	1250	1470
$fibre\_cement\_board3$	8	0.60	1925	1018

The schedule adopted for the CTU experiments and the resulting values of HLC for the test box are respectively listed in Table 5.16 and graphically represented in Figure 5.34. These experiments aimed to reach approximately a steady

Table 5.16: CTU experiment schedule.

#	$\boldsymbol{Ti}$ (^oC)	Heat Flux (W)	$\boldsymbol{Te}$ (^oC)	DURATION (days)
1	28.8	172	-17.1	7
2	27.8	148	-11.5	4
3	26.9	104	-0.5	6
4	26.9	89	3.3	6
5	28.5	130	-6.2	6
6	51.1	168	11.0	4
7	49.7	150	14.0	5
8	50.2	207	0.3	6

state situation, and then to determine the HLC as the ratio of the measured heat flux and the difference between internal and external temperatures. In particular, the red dots in Figure 5.34 indicate the HLC values calculated by assuming as internal temperature the mean internal surface temperature. Similarly the blue dots and the black dots are the HLC values derived by considering as internal temperatures respectively the mean internal temperatures, and a weighted mean of surface temperatures, air temperature and heater surface temperature.

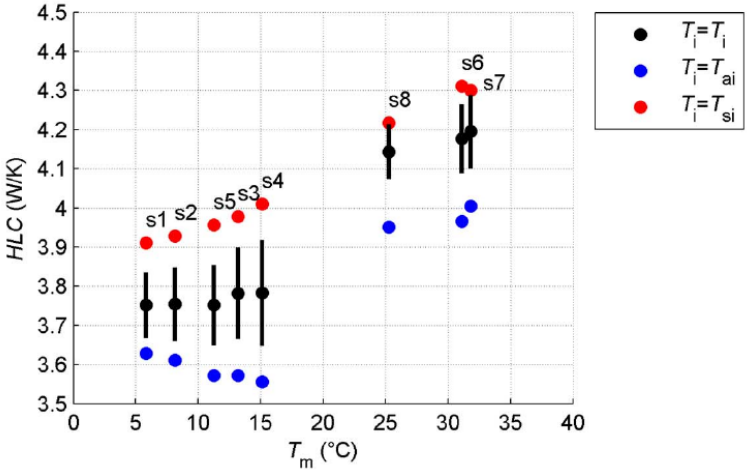


Figure 5.34: Heat loss coefficient from CTU experiment in climatic chamber.

As explained during the description of the models, the parameter $wall_{\boldsymbol{k}}$ and $wall_{\boldsymbol{\rho}}$ are not equivalent to the physical conductivity and density of the walls, but they represent the conductivity and the density of a homogeneous component having the same heat loss coefficient (i.e. including thermal bridges) and effective thermal capacity of the real test box.

A parameter comparable to $wall_{\boldsymbol{k}}$ was derived from the CTU data, represented by the red dots in Figure 5.34, which accounted only for the heat transfer due to conduction. By accordingly processing these results and by considering an error of the 3%, as suggested by the team performing the experiments, $wall_{\boldsymbol{k}}$ should have values between 0.063 $W/(mK)$ and 0.095 $W/(mK)$. Similarly, from the KUL data it was possible to calculate the conductivity and the density of an homogeneous component having same thermal resistance and heat capacity as the real wall, which respectively are 0.062 $W/(mK)$ and 590 kg/m^3. Therefore a reasonable confidence interval for $wall_{\boldsymbol{k}}$ according to the KUL and CTU data was deemed to be between 0.062 $W/(mK)$ and 0.095 $W/(mK)$.

The estimates for $wall_{\boldsymbol{\rho}}$ were considered to be compatible with the given data. In particular, the true wall construction is composed by materials with very different thermal capacities. The external and internal fibre cement boards have very high values of density and specific heat compared to the insulation layer. Especially, the internal fibre cement board will contribute the most to the effective thermal capacity of the test box. Hence, it was deemed plausible for $wall_{\boldsymbol{\rho}}$ to present relatively higher estimates respect to the analytical values obtained by averaging the densities of the different layers. Also the inferred $window_{\boldsymbol{R}}$ returned sensible values for the the U-value of the window, that is 0.947 $W/(m^2K)$ for the

BBRI calibration, and 1.207 $W/(m^2K)$ for the LECE calibration.

Although the estimates returned by the analysis on the LECE dataset were close to the derived upper bound, the inferred values of $wall_{\boldsymbol{k}}$ were substantially higher than what they should be, and outside the confidence interval derived from the provided specification and data from the environmental chamber tests. Even more the inferred values are quite different depending on the analysed dataset. In particular, while it was expected that higher $wall_{\boldsymbol{k}}$ values would have been obtained in the LECE calibration compared to the BBRI one, due to the higher temperatures to which the test box was subject, the opposite was observed. These discrepancies between the $wall_{\boldsymbol{k}}$ estimates and values derived according to the KUL and CTU data were probably due to inadequacy in the models used for calibration. The overestimation of $wall_{\boldsymbol{k}}$ could be due to the flaws of the models in adequately representing the different convection regimes influencing the BBRI and LECE experiments, since these phenomena were not considered in detail during the previously explained analysis. ESP-r calculates external and internal convection heat coefficients at each time step of the simulation according to empirical relations developed in particular experiments on real scale building components. The extrapolation of their results is recognised to involve a great uncertainty, probably increased by the particular dimensions of the test box and peculiarity of the conditions imposed by the ROLBS experiment. The internal convection is influenced by the heating device employed, by its position and by the imposed heating schedule. Even more due to the dimensions of the box, the glazing of the window represented a significant fraction of the total envelope area. Its internal surface temperature was lower than the internal air temperature, possibly causing significant convective motions and relatively high convection heat transfer coefficients. An analogous reasoning, can be done for the external convection. This process is mainly influenced by the wind speed and the wind direction. Due to its geometrical proportions, the wind velocity field in the proximity of the box, may well be different from that in the proximity of a real scale building. Thus, ESP-r probably underestimated the convection heat transfer coefficients, which, in turn, caused an overestimation of $wall_{\boldsymbol{k}}$. Also the approximation of the actual wall component with a one layer construction, may have influenced the estimation of $wall_{\boldsymbol{k}}$ as well, but it was believed that this simplification had more influence on the dynamic behaviour of the box, since it changes the distribution of the thermal mass, within the walls, floor and ceiling.

In order to test the theses just explained, two further calibration experiments were performed, considering the data from the constant temperature phase of the

BBRI experiments and from the ROLBS test of the LECE experiment. In particular, the provided datasets included external and internal surface temperatures as well as the corresponding conduction heat fluxes at the internal surfaces of the box. This made possible the analysis of the heat transfer from surface to surface instead of that from air to air, thus excluding convection. Models, similar to the one depicted by Figure 5.4, were built for the BBRI CT experiment ($\mathcal{M}_3$) and LECE ROLBS experiment ($\mathcal{M}_4$), imposing on the internal and external surfaces the relative measured temperatures. In the former, the walls of the test box were represented still, with one layer construction components, while in the latter the actual layer disposition was considered. The aim was to reproduce the measured heat fluxes. For both, these two models, the specific heats of the materials were fixed at 1600 J/kgK. Conductivities and densities were considered as calibration parameters. The achievement of results in agreement with the KUL and CTU data, proved that, indeed, the modelling of the convection and the considered different layer distribution in the test box construction components, were causes of major model inadequacies.

The $\mathcal{M}_3$ calibration yielded the estimates and confidence intervals indicated in Table 5.17 for $wall_{\boldsymbol{k}}$ and $wall_{\boldsymbol{\rho}}$. These values seem to be in agreement with

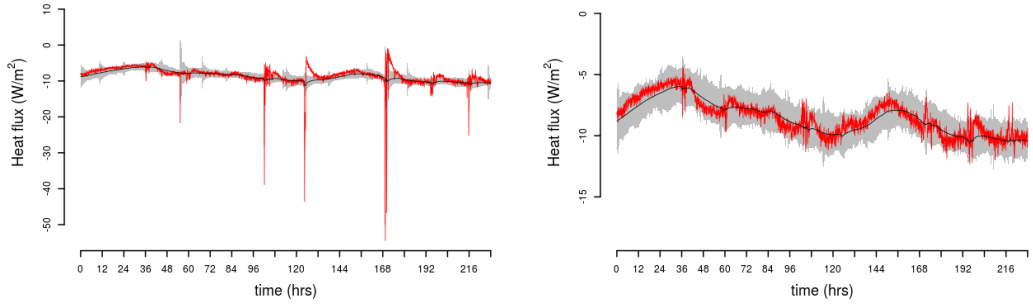
Table 5.17: Test box–*Real* BBRI: $\mathcal{M}_3$ - MAP estimates, 2.5%, 50% and 97.5% quantiles for calibration parameters

PARAMETERS	MAP	$Q2.5\%$	$Q50\%$	$Q97.5\%$
$wall_{\boldsymbol{k}}$	0.056	0.052	0.056	0.060
$wall_{\boldsymbol{\rho}}$	680.52	515.405	774.98	1073.484

the CTU and KUL data. In this case, $wall_{\boldsymbol{k}}$ does not include eventual thermal bridges and its estimate is very close to the value calculated from the KUL data (Table 5.15), that is 0.062 $W/(m^2K)$. $wall_{\boldsymbol{\rho}}$ is over estimated of the same order of magnitude as the previous calibration experiments, therefore, for the same reasons explained before, this datum is considered acceptable.

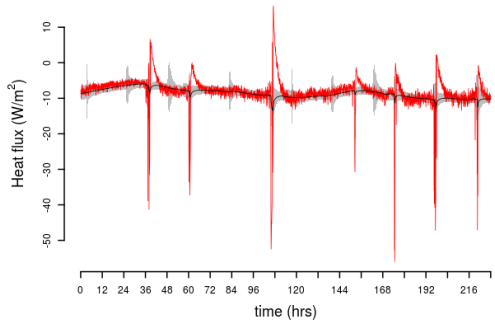
$\mathcal{M}_3$ was also able to return a particularly good fit of the target data (Figure 5.35), and the *Difference Analysis* did not highlight any of the considered variables boundary conditions as possible cause of discrepancies.

The calibration of $\mathcal{M}_4$ against the data collected during the LECE experiment returned the estimates and confidence intervals indicated in Table 5.18 for the conductivities and densities of the four layer making the walls of the test box, while the relative empirical posterior probability density distributions are represented in Figures 5.36 and 5.37.

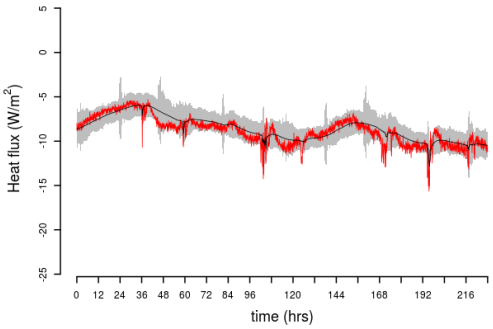


(a) left wall.

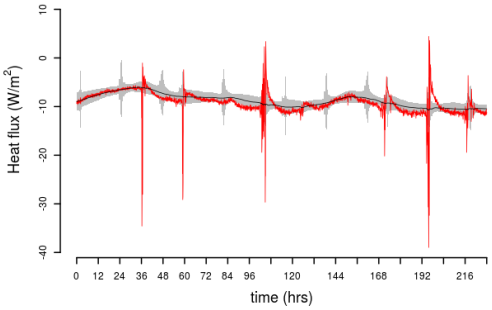
(b) back wall



(c) right wall.



(d) ceiling.



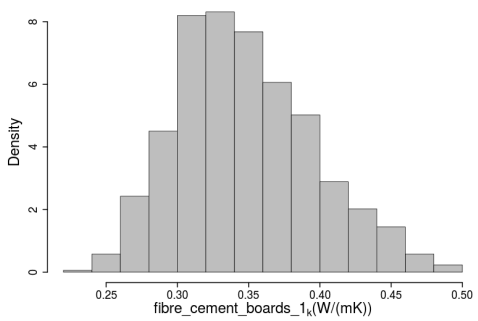
(e) floor.

Figure 5.35: Test box–*Real* BBRI: $\mathcal{M}_3$ - comparison between observed heat flux (red) and mode predictions (black). In grey are indicated the 95% c.i..

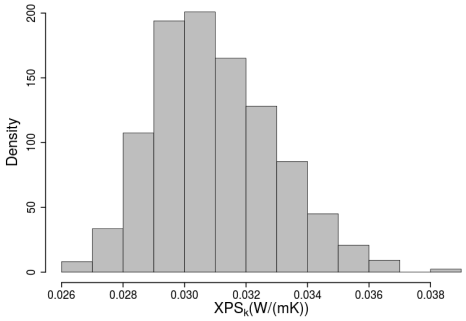
The properties of the two innermost layers were well identified. Their conductivities were very close to the value specified by the KUL data, while their densities, although showing reasonable estimates, appeared to be underestimated for $fibre\_cement\_board1$ and overestimated for XPS, also considering the different value assumed for the specific heat of the different materials. The parameters

Table 5.18: Test box–*Real* LECE: $\mathcal{M}_4$ - MAP estimates, 2.5%, 50% and 97.5% quantiles for calibration parameters

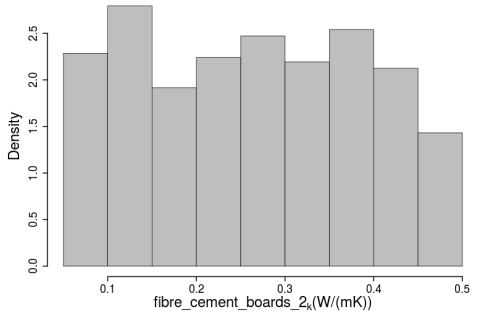
PARAMETERS	MAP	$Q2.5\%$	$Q50\%$	$Q97.5\%$
$fibre\_cement\_board1_{\boldsymbol{k}}$	0.339	0.267	0.341	0.449
$XPS_{\boldsymbol{k}}$	0.031	0.028	0.031	0.035
$fibre\_cement\_board2_{\boldsymbol{k}}$	0.199	0.066	0.266	0.480
$fibre\_cement\_board3_{\boldsymbol{k}}$	0.21	0.111	0.405	0.962
$fibre\_cement\_board1_{\boldsymbol{\rho}}$	811.322	633.317	824.121	1062.81
$XPS_{\boldsymbol{\rho}}$	337.531	107.869	323.033	473.791
$fibre\_cement\_board2_{\boldsymbol{\rho}}$	1727.033	586.186	1570.348	2413.117
$fibre\_cement\_board3_{\boldsymbol{\rho}}$	1017.329	573.970	1387.215	2424.971



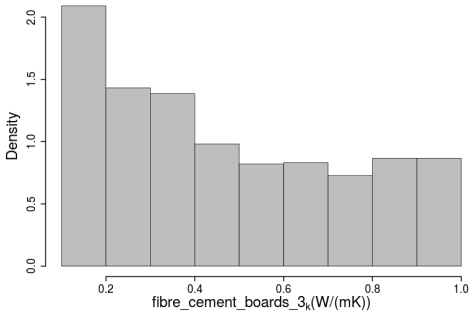
(a) left wall.



(b) back wall



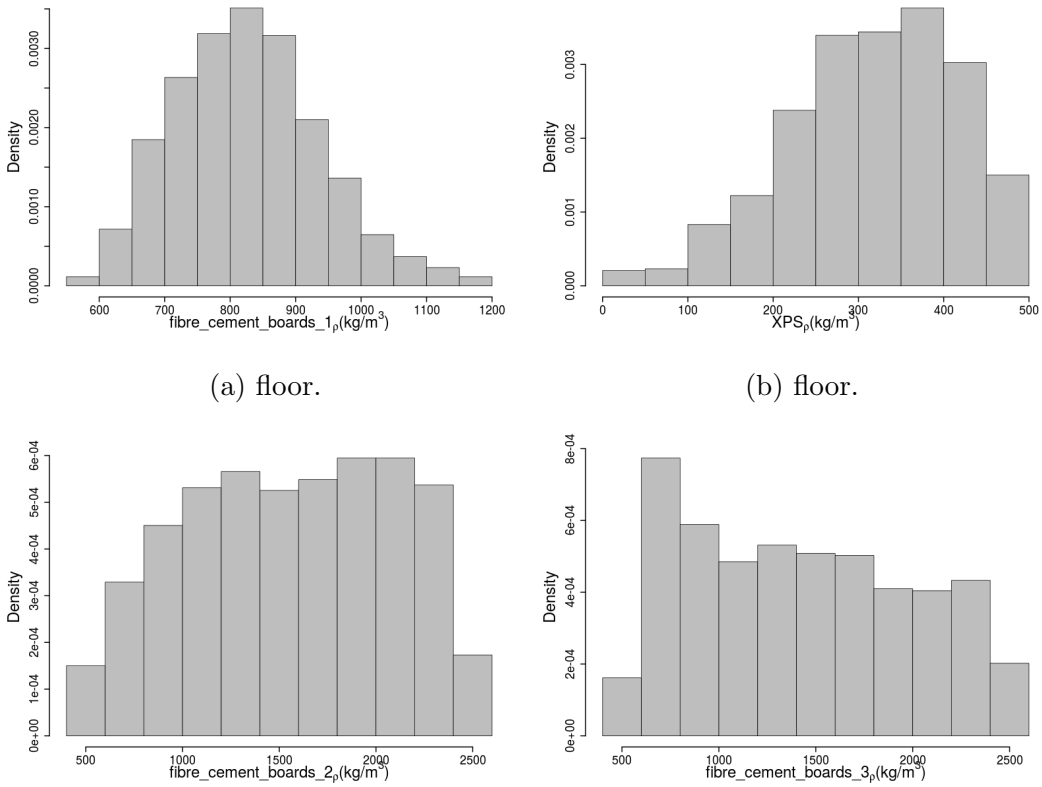
(c) right wall.



(d) ceiling.

Figure 5.36: Test box–*Real* LECE: $\mathcal{M}_3$ - calibration parameter posterior density distributions.

of $fibre\_cement\_board2$ and $fibre\_cement\_board3$ were not well identified and their estimates were not in agreement with the specifications. In particular, their posterior distributions were almost uniform. This was due to the very little effect that these variables had in determining the heat flux at the internal surfaces of the walls of the test box. Indeed, by a sensitivity analysis of $\mathcal{M}_4$, it was possible



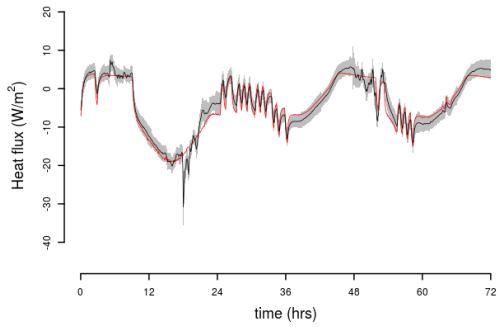
(a) floor. (b) floor.

(c) floor. (d) floor.

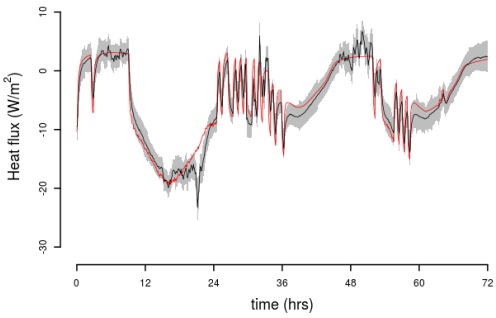
Figure 5.37: Test box–*Real* LECE: $\mathcal{M}_3$ - calibration parameter posterior density distributions.

to observe that the parameters of the two innermost layers controlled about the 82% of the model output variance, while the conductivities and densities of the two outermost layers were responsible only for about the 2% of the model output variance (the remaining fraction is due to interaction between the two groups of parameters). Therefore, probably the power and duration of the ROLBS heating pulse were not enough to excite completely the heat capacity of the box, thus making $fibre\_cement\_board2_{\boldsymbol{k}}$ and $fibre\_cement\_board2_{\boldsymbol{\rho}}$ not clearly identifiable.

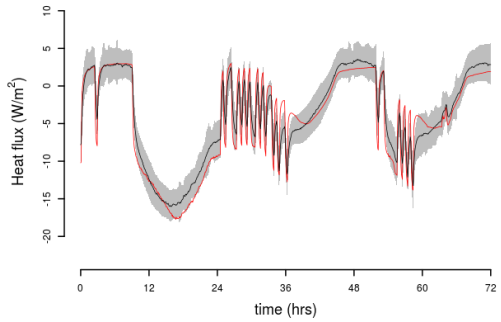
In order to investigate the differences between the specified values and the estimates of $fibre\_cement\_board1_{\boldsymbol{\rho}}$ and of $XPS_{\boldsymbol{\rho}}$, the output of the calibrated model and the output of the model having its parameters fixed to the values prescribed by the KUL data, were compared by calculating the adopted goodness of fit criteria (NMBE, CVRMSE and GOF), for the ROLBS and free float periods of the LECE experiment. It was interesting to notice that the model inputs fixed at the values inferred during the calibration, performed better in predicting the heat fluxes during the ROLBS heating sequence (Figure 5.38), while the



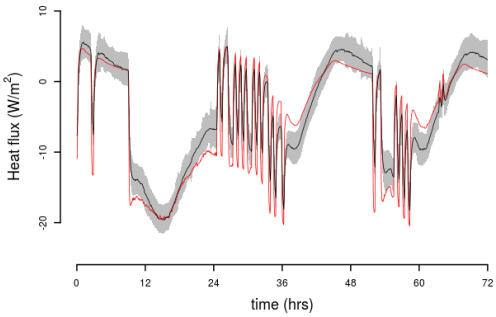
(a) left wall.



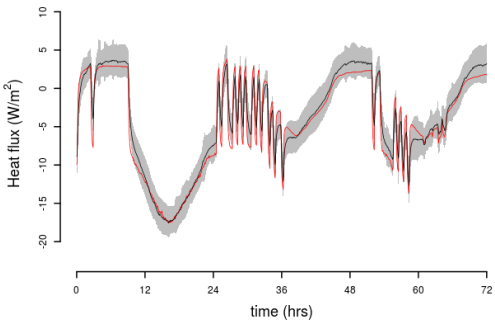
(b) back wall



(c) right wall.



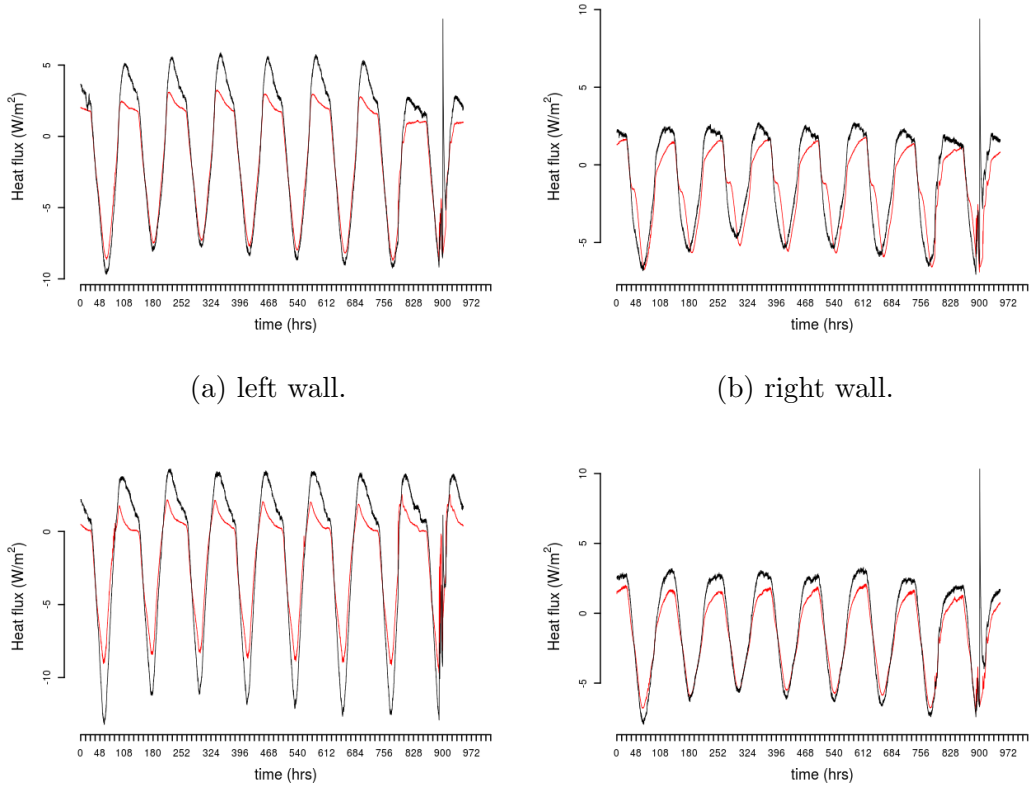
(d) ceiling.



(e) floor.

Figure 5.38: Test box–*Real* LECE: $\mathcal{M}_3$ - comparison between observed heat flux (red) and mode predictions (black) during the training period. In grey are indicated the 95% c.i..

parameters set according to the KUL data provided a better match with the data measured in the free float phase of the LECE experiment. Nonetheless, the inferred values were able to return predictions in good agreement with this latter set of measurements (Figure 5.39).



(a) left wall.

(b) right wall.

(c) ceiling.

(d) floor.

Figure 5.39: Test box–*Real* LECE: $\mathcal{M}_3$ - comparison between observed heat flux (red) and mode predictions (black) during the test period.

Thus the data provided evidence supporting a different combination of calibration parameter values from the true ones. In particular it is believed that a little over fitting of the data occurred, and that the estimation of calibration parameters mainly responsible of the model dynamics (like densities) was affected by the presence of noise in the data and minor inadequacies in the model. Especially, the short duration of the ROLBS test (just four days), might have been inadequate to give a general representation of the test box behaviour under such conditions.

In the light of these results, it is possible to conclude that the convective heat transfer was likely to be one of the major causes of discrepancies. However, the information provided in the beginning were not enough to calibrate a detailed building energy model. For the ROLBS experiment, due to the significant importance of the distribution of the different materials across the wall sections, it was not possible to achieve a good match with the measurements without having a priori information about the different layers making the walls. Even more the du-

ration of the ROLBS test, as well as the power and duration of the pulses did not seem sufficient to completely characterise the test box, thus making it difficult to clearly estimate the parameters of the model. In calibration problems involving constructions made of layers having significantly heterogeneous characteristics, it would be helpful to provide the heat fluxes at both the internal and external surfaces, so as to improve the identifiability of the material properties. This can be difficult in normal studies involving whole buildings but should become good practice in research projects.

Chapter 6

Detailed experiments

This chapter describes a series of sensitivity and calibration experiments involving two real scale small domestic buildings, Twin House N2 and Twin House O5, employed in validation and calibration exercises in the context of the IEA EBC Annex 58. The aim of this last series of case studies was to fully prove the capability of the proposed method in analysing, characterising and calibrating building energy models. Differently from the experiments depicted in the previous chapter, which were exclusively calibration studies serving mainly to develop the proposed probabilistic calibration framework, the objective of these investigations was to assist the empirical validation of ESP-r models.

Two experiments were performed by the Fraunhofer Institute of Building Physics (IBP) in Holzkirchen, Germany, in a flat and unshaded area during two different periods. The first (EXPERIMENT1) was carried out during August and September 2013, while the second (EXPERIMENT2) was undertaken between April and May 2014. Such experiments involved the monitoring of the two buildings in situations wherein the relative internal environment were kept at constant prescribed temperatures (CT phases), subject to Randomly Ordered Logarithmic Binary Sequences (ROLBSs) of heat pulses ([120]) (ROLBS phases) or Free Float (FFL) periods. For comprehensive descriptions the interested reader is referred to [115].

The analysis that will be described in the following was developed in different stages. A base model ($\mathcal{M}_0$) of Twin House N2 observed during EXPERIMENT1, was built by modellers at Energy System Research Unit (ESRU), and its predictions were compared against the measurements. This base model showed poor performances in predicting the monitored target variables, requiring improvements. Due to the complexity of the model and the large number of parameters, it was decided that a manual iterative validation approach was not suitable, and

it was taken as opportunity for using the proposed analysis framework as a tool to diagnose the base ESP-r model and identify effective upgrades.

A qualitative screening of the parameters of $\mathcal{M}_0$ was performed, by applying the Morris Method, and the retained model inputs were calibrated against the metered data. The results showed significant differences between the given specifications and many of the estimates for the calibration parameters, which were deemed not plausible. Even more, the *Difference Analysis* highlighted the possibility to achieve significant improvements by providing $\mathcal{M}_0$ with an airflow network, and including the related inputs as well as thermal bridges in the calibration parameter set.

While the effects of the latter on the model outputs were well understood, it was more difficult to foresee the influences of the implementation of the airflow network on the simulation outcomes. Thus, the sensitivity analysis procedure outlined in Chapter 3 was applied on an upgraded version of $\mathcal{M}_0$, provided with airflow network ($\mathcal{M}_2$), but depicting Twin House O5 during the ROLBS phase of EXPERIMENT2. The main reason for investigating a different building in a different experiment, was the priority given, at that time, by the Annex project in which the research was involved, to analyses regarding Twin House O5 observed during the second experiment. The result from this sensitivity analysis, revealed that the different boundary conditions under which the building was subject had negligible effects on the model outputs. Additionally, earlier investigations, showed that the two buildings were providing practically equal performances. For these reasons it was deemed not necessary to perform a second sensitivity analysis, and that it was acceptable to use the gained information in calibrating the model $\mathcal{M}_2$ configured in order to represent Twin House N2, subject to the conditions prescribed by EXPERIMENT1. The good results achieved confirmed these beliefs.

Thus, the work flow of the analysis can be summarised as follows:

- Initial parameter screening of the base model representing Twin House N2 during the constant temperature and ROLBS phases of EXPERIMENT1.
- Calibration of the retained parameters against the data from the constant temperature and ROLBS phases of EXPERIMENT1.
- Upgrade of the base model according to the results from the analysis.
- Sensitivity analysis of the upgraded model representing Twin House O5 during the ROLBS phase of EXPERIMENT2, in order to identify additional calibration parameters.

- Calibration of the upgraded model representing Twin House N2 in the constant temperature and ROLBS periods of EXPERIMENT1.

This Chapter begins with the description of the performed experiments and of the models employed during the analysis. Then the steps just outlined are explained in detail. It ends with a discussion of the main findings from the analysis. GOF measures indicating the level of agreement achieved between the actual BES models and their emulators are listed in Appendix C.

6.1 Experiments

The experiments took place in the ground floors of the two buildings (Figure 6.1). Figure 6.2 shows the main construction details.

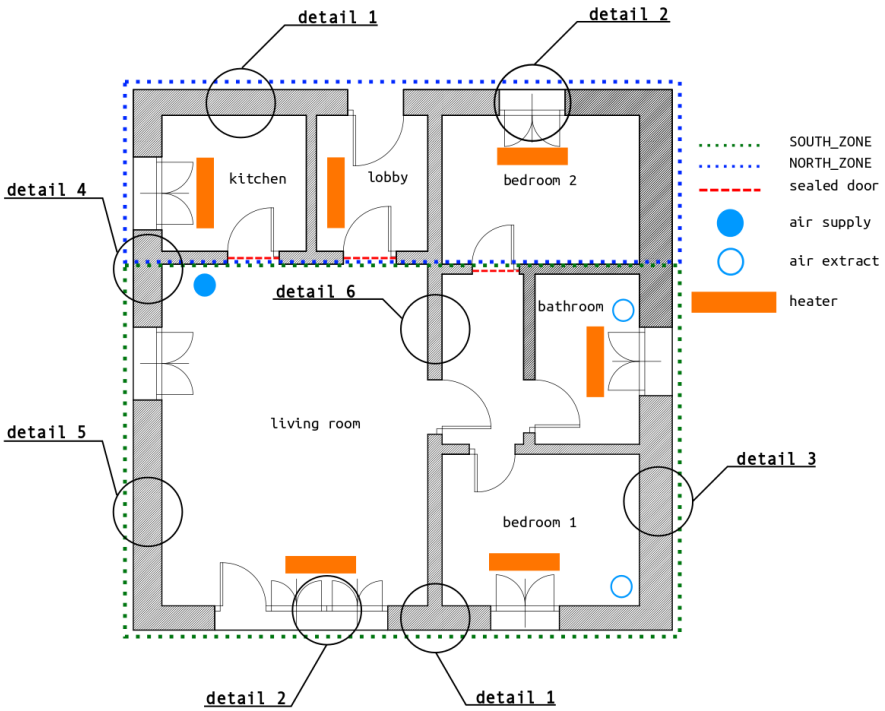


Figure 6.1: Ground floor plan.

Material conductivities were provided by the manufacturer, while the other physical properties were estimated by Fraunhofer IBP. The windows were double glazed with low emissivity coating, argon fill, and provided with insulated roller blinds, for which different configurations were tested. In both experiments, the doors connecting the *living room* to *kitchen*, *lobby* and *bedroom2* were sealed, while ventilation was allowed between the *living room*, *corridor*, *bathroom* and *bedroom1*.

EXPERIMENT1 and EXPERIMENT2 involved similar phases and configurations of the two buildings. Both alternated constant temperatures phases (CT) and periods wherein the buildings were subject to ROLBS heat pulses. Due to their construction, the heat gains from these injections are not correlated with

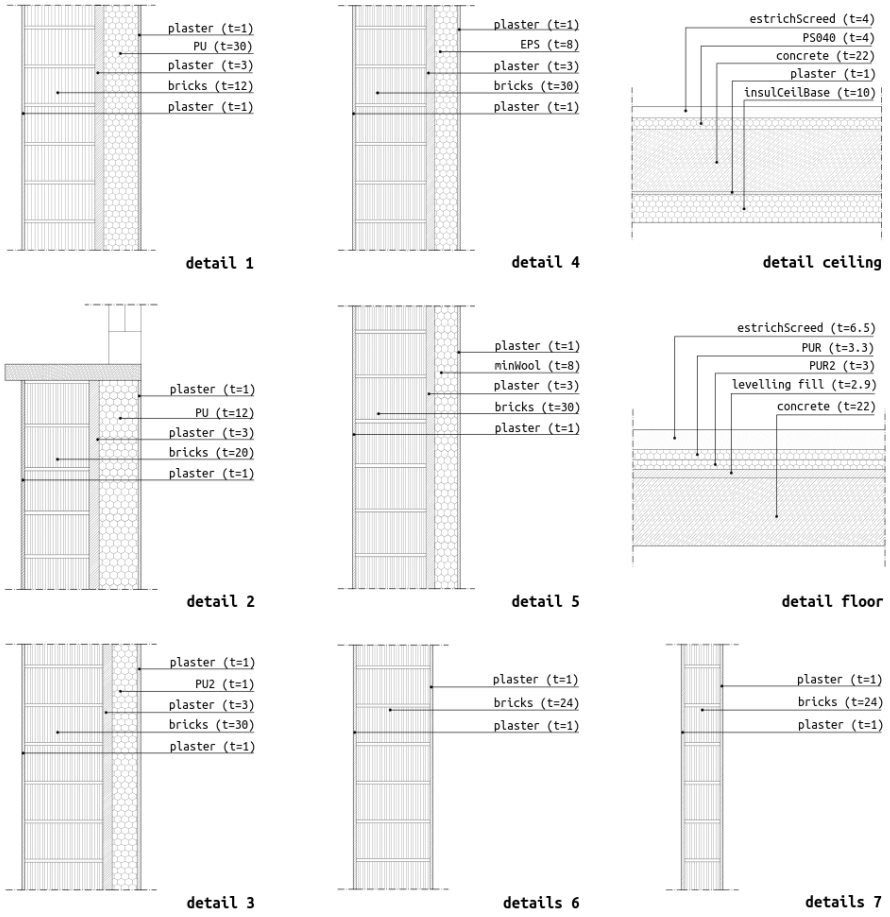


Figure 6.2: Main construction details. t indicates the thickness in centimetres.

other contributions, like solar heat gains, allowing to highlight the main thermal masses and the dynamic behaviour of the building. The heating system was composed of light weight electric heaters with fast response, having split coefficient between convective and radiative heat gains (***C/R***) of 70%/30%, according to manufacturer specifications. Their distribution is shown in figure 6.1.

In the following, further details about the two performed experiments are given, taking care to highlight the important features of each experiment.

6.1.1 Twin House N2 / EXPERIMENT 1

This first experiment was performed during August and September 2013 ([115]). Due to the difficulties in operating a cooling system with the necessary accuracy, even if not ideal for the experimental period, which is outside the cold season, a heating experiment was carried by imposing high temperature set points.

The experimental schedule was the following:

- Days 1-7: Initialization of the experiment by keeping internal constant temperatures of 30^oC.
- Days 8-14: Constant temperature test by keeping constant internal temperatures at 30^oC ($CT1$).
- Days 15-28: The house was subject to ROLBS sequence ($ROLBS1$).
- 29-35: Initialization at constant internal temperatures equal to 25^oC.
- 35-42: Free-floating period.

During the ROLBS phase, the heating pulses were injected only in *living room* and lasted from 1 to 90 hours in order to cover the expected range of time constants. The others rooms were observed under free-floating conditions. Blinds were always kept closed during the different phases in Twin House N2. The initial parameter screenings and the calibrations focused on the phases $CT1$ and $ROLBS1$.

Air leakage was investigated by performing pressurisation tests at the standard pressure of 50 Pa, returning the infiltration rates of 1.62 Ac/h and 2.2 Ac/h respectively for the whole ground floor, and for SOUTH_ZONE only.

A continuously measured ventilation rate of about 120 m^3/h was imposed through the mechanical ventilation system, in order to avoid possible overheating, due to the high interior temperatures adopted. The ventilation inlet was in the *living room* and there were two extraction points in *bathroom* and in *bedroom*1, which were configured to have the same extraction rates of 60 m^3/h each.

Uninsulated air ducts, led from the basement to the roof through the *kitchen*. The heat provided or subtracted by this duct work to or from the *kitchen* air was calculated by the experimental team.

The provided dataset was very comprehensive consisting of fifty variables observed at intervals of 1 minute, and then post-processed to ten minute and one hour averages. The former were used in the analysis. The most relevant variables employed in this analysis are external temperature ($\boldsymbol{Te}$), global horizontal solar radiation ($\boldsymbol{Gh}$), diffuse solar radiation ($\boldsymbol{Ds}$), wind speed ($\boldsymbol{Ws}$), wind direction ($\boldsymbol{Wd}$), relative humidity ($\boldsymbol{Rh}$), internal zone temperatures ($\boldsymbol{Ti}$), zone sensible heat loads ($\boldsymbol{shl}$), supplied air temperature ($\boldsymbol{Tvnt}$), heat gain from air duct in the *kitchen*, basement air temperature ($basement_{\boldsymbol{Ti}}$), attic air temperature ($attic_{\boldsymbol{Ti}}$), and the timing and magnitude of the ROLBS heating sequence (ROLBS).

