IMPACT OF ELA CALIBRATION METHODS ON BUILDING ENERGY MODEL FIDELITY AND FITNESS

A Dissertation
Presented to
The Academic Faculty

by

Mohanned M. Althobaiti

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
College of Design

Georgia Institute of Technology
May 2021

IMPACT OF ELA CALIBRATION METHODS ON BUILDING ENERGY MODEL FIDELITY AND FITNESS

Committee members:

Prof. Godfried Augenbroe, Advisor
School of Architecture
Georgia Institute of Technology

Dr. Pieter De Wilde
School of Architecture
University of Plymouth, UK

Dr. Jason Brown
School of Architecture
Georgia Institute of Technology

Dr. Cheol-Soo Park
Department of Architecture &
Architectural Engineering
Seoul National University

Ron Judkoff
Buildings and Thermal Systems Center
National Renewable Energy Laboratory

Date Approved: July 24, 2020

To my incredible father, my loving mother, my dearest wife, and my little princess Rose.

You always believed in me ♥.

ACKNOWLEDGEMENTS

This Ph.D. study has been a remarkable life-changing experience for me, and it would not have been accomplished without the help of many people. First and foremost, I would like to thank my research advisor Prof. Godfried Augenbroe for his critical role in bringing me to the Georgia Institute of Technology. His coursework truly allowed me to develop a more complete understanding of the role that physics and engineering plays in the design of architectural systems, spaces and experiences. I'm also deeply indebted to him for giving me the freedom, guidance, and encouragement to pursue my research interest. I'm sincerely grateful for his profound knowledge and sharp insights in nurturing my research perspectives and immense support to both of my academic and life pursuits.

I would like to extend my appreciation to my committee members: Mr. Ron Judkiff, Dr. Jason Brown, Dr. Pieter deWilde, and Dr. Cheolso Park for taking the time to serve on my dissertation committee, and for providing valuable advice and insightful comments to improve this work. Grateful appreciation is also extended to Dr. Jason Brown for the tremendous knowledge he provided in his coursework in my early years in Ph.D.

I gratefully acknowledge the government of Saudi Arabia and King Saud University (KSU) for the generous scholarship to obtain my Ph.D. I also wish to thank my lab mates Yifu Shi, Alya Alhashim, Mayuri, and in particular, Yun Joon for his coding aid in parts of this work. Our conversation and insight have contributed significantly to my development and understanding of the core problems addressed both by our lab and within this dissertation.

A heartfelt appreciation goes to my dearest wife, who has always stood by me with her continuous encouragement and considerate understanding throughout this journey. I am grateful to my brothers and sisters for always being there for me, and finally, I would like to express my deepest appreciation to my parents. I could not become the person I am today without their unreserved support, selfless sacrifice, and unconditional love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

As building performance is increasingly improved and building energy consumption decreases, a greater percentage of the total energy loss of a building occurs through envelope leakage. This leakage is characterized by the effective leakage area or ELA, which is a proxy parameter to what is essentially a complex flow phenomenon through cracks driven by pressure differences. Moreover, different façades and façade parts have different ELA and are typically subjected to different pressure differences in a given wind condition. This poses major challenges to building energy models.

Current building performance simulation (BPS) uses software modules that approximately calculate envelope infiltration, but the literature shows that their calibration and validation is still unsatisfactory. In fact, calibration and validation of BPS models is still an important subject of study in our quest to improve the fidelity of simulation-based predictions in various applications. The high level of interaction and subsumption* between parameters can result in a model that approximates the measurements well (and thus meets the ASHRAE auditing threshold) but whose "best estimates" of parameters are unreliable. This can be a problem in performance contracting when limits have been agreed on certain parameters such as ELA and U-value. It can also be problematic in the use of the model for certain performance assessments. This thesis exemplifies the underlying issues by comparing the results of direct and indirect calibration at different fidelities.

* Subsumption designates in this context how certain parameters can assume a different value than their best physical estimate to make up for the errors in other parameters or structural errors in the model. A model can then show good alignment with measured outcomes, although the values its parameters are (far) off from the true values.