6.1.2 Twin House O5 / EXPERIMENT 2

Similar to EXPERIMENT1, EXPERIMENT2 was performed outside the cold season, during April and May 2014 ([115]). For the same reasons explained in the previous section, a heating experiment was carried out alternating constant temperature, ROLBS and free floating periods. The experimental schedule was the following:

- Days 1-10: initialisation at constant temperature of 30^oC in SOUTH_ZONE, *bedroom*1 and *bathroom*, and 22^oC in attic, cellar and NORTH_ZONE.
- Days 11-24: ROLBS sequence in SOUTH_ZONE and constant temperature of 22^oC in attic, cellar and NORTH_ZONE ($ROLBS2$).
- Days 25-31: constant temperature of 30^oC in SOUTH_ZONE and 22^oC in attic, cellar and NORTH_ZONE.
- Days 32-40: Free-float in SOUTH_ZONE, and 22^oC in attic, cellar and NORTH_ZONE.

During all the phases, the attic and the basement were considered as boundary spaces, wherein the air temperature was kept constant at 22^oC. Roller blinds were kept closed in the NORTH_ZONE and attic, and open in the SOUTH_ZONE. The sensitivity analysis that will be presented in Section 6.5 focused on the $ROLBS2$ phase.

Air leakage was experimentally investigated by performing pressurisation tests returning rates of 1.54 Ac/h and 2.3 Ac/h for the whole ground floor and SOUTH_ZONE only respectively.

As in the previous experiment, mechanical ventilation was used during the entire experiment to avoid excessive overheating, but with lower rates. The ventilation inlet in *living room* was supplying air at 60 m^3/h, while the two extraction points in *bathroom* and *bedroom*1 were extracting air at the rate of 30 m^3/h each.

The provided data set was comprehensive, consisting of fifty variables measured with one minute time steps. Those employed in the analysis are: external temperature ($\boldsymbol{Te}$), wind speed ($\boldsymbol{Ws}$) and direction ($\boldsymbol{Wd}$), internal zone temperatures ($\boldsymbol{Ti}$), zone sensible heat loads ($\boldsymbol{shl}$), supplied air temperature ($\boldsymbol{Tvnt}$), basement air temperature ($basement_{\boldsymbol{Ti}}$), attic air temperature ($attic_{\boldsymbol{Ti}}$), and the timing and magnitude of the ROLBS sequence of heat injections (ROLBS).

6.2 Models

In this section the analysed models are presented. At first a base line model was built considering the given specifications ($\mathcal{M}_0$). This model was then calibrated against the measured data, thus creating a calibrated version of it ($\mathcal{M}_1$). Therefore, the calculated GOF measures and comparison with measurements, when regarding $\mathcal{M}_0$ are relative to the base model having its inputs fixed to the specified values, while when regarding $\mathcal{M}_1$ they are relative to the base model having its inputs set to the estimated values. Successively, due to the outcomes of the performed calibration supporting the importance to consider wind driven infiltration, the base model was upgraded with the implementation of an airflow network, resulting in model $\mathcal{M}_2$. This model was then used during the following steps of the analysis to represent Twin House N2 and Twin House O5 in EXPERIMENT1 and EXPERIMENT2 respectively. The former variant was named $\mathcal{M}_{2,N2}$, while the latter was called $\mathcal{M}_{2,O5}$.

6.2.1 Base model ($\mathcal{M}_0$) and calibrated base model ($\mathcal{M}_1$)

The base model was created, depending on the given specification and on choices and assumptions in agreement to best practice and modeller experience. It was employed to depict Twin House N2 during the $CT1$ and $ROLBS1$ experiments, and a high level of detail, allowed by the high quality information provided, was used in its creation.

The building geometry was completely respected as well as construction elements compositions. The model was composed of seven thermal zones reflecting the real room disposition. Shading blocks on the south façade were disposed in order to represent the building elements projecting shadows on the openings.

The air-gaps between window glazing and lowered blinds were assumed to be perfectly sealed and having a thermal resistance of 0.17 m^2K/W.

Two approaches were adopted in modelling thermal bridges depending on if they were between model zones or if they were between model zones and the external environment. The former were taken into account through dedicated construction components having increased thermal transmittances. The latter were considered according to the corresponding linear transmittances. The specifications provided only estimate for thermal bridges occurring at the window frames, according to the German standards. For the others, involving external walls, floor, ceiling, and internal partitions it was necessary to undertake finite elements analysis in order to estimate the consequent additional heat transfers.

Constant air flow rates of 120 m^3/h and 60 m^3/h, equal to the average of the given measurements, were used respectively as inflow rate for the mechanical ventilation system inlet in *living room* and as outflow rates for the two mechanical ventilation system extraction points in *bathroom* and in *bedroom*1. Inter-zone ventilation in the SOUTH_ZONE was simulated by imposing fixed ventilation rates of 120 m^3/h, and 60 m^3/h between *living room* and *corridor*, between *corridor* and *bathroom*, and between *corridor* and *bedroom*1, respectively. The doors separating the SOUTH_ZONE and the NORTH_ZONE were considered perfectly sealed.

Infiltrations were modelled by imposing the following constant infiltration rates, derived according to the carried out pressurisation tests:

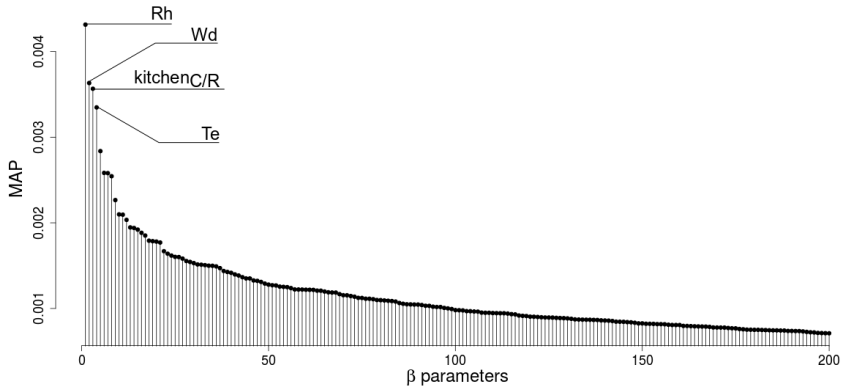
- *living room*: 0.11 Ac/h.
- *bedroom*1: 0.11 Ac/h.
- *kitchen*: 0.08 Ac/h.
- *lobby*: 0.08 Ac/h.
- *bedroom*2: 0.08 Ac/h.
- *bathroom*: 0.07 Ac/h.

The weather factors and conditions due to the particular experimental configuration were imposed on the model. In particular, the heat gains from the ROLBS pulses and from the ducts in the *kitchen* were considered as convective only.

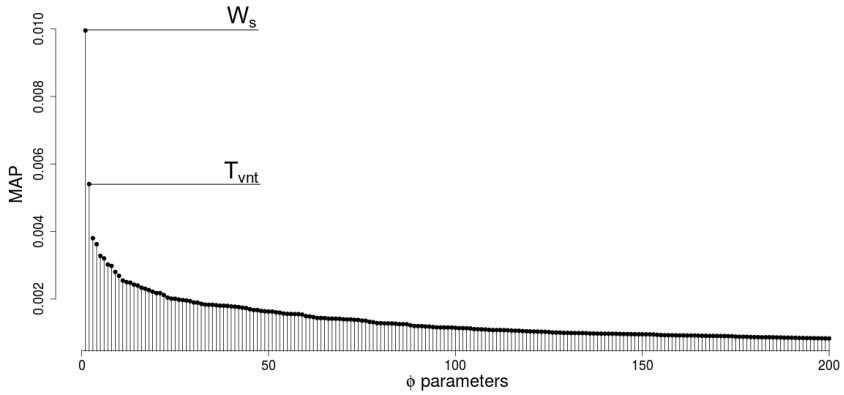
6.2.2 Airflow Model ($\mathcal{M}_2$)

$\mathcal{M}_2$ represented an upgraded version of $\mathcal{M}_0$ according to the results of the relative calibration and *Difference Analysis*. In particular the latter are represented in Figures 6.3 and 6.4.

The outcome of the calibrations performed on $\mathcal{M}_0$ showed inconsistencies for the thermal resistance of the building envelope, for the considered experimental periods. In particular, the $CT1$ calibration returned an increased thermal resistance for most of the external walls, while the $ROLBS1$ calibration yielded the opposite outcome. Similarly the parameters relative to the assumed constant infiltration rates showed sensibly different estimates depending on the analysed experiment. Internal partitions showed greatly reduced thermal resistances and thermal masses, in the $CT1$ and $ROLBS1$ calibrations, respectively. The *Difference*



(a) β parameters.

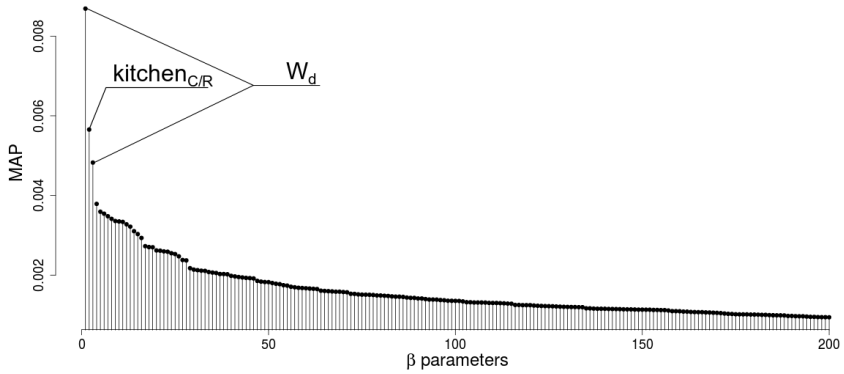


(b) ϕ parameters.

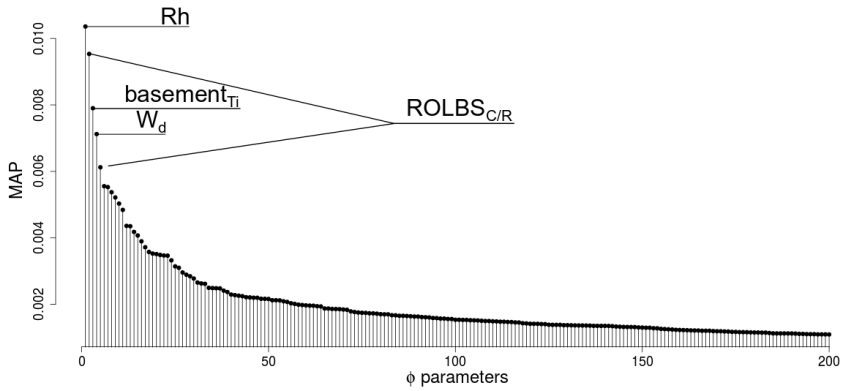
Figure 6.3: $\mathcal{M}_0$–*CT*1: results from *Difference Analysis*

Analysis showed correlation between the *Difference Vectors*, wind speed (***Ws***), wind direction (***Wd***), mechanical ventilation supplied air temperature (***Tvnt***), external temperature (***Te***), *basement* air temperature ($basement_{\boldsymbol{Ti}}$), fraction of convective sensible heat from the casual gains in the *kitchen* ($kitchen_{\boldsymbol{C/R}}$), and fraction of convective sensible heat form the ROLBS pulses in *living room* ($living\ room_{\boldsymbol{C/R}}$).

In the light of these results, the following changes were made. It was decided to include thermal bridges among the calibration parameters, in order to address the inconsistencies observed for the thermal resistance of the building envelope. The model input representing the fraction of convective heat gains from the ROLBS pulses, was found mistakenly fixed to one and considered among the calibration parameters as well. The incongruities noticed for the infiltration rate parameters might well be due to wind direction and wind velocities causing different infiltration regimes for the two examined experiments. Thus, assuming constant



(a) β parameters.



(b) ϕ parameters.

Figure 6.4: $\mathcal{M}_0$–*ROLBS*1: results from *Difference Analysis*

infiltration rates was judged inadequate and an airflow network model was implemented in $\mathcal{M}_0$, in order to simulate wind driven infiltrations and a more realistic distribution of the airflows between inside and outside environments as well as between model zones. The latter aspect could also explain the sensibly different than specified values of the properties of the internal partitions. A diagram of the airflow network model is shown in figure 6.5.

Crack components were used to represent connections between the internal and external environments, between SOUTH_ZONE and NORTH_ZONE and between *living room*, *basement* and *attic*. Constant flow rate components were adopted to model the mechanical ventilation system at the supply and extraction points. Bi-directional flow components were employed to represent the open doors between the *living room*, *corridor*, *bathroom* and *bedroom*1. In particular, the resistances of these large openings are very small compared to the others, so parameter uncertainties associated with them have been deemed negligible and

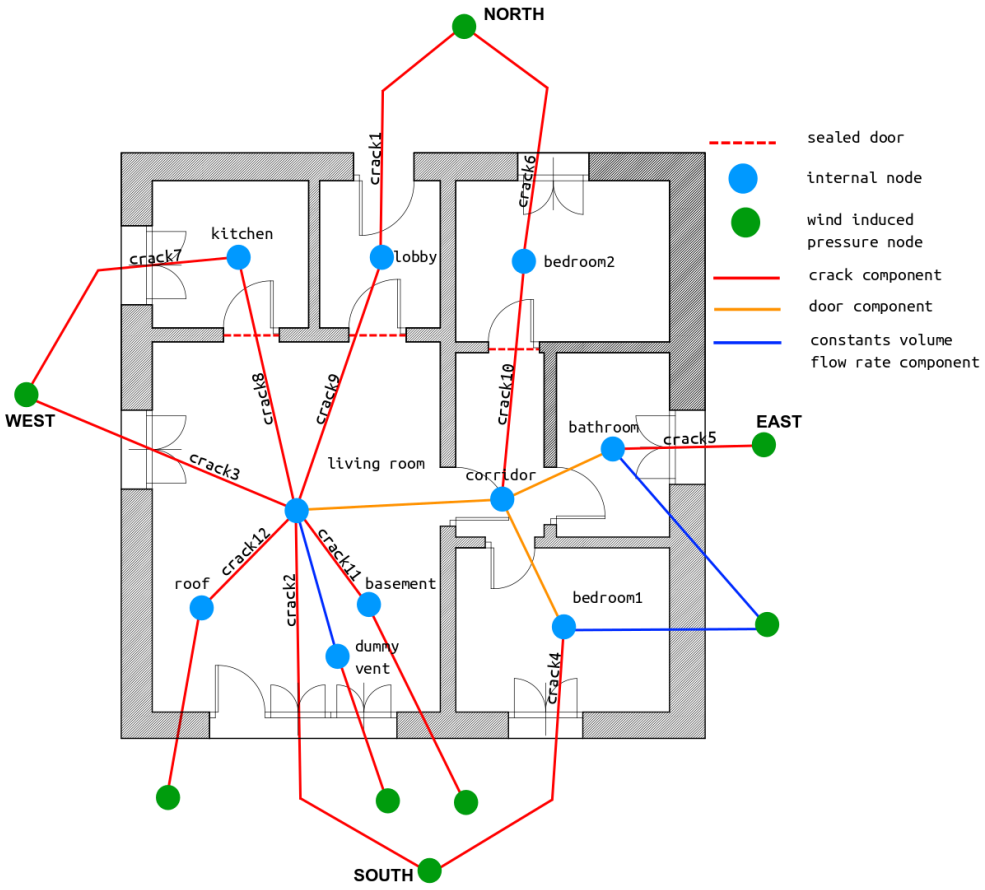


Figure 6.5: Airflow network model.

were neglected in the analysis. Crack length, air supply and extraction rates were included in the calibration parameter set. For more detail regarding the mathematical modelling of the components used to build the flow network model the reader is referred to [45].

$\mathcal{M}_2$ was employed to create two variants representing Twin House O5 observed during *ROLBS2*, and Twin House N2 observed during *CT1* and *ROLBS1*, by imposing on it the different measured boundary conditions and experimental configurations. The former was used in performing the sensitivity analysis that will presented in Section 6.5, and it will be referred as $\mathcal{M}_{2,O5}$. The latter was calibrated by considering calibration parameters according to the result discussed in Section 6.5, and it will be referred as $\mathcal{M}_{2,N2}$.

6.3 Initial parameter screening

Two independent parameter screenings were undertaken on model $\mathcal{M}_0$ depending on the considered experimental phases ($CT1$ and $ROLBS1$). The Morris Method, was adopted, and each thermal zone was considered independently.

The parameter screening was preceded by an analysis of the uncertainties regarding the model parameters considered. The necessary information, was obtained through a literature review ([70], [34] and [73]). The total number of inputs considered was 135. A summary of the estimated uncertainties is contained in Table 6.1, wherein they are listed by typologies.

Table 6.1: $\mathcal{M}_0$ parameter uncertainties by typologies.

PARAMETER CLASS	DISTRIBUTION (w.r.t initial value)
Conductivity	$\mathcal{N}(1, 0.35)$
Densities	$\mathcal{N}(1, 0.15)$
Specific heat	$\mathcal{N}(1, 0.3)$
Thermal bridges	$\mathcal{U}(0.9, 1.1)$
Infiltration	$\mathcal{N}(1, 0.33)$
Window thermal resistances	$\mathcal{N}(1, 0.35)$
Optical transmission	$\mathcal{N}(1, 0.33)$

Due to the qualitative character of the results provided by the Morris Method, parameter retention is usually done depending on empirical evaluations. In this case the ten factors having the highest $M^\star$ indexes (Equation 3.11) for each zone were considered sufficient to approximate the original model. It is on these variables that the following discussions focus.

6.3.1 Constant temperature

The outputs considered in deriving the sensitivity measures were the sensible heat loads for each zone. Figures 6.6 and 6.7 graphically describe the sensitivity analysis results. In particular they depict the ${M^\star}_p$ (Equation 3.11) scores for each thermal zone. The retained model inputs are named in the Figures, and listed in Table 6.2.

In most of the zones the thermal masses of the floor and ceiling structures ($concrete_{\boldsymbol{\rho}}$ and $concrete_{\boldsymbol{c}}$) appeared among the significant parameters. These represented most of the heat capacity of the building, and a large amount of energy had to be employed in charging them, in order to maintain the temperature set points in the different zones. This aspect influenced the initial trends of the heat loads.

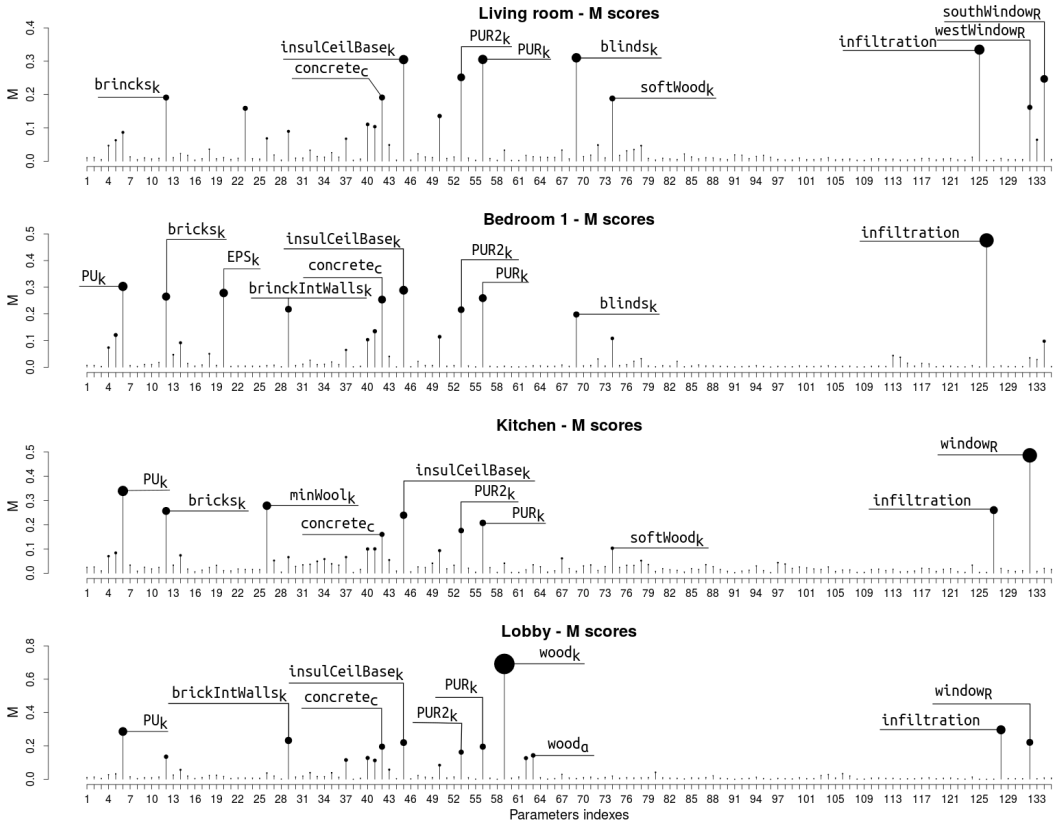


Figure 6.6: $\mathcal{M}_0$–$CT1$: $M^\star$ scores for *living room*, *bedroom1*, *kitchen* and *lobby*

Nonetheless, the sensitivity of each zone was dependent upon different sets of factors, mainly determining the heat fluxes through the building envelope. In particular, each thermal zone is dependent upon the conductivities of the materials employed and upon the infiltration rates. It is possible to identify three main sensitivity characters.

The heat requirements of *living room* were highly influenced by its infiltration rate, which was the most important parameter. This factor was followed by conductivity values of ceiling insulation ($insulCeilBase_{\boldsymbol{k}}$), floor insulations ($PUR_{\boldsymbol{k}}$ and $PUR2_{\boldsymbol{k}}$) and blinds ($blinds_{\boldsymbol{k}}$). Also window thermal resistances ($glazing_{\boldsymbol{R}}$ and $glazingBlinds_{\boldsymbol{R}}$) and the conductivity of the bricks making the external walls ($brick_{\boldsymbol{k}}$) had significant influences.

The smaller zones, namely *kitchen*, *lobby*, *bathroom* and *bedroom2*, had their sensible heat loads mostly influenced by the parameters governing heat fluxes through the openings. This is particularly true for *lobby* and *bathroom*. The former had its sensitivity largely dominated by the conductivity of the exterior door ($wood_{\boldsymbol{k}}$), and the latter had its sensitivity mainly dependent on the window thermal resistance ($glazing_{\boldsymbol{R}}$). Other parameters having significant effects are those characterizing heat conduction through the envelope and infiltrations. In partic-

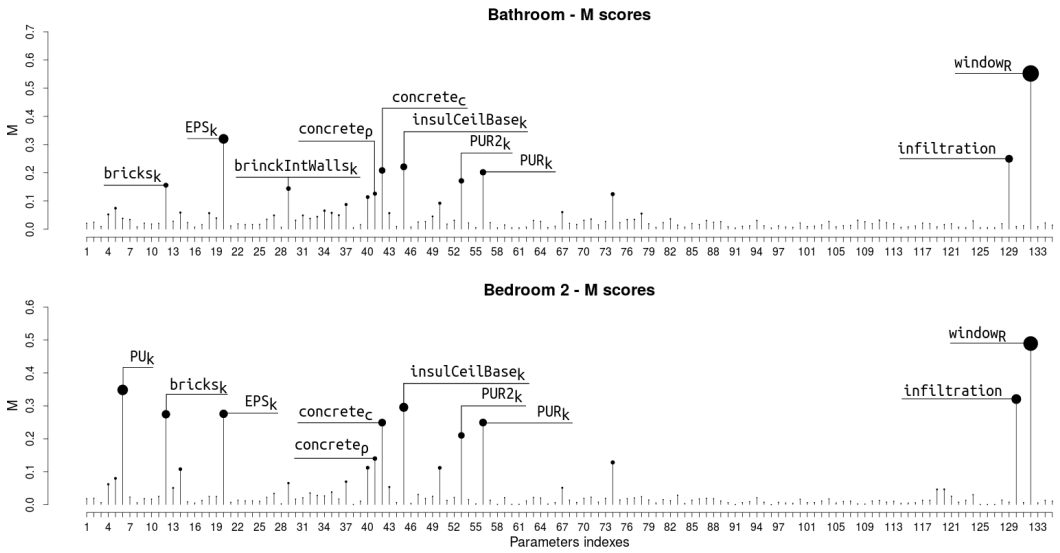


Figure 6.7: $\mathcal{M}_0$–$CT1$ - $M^\star$ scores for *bathroom* and *bedroom2*

ular these factors are different according to material locations. It is interesting to notice that, differently from the *living room*, external wall conductivities ($EPS_{\boldsymbol{k}}$, $PU_{\boldsymbol{k}}$ and $minWool_{\boldsymbol{k}}$) had higher influences than ceiling and floor conductivities.

bedroom1 presents intermediate sensitivity features between the above described two characters. Its heat requirements were mostly determined by its infiltration rate, as *living room*, but among the conductivities those relative to external walls appear to have the most relevant effects ($EPS_{\boldsymbol{k}}$, $brick_{\boldsymbol{k}}$ and $PU_{\boldsymbol{k}}$), as in the smaller zones.

6.3.2 ROLBS

The model outputs considered in the calculation of the sensitivity indexes, were the dry bulb air temperatures of the different zones. Figures 6.8 and 6.9 graphically describe the calculated $M^\star{}_p$ indexes. The model inputs retained are named in the Figures and listed in Table 6.3.

The ROLBS experiment was successful in exciting the medium and small size thermal masses ($brickIntWall_{\boldsymbol{\rho}}$, $brickIntWall_{\boldsymbol{c}}$, $estrichScreed_{\boldsymbol{\rho}}$, $estrichScreed_{\boldsymbol{c}}$, $brick_{\boldsymbol{\rho}}$ and $brick_{\boldsymbol{c}}$), but it failed in underlining the two biggest thermal capacities of the model, that are the floor and ceiling concrete slabs ($concrete_{\boldsymbol{\rho}}$ and $concrete_{\boldsymbol{c}}$). This is probably due to the particular frequencies and power of the used sequence of pulses, which probably should had been longer and more powerful.

The factors relative to the activated heat capacities (densities and specific heats) constituted a set of variables which dominated the sensitivity of the model,

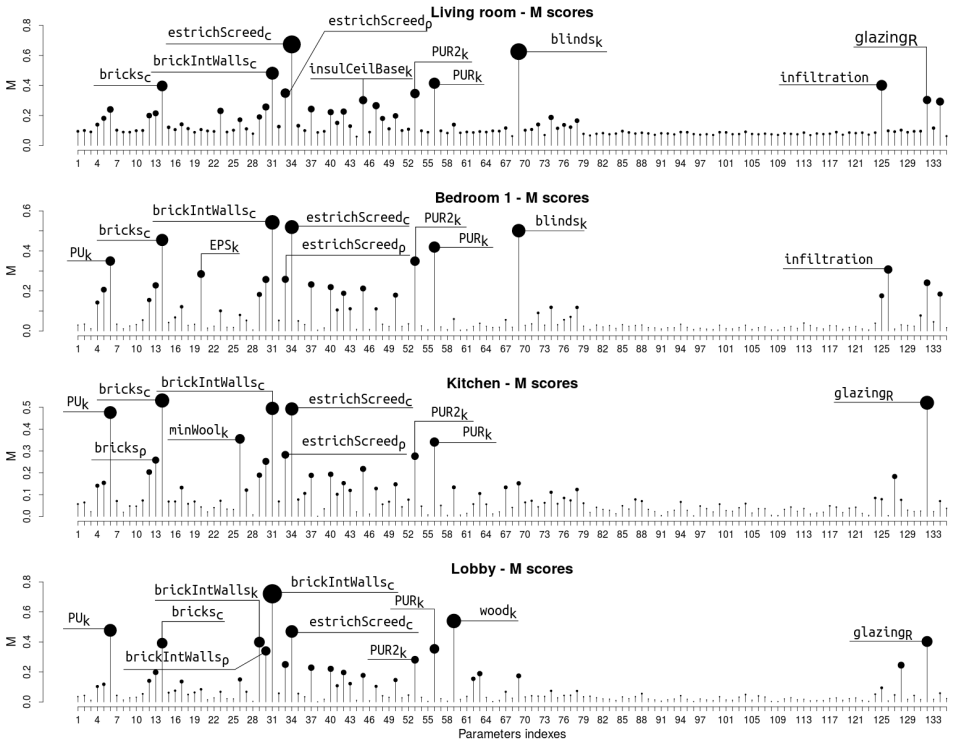


Figure 6.8: $\mathcal{M}_0$–*ROLBS*1: $M^\star$ scores for *living room*, *bedroom*1, *kitchen* and *lobby*

repeating itself, with little variations, in all the model zones. Also conductivities, window thermal resistances and infiltrations played important roles. In particular they changed from zone to zone in similar fashion as described previously.

6.3.3 Discussion

The information obtained was particularly useful in defining the assumptions upon which to build the probabilistic models used in the calibration phases.

The sensitivity results showed that during *CT*1 the different zones interacted only weakly. $\mathcal{M}_0$ did not allow for relevant interactions between model zones, since infiltration and ventilation were modelled with constant and independent flow rates. Even more, heat exchanges due to conduction between the thermal zones were made negligible by the same temperature set points being adopted. Therefore, the only important solicitations to the internal conditions, were coming from the external environment, making the envelope parameters the most important factors. Such factors were different according to the employed materials, so that thermal zones presented different MIF and sensitivity features.

With respect to *ROLBS*1, the possibility of interactions between model zones appeared more significant. During this test, the heat injected in *living room*

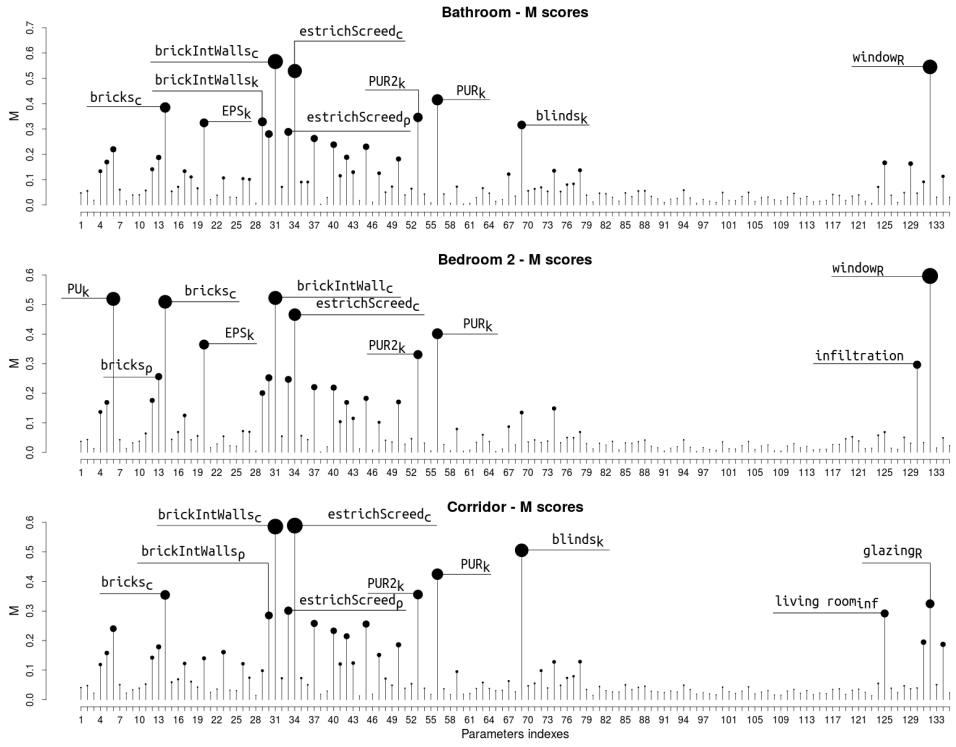


Figure 6.9: $\mathcal{M}_0$–*ROLBS*1: $M^\star$ scores for *bathroom*, *bedroom*2 and *corridor*

was in part absorbed by the zone heat capacity and in part propagated to the confining air volumes shifted in time and decreased in intensity according to the properties of the partitions. Hence, interactions between the different thermal zones occurred, and they were highlighted by the same set of model parameters dominating the sensitivity of the different model zones.

In light of these considerations, it appeared that, especially in calibrating the model against the data from the *ROLBS*1 experiment, parameter identification could benefit from the dependency of the considered target variables, since they were mainly determined by the same model inputs. In particular, the identifiability of the model inputs common to different zones should increase due to the fact the the same factors had to explain the different target variables, simultaneously observed. Nonetheless, considering all model zones separately would result in a high computational load, and therefore it was decided to divide the model in three macro-zones:

- *living room*.
- *KLB*2, grouping the sealed zones: *kitchen*, *lobby* and *bedroom*2.
- *CBB*1, representing the zones having high ventilation exchange with the *living room*: *corridor*, *bathroom*, and *bedroom*2.

The mean sensible heat loads and the mean air dry bulb temperature of these macro-zones were considered as target variables for the $CT1$ and $ROLBS1$ calibrations, respectively. Such a solution was deemed to be the best compromise between degree of detail and computational burden of the analysis.

6.4 Calibration of $\mathcal{M}_0$

Independent calibrations were performed with respect to the $CT1$ and the $ROLBS1$ experiments. The calibration framework was modified as explained in Section 4.2.2, in order to take into account the correlation and the dependency between the considered target variables.

Preliminary investigations revealed the possibility that some important factors were not retained during the screening of the model parameters. During $CT1$, $\mathcal{M}_0$ was always underestimating the sensible heat load for $KLB2$ and $CBB1$. The cause of this was deemed to be the lack of consideration of thermal bridges. During the parameter screening, such factors were considered separately from each other so that it was possible that their overall importance was underestimated. However, in ESP-r, thermal bridges, and constant infiltration rates are both treated as additions or subtractions of heat to the energy balance of the zone air nodes, and their effects are practically the same. This could result in poor identifiability for both these variable types. Therefore it was deemed appropriate to consider during the calibration only the infiltration rates, with higher uncertainties in order to include also the effects due to the thermal bridges. Similarly, during $ROLBS1$, the model showed deficiencies of thermal mass, probably due to the nonconsideration of corners and intersection elements between the walls, floor, ceiling and partitions. Therefore the uncertainties of densities and specific heats were increased as well.

A summary of the considered calibration parameters with the relative prior probability densities distributions is given in Table 6.2 and Table 6.3.

6.4.1 Constant temperature

Table 6.2 summarises the main results, consisting of MAP estimates and confidence intervals for the calibration parameters. Figure 6.10 shows a comparison between the prior and posterior probability density distribution for the calibration parameters for which specifications were provided.

Immediately, it was possible to notice that the posterior probability density distributions of $blinds_{\boldsymbol{k}}$ and $glazingBlinds_{\boldsymbol{R}}$ were particularly concentrated near

Table 6.2: $\mathcal{M}_0$–$CT1$ - prior probability density distributions, MAP estimates and 95% confidence intervals for calibration parameters.

#	PARAMETERS	SPECIFIED VALUES	PRIOR DISTRIBUTION	MAP	Q2.5	Q50.5	Q97.5
1	$PU_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.035	$\mathcal{N}(1, 0.35)$	0.026	0.007	0.033	0.065
2	$brick_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.220	$\mathcal{N}(1, 0.35)$	0.411	0.181	0.385	0.464
3	$EPS_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.022	$\mathcal{N}(1, 0.35)$	0.031	0.008	0.029	0.042
4	$minWool_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.036	$\mathcal{N}(1, 0.35)$	0.017	0.002	0.029	0.074
5	$brickIntWall_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.331	$\mathcal{N}(1, 0.35)$	0.586	0048	0.492	0.637
6	$PS040_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.040	$\mathcal{N}(1, 0.35)$	0.014	0.006	0.023	0.059
7	$concrete_{\boldsymbol{k}}$ $(\frac{W}{mK})$	2.000	$\mathcal{N}(1, 0.35)$	1.013	0.272	1.556	3.593
8	$concrete_{\boldsymbol{\rho}}$ $(\frac{kg}{m^3})$	2400.000	$\mathcal{N}(1, 0.15)$	2270.253	1597.733	2431.851	3273.494
9	$concrete_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	1000.000	$\mathcal{N}(1, 0.3)$	1065.528	665.722	1013.271	1363.956
10	$insulCeilBase_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.035	$\mathcal{N}(1, 0.35)$	0.035	0.015	0.039	0.060
11	$ausgleich_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.060	$\mathcal{N}(1, 0.35)$	0.102	0.010	0.062	0.117
12	$PUR2_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.025	$\mathcal{N}(1, 0.35)$	0.040	0.007	0.040	0.050
13	$PUR_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.023	$\mathcal{N}(1, 0.35)$	0.030	0.012	0.038	0.052
14	$wood_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.131	$\mathcal{N}(1, 0.35)$	0.114	0.057	0.159	0.245
15	$wood_{\boldsymbol{\alpha}}$ (-)	0.600	$\mathcal{U}0.8, 1.2)$	0.819	0.109	0.626	1.146
16	$blinds_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.100	$\mathcal{N}(1, 0.35)$	0.192	0.066	0.127	0.139
17	$softwood_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.130	$\mathcal{N}(1, 0.35)$	0.068	0.218	0.259	0.269
18	$living\ room_{\boldsymbol{inf}}$ (Ac/h)	0.11	$\mathcal{U}(0.001, 3.00)$	0.013	0.008	0.074	0.184
19	$bedroom1_{\boldsymbol{inf}}$ (Ac/h)	0.11	$\mathcal{U}(0.001, 15.00)$	0.496	0.001	0.021	0.108
20	$kitchen_{\boldsymbol{inf}}$ (Ac/h)	0.08	$\mathcal{U}(0.001, 12.00)$	0.832	0.190	0.368	0.653
21	$lobby_{\boldsymbol{inf}}$ (Ac/h)	0.08	$\mathcal{U}(0.001, 10.00)$	0.114	0.007	0.137	0.768
22	$bathroom_{\boldsymbol{inf}}$ (Ac/h)	0.07	$\mathcal{U}(0.001, 12.00)$	0.802	0.077	0.727	0.833
23	$bedroom2_{\boldsymbol{inf}}$ (Ac/h)	0.08	$\mathcal{U}(0.001, 10.00)$	0.574	0.066	0.486	0.785
24	$glazing_{\boldsymbol{R}}$ $(\frac{W}{m^2K})$	0.65	$\mathcal{N}(1, 0.33)$	1.191	0.260	0.934	1.245
25	$glazingBlinds_{\boldsymbol{R}}$ $(\frac{W}{m^2K})$	0.65	$\mathcal{N}(1, 0.33)$	1.356	1.247	1.350	1.375

Prior probability density distribution are scaled according the the initial values.

the upper boundaries of their variation ranges. These inputs had MAP estimates about two times their design values.

There was no clear trend in the variation of the considered conductivities. However useful insights could be drawn by analysing these parameters according to the relative construction components. $PU_{\boldsymbol{k}}$ and $minWool_{\boldsymbol{k}}$ decreased by 26% and 47% respectively and, despite $brick_{\boldsymbol{k}}$ and $EPS_{\boldsymbol{k}}$ increasing, that resulted in a higher thermal resistance in most of the external walls. Also, the ceiling and floor elements presented higher thermal resistances after calibration. In the former case, $PS040_{\boldsymbol{k}}$ decreased by 65% while $insulCeilBase_{\boldsymbol{k}}$ had a posterior estimate which is substantially equal to its design value. In the latter, $PUR_{\boldsymbol{k}}$ and $PUR2_{\boldsymbol{k}}$ increased by the 30% and 60% respectively, but $concrete_{\boldsymbol{k}}$ assumed a value about half of the given datum. Internal partitions showed an increased thermal transmittance due to the relative conductivity ($brickIntWall_{\boldsymbol{k}}$) having a MAP estimate higher than the starting value by the 77%.

The infiltration rates showed the most significant variations. All the model thermal zone, except $living\ room$ wherein the posterior infiltration rate is 12% of the initial estimate, showed important increases in their infiltration rates. In

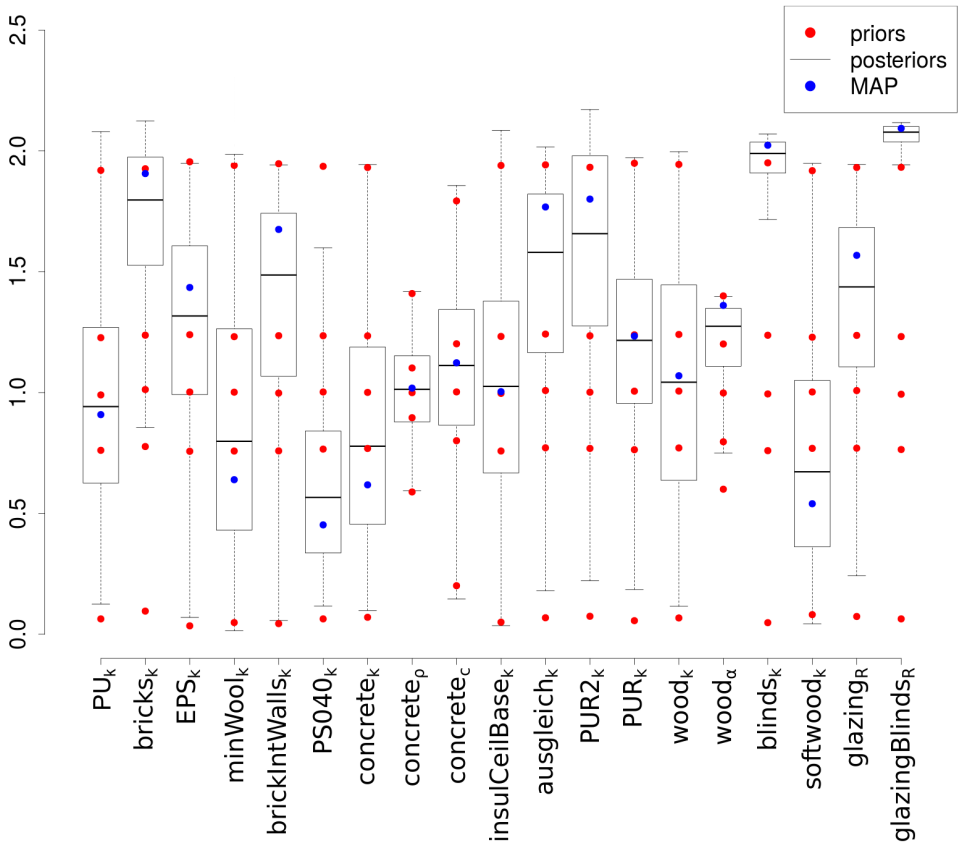


Figure 6.10: $\mathcal{M}_0$: *CT1* - prior and posterior density distributions for material parameters. Posterior samples have been scaled according to their mean to aid visualization (i.e. 1 = initial value).

particular, despite $lobby_{inf}$ which increased only by 40%, the infiltration rates of *bedroom1*, *kitchen*, *bathroom* and *bedroom2*, presented calibrated values from 5 to 10 times higher. In considering these results, it must be kept in mind that these model factors included also the thermal bridges.

6.4.2 ROLBS

In Table 6.3 are listed calibration parameter estimates and relative 95% confidence intervals, while Figure 6.11 depict a comparison between prior and posterior probability density distributions for calibration parameters for which specifications were provided.

The calibrated model ($\mathcal{M}_1$) presented a substantial increasing in the thermal mass of its envelope. $brick_\rho$ assumed a values about three times the design one, and a posterior density distribution completely squashed against the upper bound of its variation range. Differently from the results yielded by the previous calibration, most of the external walls had decreased thermal resistance, due to

Table 6.3: $\mathcal{M}_0$–*ROLBS1*: MAP and confidence intervals for calibration parameters

#	PARAMETERS	SPECIFIED VALUES	PRIOR DISTRIBUTION	MAP	Q2.5	Q50.5	Q97.5
1	$PU_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.035	$\mathcal{N}(1, 0.35)$	0.045	0.015	0.044	0.067
2	$brick_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.22	$\mathcal{N}(1, 0.35)$	0.200	0.047	0.206	0.382
3	$brick_{\boldsymbol{\rho}}$ $(\frac{kg}{m^3})$	800	$\mathcal{N}(1, 0.7)$	2387.760	1397.403	2312.927	2485.788
4	$brick_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	1000	$\mathcal{N}(1, 0.30)$	954.645	315.059	960.334	1659.353
5	$EPS_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.022	$\mathcal{N}(1, 0.35)$	0.026	0.009	0.026	0.041
6	$minWool_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.036	$\mathcal{N}(1, 0.35)$	0.032	0.005	0.032	0.060
7	$brickIntWall_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.331	$\mathcal{N}(1, 0.35)$	0.316	0.087	0.315	0.548
8	$brickIntWall_{\boldsymbol{\rho}}$	1000	$\mathcal{N}1, 0.7)$	819.415	160.241	828.716	1637.413
9	$brickIntWall_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	1000	$\mathcal{N}(1, 0.30)$	550.322	202.439	688.365	1649.049
10	$estrichScreed_{\boldsymbol{\rho}}$	2000	$\mathcal{N}(1, 0.7)$	2358.614	351.886	2661.748	5709.353
11	$estrichScreed_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	1000	$\mathcal{N}(1, 0.30)$	524.429	108.683	601.347	1635.722
12	$PS040_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.04	$\mathcal{N}(1, 0.35)$	0.037	0.007	0.037	0.068
13	$concrete_{\boldsymbol{k}}$ $(\frac{W}{mK})$	2.00	$\mathcal{N}(1, 0.35)$	2.274	0.763	2.244	3.668
14	$insulCeilBase_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.035	$\mathcal{N}(1, 0.35)$	0.025	0.005	0.026	0.052
15	$PUR2_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.025	$\mathcal{N}(1, 0.35)$	0.026	0.006	0.026	0.045
16	$PUR_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.023	$\mathcal{N}(1, 0.35)$	0.021	0.005	0.022	0.038
17	$wood_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.131	$\mathcal{N}(1, 0.35)$	0.088	0.019	0.095	0.203
18	$blinds_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.100	$\mathcal{N}(1, 0.35)$	0.038	0.005	0.046	0.156
19	$living\ room_{\boldsymbol{inf}}$ (Ac/h)	0.11	$\mathcal{N}(1, 0.33)$	0.039	0.004	0.083	0.304
20	$bedroom1_{\boldsymbol{inf}}$ (Ac/h)	0.11	$\mathcal{N}(1, 0.33)$	0.300	0.076	0..283	0.325
21	$kitchen_{\boldsymbol{inf}}$ (Ac/h)	0.08	$\mathcal{N}(1, 0.33)$	0.031	0.003	0.086	0.232
22	$bedroom2_{\boldsymbol{inf}}$ (Ac/h)	0.08	$\mathcal{N}(1, 0.33)$	0.032	0.005	0.102	0.232
23	$glazing_{\boldsymbol{R}}$ $(\frac{W}{m^2K})$	0.65	$\mathcal{N}(1, 0.33)$	1.096	0.508	1.037	1.351
24	$glazingBlinds_{\boldsymbol{R}}$ $(\frac{W}{m^2K})$	0.65	$\mathcal{N}(1, 0.33)$	0.51	0.069	0.523	1.045

Prior probability density distribution are scaled according the the initial values.

$PU_{\boldsymbol{k}}$ having an MAP value 28% higher than its design value. The other main envelope conductivities ($brick_{\boldsymbol{k}}$, $EPS_{\boldsymbol{k}}$ and $minWool_{\boldsymbol{k}}$) did not show significant variations after the calibration. The ceiling thermal resistance increased its value, due to $insulCeilBase_{\boldsymbol{k}}$ decreasing by 30%. The floor conductivities stayed almost unvaried with respect to the specified values. Both these construction elements presented lower thermal masses, mainly because of $estrichScreed_{\boldsymbol{c}}$ assuming a value about half its prior estimate. Even if with different direction and magnitude compared to the outcomes from the constant temperature calibration, $blinds_{\boldsymbol{k}}$ showed a posterior estimate sensibly different from its starting value, decreasing by 65%. $glazing_{\boldsymbol{R}}$ varied similarly as before, increasing by 69%.

The thermal mass of the internal partitions reduced sensibly, while their thermal resistances remained practically unchanged. In particular the density and the specific heat of the bricks making these elements ($brickIntWall_{\boldsymbol{\rho}}$ and $brickIntWall_{\boldsymbol{c}}$) decreased by 20% and 45% respectively.

The imposed fixed infiltration rates increased sensibly in *bedroom1*, assuming a value almost three times the initial one. In all the other zones they decreased by about 60%.

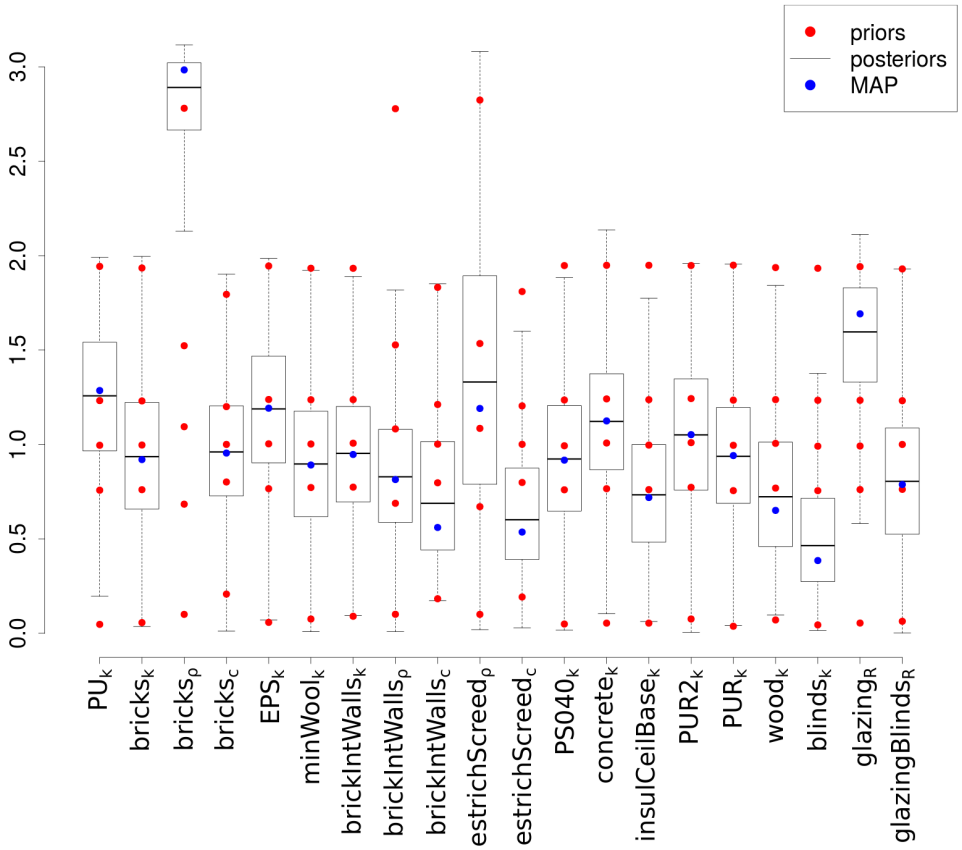


Figure 6.11: $\mathcal{M}_0$: *ROLBS*1 - prior and posterior probability distributions. Posterior samples have been scaled according to their mean to aid visualization (i.e. 1 = initial value).

6.4.3 Discussion

The result produced by the calibrations of $\mathcal{M}_0$ showed significant variations between the MAP estimates of many parameters and the corresponding given specifications. Even more such variations were inconsistent, meaning that they had different magnitudes and directions depending on the analysed experimental period. $\mathcal{M}_0$ underestimated and overestimated the heat losses through the envelope during the *CT*1 and the *ROLBS*1 phases respectively. It also presented different infiltration rates, depending on the analysed period, and sensibly different thermal masses, with respect to the given data, for external walls and partitions, especially in the case of the *ROLBS*1 calibration.

Such inconsistencies between the given data and the inferred values, and between the results of the two calibrations, were deemed to be due to deficiencies in the model. The results from the two *Difference Analyses* showed that the *Difference Vectors* were correlated to wind speed, wind direction, external temperature, mechanically supplied air temperature, basement air temperature, and

split coefficient between convective and radiative gains from the ROLBS heating sequence.

This information supported the implementation of an airflow network. In particular, given the different inferred values, the assumption of constant infiltration rates was deemed not suitable, and it was considered necessary to simulate a wind driven infiltration regime. This also allowed to distinguish between heat exchanges due to thermal bridges and heat exchanges due to infiltrations. The former were included among the calibration parameters in the following calibrations.

The implementation of an airflow network also provided the model with an alternative heat flow path, allowing for a quicker heat transmission between model zones. This could explain the reduced thermal capacity and resistance inferred for the partitions in the *ROLBS*1 calibration, and the correlation between *Difference Vectors* and basement air temperature.

Finally the split coefficient between convective and radiative gains was found mistakenly fixed to 1, proving that the method is able to spot modelling errors also in quite detailed models. It was added to the calibration parameters as well.

6.5 Sensitivity Analysis of $\mathcal{M}_{2,O5}$

This case study demonstrates the Sensitivity Analysis procedure described in Chapter 3, and it has also the objective of gaining information to help improving $\mathcal{M}_0$.

This investigation focused on model $\mathcal{M}_{2,O5}$, in particular on the additional model parameters due to the implementation of the airflow network model. The other model parameters were fixed to the specified values. It was decided to carry out the analysis, considering only the *ROLBS*2, since more significant for the airflows in the real building. In particular, it was deemed that during the ROLBS phase more significant infiltration and ventilation should occur, with respect to the CT period, because of the more dynamical trends of the internal temperatures influencing the pressures of the different thermal zones.

An initial analysis of the uncertainties due to the implementation of the airflow network was undertaken, by considering scalar as well as vectorial model inputs. Then, the three step sensitivity analysis procedure depicted in Chapter 3, consisting of Factor Screening, Factor Prioritising and Factor Fixing, and Fac-

tor Mapping, was applied. The following vector was assumed as model output:

$$\boldsymbol{T} = [\textit{living room}_{\boldsymbol{Ti}}, \textit{bedroom1}_{\boldsymbol{Ti}}, \textit{bathroom}_{\boldsymbol{Ti}}]$$

where $\textit{living room}_{\boldsymbol{Ti}}$, $\textit{bedroom1}_{\boldsymbol{Ti}}$ and $\textit{bathroom}_{\boldsymbol{Ti}}$ are the *living room*, air temperature, the *bedroom1* air temperature and the *bathroom* air temperature.

In the end of this section, insights useful in performing a future calibration of $\mathcal{M}_{2,O5}$ and useful in calibrating $\mathcal{M}_{2,N2}$ are discussed.

6.5.1 Uncertainty Analysis

Both, scalar and vectorial model inputs were considered. The former consist of variables described by time series and therefore their values change during the experiment and simulations. To adequately characterise the uncertainties of these factors it is necessary to define multidimensional probability distributions depicting time varying marginal probability and correlation patterns relative to observations at different time steps. Indeed especially for weather factors, such as wind speed and direction, monitoring conditions as well as unobserved phenomena may produce time varying magnitudes of the measurement *random errors*. The multidimensional parameters considered were:

- Wind speed, wind direction and external temperature: these factors are responsible for the main boundary conditions imposed by the exterior environment on the building affecting the airflow. In particular they determine pressure at the boundary nodes of the airflow network.

- Temperature set points for the north zones, basement and attic: non perfect control, systematic and random variability of these variables produce changes in the relative zone pressure so affecting the ventilation regime.

- ROLBS heat impulses for the south zones: as the main experimental heat inputs, it was expected that these variables had major influences on the conditions determining ventilation and infiltration.

To adequately represent the uncertainties of unidimensional model parameters, that do not change during the experiment and the simulations, it is sufficient to define univariate probability distributions. The parameters of this kind considered are:

- crack dimensions: because of the relatively low infiltration rates, only small cracks have been assumed as connections between the interior and exterior

environment. Their dimensions are sources of uncertainties since they are difficult to measure.

- wind induced pressure coefficients: these parameters together with wind speed, wind direction and ambient air temperature determine the pressure at the boundary nodes. As they were not directly observed they were subject to major uncertainties and one objective of this study was to assess their influence.

- mechanical ventilation flow rates: particular attention was paid in setting up the experiment to ensure a balance between inflow and outflow from the mechanical ventilation; nevertheless there is the possibility of imbalances due to systematic and random measuring errors, which could had a significant influence on the ventilation regime; therefore these uncertainties were included in the analysis.

- ratios between convective and radiative heat gains from ROLBS pulses ($\boldsymbol{C/R}$): besides possible inaccuracies in these ratios, their values could change because of the particular experimental conditions, zone air temperatures and velocities. For these reasons they would be better represented by multidimensional probability distributions. However the model allows only constant $\boldsymbol{C/R}$ splits, and it was necessary to approximate them with univariate probability density distributions. Variations in these parameters may influence the airflow. In particular these ratios should determine quicker or slower changes in the zone pressures.

A detailed description of the procedures and assumptions adopted in performing *Uncertainty Analysis* for the two difference kind of model inputs follows. The total number of parameters considered was 103 and a summary of the defined uncertainties is in Table 6.5.

Multidimensional variables

The measurements provided had a high sampling frequency, having 1 minute time step. Therefore it was possible to fully apply the procedure depicted in Chapter 3, in assessing the uncertainties relative to multidimensional variables.

Firstly Bootstrap was used for averaging the original time series every ten minutes and in calculating the standard errors relative to these estimates. In this way it was possible to reduce the simulation burden while keeping an adequate simulation time step, and to infer a reasonable prior probability density

distribution for the random uncertainties affecting the averaged time series. Such time series were then processed with Smoothing with Roughness Penalty in order to infer the mean vector and covariance structure for the multidimensional distributions depicting the considered vectorial model inputs. In this case the smoothing also provided for the few missing values in the data. A result example is depicted in Figure 6.12 where the assessed uncertainties relative to wind speed are represented.

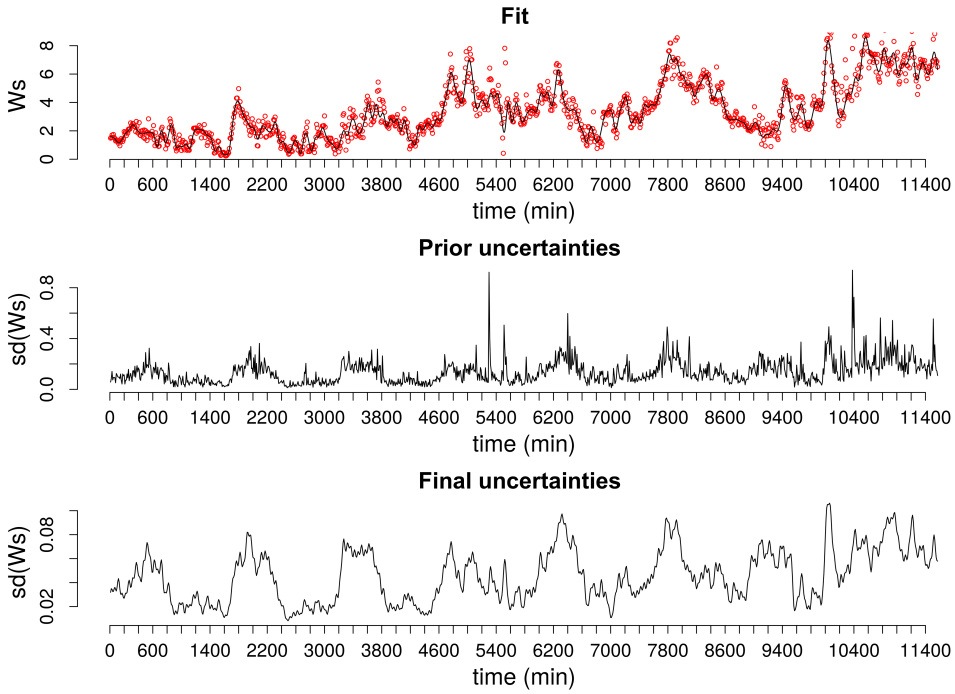


Figure 6.12: ROLBS2: Wind speed - smoothing model fit (red dots: observations, black line: model fit), prior uncertainty from Bootstrap, final uncertainties from smoothing.

In order to simulate systematic errors, values were randomly drawn from unidimensional probability distributions having zero means and with standard deviations half the estimated maximum bounds listed in Table 6.4, and then added to samples drawn from the previously defined multidimensional distributions.

Unidimensional variables

Most of the univariate variables considered have not been directly observed during the experiment. Thus it was necessary to estimate their uncertainties from indirect measurements, analyst experience and information from literature review. The defined probability distributions describing these parameters are listed in Table 6.5.

Table 6.4: EXPERIMENT2: Systematic sensor errors.

Parameter	Units	Set point	Error
Wind speed	m/s	$0-5$	± 0.1
Wind speed	m/s	>5	$\pm 10\%$
Wind direction	$degrees$	$0-360$	± 1
Temperatures	^oC	30	± 0.18
Ventilation rates	m^3/h	60	± 3.5

Ventilation flow rate Ventilation flow rates for mechanically supplied and extracted air were measured during the experiment with one minute time step. However the flow network model represents mechanical ventilation inflow and outflow with constant volume flow rate components ([45]) and it does not allow the use of time varying flow rates. Therefore univariate probability distributions were used to summarise the information relating to such variables. A model similar to the one adopted for multi-dimensional variables was considered:

$$Qvnt = \hat{Qvnt} + s + \varepsilon \tag{6.1}$$

where $Qvnt$ and $\hat{Qvnt}$ indicates a certain volume flow rate and its estimate, s is the systematic error term and ε depicts the random uncertainty. $\hat{Qvnt}$ and ε were assessed by bootstrapping the entire time series and s was defined in a similar way as for the multi-dimensional variables, that is equal to half the sensor error provided with the experimental specifications (Table 6.4). The resulting probabilistic model for $Qvnt$ is:

$$Qvnt \sim \mathcal{N}(\hat{Qvnt}, \mathrm{V}(s+\varepsilon)) \tag{6.2}$$

$$\mathrm{V}(s+\varepsilon) = (s/2)^2 + \mathrm{V}(\varepsilon) \tag{6.3}$$

where $\mathrm{V}(\varepsilon)$ is the variance of the independently and identically distributed variable ε.

Assuming the variance of $Qvnt$ as indicated in Equation 6.3 is a simplification, since *random errors* will not always be in the same direction as systematic errors. However estimating $Qvnt$ from the entire time series produces random uncertainties negligible compared to the systematic ones. In particular $\mathrm{sd}(\varepsilon)$ had estimates of 0.0032 and 0.0035 m^3/h for inflow and outflow respectively, while s was equal to 3.5 m^3/h. Thus even if it results in a slight overestimation of the uncertainties, the assumption in equation Equation (6.3) can be considered reasonable.

Table 6.5: $\mathcal{M}_{2,O5}$: Considered parameters and relative prior distributions.

#	PARAMETER	TYPE	UNITS	INITIAL ESTIMATE	DISTRIBUTION	ID
1	wind velocity	md	m/s			$\boldsymbol{Ws}$
2	wind direction	md	$degrees$			$\boldsymbol{Wd}$
3	ambient temperature	md	^oC			$\boldsymbol{Te}$
4	basement temperature	md	^oC			$basement_{\boldsymbol{Ti}}$
5	*kitchen* temperature	md	^oC			$kitchen_{\boldsymbol{Ti}}$
6	*lobby* temperature	md	^oC			$lobby_{\boldsymbol{Ti}}$
7	*bedroom2* temperature	md	^oC			$bedroom2_{\boldsymbol{Ti}}$
8	attic temperature	md	^oC			$attic_{\boldsymbol{Ti}}$
9	mech. vent. air temperature	md	^oC			$\boldsymbol{Tvnt}$
10	mech. vent. inflow rate	ud	m^3/h	60	$\mathcal{N}(60, 1.75)$	$living\ room_{Qvnt}$
11	mech. vent. outflow rate *bathroom*	ud	m^3/h	30	$\mathcal{N}(60, 0.88)$	$bathroom_{Qvnt}$
12	mech. vent. outflow rate *bedroom1*	ud	m^3/h	30	$\mathcal{N}(60, 0.88)$	$bedroom1_{Qvnt}$
13	*lobby* external door crack width	ud	mm	1	$\mathcal{U}(0.4, 1.6)$	$crack1_{\boldsymbol{width}}$
14	*lobby* external door crack length	ud	m	2.4	$\mathcal{U}(0.96, 3.84)$	$crack1_{\boldsymbol{len}}$
15	*living room* south window crack width	ud	mm	1	$\mathcal{U}(0.4, 1.6)$	$crack2_{\boldsymbol{width}}$
16	*living room* south window crack length	ud	m	7.6	$\mathcal{U}(3.04, 12.16)$	$crack2_{\boldsymbol{len}}$
17	*living room* west window crack width	ud	mm	1	$\mathcal{U}(0.4, 1.6)$	$crack3_{\boldsymbol{width}}$
18	*living room* west window crack length	ud	m	2	$\mathcal{U}(0.8, 3.2)$	$crack3_{\boldsymbol{len}}$
19	*bedroom1* window crack width	ud	mm	1	$\mathcal{U}(0.4, 1.6)$	$crack4_{\boldsymbol{width}}$
20	*bedroom1* window crack length	ud	m	2	$\mathcal{U}(0.8, 3.2)$	$crack4_{\boldsymbol{len}}$
21	*bathroom* window crack width	ud	mm	1	$\mathcal{U}(0.4, 1.6)$	$crack5_{\boldsymbol{width}}$
22	*bathroom* window crack length	ud	m	2	$\mathcal{U}(0.8, 3.2)$	$crack5_{\boldsymbol{len}}$
23	*bedroom2* window crack width	ud	mm	1	$\mathcal{U}(0.4, 1.6)$	$crack6_{\boldsymbol{width}}$
24	*bedroom2* window crack length	ud	m	2	$\mathcal{U}(0.8, 3.2)$	$crack6_{\boldsymbol{len}}$
25	*kitchen* window crack width	ud	mm	1	$\mathcal{U}(0.4, 1.6)$	$crack7_{\boldsymbol{width}}$
26	*kitchen* window crack length	ud	m	2	$\mathcal{U}(0.8, 3.2)$	$crack7_{\boldsymbol{len}}$
27	*living room-kitchen* door crack width	ud	mm	2	$\mathcal{U}(0.8, 3.2)$	$crack8_{\boldsymbol{width}}$
28	*living room-kitchen* door crack length	ud	m	1	$\mathcal{U}(0.4, 1.6)$	$crack8_{\boldsymbol{len}}$
29	*living room-lobby* door crack width	ud	mm	2	$\mathcal{U}(0.8, 3.2)$	$crack9_{\boldsymbol{width}}$
30	*living room-lobby* door crack length	ud	m	1	$\mathcal{U}(0.4, 1.6)$	$crack9_{\boldsymbol{len}}$
31	*corridor-bedroom2* door crack width	ud	mm	2	$\mathcal{U}(0.8, 3.2)$	$crack10_{\boldsymbol{width}}$
32	*corridor-bedroom2* door crack length	ud	m	1	$\mathcal{U}(0.4, 1.6)$	$crack10_{\boldsymbol{len}}$
33	*living room*-basement crack width	ud	mm	2	$\mathcal{U}(0.8, 3.2)$	$crack11_{\boldsymbol{width}}$
34	*living room*-basement crack length	ud	m	1	$\mathcal{U}(0.4, 1.6)$	$crack11_{\boldsymbol{len}}$
35	*living room*-attic crack width	ud	mm	2	$\mathcal{U}(0.8, 3.2)$	$crack12_{\boldsymbol{width}}$
36	*living room*-attic crack length	ud	m	1	$\mathcal{U}(0.4, 1.6)$	$crack12_{\boldsymbol{len}}$
37	pressure coefficient 0^o	ud	-	0.7	$\mathcal{U}(-0.75, 0.8)$	pc_0
38	pressure coefficient 22.5^o	ud	-	0.525	$\mathcal{U}(-0.65, 0.69)$	$pc_{22.5}$
39	pressure coefficient 45^o	ud	-	0.35	$\mathcal{U}(-0.60, 0.42)$	pc_{45}
40	pressure coefficient 67.5^o	ud	-	-0.075	$\mathcal{U}(-0.55, 0.04)$	$pc_{67.5}$
41	pressure coefficient 90^o	ud	-	-0.5	$\mathcal{U}(-1.02, 0.2)$	pc_{90}
42	pressure coefficient 112.5^o	ud	-	-0.45	$\mathcal{U}(-0.7, -0.12)$	$pc_{112.5}$
43	pressure coefficient 135^o	ud	-	-0.4	$\mathcal{U}(-0.75, 0.5)$	pc_{135}
44	pressure coefficient 157.5^o	ud	-	-0.3	$\mathcal{U}(-0.9, 0.375)$	$pc_{157.5}$
45	pressure coefficient 180^o	ud	-	-0.2	$\mathcal{U}(-0.8, 0.04)$	pc_{180}
46	pressure coefficient 202.5^o	ud	-	-0.3	$\mathcal{U}(-0.85, 0.04)$	$pc_{202.5}$
47	pressure coefficient 225^o	ud	-	-0.4	$\mathcal{U}(-0.6, 0.3)$	pc_{225}
48	pressure coefficient 247.5^o	ud	-	-0.450	$\mathcal{U}(-0.75, 0.15)$	$pc_{247.5}$
49	pressure coefficient 270^o	ud	-	-0.5	$\mathcal{U}(-1.03, 0.15)$	pc_{270}
50	pressure coefficient 292.5^o	ud	-	-0.075	$\mathcal{U}(-0.49, -0.04)$	$pc_{292.5}$
51	pressure coefficient 315^o	ud	-	-0.35	$\mathcal{U}(-0.51, 0.7)$	pc_{315}
52	pressure coefficient 337.5^o	ud	-	-0.053	$\mathcal{U}(-0.2, 0.7)$	$pc_{337.5}$
101	*living room* heater $\boldsymbol{C/R}$ split	ud	-	0.7	$\mathcal{N}(0.7, 0.1)$	$living\ room_{\boldsymbol{C/R}}$
102	*bedroom1* room heater $\boldsymbol{C/R}$ split	ud	-	0.7	$\mathcal{N}(0.7, 0.1)$	$bedroom1_{\boldsymbol{C/R}}$
103	*bathroom* heater $\boldsymbol{C/R}$ split	ud	-	0.7	$\mathcal{N}(0.7, 0.1)$	$bathroom_{\boldsymbol{C/R}}$

For pressure coefficients: the angle is referenced to the normal direction to the surface

For pressure coefficients: they are repeated four times (once for each boundary node considered: EAST, SOUTH and WEST)

md = multidimensional; ud = unidimensional.

Crack lengths and widths The length and width of the crack components were evaluated according to the results given by the pressurization test results at 50 Pa. Two blower door tests were performed, one for the whole ground floor and one involving only the SOUTH_ZONE:

- whole ground floor: 1.54 Ac/h.
- SOUTH_ZONE: 2.3 Ac/h.

The former should give a good picture of the total ground floor infiltration while the latter represented a mix between infiltration and ventilation between north and south zones. For the two tests the total leakage area (A) was derived according to the *orifice equation* ([45]):

$$Qm = C_d A \sqrt{2\boldsymbol{\rho}\Delta P} \tag{6.4}$$

where Qm is the mass flow rate, $C_d = 0.61$ is the discharge coefficient, $\boldsymbol{\rho} = 1.2$ kg/m^3 is the air density and $\Delta P = 50$ Pa is the pressure difference. From the result for the whole ground floor, A was decomposed relative to NORTH_ZONE and SOUTH_ZONE according to volume proportions. Then from the result regarding only SOUTH_ZONE, it was possible to assess the leakage area responsible for ventilation only, by difference. Crack lengths were estimated depending on opening characteristics and experience and consequently the widths were calculated. Uniform distributions involving ranges of ±60% the estimated values were adopted for these variables.

Wind induced pressure coefficients Wind induced pressure coefficients were possibly the most uncertain variables in the model. No information about their uncertainties came from the experiment and thus suitable probability distributions were inferred depending upon data from literature review. A complete treatment of their variability would consider the correlation between them, due to their dependence on wind speed, wind direction, location on the surface, and configuration of the surrounding area. However, with the available data it was not possible to adequately model such correlation relationships and they have been considered independent. Neglecting the correlation between these model inputs causes overestimation of their uncertainties, whereas considering their interdependence without any any specific measurements may lead to the opposite problem. The former option was selected because it was more conservative.

An extensive review of secondary sources of data for pressure coefficients can be found in [30]. This study compares pressure coefficient values from different

databases as a function of different sheltering conditions and derives plausible variation ranges depending on wind directions relative to surface normals. Such information was integrated with the data available from the ESP-r database, including different aspect ratios for walls, in order to define suitable variation ranges. In the flow network model, each boundary node is defined by 16 pressure coefficients for wind directions defined every 22.5^o relative to the surface normals. The boundary nodes considered are those named NORTH, EAST, SOUTH and WEST in Figure 6.5 for a total of 64 pressure coefficients.

Convective/radiative split for heaters (C/R) In the model only the coefficients relating to the convective part were treated as random variables, defining the remaining fractions by difference. An estimate for such variables was given by the heater manufacturer. Therefore it was considered to be substantially less uncertain than other parameters in the model, and normal probability density distributions with mean 0.7 and standard deviation 0.1 were assumed.

6.5.2 Factor Screening

The results from FS are depicted in Figure 6.13, where the first order effects calculated from the Morris Method ($M^\star{}_p$ indexes) are shown. In particular the ten most important model parameters are highlighted.

$living\ room_{\boldsymbol{C/R}}$ and $living\ room_{Qvnt}$ are the two most influential variables followed by $\boldsymbol{Ws}$, $crack3_{\boldsymbol{len}}$, $crack5_{\boldsymbol{len}}$, $EAST_{pc_{157.5}}$, $SOUTH_{pc_{90}}$, $SOUTH_{pc_{45}}$, $bedroom1_{\boldsymbol{C/R}}$ and $bathroom_{\boldsymbol{C/R}}$ which have very similar $M^\star{}_p$ indexes. These 10 factors have been labelled as MIF and grouped according to the phenomena they represent:

- $living\ room_{\boldsymbol{C/R}}$, $bedroom1_{\boldsymbol{C/R}}$, $bathroom_{\boldsymbol{C/R}}$ have been collected in the $\boldsymbol{C/R}$ parameter;
- $EAST_{pc_{157.5}}$, $SOUTH_{pc_{90}}$, $SOUTH_{pc_{45}}$ have been gathered in the PC parameter;
- $crack3_{\boldsymbol{len}}$ and $crack5_{\boldsymbol{len}}$ have been grouped in the CRACK parameter;

while $living\ room_{Qvnt}$ and $\boldsymbol{Ws}$ have been considered separately.

From the results from FS it is also possible to identify the wind direction most influencing the Twin House O5. By rotating the pressure coefficients azimuth angles, in order to refer to the same reference direction, north, the wind coming from direction within the range $[225^o, 270^o]$ seems to have the largest effects on the internal temperatures.

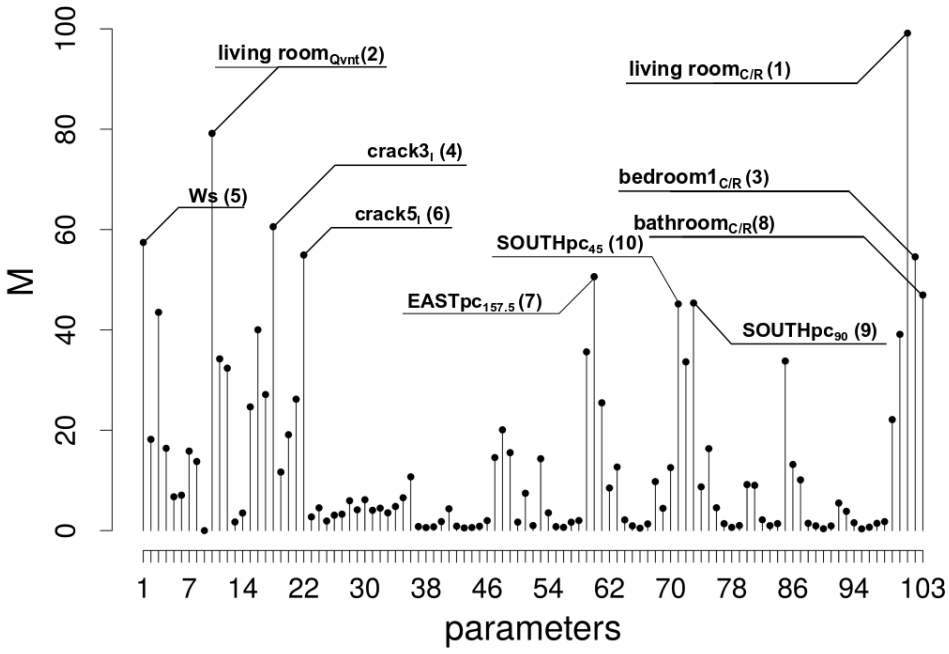


Figure 6.13: $\mathcal{M}_{2,O5}$: Main effects ($M^{\star}{}_p$ indexes) from the Morris Method for $\boldsymbol{T}$.