The study focuses on the calibration of building energy models of existing buildings. It does so by conducting calibration for different experiments, i.e., for different sources of data, and for different model fidelities. The calibration is anchored around ELA and its impact on "best estimates" of other parameters is verified. The study is done with explicit quantification of uncertainties in the experiments as well as in model parameters. The two major experiments considered are (a) direct ELA calibration through tracer gas experiments, (b) indirect ELA calibration with consumption data enhanced by spot temperature measurements. Two case studies on existing buildings are performed.

The thesis develops a new framework to address calibration and validation for different combinations of data and model fidelity, where each combination leads to probability distributions of the calibration parameter set. For each combination the ultimate aim is to determine the fitness of the resulting building energy model for given application studies such as building energy benchmarking, fault detection, unmet hour verification, etc. This requires the introduction of a novel fitness measure that determines the confidence level of a particular calibrated model for decisions in a predefined building performance assessment scenario. The thesis shows an early example of how to develop and quantify fitness.

The results will be meaningful for better understanding façade infiltration, better understanding of the limits of calibrated models, and the way this translates into fitness of the resulting model. The thesis focuses exclusively on existing buildings, but its findings may lead to large scale data sets of calibrated ELA values in existing buildings, that may find their way into better ELA quantification in energy models of new designs.

CHAPTER 1. INTRODUCTION AND BACKGROUND

Buildings are primarily designed to assure best thermal comfort with the minimum amount of energy consumption. Heating, ventilating, and air conditioning (HVAC) systems in buildings are designed to maintain thermal comfort at the desired level as well as assuring best indoor air quality (IAQ). However, the operating cost of HVAC systems is often a large percentage of the total energy cost of buildings, which on average constitutes 47% of the primary energy consumed in a building (DOE, 2017). In 2018, residential and commercial buildings consumed about 40% of total U.S. energy consumption (EIA, 2019). According to the same source, commercial floorspace grows by an average 1% per year through the projection period of 2050. This growth will increase the total demand for energy which imposes new challenges for the world and nationwide carbon emission mitigations. It calls for further research and development to explore innovative solutions in real practice. Hence, the use of energy simulation software has increased in response to the current emphasis on reducing energy consumption and greenhouse gas emissions, which asks for the investigation of different design options and their impacts on building energy use.

One option to reduce building energy use is the improvement of building envelope airtightness. This can be guaranteed by the right choice of façade technologies and proper assembly and maintenance of the façade over the lifetime of the building. Existing data show that unless extra efforts are made to the design and construction of tight building envelopes, commercial buildings are much leakier than typically assumed (Emmerich, McDowell, & Anis, 2007); (Emmerich & Persily, 2013). Therefore, the energy impacts of

uncontrolled infiltration are also greater than assumed. Nevertheless, current energy simulation software and design methods generally cannot accurately account for envelope infiltration because no good estimate of leakiness is readily available, and therefore the impacts of improving airtightness on energy may not be fully appreciated.

This dissertation investigates the effect of using various granularities of data for calibration and validation. In this study, the emphasis on data granularity and calibration method is mainly driven by the need to identify the effective leakage area (ELA) as this is the façade property that dominates air infiltration. It should however be noted that an isolated single parameter calibration is not a possible or even plausible undertaking because a model always contains a set of other parameters that are basically unknown as they cannot be measured in isolation, at least not without considerable effort in setting up experiments. So, although the prime interest is on ELA of the various facades or rather facade elements of a building, other parameters will enter the calibration process by necessity.

1.1 What is the value of calibration and what do we need to know about it?

Models are built to help us understand the reality with which we desire to interact for our advantage. Building energy predictions might lead to poor approximations compared to reality. We can relate this deficiency of predicted results to the lack of knowledge related to 1) reliable information about critical model parameters, e.g., U-value, air infiltration, lighting and plug loads; 2) experience required to analyze and schematize building specifications and derive a suitable model; 3) validity of the model and sub-models in complicated real environments that could be far outside the laboratory testing

conditions; 4) quality assurance of the simulation tool used in the calibration, e.g., modeling/coding errors.