6.5.3 Factor Fixing and Factor Prioritising

The first order (S_i) and total effects (ST_i) are listed in Table 6.6.

Table 6.6: $\mathcal{M}_{2,O5}$: First order (S_i) and total effects (ST_i) from the Sobol Methods for $\boldsymbol{T}$.

Parameter	S_i	ST_i
$\boldsymbol{C/R}$	0.42	0.46
PC	0.08	0.16
$\boldsymbol{Ws}$	0.00	0.00
$living\ room_{Qvnt}$	0.09	0.08
CRACK	0.02	0.02
LIF	0.25	0.31

The model was mainly dominated by first order effects as the small differences between S_i and ST_i indicate. In particular the sum of the first order effects was equal to the 86% of V($\boldsymbol{T}$) meaning that about 14% of V($\boldsymbol{T}$) is due to higher order effects. Most of the higher order effects could be attributed to interactions between PC and LIF. They were the only two groups having a noticeable difference between their ST_i and S_i.

In the light of this consideration the higher order effects between the defined

group of parameters were negligible. For the total variance, 61% could be attributed to the MIF (most of which is attributable to ***C/R***), 25% to the LIF and about 14% to interactions occurring especially between PC and LIF. Thus even if MIF accounted for the majority of the model variance, still one third of it was determined by less important factors and their approximation to default values should be undertaken with caution.

6.5.4 Factor Mapping

Preliminary trials were useful in determining a reasonable value for the parameter α in Equation 3.22, which was set to 8. The model did not provide a particularly good fit of the measured data. In particular it was able to provide reasonable predictions in the middle part of the ROLBS experiment, but at the beginning and at the end of the heating sequence the simulation outcomes overestimated the observed internal temperatures. This trend was noticeable especially for the *living room* (Figure 6.14a), while for the *bedroom1* (Figure 6.14b) and the *bathroom* (Figure 6.14c) the discrepancies between model outputs and measurements were less evident.

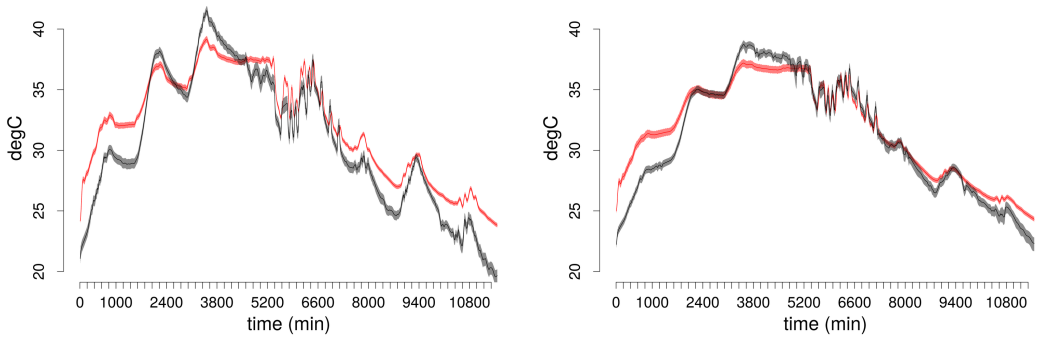
The main causes were probably model deficiencies lying in the analysed submodels or in other parts of the overall BES model. In this study they were investigated by calculating the Pearson Correlation Coefficient between the residuals and multi-dimensional model inputs showing that ROLBS sequences and residuals were moderately correlated (Table 6.7). The analysis has thus identified an aspect of the model and/or program that needs to be improved in order to get a better match with the measured data.

Table 6.7: $\mathcal{M}_{2,O5}$: Correlation between residuals and ROLBS heating sequences.

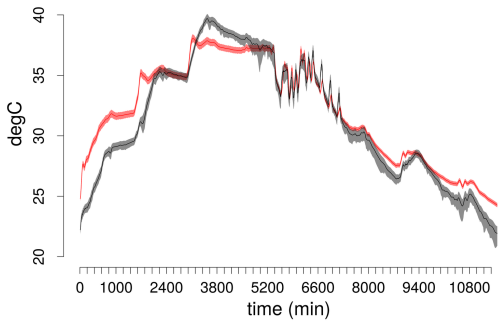
Parameter	*living room*	*bedroom1*	*bathroom*
ROLBS in *living room*	0.38	0.41	0.42
ROLBS in *bedroom1*	0.38	0.41	0.42
ROLBS in *bathroom*	0.37	0.42	0.51

Figure 6.15 shows a comparison between prior and posterior probability density distributions while Table 6.8 contains the posterior estimates and 95% confidence intervals for the MIF.

In particular the posterior distributions have been generated by sampling with replacement the simulation input vectors using as sampling weights ω_m. Even if the prior and posterior variation ranges are substantially the same there were



(a) *living room*. (b) *bedroom1*.



(c) *bathroom*

Figure 6.14: $\mathcal{M}_{2,O5}$: Comparison between model predictions (black) and observed internal air temperatures (red).

shifts between prior and posterior estimates. The three $\boldsymbol{C/R}$ ratios had estimates very close to their initial values, especially $bedroom1_{\boldsymbol{C/R}}$ and $bathroom_{\boldsymbol{C/R}}$, while $living\ room_{\boldsymbol{C/R}}$ assumes a value slightly higher. Similar considerations were drawn for the inflow ventilation rate. Its posterior value, although slightly lower, was substantially in agreement with the one inferred from the data. More significant variations between prior and posterior estimates could be observed for crack parameters and pressure coefficients, especially for the latter. Crack lengths assumed values about 4% lower than the initial model considered. Pressure coefficients moved sensibly from their initial values: $EAST_{pc_{157.5}}$ increases by 44%, $SOUTH_{pc_{45}}$ decreases by 89% and $SOUTH_{pc_{90}}$ increases by 28%.

One possible cause of significant posterior variance for crack lengths and pressure coefficients could be *over-parametrisation* of the model. This aspect has been analysed by applying the Sobol Method to the calculated weights (Table 6.9).

The variance of the weighting function is mostly due to PC and LIF. This was unexpected since the results from FF and FP were showing that $\boldsymbol{C/R}$ coefficients

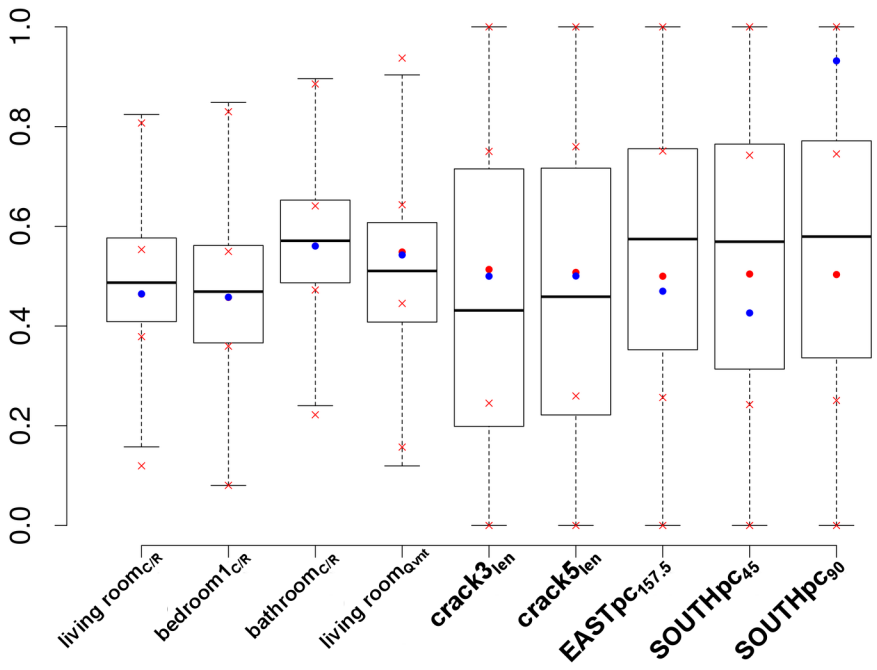


Figure 6.15: Comparison between prior (red crosses: quartiles, red dots: averages, blue dots: initial values) and posterior (box-plot) parameter distributions, for MIF. The samples have been normalized between 0 and 1.

were responsible by themselves for 42% of the model variance. Additive and linear effects accounted for the 75% of V($\boldsymbol{\omega}$), so that the remaining 25% can be attributed to higher order effects mainly due to PC and LIF.

These results indicate that among the LIF there were parameters important for model calibration and validation because of their first order effects or interactions with PC. Such variables were identified by comparing their prior and posterior distributions in the same way as was done for the MIF. LIF factors showing significant differences were likely to contribute in a relevant manner to a better match with the measurements. Such a comparison is shown in Figure 6.16 where the variables having the larger differences between their prior and posterior averages are highlighted. These parameters were $crack2_{\boldsymbol{width}}$, $crack2_{\boldsymbol{len}}$, $EAST_{pc_{135}}$, $SOUTH_{pc_{67.5}}$, $WEST_{pc_{337.5}}$.

Also scatter plots of $\boldsymbol{\omega}$ against the posterior parameter samples were analysed, and correlation between them calculated. In this way it was possible to see how the goodness of the fit changed over the range of variation of each parameter, and to break-down groups of factors in order to assess them individually. The latter aspect was particularly helpful in understanding the significantly higher

Table 6.8: $\mathcal{M}_{2,O5}$: Posterior estimates and 95% confidence intervals.

Parameter	units	posterior estimate	95% c. i.	
$living\ room_{\boldsymbol{C/R}}$	-	0.72	0.53	0.90
$bedroom1_{\boldsymbol{C/R}}$	-	0.71	0.50	0.90
$bathroom_{\boldsymbol{C/R}}$	-	0.71	0.52	0.90
$living\ room_{Qvnt}$	m^3/h	59.4	54	61.2
$crack3_{\boldsymbol{len}}$	m	1.91	0.85	3.12
$crack5_{\boldsymbol{len}}$	m	1.93	0.87	3.14
$EAST_{pc_{157.5}}$	-	-0.20	-0.84	0.32
$SOUTH_{pc_{45}}$	-	-0.04	-0.56	0.40
$SOUTH_{pc_{90}}$	-	-0.36	-0.97	0.16

Table 6.9: $\mathcal{M}_{2,O5}$: First order (S_i) and total effects (ST_i) from the Sobol Method relative to ω_i.

Parameter	S_i	ST_i
$\boldsymbol{C/R}$	0.06	0.08
PC	0.20	0.39
$\boldsymbol{Ws}$	0.00	0.00
$living\ room_{Qvnt}$	0.08	0.12
CRACK	0.04	0.07
LIF	0.37	0.49

sensitivity that model calibration showed with respect to the pressure coefficients compared to the convective/radiative ratios, which was not in agreement with the results from FP and FF. In particular this analysis showed that all the pressure coefficients in PC had relatively significant correlations with the calculated values of the weighting function and comparable with the correlation of $living\ room_{\boldsymbol{C/R}}$, whereas $bedroom1_{\boldsymbol{C/R}}$ and $bathroom_{\boldsymbol{C/R}}$ were only weakly correlated with $\boldsymbol{\omega}$. Furthermore, the value of the weighting function did not show significant changes over the range of $bathroom_{\boldsymbol{C/R}}$; therefore, varying this factor does not improve the match with the target data. The scatter plots for the parameters showing higher correlation with $\omega(\cdot)$ are in Figure 6.17.

Further investigation could neglect the $bathroom_{\boldsymbol{C/R}}$ and focus on the remaining identified factors important for model calibration, namely: $living\ room_{\boldsymbol{C/R}}$, $bedroom1_{\boldsymbol{C/R}}$, $living\ room_{Qvnt}$, $EAST_{pc_{157.5}}$, $SOUTH_{pc_{45}}$, $SOUTH_{pc_{90}}$, $SOUTH_{pc_{67.5}}$, $WEST_{pc_{337.5}}$, $crack2_{\boldsymbol{width}}$, $crack2_{\boldsymbol{len}}$ and $EAST_{pc_{135}}$, especially on crack parameters and pressure coefficients. In particular, since the posterior estimates for the $\boldsymbol{C/R}$ coefficients and $living\ room_{Qvnt}$ were quite close to their prior values, these experimental data could be considered accurate. The gained information

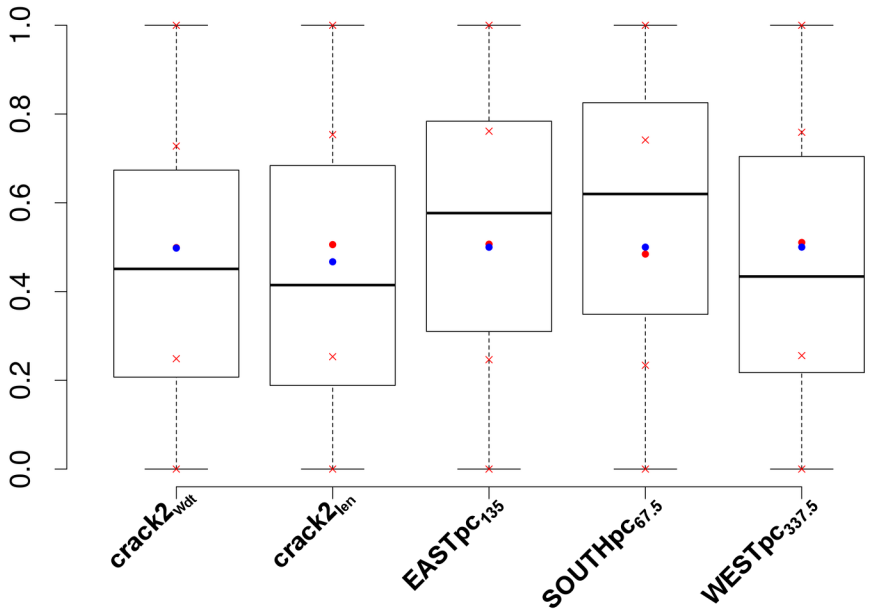
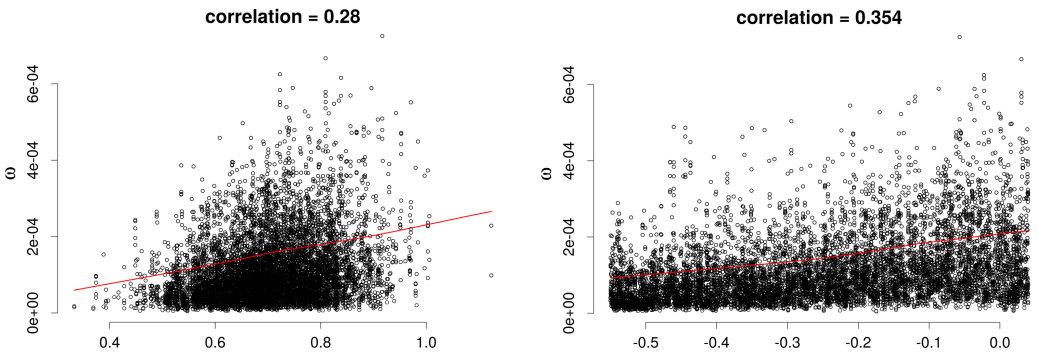


Figure 6.16: $\mathcal{M}_{2,O5}$: Comparison between prior (red crosses: quartiles, red dots: averages, blue dots: initial values) and posterior (box plot) parameter distributions for LIF. The samples have been normalized between 0 and 1.

could be used to reduce prior uncertainties, for example by replacing the uniform prior distributions of parameters showing significant shifts from their initial values with normal probability density distributions or by setting up prior density distributions favouring parameter values corresponding to higher ω_i according to the graphs in Figure 6.17. However, the latter approach should be used with caution because model *over-parametrisation* can result in more than one set of parameter values giving similar agreement with the measured data. This means that once an input vector is found to give good agreement between model outputs and measurements, moving the parameters as indicated in Figure 6.17 is likely to produce similar fits.

6.5.5 Discussion

This section has demonstrated the multi-step Sensitivity Analysis described in Chapter 3. In particular the explained methodology is particularly apt as a preparatory phase to model calibration and it unfolds in three phases, involving the application of qualitative (Morris Method) as well as quantitative (Sobol Method) techniques: Factor Screening, Factor Prioritising and Factor Fixing, and



(a) $living\ room_{\boldsymbol{C/R}}$. (b) $SOUTH_{pc_{67.5}}$.

Figure 6.17: $\mathcal{M}_{2,O5}$: $\omega(\cdot)$ trend along parameter posterior variation ranges.

Factor Mapping. During the first two steps the most influential model parameters were identified and the model variance neglected by considering only these parameters was assessed. The FM stage demonstrated how the output from GSA can be used to gain information aiding subsequent calibration or validation studies, by indicating sources of model inadequacy and the model input governing the goodness of fit with the target measured data.

Although the presence of inadequacies in $\mathcal{M}_{2,O5}$ might influence the results, it was possible to identify the parameters most responsible for producing good matches with the target data, mainly consisting of pressure coefficients. Thus future investigation should reduce model inadequacy and focus on $living\ room_{\boldsymbol{C/R}}$, $bedroom1_{\boldsymbol{C/R}}$, $living\ room_{Qvnt}$, $EAST_{pc_{157.5}}$, $SOUTH_{pc_{45}}$, $SOUTH_{pc_{90}}$, $SOUTH_{pc_{67.5}}$, $WEST_{pc_{337.5}}$, $crack2_{\boldsymbol{width}}$, $crack2_{\boldsymbol{len}}$ and $EAST_{pc_{135}}$, in order to create a model more representative of the real experiment. In particular ratios between convective and radiative heat gains from the injected ROLBS pulses appear to be particularly important and, although their values should be close to the given specification, in future experiments it may be useful to measure these variables on site. Uncertainties in wind induced pressure coefficients deserve a more rigorous treatment by accounting for their correlations. On site measurements as well as wind tunnel experiments could be helpful, although the results from the latter case will be affected by all the limitations of a scaled laboratory experiment.

The comparison between the outcomes of the two sensitivity techniques employed highlight that the Least Important Factors are not negligible since they are responsible for relevant fractions of the model and weighting function variances. Thus it is important to measure the portion of model variance considered by working only with the retained factors.

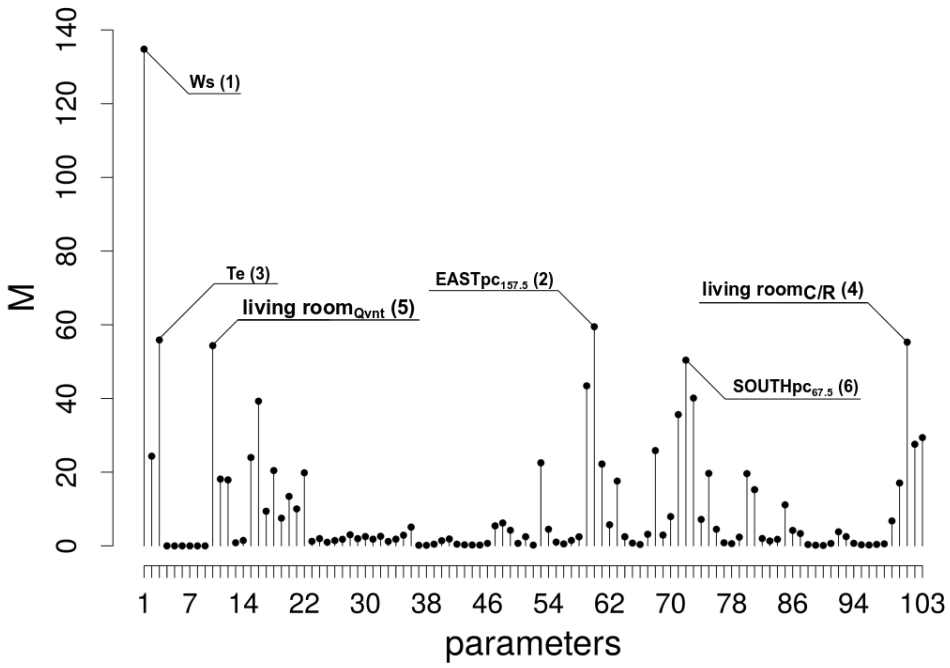


Figure 6.18: $\mathcal{M}_{2,O5}$: FS results by considering constant uncertainties for multi-dimensional variables.

Finally, the consequence of employing the depicted methodology in estimating uncertainties for vectorial model inputs, compared to adopting a more simple approach like considering constant uncertainties was investigated. In particular, FS was also performed considering constant variances equal to the average of the estimated variances using Bootstrap. The results are presented in Figure 6.18.

Comparing this with Figure 6.13, the same parameters still have the most influence, but the relative order of importance has changed, so that $\boldsymbol{Ws}$ is the dominant parameter when constant uncertainty is assumed. It is believed that the proposed approach is more sensible and rigorous than the common practice, leading to more realistic estimates of the uncertainties for multi-dimensional model inputs. Thus particularly when a substantial amount of information is available, it is advisable to employ advanced statistical techniques, like those used, in performing *Uncertainty Analysis*.

According to the results provided by the undertaken analysis, the following considerations were made in calibrating $\mathcal{M}_{2,N2}$. The results from FP and FF, showed that the only boundary condition highlighted as important for the behaviour of the $\mathcal{M}_{2,O5}$ by FS (i.e. $\boldsymbol{Ws}$), had negligible influences on the model outputs. Also in FM, $\boldsymbol{Ws}$ had a particularly low effect in improving the match with the measured data. These outcomes allowed the results from the performed

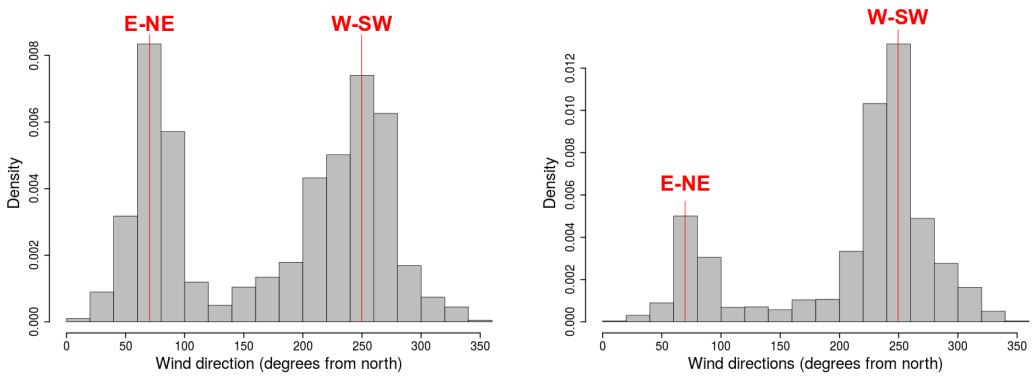
sensitivity analysis to be extrapolated to other boundary conditions, that is those imposed on $\mathcal{M}_{2,N2}$. Therefore it was deemed suitable to add to the calibration parameter set, split coefficients between convective and radiative gains for ROLBS pulses, pressure coefficients, mechanical ventilation rates, and crack dimensions.

6.6 Calibration of $\mathcal{M}_{2,N2}$

The detailed sensitivity analysis carried out on model $\mathcal{M}_{2,O5}$, provided evidence supporting the need to include in the calibration parameter set crack dimensions, pressure coefficients, mechanical ventilation flow rates and split coefficients between convective and radiative heat gains from the ROLBS pulses. However, to fully consider all these factors would have largely increased the number of calibration parameters, consequently increasing the computational burden and the complexity of the calculations. For example, the consideration of all the pressure coefficients would require the addition of 64 calibration parameters. Even more some of these model inputs, like crack widths and crack lengths, have correlated effects, thus possibly leading to identifiability problems.

Therefore, the following considerations were made. It was deemed suitable to fix the crack width, and to take into account as calibration parameters only the crack lengths considering for them increased variation ranges, in order to account for the additional uncertainties due to fixing crack widths. A simplified approach was used for pressure coefficients. In particular, to each boundary nodes of the airflow network model (NORTH, EAST, SOUTH and WEST) was attributed the same set of pressure coefficients values (Table 6.10) multiplied by an independent factor ($NORTH_{\mathcal{F}}$, $EAST_{\mathcal{F}}$, $SOUTH_{\mathcal{F}}$ and $WEST_{\mathcal{F}}$). Such factors were considered as calibration parameters and in the following they will be referred as pressure coefficient factors or $\mathcal{F}$ parameters. Such a way of considering pressure coefficients was supported also by the analysis of the dominant wind directions. Figure 6.19 shows the empirical probability density distributions of the measured wind directions during $CT1$ and $ROLBS1$. For the experimental periods the dominant wind directions were east-northeast and west-southwest, thus it was likely that only the pressure coefficients corresponding to these directions had a significant influence on the model behaviour. The thermal bridges were considered among the calibration parameters as well.

The parameters considered in the $CT1$ and $ROLBS1$ calibrations and their prior probability density distributions are displayed in Tables 6.11 and 6.12 respectively.



(a) $CT1$ (b) $ROLBS1$

Figure 6.19: Wind directions

Table 6.10: $\mathcal{M}_{2,N2}$: initial set of pressure coefficients. Wind directions are indicated in degrees from north.

directions	0	22.5	45	67.5	90	112.5	135	157.5
pressure coefficients	0.7	0.525	0.35	-0.75	-0.5	-0.450	-0.4	-0.3
directions	180	205.5	225	247.5	270	292.5	315	337.5
pressure coefficients	-0.2	-0.3	-0.4	-0.450	-0.5	-0.075	-0.350	-0.525

6.6.1 Constant temperature

In Table 6.11 are listed the MAP estimates and the confidence intervals for the considered calibration parameters. In Figures 6.20, a comparison is shown between the considered prior probability density distributions and inferred posterior probability density distributions, for the calibration parameters for which specifications were provided.

From Figure 6.20 it is possible to see that all these model inputs had estimates that can be considered in agreement with the given specifications. Indeed, most of them had MAP estimates within one quartile from the means of their prior distributions, that is the specified values.

$crack4_{\boldsymbol{len}}$, $crack6_{\boldsymbol{len}}$ and $crack11_{\boldsymbol{len}}$ had MAP values between 2 and 2.5 times their initial values. All the others decreased by amounts between 25% and 57%.

$\mathcal{F}$ parameters assumed negative values, with $NORTH_{\mathcal{F}}$ and $WEST_{\mathcal{F}}$ assuming the largest values in modulus.

$bathroom_{Qvnt}$ and $bedroom1_{Qvnt}$ had posterior estimates practically equal to the given data. $living\ room_{Qvnt}$had MAP value (0.032 m^3/h) slightly lower than the mean measured inflow rate of 0.033 m^3/h, but well within plausible uncertainty.

Table 6.11: $\mathcal{M}_{2,N2}$–$CT1$: prior probability density distributions, MAP estimates and 95% confidence intervals for calibration parameters.

#	PARAMETERS	SPECIFIED VALUES	PRIOR DISTRIBUTION	MAP	$Q2.5\%$	$Q50\%$	$Q97.5\%$
1	$PU_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.035	$\mathcal{N}(1, 0.35)$	0.041	0.019	0.041	0.063
2	$brick_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.220	$\mathcal{N}(1, 0.35)$	0.222	0.087	0.224	0.374
3	$EPS_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.022	$\mathcal{N}(1, 0.35)$	0.022	0.007	0.022	0.037
4	$minWool_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.036	$\mathcal{N}(1, 0.35)$	0.032	0.09	0.032	0.057
5	$brickIntWall_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.331	$\mathcal{N}(1, 0.35)$	0.345	0.120	0.343	0.562
6	$PS040_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.040	$\mathcal{N}(1, 0.35)$	0.042	0.016	0.041	0.066
7	$concrete_{\boldsymbol{k}}$ $(\frac{W}{mK})$	2.000	$\mathcal{N}(1, 0.35)$	1.954	0.643	1.970	3.349
8	$concrete_{\boldsymbol{\rho}}$ $(\frac{kg}{m^3})$	2400.000	$\mathcal{N}(1, 0.15)$	2464.8	1817.462	2460.832	3093.981
9	$concrete_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	1000.000	$\mathcal{N}(1, 0.3)$	942.543	392.393	942.172	1471.422
10	$insulCeilBase_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.035	$\mathcal{N}(1, 0.35)$	0.036	0.015	0.036	0.059
11	$ausgleich_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.060	$\mathcal{N}(1, 0.35)$	0.054	0.019	0.055	0.094
12	$PUR2_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.025	$\mathcal{N}(1, 0.35)$	0.025	0.009	0.025	0.042
13	$PUR_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.023	$\mathcal{N}(1, 0.35)$	0.203	0.006	0.020	0.034
14	$wood_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.131	$\mathcal{N}(1, 0.35)$	0.122	0.034	0.124	0.213
15	$wood_{\boldsymbol{\alpha}}$ (-)	0.600	$\mathcal{U}(0.8, 1.2)$	0.632	0.377	0.607	0.827
16	$blinds_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.100	$\mathcal{N}(1, 0.35)$	0.120	0.052	0.120	0.180
17	$softwood_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.130	$\mathcal{N}(1, 0.35)$	0.120	0.040	0.121	0.203
18	$glazing_{\boldsymbol{R}}$ $(\frac{m^2K}{W})$	0.65	$\mathcal{N}(1, 0.33)$	0.614	0.199	0.613	1.055
19	$glazingBlinds_{\boldsymbol{R}}$ $(\frac{m^2K}{W})$	0.65	$\mathcal{N}(1, 0.33)$	0.597	0.193	0.597	0.987
20	$living\ room_{tb}$ $(\frac{W}{K})$	4.3	$\mathcal{U}(0, 20)$	1.46	0.13	2.35	4.99
21	$bedroom1_{tb}$ $(\frac{W}{K})$	6.68	$\mathcal{U}(0, 20)$	9.49	0.68	7.74	12.92
22	$kitchen_{tb}$ $(\frac{W}{K})$	7.63	$\mathcal{U}(0, 20)$	7.32	0.57	7.34	14.57
23	$lobby_{tb}$ $(\frac{W}{K})$	1.68	$\mathcal{U}(0, 20)$	3.92	0.08	1.66	3.27
24	$bathroom_{tb}$ $(\frac{W}{K})$	6.15	$\mathcal{U}(0, 20)$	2.03	0.35	5.64	11.75
25	$bedroom2_{tb}$ $(\frac{W}{K})$	7.66	$\mathcal{U}(0, 20)$	3.56	0.26	4.80	13.5
26	$NORTH_{\mathcal{F}}$ (-)	1	$\mathcal{U}(-8, 8)$	-5.704	-7.77	-4.64	3.518
27	$EAST_{\mathcal{F}}$ (-)	1	$\mathcal{U}(-8, 8)$	-2.590	-7.301	-2.384	4.891
28	$SOUTH_{\mathcal{F}}$ (-)	1	$\mathcal{U}(-8, 8)$	-0.224	-6.499	-0.311	6.276
29	$WEST_{\mathcal{F}}$ (-)	1	$\mathcal{U}(-8, 8)$	-4.608	-7.688	-2.292	7.272
30	$crack1_{\boldsymbol{len}}$ (m)	2.4	$\mathcal{U}(0, 3)$	1.017	0.196	3.542	7.005
31	$crack2_{\boldsymbol{len}}$ (m)	7.6	$\mathcal{U}(0, 3)$	3.982	0.342	7.448	21.642
32	$crack3_{\boldsymbol{len}}$ (m)	2	$\mathcal{U}(0, 3)$	0.885	0.080	1.404	5.119
33	$crack4_{\boldsymbol{len}}$ (m)	2	$\mathcal{U}(0, 3)$	4.785	0.179	3.314	5.833
34	$crack5_{\boldsymbol{len}}$ (m)	2	$\mathcal{U}(0, 3)$	1.128	0.105	1.755	5.577
35	$crack6_{\boldsymbol{len}}$ (m)	2	$\mathcal{U}(0, 3)$	4.131	0.192	3.108	5.835
36	$crack7_{\boldsymbol{len}}$ (m)	2	$\mathcal{U}(0, 3)$	1.092	0.094	2.008	5.665
37	$crack8_{\boldsymbol{len}}$ (m)	1	$\mathcal{U}(0, 3)$	0.762	0.065	1.282	2.883
38	$crack9_{\boldsymbol{len}}$ (m)	1	$\mathcal{U}(0, 3)$	0.650	0.072	1.199	2.865
39	$crack10_{\boldsymbol{len}}$ (m)	1	$\mathcal{U}(0, 3)$	0.538	0.072	1.379	2.918
40	$crack11_{\boldsymbol{len}}$ (m)	1	$\mathcal{U}(0, 3)$	2.360	0.145	1.813	2.930
41	$crack12_{\boldsymbol{len}}$ (m)	1	$\mathcal{U}(0, 3)$	0.737	0.059	1.299	2.870
42	$living\ room_{Qvnt}$ $(\frac{m^3}{s})$	0.033	$\mathcal{N}(1, 0.1)$	0.032	0.025	0.032	0.038
43	$bathroom_{Qvnt}$ $(\frac{m^3}{s})$	0.0167	$\mathcal{N}(1, 0.1)$	0.017	0.014	0.017	0.020
44	$bedroom1_{Qvnt}$ $(\frac{m^3}{s})$	0.0167	$\mathcal{N}(1, 0.1)$	0.017	0.014	0.017	0.020
45	$kitchen_{\boldsymbol{C/R}}$ (-)	1	U(0,1)	0.728	0.037	0.531	0.977

Prior probability density distribution are scaled according the the initial values.

Thermal bridges stayed almost unchanged in the *kitchen*. *living* $room_{tb}$ and $bathroom_{tb}$ become about one third of their initial values, while $bedroom2_{tb}$ had a MAP value lower by the 54%. Increments were observed in the thermal bridges relative to *lobby* and *bedroom2*, which increased of 96% and 42% respectively. By considering the thermal bridges according to the defined macro-zones, it was

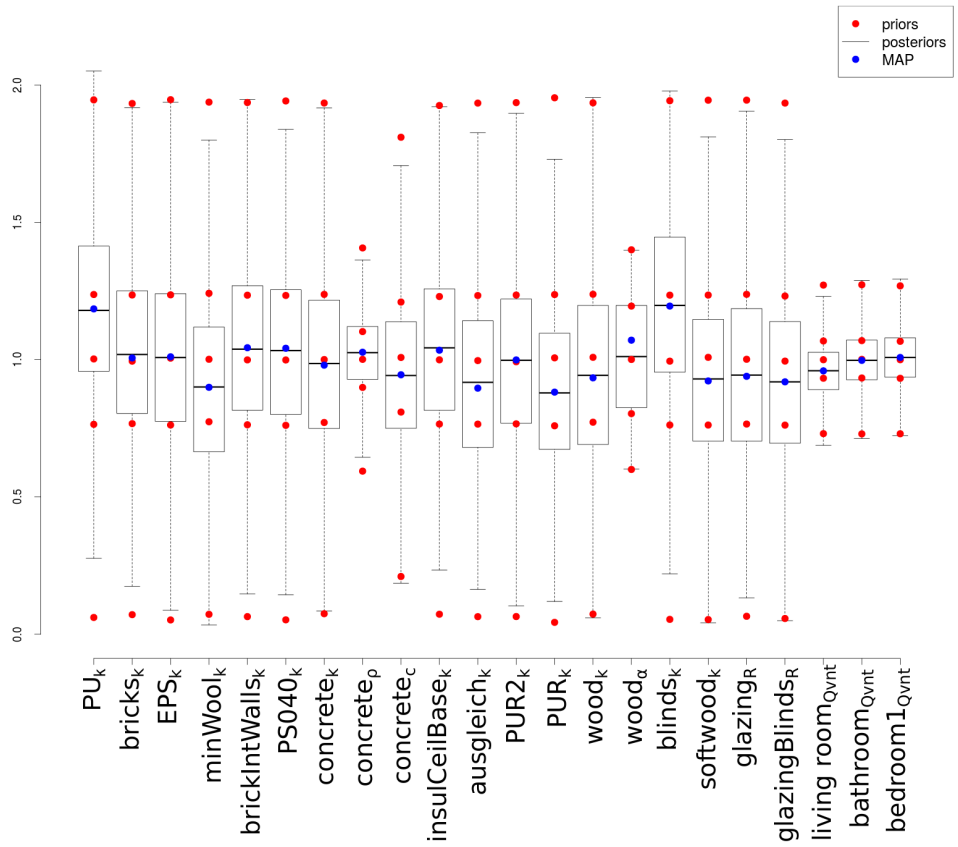


Figure 6.20: $\mathcal{M}_{2,N2}$–$CT1$: prior and posterior density distributions for model parameters with specified initial values. Samples have been scaled according to their mean to aid visualization (i.e. 1 = initial value).

possible to observe a reduction in the order of 10% in both $KLB2$ and $CBB1$.

6.6.2 ROLBS

The main results are summarised in Table 6.12 and Figures 6.21.

The outcomes of the analysis were in agreement with those returned from the calibration of $\mathcal{M}_{2,N2}$ against the $CT1$ data. All the parameters for which information was provided had posterior estimates in agreement with the given specifications. It is possible to notice only minor variations between the given values and the MAP estimates. In particular, as previously, the latter are always within one quartile from the given values.

$crack3$, $crack11$, and $crack12$ increased their lengths. In particular, they had estimates about 2.5 times their initial values, thus indicating relevant infiltration in the *living room* and ventilation between this zone, *basement* and *attic*. All other crack lengths decreased by amounts between the 60% and 80%.

Similar to the previous result, the MAP estimates inferred for $bathroom_{Qvnt}$

Table 6.12: $\mathcal{M}_{2,N2}$–ROLBS: MAP estimates and confidence intervals for calibration parameters

#	PARAMETERS	SPECIFIED VALUES	PRIOR DISTRIBUTION	MAP	$Q2.5\%$	$Q50\%$	$Q97.5\%$
1	$PU_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.035	$\mathcal{N}(1, 0.35)$	0.368	0.015	0.034	.058
2	$brick_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.22	$\mathcal{N}(1, 0.35)7)$	1094.18	221.652	971.206	1951.607
4	$brick_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	1000	$\mathcal{N}(1, 0.30)$	1023.303	507.970	952.971	1475.265
5	$EPS_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.022	$\mathcal{N}(1, 0.35)$	0.021	0.008	0.020	0.036
6	$minWool_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.036	$\mathcal{N}(1, 0.35)$	0.034	0.014	0.032	0.055
7	$brickIntWall_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.331	$\mathcal{N}(1, 0.35)$	0.319	0.121	0.293	0.513
8	$brickIntWall_{\boldsymbol{\rho}}$ $(\frac{kg}{m^3})$	1000	$\mathcal{N}(1, 0.7)$	1269.597	380.903	1135.279	2289.783
9	$brickIntWall_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	1000	$\mathcal{N}(1, 0.30)$	1022.420	494.857	942.536	1501.040
10	$estrichScreed_{\boldsymbol{\rho}}$ $(\frac{kg}{m^3})$	2000	$\mathcal{N}(1, 0.7)$	2132.435	473.469	1912.826	4050.687
11	$estrichScreed_{\boldsymbol{c}}$ $(\frac{J}{kgK})$	1000	$\mathcal{N}(1, 0.30)$	976.053	492.450	914.746	1500.233
12	$PS040_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.04	$\mathcal{N}(1, 0.35)$	0.040	0.015	0.037	0.065
13	$concrete_{\boldsymbol{k}}$ $(\frac{W}{mK})$	2.00	$\mathcal{N}(1, 0.35)$	2.127	0.885	1.974	3.342
14	$insulCeilBase_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.035	$\mathcal{N}(1, 0.35)$	0.040	0.020	0.037	0.059
15	$PUR2_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.025	$\mathcal{N}(1, 0.35)$	0.026	0.011	0.024	0.041
16	$PUR_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.023	$\mathcal{N}(1, 0.35)$	0.024	0.010	0.022	0.038
17	$wood_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.131	$\mathcal{N}(1, 0.35)$	0.128	0.050	0.116	0.209
18	$blinds_{\boldsymbol{k}}$ $(\frac{W}{mK})$	0.100	$\mathcal{N}(1, 0.35)$	0.098	0.040	0.091	0.156
19	$glazing_{\boldsymbol{R}}$ $(\frac{m^2K}{W})$	0.65	$\mathcal{N}(1, 0.33)$	0.621	0.24	0.572	0.997
20	$glazingBlinds_{\boldsymbol{R}}$ $(\frac{m^2K}{W})$	0.65	$\mathcal{N}(1, 0.33)$	0.725	0.322	0.672	1.082
21	$living\ room_{tb}$ $(\frac{W}{K})$	0.43	$\mathcal{U}(0, 2)$	0.060	0.006	0.126	0.479
22	$bedroom1_{tb}$ $(\frac{W}{K})$	0.668	$\mathcal{U}(0, 2)$	0.113	0.017	0.169	0.696
23	$kitchen_{tb}$ $(\frac{W}{K})$	0.763	$\mathcal{U}(0, 2)$	1.254	0.060	0.901	1.492
24	$lobby_{tb}$ $(\frac{W}{K})$	0.168	$\mathcal{U}(0, 2)$	0.031	0.04	0.058	0.297
25	$bathroom_{tb}$ $(\frac{W}{K})$	0.615	$\mathcal{U}(0, 2)$	1.054	0.044	0.726	1.210
26	$bedroom2_{tb}$ $(\frac{W}{K})$	0.766	$\mathcal{U}(0, 2)$	0.100	0.009	0.159	0.872
27	$NORTH_{\mathcal{F}}$ (-)	1	$\mathcal{U}(-9, 9)$	-8.701	-9.80	-7.792	3.455
28	$EAST_{\mathcal{F}}$ (-)	1	$\mathcal{U}(-9, 9)$	-7.701	-9.707	-6.913	1.319
29	$SOUTH_{\mathcal{F}}$ (-)	1	$\mathcal{U}(-9, 9)$	-7.993	-9.845	-6.443	5.473
30	$WEST_{\mathcal{F}}$ (-)	1	$\mathcal{U}(-9, 9)$	-9.100	-9.802	-6.653	8.332
31	$crack1_{\boldsymbol{len}}$ (m)	2.4	$\mathcal{U}(0, 3)$	0.503	0.049	0.822	4.834
32	$crack2_{\boldsymbol{len}}$ (m)	7.6	$\mathcal{U}(0, 3)$	2.698	0.368	9.970	22.115
33	$crack3_{\boldsymbol{len}}$ (m)	2	$\mathcal{U}(0, 3)$	5.063	0.179	3.684	5.901
34	$crack4_{\boldsymbol{len}}$ (m)	2	$\mathcal{U}(0, 3)$	0.706	0.065	1.221	5.387
35	$crack5_{\boldsymbol{len}}$ (m)	2	$\mathcal{U}(0, 3)$	0.774	0.098	1.676	5.726
36	$crack6_{\boldsymbol{len}}$ (m)	2	$\mathcal{U}(0, 3)$	0.608	0.080	1.015	5.224
37	$crack7_{\boldsymbol{len}}$ (m)	2	$\mathcal{U}(0, 3)$	0.679	0.084	1.525	5.693
38	$crack8_{\boldsymbol{len}}$ (m)	1	$\mathcal{U}(0, 3)$	0.334	0.049	0.845	2.892
39	$crack9_{\boldsymbol{len}}$ (m)	1	$\mathcal{U}(0, 3)$	0.178	0.041	0.542	2.763
40	$crack10_{\boldsymbol{len}}$ (m)	1	$\mathcal{U}(0, 3)$	0.327	0.037	0.992	2.904
41	$crack11_{\boldsymbol{len}}$ (m)	1	$\mathcal{U}(0, 3)$	2.667	0.124	2.301	2.964
42	$crack12_{\boldsymbol{len}}$ (m)	1	$\mathcal{U}(0, 3)$	2.537	0.080	1.572	2.935
43	$living\ room_{Qvnt}$ $(\frac{m^3}{s})$	0.033	$\mathcal{N}(1, 0.1)$	0.029	0.023	0.029	0.036
44	$bathroom_{Qvnt}$ $(\frac{m^3}{s})$	0.0167	$\mathcal{N}(1, 0.1)$	0.017	0.013	0.017	0.020
45	$bedroom1_{Qvnt}$ $(\frac{m^3}{s})$	0.0167	$\mathcal{N}(1, 0.1)$	0.0166	0.013	0.017	0.020
46	$ROLBS_{\boldsymbol{C/R}}$ (-)	0.7	$\mathcal{U}(0, 1)$	0.899	0.155	0.801	0.0993
47	$kitchen_{\boldsymbol{C/R}}$ (-)	1	$\mathcal{U}(0, 1)$	0.042	0.009	0.104	0.872

Prior probability density distribution are scaled according the the initial values.

and $bedroom1_{Qvnt}$ were substantially in agreement with the specified flow rates. Nonetheless $living\ room_{Qvnt}$ decreased by about 10%, indicating a small imbalance between mechanically supplied and mechanically extracted air in the model.

$\mathcal{F}$ factors all assumed similar large negative values, the highest being $NORTH_{\mathcal{F}}$

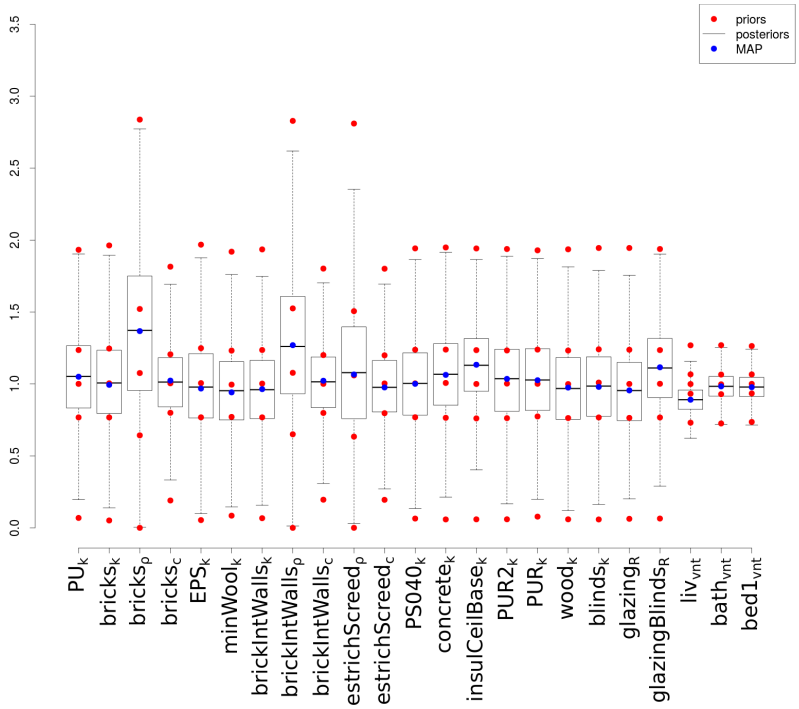


Figure 6.21: $\mathcal{M}_{2,N2}$–ROLBS: prior and posterior density distributions for envelope material parameters. Samples have been scaled according to their mean to aid visualization (i.e. 1 = initial value).

and $WEST_{\mathcal{F}}$, thus showing a similar trend with that resulting from the $CT1$ calibration.

Thermal bridges showed increments in the *kitchen* and in the *bathroom*, while they decreased in all the other zones. $living\ room_{tb}$, $bedroom1_{tb}$, $lobby_{tb}$ and $bedroom2_{tb}$ all assumed values between the 10% and 20% of their initial values. $kitchen_{tb}$ and $bathroom_{tb}$ increased of by 64% and 71% respectively. By considering the thermal bridges aggregated by macro-zones, it was possible to notice reductions of magnitudes comparable to those previously observed, that is by about 15%.

6.6.3 Discussion

$\mathcal{M}_{2,N2}$ returned estimates for material properties and mechanical ventilation flow rates in agreement with the given specifications. Only minor variations occurred between the inferred MAP values and the provided data, and they are deemed to be well within reasonable uncertainty ranges.

Therefore it seemed plausible to assume that most of the discrepancies between model predictions and observed data could be explained through the airflow network sub-model and thermal bridges.

The two undertaken calibrations returned consistent estimates with regard to crack lengths, pressure coefficient factors and thermal bridges, although these parameters are still very uncertain. Pressure coefficient factors showed, even though with different magnitudes, similar trends, assuming in both the calibrations large negative values. In particular, $NORTH_{\mathcal{F}}$ and $WEST_{\mathcal{F}}$ assumed the largest values in modulus in both the analyses. Minor differences were observed for the crack lengths, with the model presenting decreased values for most of them. However in the $CT1$ calibration, $crack4_{\boldsymbol{len}}$ and $crack6_{\boldsymbol{len}}$ increased sensibly, highlighting the possibility of significant infiltrations in the $bedroom1$ and $bedroom2$, while in the $ROLBS1$ calibration $crack3_{\boldsymbol{len}}$ estimated a larger value. The latter outcome, together with the results relative to the $\mathcal{F}$ factors, could mean a relevant infiltration flow entering from the west side of the *living room*, probably providing for the small imbalance between mechanically supplied and mechanically extracted air in the model. Particularly interesting were the congruent estimates of $crack11_{\boldsymbol{len}}$, which increased sensibly. Thus, significant heat exchanges, through ventilation, seemed to occur between the *living room* and the basement, agreeing with the outcomes of the *Difference Analysis* for $\mathcal{M}_1$, which showed correlation between the basement air temperature and the *Difference Vectors*. Also thermal bridges presented similar trends showing comparable reductions in the considered macro-zones.

6.7 Discussion of the detailed experiments

In the various steps of the analysis three different model were considered:

- $\mathcal{M}_0$: the base model having its input parameters fixed to the specified values.
- $\mathcal{M}_1$: the base model having its inputs fixed to the calibrated values.
- $\mathcal{M}_2$: the base model upgraded with an airflow network model and subsequently calibrated.

These models were compared graphically (Figures 6.22 and 6.23), and by calculating the adopted goodness of fit criteria (Tables 6.13 and 6.14).

$\mathcal{M}_1$ and $\mathcal{M}_2$ provided a significantly better match with the measurements than $\mathcal{M}_0$. Both of them had very similar predictive capabilities. The former is better in predicting the sensible heat loads for *living room* and $KLB2$, while the latter had better performances in matching the measured heat load for $CBB1$

Table 6.13: $CT1$: goodness of fit criteria

ZONE	CRITERION (%)	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_{2,N2}$
living room[1]	NMBE	-7.14	-2.37	3.63
	CVRMSE	17.67	14.09	15.65
	GOF	13.48	10.10	11.34
$KLB2$[2]	NMBE	29.71	4.40	-1.82
	CVRMSE	30.93	8.35	11.12
	GOF	30.33	6.67	7.97
$CBB1$[3]	NMBE	43.07	7.51	-3.31
	CVRMSE	43.94	10.80	12.11
	GOF	43.51	9.30	8.88
ALL	$log_{10}(p(\boldsymbol{Y}^*\|\mathcal{M}_k))$	-	37.26	41.08

(1) $\bar{\boldsymbol{y}}^* = 0.861\ kW$; (2) $\bar{\boldsymbol{y}}^* = 0.184\ kW$; (3) $\bar{\boldsymbol{y}}^* = 0.189\ kW$

Table 6.14: $ROLBS1$: goodness of fit criteria

ZONE	CRITERION (%)	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_{2,N2}$
living room[1]	NMBE	1.98	-0.93	0.59
	CVRMSE	3.74	3.03	2.44
	GOF	2.99	2.25	1.78
$KLB2$[2]	NMBE	3.28	-1.12	-0.68
	CVRMSE	4.30	2.83	2.17
	GOF	3.82	2.15	1.61
$CBB1$[3]	NMBE	1.62	0.49	0.34
	CVRMSE	2.66	2.23	1.78
	GOF	2.20	1.62	1.28
ALL	$log_{10}(p(\boldsymbol{Y}^*\|\mathcal{M}_k))$	-	40.39	41.4

(1) $\bar{\boldsymbol{y}}^* = 23.347\ ^oC$; (2) $\bar{\boldsymbol{y}}^* = 22.603\ ^oC$; (3) $\bar{\boldsymbol{y}}^* = 23.446\ ^oC$

and during the whole $ROLBS1$ phase. In particular, significant improvements in the predictive capabilities of the two models were observable relative to $KLB2$ and $CBB1$, while the agreements with the measured sensible heat load and internal temperature of *living room* were only relatively better than $\mathcal{M}_0$. Indeed the base model was already able to provide reasonably accurate predictions for *living room*. The impression is that the modelling of this thermal zone was fairly accurate since the beginning, and that the proposed upgrades had an impact especially on the capabilities of the model in describing the behaviours of $KLB2$ and $CBB1$. Nonetheless, the implementation of the airflow network allowed $\mathcal{M}_{2,N2}$ to provide a better match of the observations for inputs closer to the specified values with respect to $\mathcal{M}_1$, also for *living room*.

The differences between the calculated values of NMBE, CVRMSE and GOF of $\mathcal{M}_1$ and $\mathcal{M}_2$ were small, in the order of 2%. Bayes Factors gave more informa-

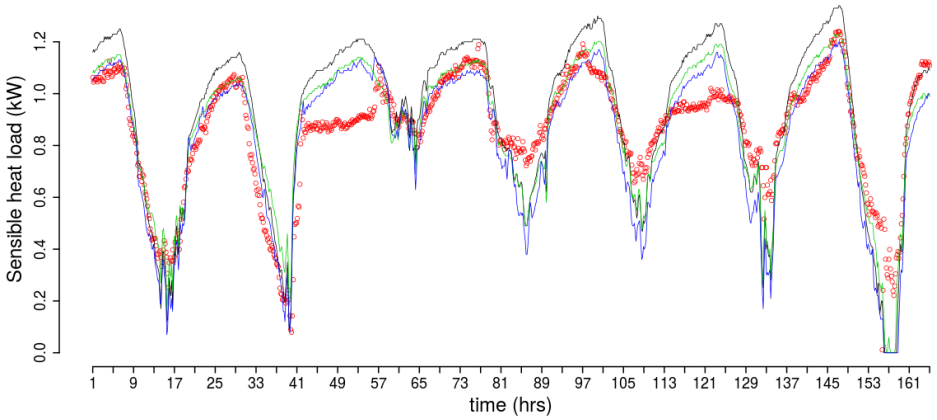
tive results about model selection. $\mathcal{M}_2$ had higher values of marginal likelihoods and $\mathcal{B}_{2,1}$ assumed the values of 3.82 and 1.01 for the $CT1$ and $ROLBS1$ calibration respectively, indicating that $\mathcal{M}_{2,N2}$ is a better representation of the real experiments than $\mathcal{M}_1$. By considering these outcomes, and the more reasonable model parameters estimates provided by $\mathcal{M}_{2,N2}$, it was decided to assume this model as the best one, despite it provided more noisy predictions, especially for $CT1$ phase. This last aspect might well be due to the particular method used to solve the airflow network model and the variability of the wind. In particular by reducing the simulation time step it should be possible to provide smother model outputs. $\mathcal{M}_{2,N2}$ was considered accurate enough for the current stage of the research.

Nonetheless, $\mathcal{M}_{2,N2}$ showed discrepancies between its predictions and the measured data, which underlined the possibility of further improvements. In particular, the *Difference Analysis* showed possible links between the observed differences, mechanical ventilation air temperature, wind speed and wind directions. The latter two boundary conditions would suggest a more detailed calibration of the airflow network model, especially through a more adequate consideration of the pressure coefficients. Such more suitable consideration of pressure coefficients in the calibration process would require to take into account the correlations between them, by setting up suitable multidimensional prior probability density distributions. A possible approach is discussed in Chapter 7.

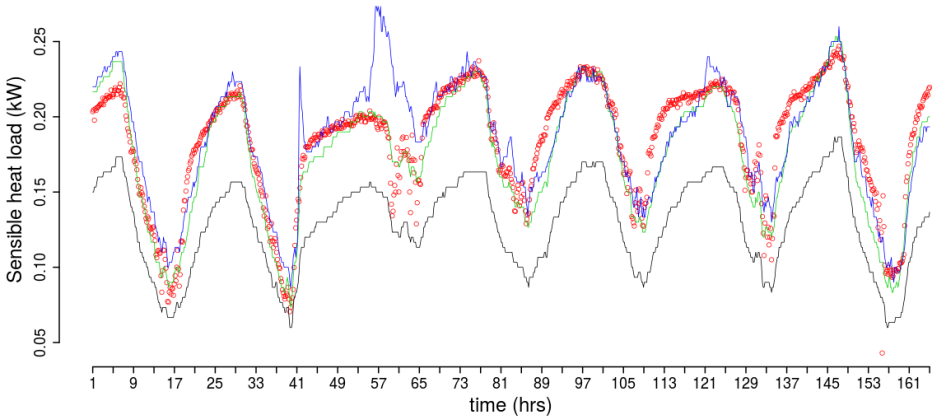
The effect of the temperature of the supplied air on the match with the field observations was investigated by performing additional calibration trials, relaxing the prior probability density distributions for $living\ room_{Qvnt}$, $bathroom_{Qvnt}$ and $bedroom1_{Qvnt}$. These attempts returned better matches between model outputs and target variables, especially for the $CT1$ experimental phase. However the resulting models were showing a large imbalance between mechanically supplied and mechanically extracted air which was judged not plausible, given the great care taken by the experimental team in assuring equality between inflow and outflow rates of the mechanical ventilation system.

These latter outcomes seemed to support the conclusions drawn from the Test Box example. Depending on data characteristics and model deficiencies, sensibly different than specified parameters may perform better in matching the given observations. In particular, this highlighted the importance, especially for complicated models, of having high quality prior information allowing the definition of adequate prior probability density distributions driving the investigations towards reasonable answers.

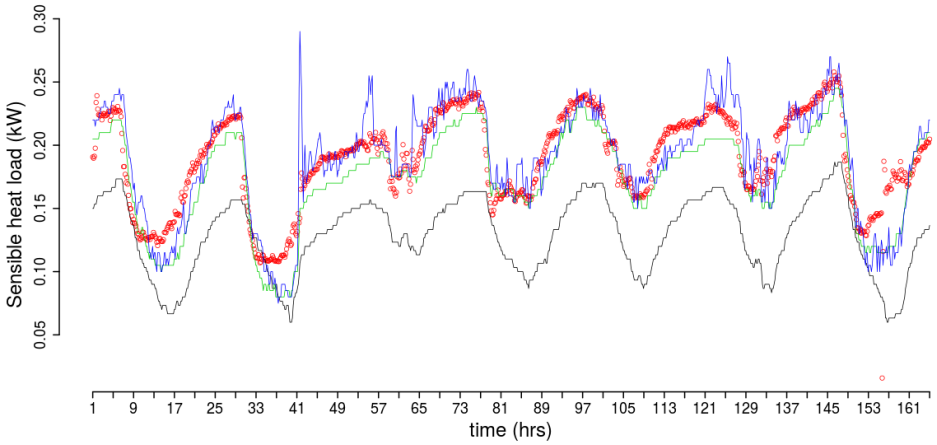
Completing the depicted analyses was the hardest challenge, because of the detail of the analysed models and,the amount of information that was necessary to process. Nonetheless the presented analysis framework for BES was successful in characterising the sensitivity of the models to the input parameters, calibrating the models and suggesting improvements for the different models.



(a) *living room*.

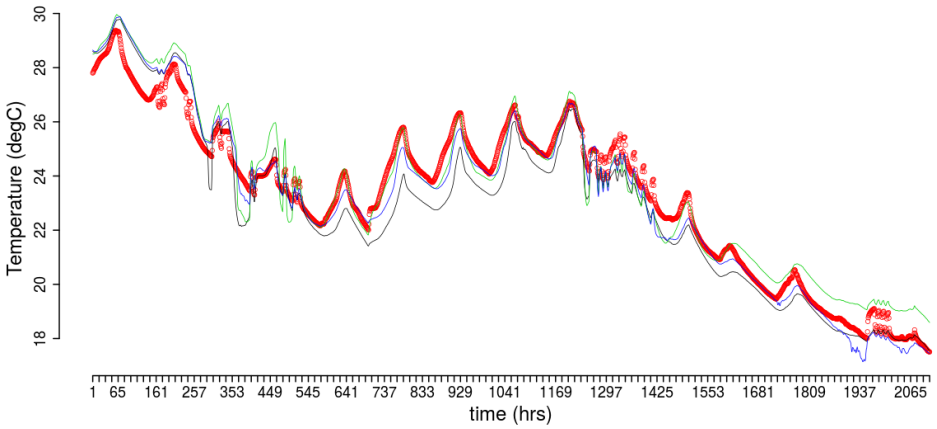


(b) *KLB2*.

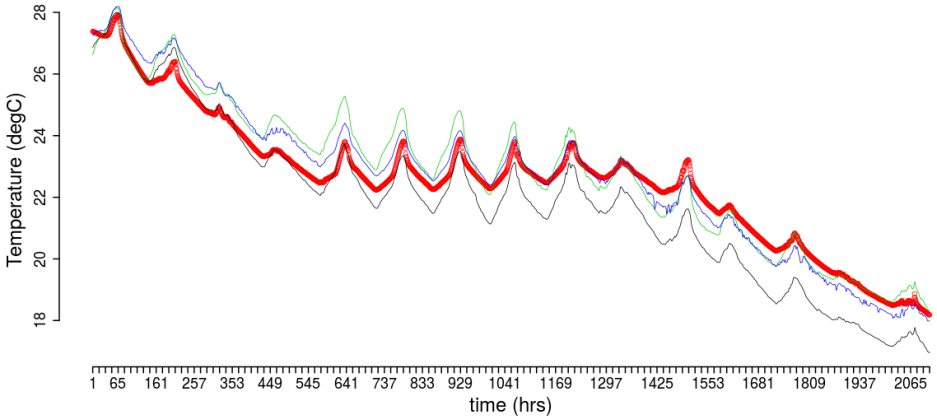


(c) *CBB1*.

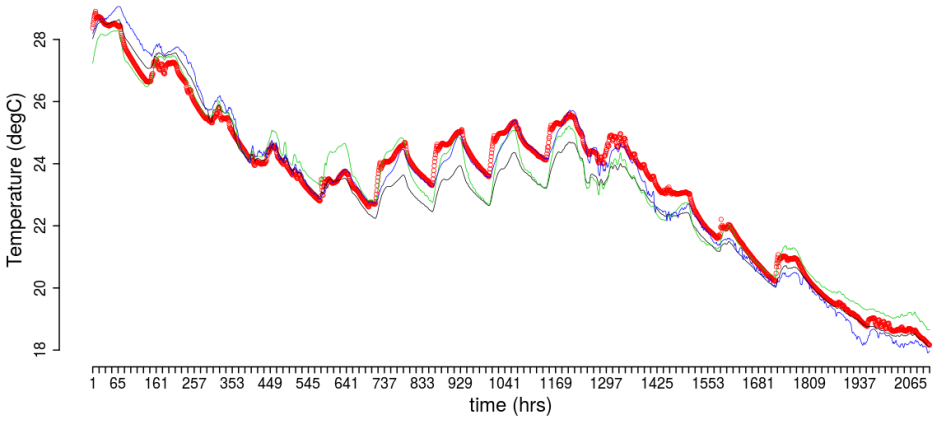
Figure 6.22: *CT1*: comparison between observation (red), $\mathcal{M}_0$ (black), $\mathcal{M}_1$ (green) and $\mathcal{M}_2$ (blue).



(a) *living room*.



(b) *KLB2*.



(c) *CBB1*.

Figure 6.23: *ROLBS1*: comparison between observation (red), $\mathcal{M}_0$ (black), $\mathcal{M}_1$ (green) and $\mathcal{M}_2$ (blue).

Chapter 7

Conclusions

In the previous chapters a framework for the analysis and diagnosis of building energy models has been described and tested. The outlined framework is composed of uncertainty analysis of the model parameters, probabilistic sensitivity analysis, Bayesian Calibration and model selection. In the *Uncertainty Analysis* phase the prior uncertainties related to model inputs and measurements are assessed and represented though adequate prior probability density distributions, so as to represent all the model parameters as random variables. The *Sensitivity Analysis* phase, consisting of Factor Screening (FS), Factor Prioritising (FP) and Factor Fixing (FF), and Factor Mapping (FM), has the objectives to reduce the dimensionality, to refine the defined prior uncertainties and gain useful information to set up the calibration of the BES model. The Morris Method was suggested for performing a qualitative preliminary screen of the model parameters, the Sobol Method was adopted as the reference technique to undertake FP and FF through quantitative variance base sensitivity measures, and GLUE was employed to perform FM. The *Calibration*, which is the core part of the framework, is divided into in *Training* and *Identification*. It employs GPR in a quasi-Bayesian framework to create emulators of the BES model allowing to effectively include uncertainties in the calculations, to infer unknown model parameters and their uncertainties, conditional on the field data. Finally, Bayes Factors is used as the main criterion for model selection in the *Model Selection* step.

However the outlined methodology is not a silver bullet but rather a sophisticated tool, which requires, by the analyst, building physics expertise and knowledge of the statistical techniques involved. This last chapter will discuss the current stage of the research. Particular focus will be given to its limitations and strengths. Finally future expansions and possible applications will be considered.

7.1 Findings

During the entire arc of the research, case studies of increasing difficulty were analysed, allowing the exploration of strengths as well as the weaknesses and limitations of the outlined analysis framework for building energy models. The main strength of the framework is the rigorous mathematical treatment of the different types of uncertainties involved in modelling and calibration problems. These are considered during the entire analysis process, starting with the *Uncertainty Analysis* phase and ending with the *Calibration* phase, and with the information coming from the field data they contribute in determining the final solution of the problem being addressed.

Novelties were introduced for investigating the uncertainties of multi-dimensional inputs, and for adequately considering multi-dimensional model outputs in the calculation of sensitivity indexes. Vectorial model inputs are rarely considered in sensitivity studies involving BES models. Usually, their uncertainties are simply represented with a constant offset from a reference measured vector. This is unrealistic, since changing monitoring conditions are likely to produce *random errors* of different magnitudes. In this work Bootstrap and Smoothing with Roughness Penalty techniques were employed to infer from the large amount of available data, plausible multi-dimensional distributions representing such uncertainties. In this way, it was possible to investigate how the random variabilities of the observed processes changed over time according to the varying monitoring conditions. It is hard to assess the correctness of uncertainty quantification; however, it is believed that the proposed approach is more sensible and rigorous than the common practice, leading to more realistic estimates of the uncertainties for multi-dimensional model inputs. Thus, when a substantial amount of information is available, it is advisable to employ advanced statistical techniques, like those used in order to perform *Uncertainty Analysis*. The effect of adopting this more rigorous approach was demonstrated on an airflow network model in Chapter 6. In this case study it was shown that the wind velocity had relatively low importance by considering its uncertainties calculated with the proposed method, while it resulted the most influential variable by defining its uncertainties according to a simplified approach.

As the literature review chapter showed, in the large majority of the cases, sensitivity analysis applied to BES consist of performing a qualitative screening of the model parameters. Qualitative screening is usually adopted because of its relatively low computational burden. Such an approach was deemed to be insufficient because it is easily subject to oversimplifications. In particular, it does

not quantify the variance of the model outputs considered by working only with the retained model inputs, which are heuristically selected. Therefore, a quantitative technique, the Sobol Method, was employed to integrate and support the qualitative model parameter screening undertaken through the Morris Method. In this way it was possible to measure the amounts of model output variance considered and neglected by working only with the subset of retained parameters, and to judge the adequacy of the qualitative screening. In particular, when the additional computational load can be sustained, measuring the amount of model output variance considered by working only with certain retained parameters, can be decisive in adequately reducing the dimensionality of the model and avoiding oversimplifications. This is particularly important if sensitivity analysis is preparatory to calibration or validation. In Chapter 6 it was shown that while the Most Important Factors certainly have the largest first order effects, the portion of model output variance attributable to them may still be only a relatively small fraction of the total, and the Least Important Factors, although having negligible first order effects, may have a combined influence decisive in defining model behaviour. The increased calculation load can be compensated by extending the analysis with Factor Mapping, which, considering the target measurements, provides complementary information with respect to the other phases, and can identify further important variables that have significant importance in improving similarity with the field observations, but which were not highlighted by Factor Screening, Factor Prioritising and Factor Fixing.

The information about the sensitivity of a certain model output to the inputs should be concise, and at the same time should be comprehensive. When the output in question is vectorial, the achievement of these two objectives is not easy. Common practice involves calculating sensitivity indexes for each time step of the simulation or the utilisation of functional-like integrals or distance measures from reference values in order to reduce the simulation vectors to scalars. While the former approach produces redundant and difficult to summarize information, the latter neglects the dynamic trends of the vectorial outputs. Here, Principal Component Analysis (PCA) has been used to complement the Morris and Sobol methods in effectively dealing with vectorial outputs and to return concise and easy to interpret information, in the form of a few significant sensitivity indexes. An expansion of the Morris Method was proposed in order to treat multidimensional model outputs. PCA was used to project the model outputs in a convenient reference space and a new sensitivity index, $M^\star$, was defined. Results carried out with this new approach were substantially in agreement with those from the Sobol

Method, and effective in screening model parameters.

It is believed that the methodology employed in this study forms a rigorous basis for undertaking sensitivity analyses, and in reducing the degrees of freedom of BES models in a more informative way, which is superior to the common practice based only on qualitative screening, especially if employed as a preparatory stage to calibration or validation. In particular it should become good practice to quantify the amount of variance attributable to the retained inputs in order to adequately justify model simplifications. FM is particular effective in complementing the result from Global Sensitivity Analysis and justifies the additional computational load required in quantifying parameter effects. The effectiveness of the presented sensitivity analysis method was demonstrated on a particular sub-model but a comprehensive analysis should consider the overall BES model.

The explained quasi-Bayesian calibration methodology was proven to be effective in dealing with models of different degrees of complexity. In particular GPR allows the building of flexible probabilistic models which can emulate diverse computer models, without changing their structures, and adequately considering model and data uncertainties. The experiments described in Chapter 5 demonstrated the ability of the method to identify model parameters. In particular, the method's capabilities of dealing with noise, model inadequacies and identifiability problems were tested. The outcomes showed that, despite the reasonable results and estimates inferred, these aspects influence the estimation of model parameters. The wall example showed that correlated noise and model deficiencies can alter input estimates, especially those determining the dynamics of the model and those having weak effects on the model outputs. The test box example demonstrated how to solve identifiability problems by considering information from different independent datasets, and showed that calibration parameters different from the specified ones may perform better in fitting the measured data, but may have mediocre performances in predicting target variables observed in diverse conditions from those used in training the model. The main reasons for this are errors in the measurements and in the modelling, or simply different conditions in which the phenomenon was observed. Furthermore, the same example underlined how the solicitations imposed during the experiments can allow or impede the estimation of model parameters. For example, the ROLBS sequence used in the LECE experiment permitted to clearly determine the properties of materials comprising the two outermost layers of the test box, but not those of the two innermost ones. Therefore, it is possible to distinguish two type of parameter identifiability: *structural identifiability* and *experimental identifiability*.

Structural identifiability is determined by the model itself, its sensitivity to the various parameters and the relationship between them. For example, deficiency in the model structure can compromise the identifiability of model parameters or lead to wrong estimates. Similarly, inputs to which the model is little or not sensitive, or which are highly correlated, are very difficult or impossible to identify. *Experimental identifiability* is dependent on the characteristic of the experimental conditions and the quality of the measured data. For example, the applied solicitations or signals can highlight only particular features of the object of the experiment and thus, allow only the estimation of the correlated parameters. Even more, the provided data can be instant values, or post-processed values (for example averages over certain time intervals), according to this aspect their correlation and their uncertainties may vary, in turn affecting parameter estimation. For these reasons, it is considered to be wrong to talk about *true value*, especially in referring to model parameter estimates, except in the case of virtual experiments wherein everything is known. Model input estimates should be interpreted as the values maximising the probability of the model of producing the observed data, and it should not be surprising that such estimates differ from the given values. All data, even measurements and given specifications, are indeed estimates, and as such they should be always associated with uncertainties and referred to the experimental conditions, influencing their monitoring. Only in this way it is possible to undertake useful comparison in order to judge analysis results. The general advice is to train the model in conditions similar to those that will affect the predictions, and, when possible, to include the largest possible variety of boundary conditions. As shown in this case study, the presented calibration method is suitable for considering multiple datasets, thus it is able to allocate this last issue well.

The most demanding tests for the presented calibration method were the series of investigations performed on models depicting the twin houses (Chapter 6), and, nonetheless, it provided satisfying performance. The calibration methodology was able to treat a large number of calibration parameters and models involving highly non-linear relationships between them. One of the main novelties proposed in these case studies was the possibility of considering multiple target variables, and benefiting from the correlation and dependency relationships between them in identifying unknown model inputs. The investigations performed in Chapter 6 also demonstrated the importance to have detailed and accurate information in analysing complicated building energy models, especially if these involve a large number of parameters. The determination of the initial uncertainties and the

definition of the search space, was facilitated by the information used in setting up the adopted prior probability density distributions, and the calculations were simplified, being able to identify a sensible solution more easily. The relaxation of the prior probability density distributions relative to mechanical ventilation flow rates produced a better fit of the field observations, but the inferred posterior distributions for the calibration parameters showed a large imbalance between mechanically supplied and mechanically extracted air which was not compatible with the experimental specifications. This also leads back to the previous considerations about differences between specified and estimated values. Therefore, the specification of appropriate probability density distributions is important, and it can be used to find a solution which is the best compromise between the best fit of the measured data and specified values. In particular, it was possible to improve significantly the match between target data and model predictions for $KLB2$ and $CBB1$, while only relatively for *living room*. Indeed, the undertaking of the calibration analysis could have been avoided, if the aim of the model was to predict only the sensible heat load injected in, and the internal temperature of this latter thermal zone. However, it is difficult to say a priori when it is necessary to proceed to model calibration, especially if it is not possible to benefit from the availability of precise specifications like in the experiments described in Chapter 6. Those were primarily validation exercises, and in ordinary calibration studies the information about the object of the modelling is limited, thus calling for the need to investigate the goodness of the model through calibration. Nonetheless, the proposed analysis framework was able to produce a better model with respect to the initial one, and able to provide more accurate predictions with values for its inputs closer to the specified ones.

All the performed experiments raised questions about the adequate degree of model detail and complexity to adopt. Since it is not possible to work considering all the degrees of freedom involved, it is believed that high model detail should be supported by high quality information, thus reducing at the outset a great part of parameter uncertainties, and to focus on as few variables as possible. When such high quality information is not available it is convenient to reduce the degrees of freedom of the model from the beginning, by assuming appropriate simplifications during the modelling phase, and to proceed by gradually augmenting the complexity of the model assessing, through adequate criteria, if the implemented model upgrades are effective in better representing the field observations. This principle was adopted in investigating all the described case studies. In particular, the *Difference Analysis*, despite the fact that correlation does not imply

dependency between two variables, was always effective in identifying suitable improvements to the analysed models. More tests have to be undertaken in order to certainly assure its capabilities, but it is believed that the practice referred to as *Difference Analysis* is a powerful and useful technique to spot model deficiencies. The different analysed models were then compared through the adopted goodness of fit criteria. The commonly used measures (Normalised Mean Bias Error (NMBE), Coefficient of Variation Root Mean Squared Errors (CVRMSE) and Goodness Of Fit (GOF)) are simple and easy to calculate, but subject to model *equifinality*. Even more they are poorly suited to consider multiple target variables and datasets, and it is not clear how they should be aggregated in these cases. Bayes Factors are better for addressing this issues. This criterion is less subject to model *equifinality*, and the probabilistic models employed for calibration naturally allow their calculation, also taking into account different target variables and datasets. In particular they were more informative than NMBE, CVRMSE and GOF in performing model selection, and are suggested as the model selection criterion of reference for the developed BES analysis framework.

It is believed that the proposed analysis framework for BES models is a substantial improvement with respect to the current practice. Especially the outlined calibration methodology is considered a great advancement respect to previous approaches, since it allows to probabilistically treat calibration and validation problems, while benefiting from the robustness and reliability of state of the art dynamic simulation programs. In order to quantify such improvement it is useful to make a comparison with the levels of agreement with measured data obtained in previous calibration studies. In most of these investigations, consisting of manual iterative and mathematical analytical calibrations, BES models were considered calibrated against hourly data if they had a NMBE and a CVRMSE respectively lower than 10% and 30%, as indicated in [1]. No information is given about possible threshold values for these two GOF criteria when measurements sampled at higher frequencies are employed, but it seems reasonable to assume increased values for both of them. In the carried out calibration experiments data were used with time steps ranging from 0.5 hour to 5 minutes and the calculated NMBE and CVRMSE were always significantly lower than the above mentioned values. The former had values between 0.34% and 3.48%, while the latter were between 1.78% and 15.4%. The author is aware of the substantial differences between the character of many studies cited in the literature review and that of the described examples, that could lessen the value of this comparison. In particular the former treated whole occupied buildings and did not have the high

monitoring standards adopted in the latter, which involved unoccupied research experimental facilities. Nonetheless this is an excellent premise that supports the interest of testing the proposed framework in more common problems, and the confidence that it will keep providing high quality calibrated models.

7.2 Future research

Despite the degree of reliability reached by the proposed analysis framework, it still presents limitations requiring future research and improvements. Especially sensitivity analysis methodology and calibration method presented are suitable for improvements and expansions. The priority of future researches will be to substantially reduce the computational time required that, for the more complex investigations and without the employment of a high performance computing facility, may reach duration of 3-4 days.

The proposed sensitivity analysis method is quite robust, and the coupling of GSA and GLUE is particularly apt in adequately analysing complicated BES models. However sensitivity analysis techniques, different from the Morris Method and Sobol Method, should be investigated in order to compare relative performances and calculation times. Interesting approaches to the fast determination of the sensitivity indexes S_i and ST_i can be found in [95] and in [69]. The application of these methodologies will avoid the need to apply the Morris Method in the initial Factor Screening, and they will allow to perform the entire analysis with the same Monte Carlo sample of model simulations, thus significantly reducing the number of model runs required. Additionally, the consideration of multiple model outputs must be improved. In more practical case studies it is necessary to consider multiple performance indicators since design problems are often multi-objective. This issues was already partially addressed in Chapter 6, by considering the concatenation of the vectors representing the zones air temperatures as model output. However, especially when the considered model outputs represent different performance indicators (e.g. heat loads and temperatures), it may be convenient to define a hierarchy among them. For example, a weighting function of the sensitivity indexes calculated for the different performance metrics can serve this purpose. This should be explored in future work, and comparison between different approaches made.

The described *Calibration* framework involves Bayesian techniques and Markov Chain Monte Carlo (MCMC) methods, which have a steep learning curve and are computationally demanding. In particular, the latter aspect required the adop-

tion of meta-model based upon Gaussian Process Regression in order to accelerate the calculations. This introduced further uncertainties, due to the inevitable error of the meta-model in approximating the BES model. This could be avoided by implementing the explained probabilistic and stochastic techniques in dynamic simulation programs. Because of the non-negligible running time of generally employed BES models, at the moment this seems not convenient. However, as computational power becomes more available it will be necessary to think of ways to include the described methods in the main computer codes.

Different optimization and MCMC algorithms should be tested, in order to build a library of methods and guidelines advising on the best techniques facilitating and accelerating the calculations calibrating a particular model. It is strongly believed that the applied algorithms based on Markov Chain Monte Carlo methods perform better in exploring high dimensional parameter spaces, and provide more information respect to the more commonly employed optimisation routines like genetic and particle swarm algorithms. In particular, MCMC methods allow naturally the estimation of parameter uncertainties, which instead appears more difficult to perform with the latter king of optimisation technique. For example, in the case of a non identifiable parameter the applied algorithms will return a posterior probability density distribution practically equal to the prior one, thus highlighting that it is not possible to improve its estimate. By applying genetic, particle swarm or alike algorithms, it would appear that the parameter has been perfectly estimated, since they return point solutions. This issue could be somewhat circumvent by restarting the optimisation multiple times, but then this class of methods would loose the advantage, in terms of less computational time, which is the only reason that would make them more appealing that MCMC based algorithms. That being said, to undertake a comparison study would be interesting and would allow to draw more robust conclusions.

As a general consideration, it is useful to say that the computational time required is highly dependent on the number of calibration parameters. Therefore the complexity of the model should be reduced at the beginning, by including in it only the aspects that according to the modeller knowledge are important in representing the observed phenomena, and avoiding the introduction of unnecessary uncertainties. Then, the depicted analysis framework allows to test the initial set of assumptions and hypothesis against the field data, and eventually to refine accordingly the detail of the model.

As mentioned in Section 4.2.2 the proposed calibration framework is effective in considering multiple target variables when the vectors representing the

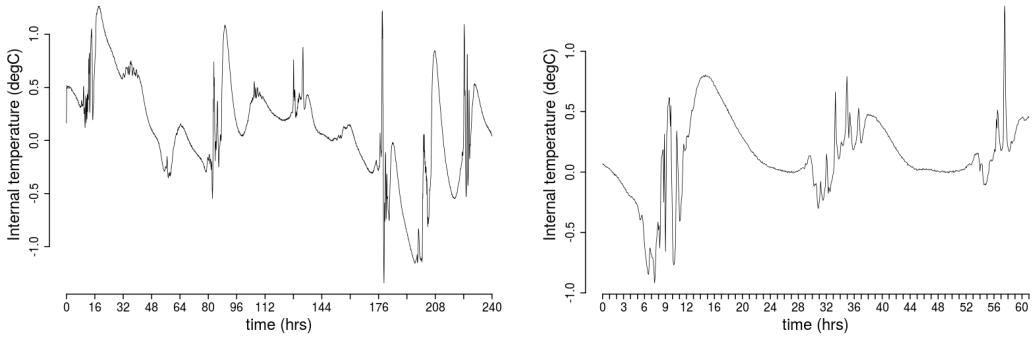
measurements of such variables show similar shapes and trends. In this way, it is possible to approximate with enough accuracy the different target variables employing the same few basis vectors derived through PCA. Otherwise, despite the fact that PCA can be used to derive a linear combination approximating the measurements, it may be necessary to consider several basis vectors, thus lessening the effectiveness of the dimensionality reduction of the datasets. Even more, a linear combinations based upon the bases defined according to the given observed variables, will have good performance in approximating the training dataset, but may have poor performance in representing other datasets relative to different boundary conditions, thus reducing the predictive power of the GPR model. Thus effort should be put into developing methods, which have the benefit of representing the data through basis expansions, while avoiding as much as possible the drawbacks. Useful hints can be found in recent contributions investigating Sparse Representation and Reduced Rank Approximation of GPR models. In the future, efforts will be put into expanding the developed methodology with these concepts. Furthermore, the GPR models adopted, are very good interpolators, but may have bad performances in extrapolating predictions for conditions and input very different from those included in the training set. In these cases, predictions will have mean close to zero and large variances. At the moment this is considered to be a secondary problem because the BES model can always be used to make predictions. It is reasonable that the uncertainties will be relatively high for boundary conditions not previously observed. Nonetheless, the sensibleness of the predictive uncertainties must be assessed in future research, through the systematic comparison of model predictions and measured data, and more complex models for the measurement errors will be tested if necessary. For what concern the inferred uncertainties for the calibration parameters, the employment of statistical tests for comparing the results from multiple calibrations against the same experimental dataset, can provide useful insight about their sensibleness. The application of such tools will be investigated in future studies.

Through the presented calibration framework it was possible to treat and calibrate several kinds of BES model inputs. In particular, the depicted experiments considered material properties, constant infiltration flow rates, airflow network model parameters (like crack dimensions and pressure coefficients) and split coefficients between convective and radiative heat gains. The method is flexible and able to calibrate every aspect of a computer model, even specific components like sub-models. Indeed it could be use in future investigation to infer the parameters of empirical models, like convection heat transfer models, serving the

main BES model. However, at the current stage of the research, the consideration of vectorial calibration parameters like occupancy patterns, or correlated calibration parameters like pressure coefficients is difficult, and the method must be improved in order to provide for this. This is necessary in order to be able to calibrate all the variety of parameters involved in BES models. Future research will investigate the creation of hierarchical GPR models, wherein GPR sub-models depicting these kind of calibration parameters will be connected to the main GPR model. Elliptical Slice Sampling ([77]) could, then, be used to draw correlated samples from these sub-models, and pass them to the main GPR model. Similar considerations can be applied to boundary conditions. None of the treated examples required the consideration of the measurement uncertainties of the boundary conditions. However, as studies with less detailed measurements are undertaken, it will be necessary to account for these aspects, which may have influences on the estimation and posterior uncertainties of the calibration parameters. The calibration methodology is already able to consider uncertainty for unidimensional boundary conditions, and can be easily extended, by implementing the means to use multidimensional prior probability distributions, in order to be able to include in the analysis the uncertainties of multidimensional boundary conditions as well.

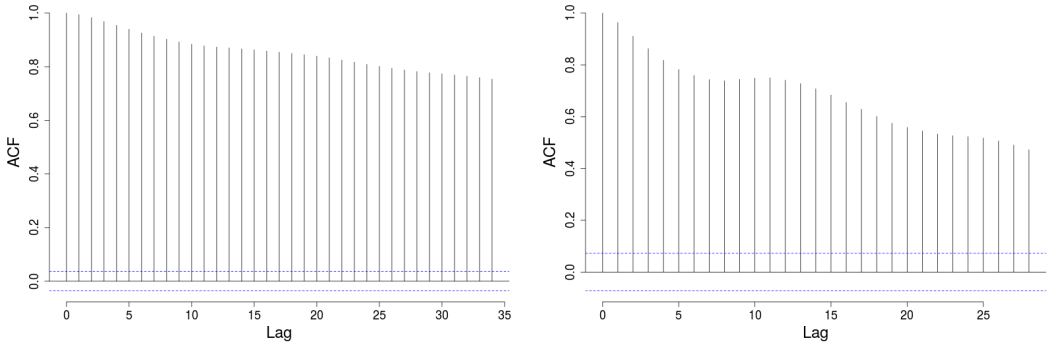
Another improvement to the methodology that would be of great benefit for the analysis of BES model in calibration and validation exercises, would be the possibility to use means for model validation similar to those used for grey-box models. In particular, as discussed in Chapter 2, the employment of this models and of the Kalman Filter for the calculation of the likelihood of the innovations allows to validate the model by analysing the autocorrelation of the residuals between predictions and measurements and testing their similarity with white noise. In the current framework, since the calculation of the likelihood is performed over the entire time series at once, it is not possible use these tools because the residuals will result highly correlated. For example, Figure 7.1 shows the residuals between the model $\mathcal{M}_2$ predictions and the measurements during the training and test periods of the experiment on the test box performed by the Belgian Building Research Institute (BBRI) (Section 5.2.5). It is evident that they are far different from white noise and non-stationary, as also highlighted by their autocorrelation functions, depicted in Figure 7.2. Therefore, if on one hand the developed methodology allows the analysis of complex BES models and to benefit from reliable robust state of the art dynamic simulation programs, on the other hand it lacks the means to statistically validate these models. To provide

model validation methods similar to those adopted for grey-box models, within the presented framework, will be a priority of future research efforts.



(a) training period. (b) test period.

Figure 7.1: Test box–*Real* BBRI:residual between $\mathcal{M}_2$ predictions and measured data during the training and test periods.



(a) training period. (b) test period.

Figure 7.2: Test box–*Real* BBRI:autocorrelation of the residual between $\mathcal{M}_2$ predictions and measured data during the training and test periods.

The previously identified improvements to the current research are deemed to be the most important in order to complete the explained analysis framework for detailed dynamic building energy models. However, as building simulation methods and programs will evolve, it will be necessary to adapt the explained methodologies to the new needs. It is believed that the proposed analysis framework has the modularity and flexibility necessary to adequately answer most of the problems that the calibration and validation of building energy model will present.

7.3 Future applications

Building emission reduction is a central topic in the politics of the government of most developed countries, which is being achieved mainly through better design of buildings in order to improve energy performances and by increasing the usage of renewables. Thus the mismatch between measured and predicted building performance and the high variability in the energy supply offered by renewable energy sources are major concerns in the drive to cut such emissions. Solutions to these problems have to be sought at both single building and community levels. On one hand, poor predictions of building energy models are mainly due to wrong initial knowledge upon which models are built, erroneous operation in the post-occupancy phase and delayed detection and correction of malfunctions in building services. Actions at building level are required in order to provide the means to correct incorrect design models, thus reconciling predictions with field observations, and to implement smart control strategies, and effective methods to promptly spot and effectively correct faults in building services. On the other hand, effective demand shifting in order to better match the irregular energy supply provided by renewables can be achieved only by coordinating the performances of ensembles of buildings. Finally, the advent of widespread deployment of smart meters and increased monitoring activities will provide data and information necessary to complement detailed building energy models with probabilistic data driven models in an increasingly more effective way. The depicted analysis framework, and in particular the explained calibration methodology, can be already used to answer several of these questions.

The natural evolution of the current research will be to serve as a base for developing an automatic on-line calibration system, in order to dynamically calibrate and upgrade BES models according to the monitored data. This will lead to opportunities for the implementation of model predictive controls (especially for energy services incorporating renewables), fault detection, retrofit analysis and identification of energy saving opportunities, and feedback to users and building owners on their energy demand. The first two of these applications seems to be the most novel and interesting. In order to undertake this task successfully, it will be necessary to treat a constantly growing volume of data, and methods to select and retain only useful information are needed. In particular, the current calibration method could be extended by investigating novel on-line learning techniques involving Sparse Representation and Reduced Rank Approximation of GPR models ([32] and [92]), in order to deal with the constantly growing inflow of data, and complemented with algorithms able to solve classification problems,

in order to perform fault detection.

The resulting system can be applied to single buildings, but, due to the probabilistic character of its solutions, it could also be effectively used to build a hierarchical system working at community level. The underlying concept will be the creation of a hierarchical control algorithm so that decisions and actions taken locally influence those taken at higher levels, in order to achieve a perfect balance between building level, and community level objectives. This could be achieved by including in the system a high level model of the building community taking as inputs the probability density distributions inferred in the local calibrations, and averaging them in order to infer the best solution at community level. In this way the single building actions would be orchestrated as functions of the optimal compromise between local and global performances.

The subjects benefiting from the development of such systems will be many. The main initial impacts will be on small isolated communities that cannot rely on conventional energy sources, but instead employ hybrid energy systems wherein renewables play important roles. Examples of such communities can be found in the north of Scotland and on its islands where wind and tidal energy is being increasingly adopted. However, there are wider beneficiaries of the project due to the generality of its application: building occupants, building owners, architects, energy managers, consulting engineers, energy utilities and building energy software developers. In particular, it has the potential to rationalize the energy consumption of communities without penalising comfort conditions at single building level, leading to advantages for building users and managers as well as energy providers. The former will benefit from energy savings, emission reductions, and decreasing maintenance activity of their building services. The latter will see the reduction in their revenues from energy sales overcome by the savings from management of the grid network, resulting from a more even energy demand profile. Finally, the academic impact will be developments in combining deterministic and probabilistic modelling techniques. This will lead to increase applicability and accuracy of Building Energy Simulation models and will open new research directions.

The application of such a paradigm to complex problems like the operation and control of buildings and the tight coupling of probabilistic modelling techniques from Machine Learning with BES, represent novel, ambitious but nonetheless realistic objectives. In particular the recent development in dynamic simulation programs, the improvement of data acquisition system and Data Analysis methods, as well as the constantly increasing computational power, make the

challenges involved in its realisation approachable.

Appendix A

Further Math

In this Chapter few more mathematical complements are given in order to fully clarify the mathematics behind the presented analysis framework for BES models. In particular the mathematical operation defined as *completing the square*, and the method adopted in this work for performing Principal Component Analysis are explained.

A.1 Completing the square

With the term *completing the square* it is indicated a common algebraic manipulation of quadratic forms, that is particularly useful in calculations involving Gaussian probability density distributions. The objective of this operation is to express the following quadratic form:

$$(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{w})^T \boldsymbol{C}(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{w}) \tag{A.1}$$

as

$$(\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{G}(\boldsymbol{w} - \hat{\boldsymbol{w}}) + constant \tag{A.2}$$

where $\boldsymbol{C}$ and $\boldsymbol{G}$ are symmetric positive defined matrices.

Developing Equation (A.1) it results:

$$\boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y} - 2\boldsymbol{w}^T\boldsymbol{K}^T\boldsymbol{C}\boldsymbol{y} + \boldsymbol{w}^T\boldsymbol{K}^T\boldsymbol{C}\boldsymbol{K}\boldsymbol{w} \tag{A.3}$$

Similarly, developing Equation A.2 it results:

$$\boldsymbol{w}^T\boldsymbol{G}\boldsymbol{w} - 2\boldsymbol{w}^T\boldsymbol{G}\hat{\boldsymbol{w}} + \hat{\boldsymbol{w}}^T\boldsymbol{G}\hat{\boldsymbol{w}} + constant \tag{A.4}$$

By equating the terms in Equations A.3 and A.4 the following is achieved:

$$\begin{aligned}
&\boldsymbol{w}^T\boldsymbol{K}^T\boldsymbol{C}\boldsymbol{K}\boldsymbol{w} = \boldsymbol{w}^T\boldsymbol{G}\boldsymbol{w} \Rightarrow \boldsymbol{G} = \boldsymbol{K}^T\boldsymbol{C}\boldsymbol{K} \\
&2\boldsymbol{w}^T\boldsymbol{K}^T\boldsymbol{C}\boldsymbol{y} = 2\boldsymbol{w}^T\boldsymbol{G}\hat{\boldsymbol{w}} \Rightarrow \hat{\boldsymbol{w}} = \boldsymbol{G}^{-1}\boldsymbol{K}^T\boldsymbol{C}\boldsymbol{y} \\
&\boldsymbol{y}^T\boldsymbol{G}\boldsymbol{y} = \hat{\boldsymbol{w}}^T\boldsymbol{G}\hat{\boldsymbol{w}} + constant \Rightarrow \\
&constant = \boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y} - \hat{\boldsymbol{w}}^T\boldsymbol{G}\hat{\boldsymbol{w}} = \boldsymbol{y}^T(\boldsymbol{C} - \boldsymbol{C}\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{K}^T\boldsymbol{C})\boldsymbol{y}
\end{aligned}$$

A.2 Principal Component Analysis

Principal Component Analysis (PCA) ([63]) is a general purpose method which can be used to decompose and analyse a dataset variance while reducing at the same time its dimensionality, with the least loss of information. In particular,PCA projects the data in a space of lower dimensionality wherein only few variables are needed to represent the majority of their variability. For a comprehensive discussion of PCA the reader is referred to [91] and [63], in the following only the aspects of interest for this work will be underlined.

HerePCA is used primary to reduce the dimensionality of the datasets in hand, in order to decrease the computational burden, and to summarise multidimen-

sional model outputs in calculating sensitivity indexes. These two objective are achieved by representing the original time series with basis expansions built upon an optimal empirical sets of orthogonal basis vectors, which are tailored on the data in hand. PCA allows to define such orthogonal bases, by decomposing the original dataset in the most possible efficient way, meaning that the derived basis system system always explains an higher fraction of the original data variance than any other of the same dimension. The main result is an accurate approximation of the original time series through a small number of variables. This is particularly important in the analysis of BES models in order to effectively reduce the computational cost. In particular due to the variety of frequencies contained in the signals involved, other basis expansions would require a large number of basis vectors to adequately approximate the original time series.

PCA can be achieved through eigen-decomposition of the data empirical covariance matrix. The following approach based on Singular Value Decomposition (SVD) was adopted. Let $\boldsymbol{Y}$ be the $N \times M$ matrix having as columns the realisations of a certain multidimensional variable, and let $\boldsymbol{Y}$ be centred in such a way that each row as zero mean, then by SVD, $\boldsymbol{Y}$ can be expressed like:

$$\boldsymbol{Y} = \boldsymbol{\Xi}\boldsymbol{D}\boldsymbol{\Omega}^T \tag{A.5}$$

where $\boldsymbol{\Xi}$ and $\boldsymbol{\Omega}$ are respectively matrices whose columns are orthonormal vectors and $\boldsymbol{D}$ is a diagonal matrix. The diagonal elements of $\boldsymbol{D}$ are called singular values (d_q). Consequently the empirical covariance matrix of $\boldsymbol{Y}$ ($\boldsymbol{\Sigma}$) can be expressed as:

$$\boldsymbol{\Sigma} = \frac{1}{M}\boldsymbol{Y}\boldsymbol{Y}^T = \frac{1}{M}\boldsymbol{\Xi}\boldsymbol{D}\boldsymbol{\Omega}^T\boldsymbol{\Omega}\boldsymbol{D}\boldsymbol{\Xi}^T = \frac{1}{M}\boldsymbol{\Xi}\boldsymbol{D}^2\boldsymbol{\Xi}^T$$

so that the columns of $\boldsymbol{\Xi}$ are eigenvectors for $\boldsymbol{\Sigma}$ and the relative eigenvalues are the diagonal element of the matrix $\frac{1}{M}\boldsymbol{D}^2$. Defining the total variance of $\boldsymbol{Y}$ as the trace of its empirical covariance matrix:

$$\mathrm{V}(\boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{\Sigma}) \tag{A.6}$$

the eigenvectors contained in $\boldsymbol{\Xi}$ are ordered so as to explain decreasing fractions of $\mathrm{V}(\boldsymbol{Y})$. Since the trace of the empirical covariance matrix is equal to the sum of its eigenvalues, the $q-th$ fraction can be calculated as the ratios of the $q-th$ eigenvalue and their total sum.

The columns of the matrix $\boldsymbol{Y}$ can, then, be expanded in the following way:

$$\boldsymbol{y}_m = \sum_{q=1}^{Q} \boldsymbol{k}_q w_{m,q} + \boldsymbol{\varepsilon} = \boldsymbol{K}\boldsymbol{w}_m + \boldsymbol{\varepsilon} \tag{A.7}$$

According to Equation A.5 the entities in Equation A.7 can be defined in different manners. In this study the following two conventions have been adopted. WhenPCA is used to summarise multidimensional model outputs in the calculation of sensitivity indexes:

$$\boldsymbol{K} = \boldsymbol{\Xi}$$
$$\boldsymbol{W} = \boldsymbol{\Omega}\boldsymbol{D}$$

where $\boldsymbol{W}$ is the matrix having as rows the vectors $\boldsymbol{w}_m$. In this way the data set consisting of the sum of the columns of $\boldsymbol{W}$, $\boldsymbol{w}_q$ ($\sum_q^Q \boldsymbol{w}_q$), has the same variance as the original data set ($\boldsymbol{Y}$). Since it is composed by unidimensional variables it can be used in the calculation of the needed sensitivity measures ([67]).

WhenPCA is used in the *Calibration* step for reducing the dimensionality of the employed datasets:

$$\boldsymbol{K} = \boldsymbol{\Xi}\boldsymbol{D}$$
$$\boldsymbol{W} = \boldsymbol{\Omega}$$

so that it is easier to specify the GP prior probability density distributions for the vectors $\boldsymbol{w}_q$.

However these are just conventions and many other can be found in literature ([50]). As long as they are correctly considered during the calculation they do not change the final results.

Only the first Q principal components ($\boldsymbol{k}_q$ and $\boldsymbol{w}_q$ pairs) are used in approximating the original time series, and iid Gaussian noise ($\boldsymbol{\varepsilon}$) is added to underline the degree of uncertainty due to this approximation. Several tools are available for identifying the right amount of principal components to retain ([42]). Here, due to the character of the data, which are mostly synthetic, or for which often is available a limited amount of realisations, a threshold approach was used. In particular, retaining the first Q principal components (pairs of $\boldsymbol{k}_q$ and $\boldsymbol{w}_q$) explaining the 99% of the original dataset variance was found to give suitable approximations.

Appendix B

Algorithms

During the research significant effort was put in searching suitable algorithms able to carry out the needed calculation effectively. The two main requirements were the computation time and the capability of the algorithm to adequately explore the complicated probability density functions to maximise or integrate. Deterministic algorithms, like Conjugate Gradients methods, Variable Metric methods and the Downhill Simplex method were tested, as well as stochastic ones, like Markov Chain Monte Carlo (MCMC) algorithms and Simulate Annealing (SAN).

In the treated problems, the latter typology performed significantly better than the former. In particular, Adaptive Metropolis within Gibbs (AMG) algorithm was adopted as MCMC method of reference for its good performances in high dimensional spaces. Upon it, SAN and Annealed Importance Sampling (AIS) were developed. The former, was used in estimating the hyper parameter of the *Training Model*, while the latter was employed during the *Calibration* phase, for estimating calibration parameters, and for calculating marginal likelihoods needed for the calculation of Bayes Factors.

A discussion of Monte Carlo and MCMC methods is beyond the purpose of this work. The interested reader is referred to [10] and [100]. A general overview of optimisation methods can be found in [85] and [86]. In this section, the adopted algorithms are explained, and their implementation in computational routines described through pseudo code. The goodness of the developed routines were checked by repeating examples contained in previous studies, and by comparing the results from the virtual experiments in Chapter 5 with the known exact solutions.

In the following $p(\cdot)$ refers to a probability density distribution or to a function proportional to a certain probability density distribution.

B.1 Adaptive Metropolis Within Gibbs

The Adaptive Metropolis within Gibbs algorithm ([103]), combines the sampling cycle of a Gibbs Sampler ([40]) with the proposing-accept/reject step of a Metropolis Hastings Algorithm ([74]). As a Gibbs Sampler, AMG decompose a complicated sampling problem, in a series of samplings from conditional probability density distributions. In particular, AMG achieves parameters samples from the joint posterior probability density distribution of the parameters ($\boldsymbol{z}$), by trying to orthogonally sample from each dimension of the function to integrate.

For example let be $\boldsymbol{z}^{(k)}$ a vector of length, P defining the $k-th$ state of the Markov chain, $z_p^{(k)}$ its $p-th$ element. Similarly let be $\boldsymbol{z}^{(l)}$ the previous state of the Markov chain, and $\boldsymbol{z}_{-p}^{(l)}$ the subset of $\boldsymbol{z}^{(l)}$ consisting in all its elements but $z_p^{(l)}$. Then, a normal Gibbs Sampler would proceed as in Algorithm 1.

Algorithm 1 GIBBS SAMPLER

1: **procedure** GIBBS($\boldsymbol{z}$)
2: **for** $p = 1$ to P **do**
3: generate $z_p^{(k)} \sim p(z_p^{(k)}|\boldsymbol{z}_{-p}^{(l)})$
4: write($\boldsymbol{z}^{(k)}$)
5: repeat

However, in order to effectively draw random values from the conditional probability density distributions $p(z_p^{(k)}|\boldsymbol{z}_{-p}^{(l)})$, these must have simple forms, and this is seldom the case. The proposing-accept/reject step of the Metropolis Hastings Algorithm helps to overcome this issue. Let be $\boldsymbol{z}_p^{(k)}$ the $k-th$ state of the chain where only the $p-th$ element has been sampled. AMG uses Normal proposal probability density distributions, so that the transition kernel of the Markov chain reduces to:

$$T(\boldsymbol{z}^{(l)} \rightarrow \boldsymbol{z}_p^{(k)}) = \min\left(1, \frac{p(\boldsymbol{z}_p^{(k)})}{p(\boldsymbol{z}^{(l)})}\right) \tag{B.1}$$

The AMG algorithm is described in Algorithm 2.

Orthogonal variations are applied to the vector $\boldsymbol{z}$ according to Normal proposal probability density distributions. The transition probability is evaluated and the new vector accepted or rejected depending on $T(\boldsymbol{z}_p^{(l)} \rightarrow \boldsymbol{z}^{(k)})$ being greater than a number randomly drawn within [0,1]. In this way, the algorithm always accepts $\boldsymbol{z}_p^{(k)}$ increasing the value of $p(\cdot)$, but, it can also move in $\boldsymbol{z}_p^{(k)}$ decreasing the value of $p(\cdot)$; thus potentially escaping local modes and exploring the parameter space more efficiently than deterministic methods. The set of instructions in Algorithm 2 is repeated a number of time sufficient to achieve convergence,

Algorithm 2 ADAPTIVE METROPOLIS WITHIN GIBBS (AMG)

procedure AMG($\boldsymbol{z}^{(l)}$, sd(z_p) ($p = 1, \ldots, P$))
  for $p = 1$ to P **do**
    generate $z_p^{(k)} \sim N(z_p^{(l)}, \text{sd}(z_p))$
    if $\frac{p(\boldsymbol{z}_p^{(k)})}{p(\boldsymbol{z}^{(l)})} > U(0,1)$ **then**
      $z_p^{(l)} \leftarrow z_p^{(k)}$
    else
      $z_p^{(k)} \leftarrow z_p^{(l)}$
  write($\boldsymbol{z}^{(l)}$)
  repeat

meaning that marginal stationary probability density distributions for all the parameters contained in the vector $\boldsymbol{z}$ are reached. The convergence of the Markov chains was evaluated according to the methods outline in the R package CODA ([84]), in particular with Batch Means ([38] and [39]).

During the simulation, AMG adapts the proposal probability density distributions in order to maximise the efficiency of the mixing of the Markov chain. The adapting is done according to the the acceptance rate of each model parameter, which is evaluated periodically after a certain number of iterations (I):

$$\frac{nuber\ of\ accepted\ samples\ for\ z_p\ in\ I\ iterations}{I} \tag{B.2}$$

Since the algorithm proceeds by moving independently each z_p, the standard deviation of the proposal element of $\boldsymbol{z}$ are varied in order to achieve acceptance rates (Equation B.2) equal to 0.44, which is the optimum for the one dimensional case ([101]). If the acceptance rate of z_p is lower than 0.44, sd(z_p) is decreased so as to evaluate points closer to the actual one. Diversely, if the acceptance rate of z_p is higher than 0.44, sd(z_p) is increased so as to explore a larger portion of the search space. In particular the adapting of the sd(z_p) is performed by adding or subtracting the following amount to log(sd(z_p)):

$$\min(0.01, \frac{1.0}{\sqrt{total\ number\ of\ iteratations}})$$

It is important to notice that as the number of iteration approaches ∞, the amount of which the proposal distributions are adapted tend to zero. Therefore, as AMG approaches the stationary probability density distributions for all the z_p also the relative proposal distributions reach stationary states.

B.2 Simulate Annealing

Simulate Annealing (SAN) is a general empirical procedure for achieving stochastic maximisation of functions which does not require the evaluation of derivatives. The definition of SAN and a deep study of its convergences properties are contained in [41]. The basic concept, is to build a sequence of functions $(g_i(\boldsymbol{z}))$ having the same maxima as the function to maximise $(p(\boldsymbol{z}))$, and which gradually concentrates around the absolute maximum of $p(\boldsymbol{z})$. A convenient way to built such sequence of functions is:

$$g_i(\boldsymbol{z}) = exp\{\frac{p(\boldsymbol{z})}{\mathcal{T}_i}\} \tag{B.3}$$

where $\mathcal{T}_i$ is a decreasing sequence of positive real numbers, generally called temperature parameters. As the simulation progress and $\mathcal{T}_i$ gradually approaches zero, $g_i(\boldsymbol{z})$ will be increasingly peaked around the absolute maximum of $p(\boldsymbol{z})$. In this way the maximisation of the function $p(\boldsymbol{z})$ is reduced to drawing points from the $g_i(\boldsymbol{z})$ and estimating their modes.

Therefore, SAN requires the setting up a temperature schedule and the drawing of points from the functions $g_i(\boldsymbol{z})$. The former issue is highly dependent on the complexity of the function to be maximised. It can be shown that it exists an optimal temperature schedule assuring the convergence of the algorithm ([41]). However its determination is complicated and as general principle it can be assumed that the probability to correctly identify the absolute maximum of $p(\boldsymbol{z})$ increases for slowly changing temperature schedules. Linearly, logarithmically or geometrically decreasing trends are all valid strategies in defining suitable sequences of $\mathcal{T}_i$ schedules, and it is advised to run SAN with increasingly refined temperature schedules in order to check its convergence. The latter issue can easily be solved adopting a MCMC strategy to sample from the functions $g_i(\boldsymbol{z})$. In particular, AMG was used. The implementation of SAN adopted in this research is outlined in Algorithm 3, wherein the notation as the same meaning as in Section B.1.

During the I-cycle $\mathcal{T}_i$ is decreased according to the given schedule. At each iteration of the I-cycle, the algorithm evaluates $g_i(\boldsymbol{z})$ J times looking for possible maxima, and it adapts the $\mathrm{sd}(z_p)$ similarly as described in Section B.1, in order to facilitate the next search.

For high values of $\mathcal{T}_i$, the corresponding $g_i(\boldsymbol{z})$ will be quite flat, and its local maxima will be confounded with it global maximum. This make, SAN particularly capable of escaping local modes in the first iterations of the I-cycle, and

Algorithm 3 SIMULATE ANNEALING (SAN)

procedure SAN($\boldsymbol{z}^{(l)}$, sd(z_p) ($p = 1, \ldots, P$), $\mathcal{T}_i$ ($i = 1, \ldots, I$))
  for $i = 1$ to I **do** ▷ I-cycle
    for $j = 1$ to J **do**
      for $p = 1$ to P **do**
        generate $z_p^{(k)} \sim N(z_p^{(l)}, \text{sd}(z_p))$
        if $\frac{g_i(\boldsymbol{z}_p^{(k)})}{g_i(\boldsymbol{z}^{(l)})} > U(0,1)$ **then**
          $z_p^{(l)} \leftarrow z_p^{(k)}$
        else
          $z_p^{(k)} \leftarrow z_p^{(l)}$
    write($\boldsymbol{z}^{(l)}$)
    write(max($g_i(\boldsymbol{z}^{(j)}); j = 1, \ldots, l$))
    write($\boldsymbol{z}^{(j)}$ for which $g_i(\boldsymbol{z})$ is maximised)
    adapt all the sd(z_p)

effective in adequately exploring the domain of the function. In particular, it may be useful to consider the following analogy. For large value of $\mathcal{T}_i$, SAN has large energy and a very well defined maximum is needed to trap it. As $\mathcal{T}_i$ decreases, so it does the energy of SAN, which gradually is captured, hopefully, by the global maximum of $p(\boldsymbol{z})$.

Although SAN uses AMG to sample the different $g_i(\boldsymbol{z})$, it does not have the properties of MCMC algorithms since it does not converge, because not required, to stationary probability density distributions. It only returns a point estimated to be the global maximum of the function in object. Therefore the same method adopted in checking the convergence of AMG cannot be used in this case. In particular, as stopping rule for SAN the following criterion as been adopted. At each iteration of the I-cycle, Algorithm 3, returns the maximum value of $g_i(\boldsymbol{z})$ and the relative input vectors ($\boldsymbol{z}_i$). The values of the original function can be retrieved as follows:

$$p(\boldsymbol{z}_i) = g_i(\boldsymbol{z}_i)^{\mathcal{T}_i}$$

Because of the properties of SAN, the values $p(\boldsymbol{z}_i)$ will be a non decreasing sequence having decreasing differences that will tend to zero as the algorithm approaches the global maximum. There fore as stopping rule the following criterion was adopted:

$$\frac{\frac{1}{I-1}\sum_{i=2}^{I} p(\boldsymbol{z}_i) - p(\boldsymbol{z}_{i-1})}{\frac{1}{I}\sum_{i=1}^{I} p(\boldsymbol{z}_i)} < \epsilon$$

where ϵ was set to the fourth root of the machine precision.

B.3 Annealed Importance Sampling

Annealed Importance Sampling (AIS) [80] blends together principles coming from Importance Sampling [100] and the previously described Simulate Annealing.

Importance Sampling is a Monte Carlo integration method based on the concepts of importance functions. In particular, having the following integral:

$$\int h(\boldsymbol{z})p(\boldsymbol{z})d\boldsymbol{z} \tag{B.4}$$

the normal Monte Carlo estimator would approximate it as:

$$\mathrm{E}(h(\boldsymbol{z})) = \frac{1}{J}\sum_{j=1}^{J} h(\boldsymbol{z}_j)$$

where $\boldsymbol{z}_i$ are generated from $p(\boldsymbol{z})$. However, unless $p(\boldsymbol{z})$ is simple, it can be difficult or even not possible to directly generate values from it. Importance Sampling answers to this problem by reformulating Equation (B.4) as indicated below:

$$\int h(\boldsymbol{z})p(\boldsymbol{z})d\boldsymbol{z} = \int \frac{h(\boldsymbol{z})p(\boldsymbol{z})}{g(\boldsymbol{z})} g(\boldsymbol{z})d\boldsymbol{z}$$

therefore the estimator of the integral becomes:

$$\mathrm{E}(h(\boldsymbol{z})) = \frac{1}{J}\sum_{j=1}^{J} h(\boldsymbol{z}_j)\frac{p(\boldsymbol{z}_j)}{g(\boldsymbol{z}_j)} \tag{B.5}$$

where $\boldsymbol{z}_j$ are generated from $g(\boldsymbol{z})$. $g(\boldsymbol{z})$ can be chosen simple enough so that it is possible to directly generate values from it. Nonetheless, in order for Importance Sampling to be efficient, $g(\boldsymbol{z})$ should good approximation of $p(\boldsymbol{z})$, so as to have the same zones of high probability.

The function $g(\boldsymbol{z})$ is called importance function and the ratios $\frac{p(\boldsymbol{z}_j)}{g(\boldsymbol{z}_j)}$ are called importance weights (ω_j). By normalizing such weights so that they sum up to one it is possible to derive the self-normalized importance sampling estimator:

$$\mathrm{E}(h(\boldsymbol{z})) = \sum_{j=1}^{J} \frac{\omega_j}{\sum_{j=1}^{J}\omega_j} h(\boldsymbol{z}_j) \tag{B.6}$$

which is usually employed in Importance Sampling. It is interesting to notice that the average of ω_j provides and estimate for the ratio of the normalizing constants of $p(\boldsymbol{z})$ and $g(\boldsymbol{z})$. The latter, being chosen a priory, is often a proper probability density distribution (i.e. normalized), therefore the average of ω_j can be used to

estimate the normalizing constant of $p(\boldsymbol{z})$.

One of the major issue related to normal Importance Sampling is the choice of the importance function. AIS proposes a solution by defining a sequence of importance functions $(g_j(\boldsymbol{z}))$ smoothly connecting a starting probability density distribution $(g_0(\boldsymbol{z}))$, from which is trivial to simulate values, and the density probability distribution of interest $(p(\boldsymbol{z}))$. MCMC methods are used to sample from, and move across this sequence of importance functions. The sequence $g_i(\boldsymbol{z})$ is defined as follows:

$$g_i(\boldsymbol{z}) = p(\boldsymbol{z})^{\mathcal{T}_i} g_0(\boldsymbol{z})^{1-\mathcal{T}_i} \tag{B.7}$$

where the temperature parameters $\mathcal{T}_i$ $(i = 1, \ldots, I)$ are an increasing sequence of real numbers in $[0, 1]$ such as $\mathcal{T}_0 = 0$ and $\mathcal{T}_I = 1$. The pseudo code describing AIS is depicted in Algorithm 4, were the starting $\boldsymbol{z}^{(l)}$, given as input, is generated from $g_0(\boldsymbol{z})$.

Algorithm 4 ANNEALED IMPORTANCE SAMPLING (AIS)

procedure AIS($\boldsymbol{z}^{(l)}$, sd(z_p) $(p = 1, \ldots, P)$, $\mathcal{T}_i$ $(i = 1, \ldots, I)$)
  $\omega \leftarrow 1$
  for i = 1 to I **do** ▷ I-cycle
    $\omega \leftarrow \omega \cdot \frac{g_i(\boldsymbol{z}_i^{(l)})}{g_{i-1}(\boldsymbol{z}_i^{(l)})}$
    for $j = 1$ to J **do** ▷ J-cycle
      for $p = 1$ to P **do**
        generate $z_p^{(k)} \sim N(z_p^{(l)}, \text{sd}(z_p))$
        if $\frac{g_i(\boldsymbol{z}_p^{(k)})}{g_i(\boldsymbol{z}^{(l)})} > U(0, 1)$ **then**
          $z_p^{(l)} \leftarrow z_p^{(k)}$
        else
          $z_p^{(k)} \leftarrow z_p^{(l)}$
  write($\boldsymbol{z}^{(l)}$)
  write(ω)

By repeating Algorithm 4, it is possible to go through the sequence of $g_i(\boldsymbol{z})$ multiple times, thus generating a sequence of points drawn from $g_{I-1}(\boldsymbol{z})$. The importance weight of each one of these points can be calculated as in Equation (B.8).

$$\omega = \prod_{i=1}^{I} \frac{g_i(\boldsymbol{z}_i)}{g_{(i-1)}(\boldsymbol{z}_i)} \tag{B.8}$$

It is then possible to employ the estimator in Equation (B.6) to evaluate the quantity of interest.

As for Importance Sampling, the average of the importance weights is an

estimate of the ratio of the normalizing constants of $g_0(\boldsymbol{z})$ and $p(\boldsymbol{z})$. In case of Bayesian calculation, as in this research, it is possible to set $g_0(\boldsymbol{z})$ to be the joint prior probability distribution of the parameters to estimate and $p(\boldsymbol{z})$ their not-normalised joint posterior probability density distribution. Thus the average of the importance weights provides an estimate of the marginal likelihood of $p(\boldsymbol{z})$ which can be used for model comparison, in particular in the calculation of Bayes Factors.

It is possible to generate points directly from $p(\boldsymbol{z})$ by leaving AIS to draw from $g_I(\boldsymbol{z})$ for a certain number of time in a MCMC fashion, for example by increasing the number of iteration of the J-cylce for $i{=}I$. By taking only the last point for each repetition of Algorithm 4 it is possible to obtain a sample of independent point generated from $p(\boldsymbol{z})$. Also for AIS it was decided to adapt the proposal probability density distributions as described in Section B.1, but in this case the sd(z_p) are adapted separately for each temperature parameter ($\mathcal{T}_i$). In this way at each repetition, AIS improves the drawing from each $g_i(\boldsymbol{z})$. The convergence of AIS was assessed according to the methods outline in the R package CODA ([84]), in particular Batch Means were employed ([38] and [39]).

Appendix C

GOF between BES models and GPR emulators

In this Chapter are listed Goodness Of Fit criteria representing the agreement between the analysed BES models and the created GPR emulators. The adopted GOF criteria are:

- The Coefficient of Determination (R^2):

$$R^2 = \frac{\sum_m (f(\boldsymbol{z}_m) - \bar{f(\boldsymbol{z})})^2}{\sum_m (\boldsymbol{y}_m - \bar{\boldsymbol{y}})^2}$$

  where $f(\boldsymbol{z}_m)$ is the output of the GPR model for the input vector $\boldsymbol{z}_m$, $\bar{f(\boldsymbol{z})}$ is the empirical mean of the GPR model output, $\boldsymbol{y}_m$ is the $m-th$ BES model simulation and $\bar{\boldsymbol{y}}$ is the empirical mean of the BES model output. It could be seen as the fraction of variance of the original model explained by the emulator.

- The Root Mean Squared Error.

- The 95% quantile of the residuals ($\boldsymbol{res}$) between the output of the BES model and the GPR emulator ($Q(\boldsymbol{res}, 95\%)$).

C.1 For the wall experiment

Table C.1: The Wall: R^2 the RMSE and $Q(\boldsymbol{res}, 95\%)$ between the actual BES model and the GPR emulator.

R^2	RMSE $(\frac{W}{m^2})$	$Q(\boldsymbol{res}, 95\%)$ $(\frac{W}{m^2})$
0.992	0.087	0.189

C.2 For the Test Box experiments

Table C.2: Test Box: R^2 the RMSE and $Q(\boldsymbol{res}, 95\%)$ between the actual BES model and the GPR emulator.

MODEL	DATASET	R^2	RMSE (kW)	$Q(\boldsymbol{res}, 95\%)$ (kW)
$\mathcal{M}_0$	BBRI-CT	0.987	0.003	0.006
$\mathcal{M}_1$	BBRI-CT	0.991	0.003	0.006
$\mathcal{M}_2$	BBRI-CT	0.988	0.004	0.007
MODEL	DATASET	R^2	RMSE (^oC)	$Q(\boldsymbol{res}, 95\%)$ (^oC)
$\mathcal{M}_0$	BBRI-FF	0.998	0.230	0.455
$\mathcal{M}_1$	BBRI-FF	0.991	0.371	0.765
$\mathcal{M}_2$	BBRI-FF	0.994	0.391	0.775
MODEL	DATASET	R^2	RMSE $(\frac{W}{m^2})$	$Q(\boldsymbol{res}, 95\%)$ $(\frac{W}{m^2})$
$\mathcal{M}_3$	BBRI-CT	0.994	1.072	2.317
MODEL	DATASET	R^2	RMSE (^oC)	$Q(\boldsymbol{res}, 95\%)$ (^oC)
$\mathcal{M}_1$	LECE	0.992	0.567	1.190
$\mathcal{M}_2$	LECE	0.995	0.346	0.734
MODEL	DATASET	R^2	RMSE $(\frac{W}{m^2})$	$Q(\boldsymbol{res}, 95\%)$ $(\frac{W}{m^2})$
$\mathcal{M}_4$	LECE	0.992	0.735	1.527

C.3 For Twin House N2

Table C.3: Twin House N2: R^2 the RMSE and $Q(\boldsymbol{res}, 95\%)$ between the actual BES model and the GPR emulator.

MODEL	DATASET	R^2	RMSE (kW)	$Q(\boldsymbol{res}, 95\%)$ (kW)
$\mathcal{M}_1$	$CT1$	0.990	0.091	0.206
$\mathcal{M}_2$	$CT1$	0.953	0.101	0.196
MODEL	DATASET	R^2	RMSE (^oC)	$Q(\boldsymbol{res}, 95\%)$ (^oC)
$\mathcal{M}_1$	$ROLBS1$	0.990	0.099	0.195
$\mathcal{M}_2$	$ROLBS1$	0.991	0.085	0.169

Glossary

Bayesian Calibration In this research the term Bayesian Calibration is referred to the employment of meta-models based upon GPR in a quasi-Bayesian framework in the calibration of complicated computer models. 25, 26, 28–30, 32, 55, 230

Calibration Model GPR model describing the BES model evaluated at the unknown vector of calibration parameters. It explains the variability of the observed measurement that can be explained by the BES model. 76, 82, 87, 91, 94, 98, 99, 228, 229, 236, 237

Calibration Second phase of the presented methodology for analysis and diagnosis of BES models. It involves *Training* and *Identification*. 4, 18, 20, 58, 199, 200, 206, 217, 218, 228, 230

Difference Analysis It indicates the analysis and interpretation of the results from the training of the *Difference Model*. In particular according to the MAP of the hp of such GPR models possible causes of discrepancy between model predictions and field observations are identified as well as model improvements. 10, 11, 97, 98, 109, 111, 115, 116, 123–127, 142, 150, 157–159, 170, 193, 195, 204, 205

Difference Model GPR model describing the variability of the observed measurement that cannot be explained by the BES model. It is defined so as to be complementary to the *Calibration Model*, and it depends only on the variable boundary conditions. 72, 76, 82, 83, 87, 91, 93, 95, 97–99, 111, 113, 123, 228, 229, 236, 245, 246

Difference Vectors Residuals of the field observations after least square fit based on the basis vectors $\boldsymbol{k}_q$. 82, 89, 93, 98, 113, 123, 158, 170, 171, 193, 235, 237–239, 241, 244

Identification Second phase of the *Calibration* framework. A GPR model depicting the observed measurements is built and linked to the *Training Model*

in order to infer unknown model parameters according to the BES model behaviour. This model is composed of *Calibration Model* representing the variability of the data that the BES model can explain, and the *Difference Model* representing the variability of the data the the BES cannot explain. 76, 77, 83, 86, 91, 94, 97, 99, 199, 228

Manual Iterative Calibration Ad hoc iterative series of steps wherein the analyst gathers information and data about the building object of the study, and according to its judgement, experience and intuition tries to identify a suitable model by tuning its inputs in order to achieve a satisfying match with the measured target data. 41, 42, 44, 46, 54

Mathematical Analytical Calibration It employs sensitivity analysis and optimization techniques to determine possible solutions to the calibration problem. It often uses statistical tools like Monte Carlo Simulation and statistical tests, but differ sensibly from a fully probabilistic/stochastic approach since calibration is treated as a standard optimization problem. 46, 54

Model Inadequacy Uncertainty due to errors in the modelling of the real processes. Thus even if parameter uncertainty is negligible, the observed process can exhibit variability which the computer model is not able to explain.. 21, 100

Model Selection It is the comparison and ranking of different models in order to select the best one. In this research the adopted criterion of reference is Bayes Factors. 4, 18, 20, 91, 199

Observation Errors Uncertainty due to random and systematic measurements errors. 21, 87, 97, 100

Parameter Uncertainty Uncertainty in model inputs which might be unknown, vary from context to context or are dependent on other variables. 21, 22

Parametric Variability Uncertainty involved in using the model to make predictions in unknown, uncontrolled or unspecified conditions. In BES models this kind of uncertainty is related mainly to the variability of the weather factors determining the boundary conditions.. 21

Probabilistic/Stochastic Calibration It involves using grey-box models or supportive black-box probabilistic models in order to perform parameter

identification by likelihood maximization or Maximum A Posteriori (MAP) estimation, when a Bayesian framework is used. In particular, the use of meta-model based on GPR in a quasi-Bayesian framework to calibrate computer models is referred as *Bayesian Calibration*. 50, 54, 55

Representation Uncertainty Complicated computer models, like BES models, involve complicated relations among the different part of the calculation. Thus even though the computer code, acting as a deterministic function, is actually known, its complexity as well as possible errors make its output, with respect to a particular input configuration, unknown until the simulation is actually run. Running the code for each possible input configuration is not feasible; thus there is additional uncertainty to consider. In [65] this was referred to as *Code Uncertainty*. This therm was deemed to be ambiguous in this context since it could have been interpreted as uncertainties in the coding activity itself. In order to avoid misunderstanding it was changed as here indicated. 21, 84, 97, 99

Residual Variability Uncertainty due to the property of a process to assume different values for different observations in the same conditions. It may be due to actual unpredictability of the phenomenon or to some unobserved conditions acting on the process. 21, 87, 97, 100

Sensitivity Analysis It is the study of the sensitivity of a output of a model to its inputs. In this study it is referred to the procedure explained in Chapter 3, and its objective is to identify a suitable set of calibration parameters. 4, 18, 20, 40, 58, 199

Training Model GPR model created in the *Training* phase. It emulates the original BES model within the inputs parameter space defined during the *Uncertainty Analysis* phase. 76, 94, 98, 218, 228, 237, 240, 246

Training First phase of the *Calibration* framework. A GPR model is trained on a significant sample of the BES model output in order to obtain a fast running emulator. 76, 77, 83, 87, 89, 91, 94, 97, 199, 228, 230

Uncertainty Analysis It is defined as the study of the variability of all the free parameters in a model. In this research it is the first step of the outlined procedure for BES models analysis. A probabilistic approach is adopted, meaning that probability density distributions representing the initial beliefs about model parameter uncertainties are attributed to each model input. 4, 18, 20, 22, 58, 173, 186, 199, 200, 230

equifinality In this context the term is referred to complex *over-parametrised* computer models and underlines the fact that models of different structures and for different combinations of their inputs can provide substantially equally good representation of the observed phenomena. 17, 27, 28, 38, 126, 127, 138, 205, 231

experimental identifiability It is the parameter identifiability determined by the experiment, especially by the solicitations and signal applied. For example, the applied solicitations or signals can highlight only particular features of the object of the experiment and thus allow only the estimation of the correlated parameters.. 202, 203, 231

over-parametrisation Characteristic of a model having a large number of parameters which are not clearly identifiable with the information and the data provided due to *structural identifiability* and *experimental identifiability*. 17, 22, 27, 29, 56, 71, 181, 184, 231

random error Measurement error that in replicate measurements varies in a unpredictable manner. 58–60, 172, 175, 200, 238, 241

structural identifiability It is the parameter identifiability determined by the model itself, independently from the applied solicitation or signals during the experiments. For example, the identifiability of a parameter is greatly dependent from its influence on the model behaviour. It is impossible to identify inputs to which the model is not sensitive. Also correlations between model inputs may cause *equifinality* and therefore identifiability problems. 202, 203, 231

systematic error Measurement error that in replicate measurements remains constant or varies in a predictable manner. 58, 59, 62, 66

true value It can be defined as the result of a perfect measurement of a quantity, meaning that no errors occurred during the measurement. Besides virtual experiment were measurement errors can be actually avoided the actual *true value* of a quantity is unknown and it is usually replaced with the *conventional true value*, which is defined as value attributed to a particular quantity and accepted, as having an uncertainty appropriate for a given purpose ([57]). In this work, this term and more generally with the attribute *true* given specifications and highly reliable provided data. 59, 67, 107, 110, 111, 115, 118, 203, 243

Acronyms

AHU Air Handling Unit. 44

AIS Annealed Importance Sampling. 72, 77, 91, 101, 127, 218, 223–225, 240

AMG Adaptive Metropolis within Gibbs. 218–222

ARD Automatic Relevance Determination. 96, 98

BBRI Belgian Building Research Institute. 10–14, 119–121, 124–136, 138, 141–143, 209, 210, 227

BES Building Energy Simulation. 4, 14, 15, 17–21, 23–27, 29–33, 35, 36, 39–41, 49, 52, 53, 56, 58, 63, 72, 74, 76–79, 82–84, 86, 87, 89, 102, 151, 180, 196, 199, 200, 202, 205–209, 211, 212, 214, 216, 226–230, 236, 239, 246

CT Constant Temperature. 10, 13, 120, 129, 130, 132, 142, 149, 152, 171, 227

CTU Czech Technical University. 11, 13, 139–142

CVRMSE Coefficient of Variation Root Mean Squared Errors. 27, 28, 46–48, 111, 115, 116, 132, 135, 136, 145, 194, 205

EBC Energy in Buildings and Communities Programme. 31, 102, 119, 149

ESRU Energy System Research Unit. 149

FF Factor Fixing. 57, 63, 66–68, 171, 181, 183, 184, 186, 199, 201, 227

FFL Free Float. 10, 13, 120, 121, 127–130, 132, 149

FM Factor Mapping. 57, 63, 68, 171, 185, 186, 199, 201, 202

FP Factor Prioritising. 57, 63, 66–68, 171, 181, 183, 184, 186, 199, 201

FS Factor Screening. 12, 57, 62, 63, 68, 171, 178, 184, 186, 199, 201, 206

GCV Generalised Cross Validation criterion. 61, 62

GLUE Generalised Likelihood Uncertainty Estimation. 37–40, 50, 63, 68, 69, 71, 199, 206, 240, 245

GOF Goodness Of Fit. 27, 28, 48, 49, 102, 116, 121, 132, 135, 136, 145, 151, 156, 194, 205, 226

GP Gaussian Process. 72–75, 86, 89, 90, 93–96, 217, 246

GPR Gaussian Process Regression. 4, 14, 18, 26, 51, 52, 72–77, 83–86, 88, 94, 96, 199, 202, 207–209, 211, 226–228, 230, 236, 238

GSA Global Sensitivity Analysis. 4, 18, 24, 37, 39, 40, 71, 185, 202, 206

HLC Heat Loss Coefficient. 138, 139

hp Hyper Parameters. 75, 76, 84, 86, 89, 90, 94–98, 228, 238

IBP Fraunhofer Institute of Building Physics. 149, 152

IEA International Energy Agency. 31, 102, 119, 149

iid independently and identically distributed. 59, 217

KUL Katholieke Universiteit Leuven. 13, 119, 139–143, 145, 146

LECE Building Component Test Laboratory. 10, 11, 13, 14, 51, 120–123, 131, 135–138, 141, 142, 144–147, 202, 227

LHS Latin Hyper Cube Sampling. 48, 58, 76, 79, 238

LIF Least Important Factors. 12, 63, 66, 71, 179–185, 201

MAP Maximum A Posteriori. 13, 14, 50, 86, 90, 101, 107, 109–111, 113, 115, 116, 127, 128, 130–132, 135, 142, 144, 166, 167, 169, 188–192, 228, 230

MCMC Markov Chain Monte Carlo. 4, 52, 69, 72, 206, 207, 218, 221, 222, 224, 225

MIF Most Important Factors. 12, 63, 66, 164, 178, 180, 182, 201

NMBE Normalised Mean Bias Error. 27, 28, 46–48, 116, 132, 135, 136, 145, 194, 205

NZE Near Zero Energy. 36, 37

PCA Principal Component Analysis. 9, 24, 39, 53, 63–65, 68, 72, 78, 79, 82, 201, 208, 214–217

PRBS Pseudo Random Binary Sequance. 50

RMSE Root Mean Squared Error. 14, 226, 227

ROLBS Randomly Ordered Logarithmic Binary Sequence. 12, 14, 120–123, 131, 135, 136, 138, 141, 142, 145, 147–152, 154, 155, 157, 158, 163, 171–173, 180, 185, 187, 191, 192, 202, 248

SAN Simulate Annealing. 72, 76, 86, 218, 221–223, 240

SE Square Exponential function. 94

SM Spectral Mixture kernels. 95, 115, 236, 238

SSE Sum of Squared Errors. 61, 69, 96, 97

SVD Singular Value Decomposition. 216, 236, 238, 241

List of Mathematical Symbols

C Number of basis used in approximating a certain boundary condition. In particular $C^{(s)}$ is the number of basis vectors employed for approximating the $s-th$ boundary condition. 60, 79, 95, 239

D Number of basis vectors used for approximating the *Difference Vectors*. 80–83, 88, 89, 91, 98, 236, 239

EE Generalized elementary effect for multidimensional model outputs. 65, 241

I Generic integer. 95, 220, 222, 224, 225

J Generic integer. 221–224

L Number of elementary effects calculated for each model input. 64, 65, 240, 241

M^* Number of periods in which the field observations have been divided. 79, 81, 83, 88–93, 99, 239

$M^\star$ Generalisation of $\mu^\star$ for multidimensional model outputs. 11, 65, 66, 161–165, 178, 179, 201

M Total number of model simulation (i.e. after the division in periods). 64, 68, 70, 71, 79, 80, 82, 83, 85, 86, 92, 97, 99, 216, 238, 239

N Length of the simulation and observation vectors. 27, 60, 62, 64, 68, 69, 79, 80, 82–85, 87, 88, 97, 216

P Total number of model inputs, including variable boundary conditions and calibration parameters. 64, 83, 93, 94, 219, 220, 222, 224

$Q2.5\%$ 2.5 % quantile. 107, 111, 113, 116, 130, 132, 135, 142, 144, 189, 191

$Q50\%$ 50 % quantile. 107, 111, 113, 116, 130, 132, 135, 142, 144, 189, 191

$Q97.5\%$ 97.5 % quantile. 107, 111, 113, 116, 130, 132, 135, 142, 144, 189, 191

Q^* $Q + D$. 88

Q Number of basis vectors used for approximating the BES model output. 64, 65, 79–81, 83, 85, 86, 88–91, 97, 217, 236, 238, 239

R^2 Coefficient of determination. 14, 226, 227

R Number of considered calibration target variables, and, consequently, number of BES model outputs. 78–83, 91–93

SD Generalisation of ς for multidimensional model outputs. 65, 66

ST_i $\boldsymbol{z}_i$ total effects calculated with the Sobol Method. 14, 67, 71, 179, 183, 206

S_i $\boldsymbol{z}_i$ first order effects calculated with the Sobol Method. 14, 67, 71, 179, 183, 206

S Number of known variable boundary conditions influencing the observation of a real process. 78, 80, 81, 83, 95

T Number of calibration parameters. 78, 83

α Parameter controlling the power of the weighting function ($\omega(\cdot)$). 69, 70, 180

β Rate parameter or characteristic length parameter in the used covariance functions. In the *Calibration Model* $\beta_{p,q}^{(j)}$ (where $j = r$, l or k) refers to the β parameter of the $p-th$ input of the $q-th$ GPR model, relative to the $r-th$ model output. In the *Difference Model*, wherein SM kernels are adopted, $\beta_{i,c,d}^{(r,s)}$ refers to the β parameter of the $c-th$ coefficient of the $s-th$ boundary condition, used as input of the $d-th$ GPR model relative to the $r-th$ model output. 9, 94–99, 123, 236, 240

$\boldsymbol{A}$ Simmetric positive defined matrix, containing β parameters for the functions $\varrho_q(\cdot)$. 93, 94, 236

$\boldsymbol{B}$ Symmetric positive defined matrix containing β parameters for the functions $\varrho_q(\cdot)$ modelling the covariance between the two model outputs. $\boldsymbol{B}^{(l,k)} = \boldsymbol{A}_q^{(l)}(\boldsymbol{A}_q^{(l)} + \boldsymbol{A}_q^{(k)})^{-1}\boldsymbol{A}_q^{(k)}$. 93

$\boldsymbol{C}$ Generic symmetric positive defined matrix. 86, 89, 90, 92, 215

$\boldsymbol{D}$ Diagonal matrix containing the singular values from SVD decomposition. 216, 217

$\boldsymbol{G}$ Generic symmetric positive defined matrix. 89, 91, 92, 215

$\boldsymbol{H}$ Matrix having as columns the basis vectors $\boldsymbol{h}_d$. 80–82, 238

$\boldsymbol{I}$ Identity matrix. 62, 85, 88

$\boldsymbol{K}$ Matrix having as columns $\boldsymbol{k}_q$. 64, 79–82, 84–86, 97, 215, 217, 238

$\boldsymbol{R}$ Matrix quantifying the roughness of a function. 61, 237

$\boldsymbol{S} = (\eta\boldsymbol{R} + \boldsymbol{\Psi}^T\boldsymbol{\Lambda}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^T\boldsymbol{\Lambda}$. It projects a vector in the space spanned by the basis functions $(\psi_c(t))$ defining the columns of $\boldsymbol{\Psi}$. 61, 62

$\boldsymbol{U}$ Matrix having as rows $\boldsymbol{u}_{m^*}$, and columns $\boldsymbol{u}_d$. 81, 83, 89–91, 101

$\boldsymbol{V}$ Matrix having as rows $\boldsymbol{v}_m$, and columns $\boldsymbol{v}_c$. Similarly $\boldsymbol{V}^*$ is the matrix having as rows $\boldsymbol{v}_m^*$, and columns $\boldsymbol{v}_c^*$. The superscript (s) indicates that the coefficients in the matrix are relative to the $s-th$ boundary condition. When the superscript $\hat{}$ is present the matrix collects the estimates such coefficients. 79–81, 89–91, 93, 95

$\boldsymbol{W}$ Matrix having as rows $\boldsymbol{w}_m$, and columns $\boldsymbol{w}_q$. Similarly $\boldsymbol{W}^*$ is the matrix having as rows $\boldsymbol{w}_m^*$, and columns $\boldsymbol{w}_q^*$. The superscript (r) indicates that the coefficients in the matrix are relative to the $r-th$ model output or target variable. When the superscript $\hat{}$ is present the matrix collects the estimates such coefficients. 79–81, 83, 85, 86, 88–91, 101, 217

$\boldsymbol{X}$ Matrix having as columns the measurements of the boundary conditions. In particular $\boldsymbol{X}^{(s)}$ has as columns the measurements of the $s-th$ boundary condition only. 79

$\boldsymbol{Y}$ Matrix having as columns the realisation of the model output. $\boldsymbol{Y}^*$ is the matrix having as columns the observation of the real process (target variable). The superscript (r) indicates that only the $r-th$ model output or target variable are collected in the matrix. 68, 79–81, 85, 86, 88, 89, 100, 101, 136, 194, 216, 217

$\boldsymbol{Z}$ Input matrix for the *Training Model*. It is the matrix having as rows the vector $\boldsymbol{z}_m$. Similarly, $\boldsymbol{Z}^*$ is the input matrix of the *Calibration Model*, having as rows the vectors $\boldsymbol{z}_{m^*}^*$. 74, 75, 83, 85, 86, 88–90, 92, 93

$\boldsymbol{\Delta}$ Matrix having as columns the *Difference Vectors* $\boldsymbol{\delta}_{m^*}$. The superscript (r) indicated that $\boldsymbol{\Delta}$ collects only the *Difference Vectors* for the $r-th$ model output $(\boldsymbol{\delta}_{m^*}^{(r)})$. 82

$\boldsymbol{\Lambda}$ Precision matrix. 60–62, 237

$\boldsymbol{\Omega}$ Matrix having as columns orthonormal vectors from SVD decomposition. 216, 217

$\boldsymbol{\Phi} = [\boldsymbol{K}, \boldsymbol{H}]$. 80, 81, 88

$\boldsymbol{\Psi}$ Matrix having as columns the basis vectors used to expand a certain boundary condition. In particular, $\boldsymbol{\Psi}^{(s)}$ collects only the basis vectors relative to the $s-th$ boundary condition ($\boldsymbol{\psi}_c^{(s)}$). 60–62, 79, 80, 237

$\boldsymbol{\Sigma}$ Covariance matrix. 68, 74, 90, 91, 216

$\boldsymbol{\Xi}$ Matrix having as columns orthonormal vectors from SVD decomposition. 216, 217

$\boldsymbol{\delta}$ *Difference Vectors*. In particular, $\boldsymbol{\delta}_{m^*}^{(r)}$ is the $m^* - th$ *Difference Vector* relative to the $r-th$ model output. 82, 87, 237, 238

$\boldsymbol{\lambda}$ Vector of precision parameters. 88–91, 99–101

$\boldsymbol{\nu}$ Vector containing the hp of $\varphi_d(\cdot)$. 90, 91, 99–101

$\boldsymbol{\omega}$ Vector collecting the M weights (ω_m) from the function $\omega(\cdot)$. 71, 182, 183

$\boldsymbol{\psi}$ Basis vectors used for approximating a certain boundary condition. In particular $\boldsymbol{\psi}_c^s$ is the $c-th$ basis vector for the $s-th$ boundary condition. 60, 79–81, 238

$\boldsymbol{\theta}$ Calibration parameters. In particular $\boldsymbol{\theta}_m$ are the vectors of known calibration parameters resulting form sampling the input parameter space with LHS. $\boldsymbol{\theta}^*$ is the vector of unknown calibration parameters characterising the observed processes. 79, 83, 89–91, 99, 101

$\boldsymbol{\varepsilon}$ Vector indicating noise or *random errors*. 59, 60, 64, 79–81, 84, 87, 217

$\boldsymbol{\vartheta}$ Vector containing the hp of the Q $\varrho_q(\cdot)$. 86, 89, 90, 99

$\boldsymbol{a}$ Vector of mean parameters in SM kernels. In particular $\boldsymbol{a}_q^{(r)}$ are the mean parameters of the $r-th$ model output for the $q-th$ GPR model, and $\boldsymbol{a}^{(l,k}$ is the difference between $\boldsymbol{a}_q^{(l)}$ and $\boldsymbol{a}_q^{(k)}$. 93, 94, 241

$\boldsymbol{b} = [(\boldsymbol{w}_{m^*}^*)^T, (\boldsymbol{u}_{m^*}^*)^T]^T$. If the superscript (r) is present the coefficients are relative to the $r-th$ model output and target variable only. The superscript $\hat{}$ indicates that the vector elements are the coefficient estimates. 80, 81, 88

$\boldsymbol{e}_p$ Vector having all its elements zero except the $p-th$ one which is equal to 1. 64

$\boldsymbol{h}$ Basis vectors used for approximating the *Difference Vectors*. $\boldsymbol{h}_d$ is the $d-th$ basis vector. 80–82, 89, 92, 98, 237, 239

$\boldsymbol{k}$ Basis vectors used for approximating the model outputs. $\boldsymbol{k}_q$ is the $q-th$ basis vector. 64, 65, 79, 80, 82, 83, 86, 89, 92, 217, 228, 237, 239

$\boldsymbol{s}$ Vector representing a series of systematic errors. 59

$\boldsymbol{u}$ Vector of u coefficients. In particular, $\boldsymbol{u}_{m^*} = [u_{m^*,1}, \ldots, u_{m^*,d}, \ldots, u_{m^*,D}]^T$ and $\boldsymbol{u}_d = [u_{1,d}, \ldots, u_{m^*,d}, \ldots, u_{M^*,d}]^T$. When the superscript (r) is present the vectors are relative to the $r-th$ model output only. The superscript $\hat{}$ indicates that the vector collects estimates. 80–83, 89–92, 237, 238

$\boldsymbol{v}$ Vector of v coefficients. In particular, $\boldsymbol{v}_m = [v_{m,1}, \ldots, v_{m,c}, \ldots, v_{m,C}]^T$, and $\boldsymbol{v}_c = [v_{1,c}, \ldots, v_{m,c}, \ldots, v_{M,c}]^T$. When the superscript (s) is present the vectors are relative to the $s-th$ boundary condition only. The superscript $\hat{}$ indicates that the vector collects estimates. 60–62, 79–81, 83, 95, 237

$\boldsymbol{w}$ Vector of w coefficients. In particular, $\boldsymbol{w}_m = [w_{m,1}, \ldots, w_{m,q}, \ldots, w_{m,Q}]^T$, $\boldsymbol{w}_q = [w_{1,q}, \ldots, w_{m,q}, \ldots, w_{M,q}]^T$, $\boldsymbol{w}^*_m = [w^*_{m,1}, \ldots, w^*_{m,q}, \ldots, w^*_{m,Q}]^T$ and $\boldsymbol{w}^*_q = [w^*_{1,q}, \ldots, w^*_{m,q}, \ldots, w^*_{M,q}]^T$. If the superscript (r) is present the vectors are relative to the $r-th$ model output and target variable only. If the superscript $\hat{}$ is present the vectors contain coefficient estimates. 64, 79–81, 83–86, 89–92, 97, 215, 217, 237, 238

$\boldsymbol{x}$ Variable boundary conditions influencing the real experiment, and imposed on the BES model. In particular, $\boldsymbol{x}^{(s)}$ is the $s-th$ boundary condition.. 59–62, 79–81, 87

$\boldsymbol{y}$ Model outputs and target variables. In particular, $\boldsymbol{y}^{(r)}_m$ is the $m-th$ realisation of the $r-th$ model output, and $\boldsymbol{y}^{(r)*}_{m^*}$ is the m^*-th observation of the $r-th$ target variable . 64, 66, 67, 69, 70, 74, 75, 79–82, 84, 85, 87, 88, 96, 97, 116, 132, 136, 194, 215, 217, 226, 239

$\boldsymbol{z}_i$ $i-th$ model input or set of inputs. 66, 67, 236, 239

$\boldsymbol{z}_{-i}$ complement of $\boldsymbol{z}_i$. 66, 67

$\boldsymbol{z}$ Model input vectors. In particular, $\boldsymbol{z}_m$ is the input vector for the $m-th$ simulation, and $\boldsymbol{z}^*_{m^*}$ are the model input vectors for the field observations. 64, 69, 70, 74, 75, 83–85, 87–89, 93, 94, 219–226, 237, 240, 242, 245

$\boldsymbol{res}$ Vector of residuals. 14, 226, 227

$\boldsymbol{rt}$ Vector containing *rt* parameters. 88, 90, 91

$\boldsymbol{shp}$ Vector containing *shp* parameters. 88, 90, 91

ϵ Arbitrary small real positive number.. 222

η Parameter controlling the power of the smoothing in the Smoothing with Roughness Penalty framework. 61, 62, 237

γ Parameters of $\varrho_q(\cdot)$. $\gamma_{p,q}^{(l,k)} = \frac{\beta_{p,q}^{(l)}\beta_{p,q}^{(k)}}{\beta_{p,q}^{(l)}+\beta_{p,q}^{(k)}}$. 94

λ Precision parameters for the *Training Model*. The superscript (r) identifies the precision parameter for the $r-th$ model output. 84–90, 92, 97, 99

$\mathcal{B}$ Bayes Factor. In particular, $\mathcal{B}_{j,i}$ is the Bayes factor between $\mathcal{M}_j$ and $\mathcal{M}_i$. 100, 101, 131, 136, 195

$\mathcal{M}$ Model. In particular $\mathcal{M}_k$ is the $k-th$ analysed model. 10–12, 14, 86, 87, 89–91, 98, 100, 101, 121–127, 131–138, 142–147, 149, 150, 156–172, 176, 179–181, 183–195, 197, 198, 209, 210, 227, 240

$\mathcal{T}$ Temperature parameters in the SAN and AIS algorithms. In particular, $\mathcal{T}_i$ is the $i-th$ temperature. 221, 222, 224, 225

$\mu^\star$ Empirical means of the absolute values of the *ee*. In particular, $\mu_p^\star$ is the empirical mean of the *L ee* relative to the $p-th$ model input. 64–66, 235

ω Importance weights. In particular, in GLUE ω_m is the weight attributed to the $m-th$ model simulation. 70, 71, 180, 184, 223, 224, 238

ϕ Frequency parameter in SM covariance functions. In particular, $\phi_{i,c,d}^{(r,s)}$ is the frequency parameter of $v_{c,d}^{(s)*}$ in the $i-th$ component of the covariance function $\varphi_d^{(r,r)}(\cdot)$. 95, 96, 98, 123

σ Additive variance parameters in SM covariance function. In particular $\sigma_d^{(r)}$ is the additive variance parameter of the covariance function $\varphi_d^{(r,r)}(\cdot)$. 95, 96, 98, 111

τ Amplitude or marginal variance parameters in the adopted covariance functions. 93–96, 98

ε Unidimensional *random error*. 175

ς Empirical standard deviations of the *ee*. In particular, ς_p is the empirical standard deviation of the L *ee* relative to the $p-th$ model input. 64–66, 236

ζ Applied variations in the calculation of *ee* (or *EE*). In particular, $\zeta_{p,l}$ is the applie variation in the calculation of the $l-th$ elementary effect relative to the $p-th$ model parameter. 64, 65

a Elements of the vector $\boldsymbol{a}$. In particular, $a_q^{(l,k)} = a_q^{(l)} - a_q^{(k)}$ describe constant offset between the $l-th$ and the $r-th$ model ouputs and target variables. 94, 96, 98, 241

d Singular values from SVD decomposition. In particular d_q is the $q-th$ singular value. 216

ee Elementary effects calculated with the Morris Method. In particular, $ee_{l,p}$ is the $l-th$ elementary effect of the $p-th$ model input. 63, 64, 240, 241

rt Rate parameter of the Gamma distribution. 84–88, 97, 240

se Empirical standard errors. In particular, se_i is the $i-th$ estimated empirical standard error. 60, 61

shp Shape parameter of the Gamma distribution. 84–88, 96, 97, 240

s Unidimensional systematic error. 175

u Coefficients of the basis vector used for approximating the *Difference Vectors*. In particulr, $u_{m^*,d}$ is the coefficient of the $d-th$ basis vector for the m^*-th *Difference Vector*. 80, 239

v Coefficients of the basis vectors used for approximating a certain boundary condition. In particular, $v_{m,c}^{(s)}$ is the coefficient of the $c-th$ basis for the $s-th$ boundary condition reltive to the $m-th$ model simulation. Similarly it is possible to define $v_{m^*,c}^{(s)*}$. 60, 79, 80, 95, 239, 240

w Coefficients of the basis vectors used for approximating the model outputs. In particular, $w_{m,q}^{(r)}$ is the coefficient of the $q-th$ basis vector for the $m-th$

realisation of the $r-th$ model output. Similarly it is possible to define $w^{(r)*}_{m^*,q}$. 64, 65, 79, 80, 217, 239

y One point of a certain model simulation output or measured target variable. In particular, $y_{m,i}$ is the $i-th$ point of the $m-th$ model simulation, and y^*_i is the $i-th$ measurement of the considered target variable. 27, 69

z Elements of the vector $\boldsymbol{z}$. In particular, $z_{g,p}$ is the $p-th$ element of the $g-th$ input vector (where $g = l$ or k). 94, 219–222, 224, 225

List of Mathematical Superscripts

$*$ Refers to the measurements of a target variable. 27, 69, 70, 75, 80–83, 87–93, 95, 99–101, 116, 132, 136, 194, 237–242

$\bar{}$ Indicates empirical means. 64, 65

$\hat{}$ Indicates estimates. 61, 69, 70, 237–239

$\prime$ Refers to posterior or updated parameters or probability density distributions. 69, 70, 85, 86, 88, 90, 91, 97

$\tilde{}$ Indicates *true value*. 59

j Generic superscript. 94, 96, 222, 236

k Generic superscript. 92–94, 96, 98, 219, 220, 222, 224, 236, 238, 240, 241, 245

l Generic superscript. 92–94, 96, 98, 219, 220, 222, 224, 236, 238, 240, 241, 245

r Refers to the $r-th$ model output or target variable. 79–83, 88, 92, 94–96, 98, 111, 236–242

s Refers to the $s-th$ boundary condition. 79–81, 83, 87, 95, 96, 98, 235–241

List of Mathematical Subscripts

List of Mathematical Operators

$T(\cdot)$ Transition probability for a Markov chain. In particular $T(\boldsymbol{z}^{(l)} \rightarrow \boldsymbol{z}_p^{(k)})$ is the probability that the chain goes from state $\boldsymbol{z}^{(l)}$ to state $\boldsymbol{z}^{(k)}$. 219

$\Delta(\cdot)$ Function representing the *Difference Model*. 87, 88

$\mathcal{E}xp(\cdot)$ Operator indicating an Exponential probability density distribution according to the given arguments. 96

$\mathcal{N}(\cdot)$ Operator indicating a Normal unidimensional or multidimensional probability density distribution according to the given arguments. 60, 161, 167, 169, 176, 189, 191

$\mathcal{U}(\cdot)$ Operator indicating a Uniform unidimensional probability density distribution according to the given arguments. 106, 122, 167, 169, 176, 189, 191

$\omega(\cdot)$ Weighting function used in GLUE. 12, 69, 70, 183, 185, 236, 238

$\mathrm{E}(\cdot)$ Estimate operator. 67

$\mathrm{V}(\cdot)$ Variance operator. 67, 68, 71

$\mathrm{cov}(\cdot)$ Covariance operator. 69, 70

$\mathrm{diag}(\cdot)$ Operator creating a diagonal (or block diagonal) matrix according to the given arguments. 60

$\mathrm{d}(\cdot)$ Operator returning the Euclidean distance between two points or lines. 65

$\mathrm{tr}(\cdot)$ Trace operator. 62, 68

$f(\cdot)$ Function representing a certain computer model. 64, 69, 84, 87, 88

$g(\cdot)$ Generic function. 221, 224

$h(\cdot)$ Generic function. 223

$p(\cdot)$ Probability density function. 84, 85, 87, 88, 218, 219

$\mathrm{df}(\cdot)$ Degree of freedom operator. 62

$\max(\cdot)$ Maximum operator. 221

$\min(\cdot)$ Minimum operator. 219, 220

$\mathrm{sd}(\cdot)$ Standard deviation operator. 219, 221, 224

$Q(x, q)$ Quantile function. It returns the q quantile of the sample x. 226

$\Gamma(\cdot)$ Gamma function. 84, 85, 87

$\mathcal{G}(\cdot)$ Operator indicating a Gamma probability density distribution according to the given arguments. 84, 85, 87, 88

$\mu(\cdot)$ Mean function of a certain GP. Returns the mean vector of a GP according to the given arguments. 74

$\psi_c(t)$ Function depicting a basis expanding a certain boundary condition. 60, 61, 237

$\varphi_d(\cdot)$ $q-th$ covariance function used to built the *Difference Model*. 90, 97, 98, 238

$\varrho_q(\cdot)$ $q-th$ covariance function used to built the *Training Model*. 86, 94, 236, 238, 240

$cf(\cdot)$ Covariance function of a certain GP. Returns the covariance matrix a GP according to its inputs and hyper parameters. 74, 75

$x(t)$ Function depicting a measured variable condition input for the BES model. 59–61

List of Modelling Symbols

A Area. 177

$CBB1$ Macro-zone comprising *corridor*, *bathroom*, and *bedroom2*. 165, 166, 190, 193, 194, 197, 198, 204

$CT1$ Constant temperature phase in EXPERIMENT1. 11, 12, 14, 154, 156–158, 160–164, 166–168, 170, 187–190, 192–195, 197, 227

C_d Discharge coefficient. 177

EPS Insulation material in the external walls of Twin House N2 and Twin House O5 (Figure 6.2). In particular, $EPS_{\boldsymbol{k}}$ is the relative conductivity. 163, 167, 169, 189, 191

FB Glass fibre boards constituting the internal and external layers of wall analysed in Section 5.1. In particular, $FB_{\boldsymbol{c}}$ is the relative specific heat. 9, 103, 106–108, 110–117

GC Gas concrete blocks constituting the core of the wall component analysed in Section 5.1. In particular, $GC_{\boldsymbol{k}}$ and $GC_{\boldsymbol{c}}$ are the relative conductivity and specific heat respectively. 103, 104, 106–108, 110–117

$KLB2$ Macro-zone grouping *kitchen*, *lobby* and *bedroom2*. 165, 166, 190, 193, 194, 197, 198, 204

$PS040$ Insulation material used in the ceiling of Twin House N2 and Twin House O5 (Figure 6.2). In particular, $PS040_{\boldsymbol{k}}$ is the relative conductivity. 167, 169, 189, 191

$PUR2$ Insulation material in the floor of Twin House N2 and Twin House O5 (Figure 6.2) In particular, $PUR2_{\boldsymbol{k}}$ is the relative conductivity. 162, 167, 169, 189, 191

PUR Insulation material in the floor of Twin House N2 and Twin House O5 (Figure 6.2). In particular, $PUR_{\boldsymbol{k}}$ is the relative conductivity. 162, 167, 169, 189, 191

PU Insulation material in the external walls of Twin House N2 and Twin House O5 (Figure 6.2) In particular, $PU_{\boldsymbol{k}}$ is the relative conductivity. 163, 167, 169, 189, 191

Qm Mass flow rate. 177

$Qvnt$ Mechanical ventilation flow rate. 175, 176, 178, 179, 183, 185, 188–191, 195

$ROLBS1$ ROLBS phase in EXPERIMENT1. 11, 12, 14, 154, 156, 157, 159–161, 164–166, 169–171, 187, 188, 193–195, 198, 227

$ROLBS2$ ROLBS phase in EXPERIMENT2. 155, 160, 171

XPS Insulation layer employed in the Test Box envelope (Table 5.15). In particular, $XPS_{\boldsymbol{k}}$ and $XPS_{\boldsymbol{\rho}}$ are the relative conductivity and density respectively. 139, 143–145

ΔP Pressure difference. 177

$\boldsymbol{R}$ Thermal resistance. 122, 123, 128, 130–132, 134, 135, 137, 140, 162, 166, 167, 169, 189, 191, 251, 252

$\boldsymbol{T} = [living\ room_{\boldsymbol{Ti}}, bedroom1_{\boldsymbol{Ti}}, bathroom_{\boldsymbol{Ti}}]$. 11, 14, 172, 179

$\boldsymbol{\alpha}$ Absorptivity. 122, 123, 128, 130, 132, 134, 135, 137, 167, 189, 252

$\boldsymbol{\rho}$ Density. 10, 105, 122, 123, 127–130, 132, 134, 135, 137, 140, 142, 144, 145, 161, 163, 167–169, 177, 189, 191, 248, 250–252

$\boldsymbol{c}$ Specific heat. 9, 103, 104, 106–108, 110–117, 161, 163, 167, 169, 189, 191, 247, 250, 251

$\boldsymbol{e}$ Emissivity. 122, 124, 132, 134, 135, 137

$\boldsymbol{k}$ Conductivity. 10, 103, 105–108, 110–117, 122, 123, 127–132, 134, 135, 137, 140–142, 144, 145, 162, 163, 166, 167, 169, 189, 191, 247, 248, 250–252

C/R Ratio between convective and radiative sensible heat gains from a certain heat source. In calibration experiments only the convective fraction is considered as parameter since the radiative part can always be derived by difference. 98, 122, 123, 135, 153, 158, 173, 176, 178–181, 183, 185, 189, 191

Ds Diffuse solar radiation. 122, 123, 154

Gh Global horizontal solar radiation. 122, 123, 154

Gv Global vertical solar radiation. If not specified is referred to the plane perpendicular to the south direction. 122, 123

Rh Relative Humidity. 98, 122, 123, 154

Te External temperature. 98, 113, 115, 122, 125, 139, 154, 155, 158, 176

Ti Internal temperature. 139, 154, 155, 158, 172, 176, 248

Tvnt Mechanically supplied air temperature in EXPERIMENT1 and EXPERIMENT2 (Chapter 6). 154, 155, 158, 176

Wd Wind direction. 98, 122, 123, 125, 154, 155, 158, 176

Ws Wind speed. 122, 123, 125, 154, 155, 158, 176, 178, 179, 183, 186

inf Infiltration. 167–169

len Length. 122, 126, 131, 132, 134–136, 138, 176, 178, 182, 183, 185, 188, 189, 191, 193, 251, 252

shl Sensible heat loads necessary to keep the prescribed internal temperature set points. 154, 155

tr 90^o optical transmission factor of glazing components. 122, 123, 128, 130, 132, 134, 135, 137, 251

width Width. 176, 182, 183, 185

$\mathcal{F}$ Factor multiplying a set of pressure coefficient values (Section 6.6). In particular, $node_{\mathcal{F}}$ (where *node* can be NORTH, EAST, SOUTH or WEST) is the factor multiplying the pressure coefficient set of *node*. 187–189, 191–193

attic Thermal zone representing the attic of Twin House N2 and Twin House O5. 154, 155, 159, 176, 190

ausgleich Levelling fill used in the floor of Twin House N2 and Twin House O5 (Figure 6.2) In particular, $ausgleich_{\boldsymbol{k}}$ is the relative conductivity. 167, 189

basement Thermal zone representing the basement of Twin House N2 and Twin House O5. 154, 155, 158, 159, 176, 190

bathroom Thermal zone representing the bathroom of Twin House N2 and Twin House O5. 11, 152, 154, 155, 157, 159, 162, 163, 165, 167, 168, 172, 176, 178, 180, 181, 183, 188–192, 195, 247, 248, 253

*bedroom*1 Thermal zone representing the bedroom 1 of Twin House N2 and Twin House O5. 11, 152, 154, 155, 157, 159, 162–164, 167–169, 172, 176, 178, 180, 181, 183, 185, 188, 189, 191–193, 195, 248, 253

*bedroom*2 Thermal zone representing the bedroom 2 of Twin House N2 and Twin House O5. 11, 152, 157, 162, 163, 165, 167–169, 176, 189, 191–193, 247, 253

blinds Material used to represent the roller blinds of Twin House N2 and Twin House O5. In particular, $blinds_{\boldsymbol{k}}$ is the relative conductivity. 162, 166, 167, 169, 189, 191

brickIntWall Material used to represent bricks in the internal walls of of Twin House N2 and Twin House O5 (Figure 6.2) In particular, $brickIntWall_{\boldsymbol{k}}$, $brickIntWall_{\boldsymbol{\rho}}$ and $brickIntWall_{\boldsymbol{c}}$ are the relative conductivity, density and specific heat respectively. 163, 167, 169, 189, 191

brick Material used to represent bricks in the external walls of of Twin House N2 and Twin House O5 (Figure 6.2). In particular, $brick_{\boldsymbol{k}}$, $brick_{\boldsymbol{c}}$ and $brick_{\boldsymbol{\rho}}$ are the relative conductivity, specific heat ans density respectively. 162, 163, 167–169, 189, 191

concrete Structural material used in the floor and ceiling of Twin House N2 and Twin House O5 (Figure 6.2). In particular, $concrete_{\boldsymbol{\rho}}$ and $concrete_{\boldsymbol{c}}$ are the relative density and specific heat. 161, 163, 167, 169, 189, 191

corridor Thermal zone representing the corridor of Twin House N2 and Twin House O5. 11, 152, 157, 159, 165, 176, 247, 253

crack Indicates a crack component of the airflow network model in Chapter 6. The different cracks are distinguished by applying an numeric index at the end (Figure 6.5). 176, 178, 182, 183, 185, 188–191, 193

estrichScreed Levelling material used in the floor and ceiling of Twin House N2 and Twin House O5 (Figure 6.2). In particular, $estrichScreed_{\boldsymbol{\rho}}$ and $estrichScreed_{\boldsymbol{c}}$ are the relative density and specific heat. 163, 169, 191

$fibre\_cement\_board1$ Internal fibre cement board layer of the envelope of the test Box envelope (Table 5.15). In particular, $fibre\_cement\_board1_{\boldsymbol{k}}$ and $fibre\_cement\_board1_{\boldsymbol{\rho}}$ are the relative conductivity and density respectively. 139, 143–145

$fibre\_cement\_board2$ Fibre cement board layer of the envelope of the test box (Table 5.15). In particular, $fibre\_cement\_board2_{\boldsymbol{k}}$ and $fibre\_cement\_board2_{\boldsymbol{\rho}}$ are the relative conductivity and density respectively. 139, 144, 145

$fibre\_cement\_board3$ External fibre cement board layer of the envelope of the test box envelope (Table 5.15). In particular, $fibre\_cement\_board3_{\boldsymbol{k}}$ and $fibre\_cement\_board3_{\boldsymbol{\rho}}$ are the relative conductivity and density respectively. 139, 144

$floor\_crack$ Component used for representing a crack in the floor of the test box in Section 5.2. In particular, $floor\_crack_{\boldsymbol{len}}$ is the relative length. 122, 126, 132, 134, 135, 138

glass Construction component used in representing the window glass of the test box in Section 5.2. In particular, $glass_{\boldsymbol{tr}}$ is the relative 90^o optical transmission factor. 122, 123, 128, 130, 132, 134, 135, 137

glazingBlinds Material used to represent the glazing components of the openings with blinds of Twin House N2 and Twin House O5. In particular, $glazingBlinds_{\boldsymbol{R}}$ is the relative thermal resistance. 162, 166, 167, 169, 189, 191

glazing Material used to represent the glazing components of the openings with no blinds of Twin House N2 and Twin House O5. In particular, $glazing_{\boldsymbol{R}}$ is the relative thermal resistance. 162, 167, 169, 189, 191

insulCeilBase Insulation material in the ceiling of Twin House N2 and Twin House O5 (Figure 6.2). 162, 167, 169, 189, 191

kitchen Thermal zone representing the kitchen of Twin House N2 and Twin House O5. 11, 98, 152, 154, 157, 158, 162, 164, 165, 167–169, 176, 189, 191, 192, 247, 253

living room Thermal zone representing the living room of Twin House N2 and Twin House O5. 11, 152, 154, 155, 157–159, 162–165, 167, 169, 172, 176, 178–181, 183, 185, 188–195, 197, 198, 204, 248, 253

lobby Thermal zone representing the lobby of Twin House N2 and Twin House O5. 11, 152, 157, 162, 164, 165, 167, 168, 176, 189, 191, 192, 247, 253

minWool Insulation material in the external walls of Twin House N2 and Twin House O5 (Figure 6.2). In particular, $minWool_{\boldsymbol{k}}$ is the relative conductivity. 163, 167, 169, 189, 191

pc Pressure coefficients relative to wind induced pressure boundary nodes in air flow network models. In particular pc_{angle} is the pressure coefficient for the direction indicated by *angle* which is the azimuth in degrees from the reference direction (north). The notation $node_{pc_{angle}}$ refers pc_{angle} of the boundary node indicated by *node* (*node* can be NORTH, EAST, SOUTH, WEST (Figure 6.5)). 176, 178, 181–183, 185

softwood Material used to represent the window frames of Twin House N2 and Twin House O5. In particular, $softwood_{\boldsymbol{k}}$ is the relative conductivity. 167, 189

tb Thermal bridges. In particular, $zone_{tb}$ refers to the thermal bridges of a certain thermal zone. 189, 191, 192

wall Construction component used in representing the walls of the test box in Section 5.2. In particular, $wall_{\boldsymbol{k}}$ and $wall_{\boldsymbol{\rho}}$ are the relative conductivity and density. 10, 122–124, 127–132, 134, 135, 137, 140–142

window_crack Component used for representing a crack in the window frame of the test box in Section 5.2. In particular $window\_crack_{\boldsymbol{len}}$ is the relative length. 122, 126, 131, 132, 134–136, 138

window Construction component used in representing the window of the test box in Section 5.2. In particular, $window_{\boldsymbol{R}}$ are the relative thermal resistance. 122, 123, 128, 130–132, 134, 135, 137, 140

wood Material used to represent the doors of Twin House N2 and Twin House O5. In particular, $wood_{\boldsymbol{\alpha}}$ and $wood_{\boldsymbol{k}}$ are the relative absorptivity and conductivity respectively. 162, 167, 169, 189, 191

CRACK Macro-parameter representing a group of crack lengths (Section 6.5). 178, 179, 183

EAST Wind induced pressure boundary node at the east walls of Twin House N2 and Twin House O5 (Figure 6.5). 176, 178, 181–183, 185, 187, 189, 191, 249, 252

EXPERIMENT1 First experiment involving Twin House N2 and Twin House O5 (Chapter 6). 149–152, 155, 156, 247–249, 253

EXPERIMENT2 Second experiment involving Twin House N2 and Twin House O5 (Chapter 6). 14, 59, 149, 150, 152, 155, 156, 175, 248, 249, 253

NORTH Wind induced pressure boundary node at the north walls of Twin House N2 and Twin House O5 (Figure 6.5). 178, 187–189, 191, 193, 249, 252

NORTH_ZONE Macro-zone including the non-sealed zones (*living room*, *corridor*, *bedroom1* and *bathroom*) in Twin House N2 and Twin House O5, in EXPERIMENT1 and EXPERIMENT2 (Chapter 6). 155, 157, 159, 177

PC Macro-parameter representing a group of pressure coefficients (Section 6.5). 178–183

SOUTH Wind induced pressure boundary node at the south walls of Twin House N2 and Twin House O5 (Figure 6.5). 176, 178, 181–183, 185, 187, 189, 191, 249, 252

SOUTH_ZONE Macro-zone including the sealed zones (*kitchen*, *lobby* and *bedroom2*) in Twin House N2 and Twin House O5, in EXPERIMENT1 and EXPERIMENT2 (Chapter 6). 154, 155, 157, 159, 177

WEST Wind induced pressure boundary nodes at the west walls of Twin House N2 and Twin House O5 (Figure 6.5). 176, 178, 182, 183, 185, 187–189, 191–193, 249, 252

References

[1] ASHRAE. Ashrae guide 14-2002: Measurement of energy and demand savings, 2002.

[2] P. Aude, L. Tabary, and P. Depecker. Sensitivity analysis and validation of buildings' thermal models using adjoint-code method. *Energy and Buildings*, 31(3):267–283, Apr 2000.

[3] Peder Bacher and Henrik Madsen. Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings*, 43(7):1511–1522, Jul 2011.

[4] P. Bannister. Why good buildings go bad while some are just born that way. *Ecolibrium*, 2009.

[5] M. J. Bayarri, J. O. Berger, J. Cafeo, G. Garcia-Donato, F. Liu, J. Palomo, R. J. Parthasarathy, R. Paulo, J. Sacks, and D. Walsh. Computer model validation with functional output. *The Annals of Statistics*, 35(5):1874–1906, Oct 2007.

[6] M. J. Bayarri, J. O. Berger, D. Higdon, M. C. Kennedy, A. Kottas, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C. H. Lin, and J. Tu. A framework for validation of computer models. Technical report, National Institute of Statistical Sciences, 19 T. W. Alexander Drive, PO Box 14006, Research Triangle Park, NC 27709-4006, www.niss.org, 2005.

[7] M. J. Bayarri, J. O. Berger, M. C. Kennedy, A. Kottas, R. Paulo, J. Sacks, J. A. Cafeo, C. H. Lin, and J. Tu. Bayesian validation of a computer model for vehicle collision. Technical report, National Institute of Statistical Sciences 19 T. W. Alexander Drive PO Box 14006 Research Triangle Park, NC 27709-4006 www.niss.org, 2005.

[8] K. Beven and A. Binley. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3):279–298, Jul 1992.

[9] K. Beven, J. Freer, B. B Hankin, and K. Schultz. The use of generalised likelihood measures for uncertainty estimation in high order models of environmental systems. In Fitzgerald W. J., Smith R. L., Walden A. T., and Young P. C., editors, *Nonlinear and Nonstationary Signal Processing*. Cambridge University Press, 2000.

[10] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[11] A.T. Booth and R. Choudhary. Decision making under uncertainty in the retrofit analysis of the uk housing stock: Implications for the green deal. *Energy and Buildings*, 64:292–308, Sep 2013.

[12] B. Bordass, R. Cohen, M. Standeven, and A. Leaman. Assessing building performance in use 3: energy performance of the probe buildings. *Building Research & Information*, 29:114–128, 2001.

[13] Phillip Boyle and Marcus Frean. Dependent gaussian processes. In *In Advances in Neural Information Processing Systems 17*, pages 217–224. MIT Press, 2005.

[14] H. Brohus, P. Heiselberg, A. Hesselholt, and H. Rasmussen. Application of partial safety factors in building energy performance assessment. In *Building Performance Simulation Association (IBPSA) conference*, pages 27–30, 2009.

[15] S. Bucking, R. Zmeureanu, and A. Athienitis. A methodology for identifying the influence of design variations on building energy performance. *Journal of Building Performance Simulation*, 7(6):411–426, Dec 2013.

[16] E. Burman, D. Rigamonti, J. Kimpain, and D. Mumovic. Performance gap and thermal modelling: A comparison of simulation results and actual energy performance for an academy in north west england. In *First Building Simulation and Optimization Conference Loughborough, UK*, 2012.

[17] K. Campbell, M.D. McKay, and B.J. Williams. Sensitivity analysis when model outputs are functions. *Reliability Engineering & System Safety*, 91(10-11):1468–1472, Oct 2006.

[18] F. Campolongo and R. Braddock. The use of graph theory in the sensitivity analysis of the model output: a second order screening method. *Reliability Engineering & System Safety*, 64(1):1–12, 1999.

[19] F. Campolongo, J. Cariboni, and A. Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10):1509–1518, October 2007.

[20] S. Cho and J. Haberl. Validation of the ecalc commercial code-compliant simulation versus measured data from an office building in a hot and humid climate. In *Proceedings of the 16th Symposium on Improving Building Systems in Hot and Humid Climates, Texas A&M University, Dallas, Texas, published on CD ROM (December)*, 2008.

[21] S. Cho and J.S. Haberl. Development of a simulation toolkit for the selection of high performance systems for office buildings in hot and humid climates. In *Third National Conference of IBPSA-USA Berkeley, California*, 2008.

[22] J. Cipriano, G. Mor, D. Chemisana, D. Pérez, G. Gamboa, and X. Cipriano. Evaluation of a multi-stage guided search approach for the calibration of building energy simulation models. *Energy and Buildings*, 87(0):370 – 385, 2015.

[23] J Clarke. *Energy Simulation in Building Design*. Butterworth-Heinemann, 2001.

[24] J A Clarke, P A Strachan, and C Pernot. An approach to the calibration of building energy simulation models. *ASHRAE Trans*, pages 917–927, 1993.

[25] J.A. Clarke and J.L.M. Hensen. Integrated building performance simulation: Progress, prospects and requirements. *Building and Environment*, 91:294–306, Sep 2015.

[26] D. Coakley, Raftery P., and Molloy P. Calibration of whole building energy simulation models: detailed case study of a naturally ventilated building using hourly measured data. In *First Building and Optimisation Conference, Loughborough, UK, 10 - 11 September*, 2012.

[27] D. Coakley, P. Raftery, P. Molloy, and G White. Calibration of a detailed bes model to measured data using an evidence-based analytical optimisation approach. In *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association, Sydney*, 14-16 November 2011.

[28] Daniel Coakley, Paul Raftery, and Marcus Keane. A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews*, 37(0):123 – 141, 2014.

[29] V. Corrado and H. E. Mechri. Uncertainty and sensitivity analysis for building energy rating. *Journal of Building Physics*, 33(2):125–156, June 2009.

[30] D. Costola, B. Blocken, and J.L.M. Hensen. Overview of pressure coefficient data in building energy simulation and airflow network programs. *Building and Environment*, 44(10):2027–2036, Oct 2009.

[31] P. Craven and G. Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4):377–403, 1978.

[32] Lehel Csató. Sparse on-line gaussian processes. *Neural Computation*, 2002.

[33] T. Cui, C. Fox, and M. J. O'Sullivan. Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance metropolis hastings algorithm. *Water Resources Research*, 47(10):W10521, Oct 2011.

[34] F. Dominguez-Munoz, B. Anderson, J. M. Cejudo-Lopez, and A. Carrillo-Andres. Uncertainty in the thermal conductivity of insulation materials. *Energy and Buildings*, 42(11):2159–2168, November 2010.

[35] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.

[36] A. M. Egan. Three case studies using building simulation to predict energy performance of australian office buildings. In *Eleventh International IBPSA Conference Glasgow, Scotland*, 2009.

[37] B. Eisenhower, Z. O'Neill, V.A. Fonoberov, and I. Mezic. Uncertainty and sensitivity decomposition of building energy models. *Journal of Building Performance Simulation*, 5(3):171–184, May 2012.

[38] James M. Flegal, Murali Haran, and Galin L. Jones. Markov chain monte carlo: Can we trust the third significant figure. Technical report, University of Minnesota, School of Statistics, 2007.

[39] James M. Flegal and Galin L. Jones. Batch means and spectral variance estimators in markov chain monte carlo. *The Annals of Statistics*, 38(2):1038–1070, November 2010.

[40] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-6(6):721–741, Nov 1984.

[41] Heikki Haario and Eero Saksman. Simulated annealing process in general state space. *Advances in Applied Probability*, 23(4):866, Dec 1991.

[42] J.C. Hayton, D.G. Allen, and V. Scarpello. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2):191–205, 2004.

[43] Katrin Heitmann, David Higdon, Charles Nakhleh, and Salman Habib. Cosmic calibration. *Astrophys.J.*, 646:L1–L4, 2006.

[44] D. A. Henderson, R. J. Boys, and D. J. Wilkinson. Bayesian calibration of a stochastic kinetic computer model using multiple data sources. *Biometrics*, 66(1):249–256, Apr 2009.

[45] J.L.M. Hensen. *On the Thermal Interaction of Building Structure and Heating and Ventilating System*. PhD thesis, Technische Universiteit Eindhoven, 1991.

[46] Y. Heo. *Bayesian Calibration of Building Energy Models for Energy Retrofit Decision-Making under Uncertainties*. PhD thesis, Georgia Institute of Technology, 2011.

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

[48] Y. Heo, R. Choudhary, and G.A. Augenbroe. Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, 47:550–560, 2012.

[49] Yeonsook Heo, Diane J. Graziano, Leah Guzowski, and Ralph T. Muehleisen. Evaluation of calibration efficacy under different levels of uncertainty. *Journal of Building Performance Simulation*, 8(3):135–144, 2015.

[50] D. Higdon, J. Gattiker, B. Williams, and M. Rightley. Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103(482):570–583, 2008.

[51] Dave Higdon. Space and space-time modeling using process convolutions. In CliveW. Anderson, Vic Barnett, PhilipC. Chatwin, and AbdelH. El-Shaarawi, editors, *Quantitative Methods for Current Environmental Issues*, pages 37–56. Springer London, 2002.

[52] Dave Higdon, Marc Kennedy, James C. Cavendish, John A. Cafeo, and Robert D. Ryne. Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26(2):448–466, January 2004.

[53] C.J. Hopfe and J.L.M. Hensen. Uncertainty analysis in building performance simulation for design support. *Energy and Buildings*, 43(10):2798–2805, Oct 2011.

[54] M.F. Hutchinson and F.R. de Hoog. Smoothing noisy data with spline functions. *Numerische Mathematik*, 47(1):99–106, 1985.

[55] P. Im and J.S. Haberl. Analysis of the energy savings potential in k-5 schools in hot and humid climates: Application of high performance measures and renewable energy systems. In *Proceedings of the 16th Symposium on Improving Building Systems in Hot and Humid Climates, Texas A&M University, Dallas, Texas, published on CD ROM (December).*, 2008.

[56] P. Im and J.S. Haberl. Development of a simplified simulation tool for high performance k-5 schools in hot and humid climates. In *Third National Conference of IBPSA-USA Berkeley, California*, 2008.

[57] JCGM. Evaluation of measurement data – guide to the expression of uncertainty in measurement, September 2008.

[58] M. J. Jimenez, L. Castano S., de La Torre, H. Madsen, G. Flamant, G. Lethe, G. Bauwens, and S. Roels. Iea ebc annex 58 subtask 3, 4th common exercise on data analysis. instruction document. Technical report, IEA EBC Annex 58 internal report, 2013.

[59] M. J. Jimenez, H Madsen, G. Flamant, G. Lethe, G. Bauwens, and S. Roels. Iea ebc annex 58, subtask 3, 3rd common exercise on data analysis. instruction document. Technical report, IEA EBC Annex 58 internal report, 2013.

[60] M.J. Jiménez, H. Madsen, and K.K. Andersen. Identification of the main thermal characteristics of building components using matlab. *Building and Environment*, 43(2):170–180, Feb 2008.

[61] N. Johnson and H. Hu. Developing an effective and reliable approach for building energy audit in support of retrofit decision making. In *International High Performance Building Conference, Purdue*, 2012.

[62] I.M. Johnstone and B.W. Silverman. Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59(2):pp. 319–351, 1997.

[63] I.T. Jolliffe. *Principal Component Analysis (Springer Series in Statistics)*. Springer, 2002.

[64] Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, Jun 1995.

[65] M.C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society*, Series B 63:425–464., 2001.

[66] Y.J. Kim, K.C. Kim, C.S. Park, and I.H. Kim. Deterministic vs. stochastic calibration of energy simulation model for an existing building. In *ASim 2014, IBPSA Asia Conference, Nasoga Japan, 2014*, November 2014.

[67] M. Lamboni, H. Monod, and D. Makowski. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, 96(4):450–459, Apr 2011.

[68] Philippe Lauret, Frédéric Miranville, Harry Boyer, Francois Garde, and Laetitia Adelard. Bayesian parameter estimation of convective heat transfer coefficients of a roof-mounted radiant barrier system. *Journal of Solar Energy Engineering, Transactions of the ASME*, 128(2):213–225, 2006.

[69] Genyuan Li, Sheng-Wei Wang, Herschel Rabitz, Sookyun Wang, and Peter Jaffé. Global uncertainty assessments by high dimensional model representations (HDmr). *Chemical Engineering Science*, 57(21):4445–4460, Nov 2002.

[70] I. Macdonald and P. Strachan. Practical application of uncertainty analysis. *Energy and Buildings*, 33(3):219–227, Feb 2001.

[71] M.A. MacFarlan and F.A. Graybill. *Introduction to the theory of statistics*. McGraw-Hill, New York [u.a.], 2. ed edition, 1963.

[72] H. Madsen and J. Holst. Estimation of continuous-time models for the heat dynamics of a building. *Energy and Buildings*, 22(1):67–79, Mar 1995.

[73] K. Martin, A. Campos-Celador, C. Escudero, I. Gomez, and J.M. Sala. Analysis of a thermal bridge in a guarded hot box testing facility. *Energy and Buildings*, 50:139–149, July 2012.

[74] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1953.

[75] J. Milnor. *Games against nature*, chapter Games agaist nature. John Wiley & Sons, London, UK, 1954.

[76] M.D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, April 1991.

[77] Iain Murray, Ryan Prescott Adams, and David J.C. MacKay. Elliptical slice sampling. *Journal of Machine Learning Research: Workshop and Conference Proceedings (AISTATS)*, 9:541–548, 05/2010 2010.

[78] I. Naveros, P. Bacher, D.P. Ruiz, M.J. Jiménez, and H. Madsen. Setting up and validating a complex model for a simple homogeneous wall. *Energy and Buildings*, 70:303–317, Feb 2014.

[79] Radford M. Neal. *Bayesian Learning for Neural Networks*. PhD thesis, University of Toronto, Secaucus, NJ, USA, 1996.

[80] RadfordM. Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001.

[81] D.J. Nott, L. Marshall, and J. Brown. Generalized likelihood uncertainty estimation (glue) and approximate bayesian computation: What's the connection? *Water Resources Research*, 48:W12602, 2012.

[82] A. Pedrini, F.S. Westphal, and R. Lambertsb. A methodology for building energy modelling and calibration in warm climates. *Building and Environment*, 37:903–912, 2002.

[83] I.M. Pegg, A. Cripps, and M. Kolokotroni. Post-occupancy performance of five low-energy schools in the uk. *ASHRAE Transactions*, 2007.

[84] Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11, 2006.

[85] W.H. Press. *The Art of Scientific Computing*. Numerical recipes in FORTRAN. Cambridge University Press, 1992.

[86] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in Fortran 90 (2Nd Ed.): The Art of Parallel Scientific Computing*. Cambridge University Press, New York, NY, USA, 1996.

[87] Paul Raftery, Marcus Keane, and Andrea Costa. Calibrating whole building energy models: Detailed case study using hourly measured data. *Energy and Buildings*, 43(12):3666 – 3679, 2011.

[88] Paul Raftery, Marcus Keane, and James O'Donnell. Calibrating whole building energy models: An evidence-based methodology. *Energy and Buildings*, 43(9):2356 – 2364, 2011.

[89] Rubina Ramponi, Adriana Angelotti, and Bert Blocken. Energy saving potential of night ventilation: Sensitivity to pressure coefficients for different European climates. *Applied Energy*, 123:185–195, Jun 2014.

[90] J. O. Ramsay, G. Hooker, and S. Graves. *Functional Data Analysis with R and MATLAB*. Springer Publishing Company, Incorporated, 1st edition, 2009.

[91] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, 2nd edition, 2005.

[92] Carl Edward Rasmussen and Joaquin Quiñonero Candela. Healing the relevance vector machine through augmentation. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05, pages 689–696, New York, NY, USA, 2005. ACM.

[93] C.E. Rasmussen and C.K.I. Williams. *Gaussian Process for Machine Learning*. the MIT Press, 2006.

[94] M. Ratto. Sensitivity analysis in model calibration: Gsa-glue approach. *Computer Physics Communications*, 136(3):212–224, May 2001.

[95] Marco Ratto, Andrea Pagano, and Peter Young. State dependent parameter metamodelling and sensitivity analysis. *Computer Physics Communications*, 177(11):863–876, Dec 2007.

[96] T. Reddy. Literature review on calibration of building energy simulation programs: Uses, problems, procedures, uncertainty, and tool. *ASHRAE Transactions*, 2005.

[97] T.A. Reddy, I. Maor, and C. Panjapornpon. Calibrating detailed building energy simulation programs with measured data–part ii: application to three case study office buildings (rp-1051). *HVAC & R Research*, 13(2):221–241, 2007.

[98] T.A. Reddy, I. Maor, I., and C. Panjapornpon. Calibrating detailed building energy simulation programs with measured data–part i: general methodology (rp-1051). *HVAC & R Research*, 13(2):243–265, 2007.

[99] M. Riddle and R.T. Muehleisen. A guide to bayesian calibration of building energy models. In *2014 ASHRAE/IBPSA-USA Building Simulation Conference Atlanta, GA*, September 2014.

[100] C. Robert and G. Casella. *Introducing Monte Carlo Method with R.* Springer, 2010.

[101] Gareth O. Roberts and Jeffrey S. Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, Jan 2009.

[102] Renata Romanowicz, Helen Higson, and Ian Teasdale. Bayesian uncertainty estimation methodology applied to air pollution modelling. *Environmetrics*, 11(3):351–371, 2000.

[103] Jeffrey S. Rosenthal. AMcmc: An r interface for adaptive MCMC. *Computational Statistics & Data Analysis*, 51(12):5467–5470, Aug 2007.

[104] David Ruppert. *Statistics and Data Analysis for Financial Engineering (Springer Texts in Statistics)*. Springer, 2010.

[105] A.M. Rysanek and R. Choudhary. Optimum building energy retrofits under technical and economic uncertainty. *Energy and Buildings*, 57:324–337, Feb 2013.

[106] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, May 2002.

[107] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259 – 270, 2010.

[108] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley-Blackwell, Feb 2002.

[109] D. Garcia Sanchez, B. Lacarrière, M. Musy, and B. Bourges. Application of sensitivity analysis in building energy simulations: Combining first- and second-order elementary effects methods. *Energy and Buildings*, 68, Part C:741 – 750, 2014.

[110] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Math. Comput. Simul.*, 55(1-3):271–280, February 2001.

[111] S. Song. *Development of New Methods for Evaluating the Energy Performance of New Commerciao Buildings*. PhD thesis, Texas A&M University, 2006.

[112] S. Song and J. Haberl. A procedure for the performance evaluation of a new commercial building: Part ii – overall methodology and comparison of results. *ASHRAE Transactions-Research*, 114:389 – 403, 2008.

[113] S. Song and J.S. Haberl. A procedure for the performance evaluation of a new commercial building: Part i – calibrated as-built simulation. *ASHRAE Transactions-Research*, 114:375–388, 2008.

[114] C. Spitz, L. Mora, E. Wurtz, and A. Jay. Practical application of uncertainty analysis and sensitivity analysis on an experimental house. *Energy and Buildings*, 55:459–470, Dec 2012.

[115] P. Strachan, K. Svehla, I. Heusler, and M. Kersken. Whole model empirical validation on a full-scale building. *Journal of Building Performance Simulation*, page 1–20, Sep 2015.

[116] J. Sun and T.A. Reddy. Calibration of building energy simulation programs using the analytic optimization approach (rp-1051). *HVAC&R RESEARCH*, 12(1):177–196, 2006.

[117] Wei Tian. A review of sensitivity analysis methods in building energy analysis. *Renewable and Sustainable Energy Reviews*, 20(0):411–419, Apr 2013.

[118] N.H. Timm. *Applied Multivariate Analysis*. Springer Texts in Statistics. Springer, 2002.

[119] C. Turner and M Frankel. Energy performance of leed® for new construction buildings. Technical report, U.S. Green Building Council, 2008.

[120] HAL van Dijk and FM Tellez. Measurement and data analysis procedures, final report of the joule ii compass project (jou2-ct92-0216). Technical report, 1995.

[121] V. Vyshemirsky and M.A. Mark A. Girolami. Bayesian ranking of biochemical system models. *BIOINFORMATICS*, 24(6):833–839, 2008.

[122] G Wei, M. Liu, and D.E. Claridge. Signatures of heating and cooling energy consumption for typical AHUs. In *Proceedings of the Eleventh Symposium on Improving Building Systems in Hot and Humid Climates, Fort Worth, TX*, 1998.

[123] F. S. Westphal and R. Lamberts. Building simulation calibration using sensitivity analysis. In *Ninth International IBPSA Conference Montréal, Canada August 15-18*, 2005.

[124] Andrew Gordon Wilson and Ryan Prescott Adams. Gaussian process kernels for pattern discovery and extrapolation. *International Conference on Machine Learning (ICML), JMLR W&CP*, 28(3):1067–1075, February 2013.

[125] M. Wood, M. Eames, and P. and Challenor. A comparison between gaussian process emulation and genetic algorithms for optimising energy use of buildings. In *14th International Building Simulation Conference BS2015, Hyderabad, India, Dec 7-9th, 2015*, 2015.

[126] J. Yoon, E. Lee, and D.E. Claridge. Calibration procedure of energy performance simulation model for a commercial building. *Journal of Solar Energy Engineering*, 125(August):251–257, 2003.