Given the aforementioned causes of deficient building energy models for existing buildings, calibration techniques can play a significant role to mitigate them, as it fills in for the lack of prior knowledge. Calibration is commonly applied to building performance simulation (BPS) to benchmark buildings under normative usage scenarios, perform diagnostics and determine post retrofit savings predictions, as part of energy savings performance contracts and related PM&V methods. Some BPS software offers an extension to embed a calibration technique in the simulation environment itself. The basic approach is always to tune selected unknown parameters to reconcile the model predictions with monitored data. However, while it is logical to use monitored data (the observations) and in particular building energy consumption data to calibrate a building simulation model, the calibration technique may not always result in a more reliable use in above mentioned applications. When calibrating a large number of model parameters to a limited number of observations (also known as under-determined or over-parameterized problem in mathematics), there can be many combinations of model parameters that will result in a close match to the observations (e.g. utility consumption data). Hence, a close match is not in itself proof of a good calibration as the calibrated model may show a good match with the pre-retrofit observations but may fail to predict accurate outcomes of the post retrofit case (Judkoff, Polly, & Neymark, 2016), (Reddy T. A., Maor, Jian, & Panjapornpon, 2006). Therefore, the probability that the calibration actually improves the model and associated energy savings predictions after the retrofit is correlated with the observation frequency of the building performance data, or the informational content of the data.

Therefore, it is useful to have a test method for calibration techniques that provides at least the following three primary metrics (Judkoff, Polly, & Neymark, 2016): 1) the goodness of fit between the calibrated model generated data and utility consumption observations, 2) the closeness of the calibrated parameter values and the "true" parameter values, and 3) the accuracy of the savings prediction or other application. Not all calibration techniques can use all three metrics. Unfortunately, most of the calibration studies are only limited to metric 1 and 3, which only tests different aspects of "correctness" of the calibration algorithm. Although this is useful, it is only a weak test of the value of the calibrated model. As argued above, metric 2 may in fact play the most significant role to determine the goodness of the model for a specific application.

The true complexity of calibration which cannot be tested or standardized lies in the fact that most calibration parameters are surrogate parameters that moreover appear in simplified physical relationships at the core of the building energy model. As a consequence, there is an unavoidable and significant model discrepancy, usually referred to as model form uncertainty (Sun Y. , 2014). As such any calibration will only lead to "best estimates" of calibration parameters that subsume that part of building behavior that is fundamentally not addressed in the model. It is significant to note that enlarging the pool of calibration parameters or enlarging the amount and resolution of real observations will NOT remove this bias. The reason is that "subsumption" is unavoidable, so structural model discrepancy will always show up in the deviation of calibration parameter values from their true values. Only a joint effort to both increase the fidelity of the model (thereby decreasing the structural error) and increase the parameter pool may remove the bias. However, this will potentially lead to vastly over-engineered and run-intensive simulation

models. Before undertaking such an effort, it needs to be established first whether (in spite of the described bias) the calibrated model is fit enough for anticipated use. The corollary is that if sufficient fitness can be proven for the low fidelity model with a certain data resolution, the extra effort would be wasted. It is the core objective of this thesis to make a fundamental contribution to this longer term aspiration.

1.2 Literature review

This section presents an extensive literature review of current infiltration modeling and building performance simulation (BPS) modeling and calibration methods that are widely considered in this research arena, e.g. (Reddy T. A., 2006); (Price, Chang, & Sohn, 2004); (Judkoff & Neymark, 2006). We will also review infiltration testing techniques.

1.2.1 Infiltration modeling method

In an existing building, there are many leakage spots on the building skin, for simplicity regarded as a collection of façade parts, each of which can be assigned an effective leakage area (ELA) (Sherman & Grimsrud, 1980). Currently, available methods for evaluating the building's ELA (we will use this term from now on as a vector of ELA values for the distinguished facade parts) and associated infiltration rates range from simple air change methods to complex physical modeling methods. Many studies use different sources of the infiltration rates in BPS. Infiltration rates could be set up as certain air change rates according to an estimation of building envelope tightness (Thormark, 2002); (Niu, Zhang, & Zuo, 2002). Infiltration rates can also be calculated according to numerical equations and correlations with wind speed (Kurnitski, et al., 2011), where wind speed is used as a proxy to wind pressure which is the real driver of air infiltration. Multi-zone

modeling simulation is employed in building performance analyses to consider detailed building configurations and provide relatively accurate infiltration rates (Emmerich & Persily, 1998). In BPS tools, such as EnergyPlus, the energy calculation methods typically use default infiltration rates depending on different leakage properties of the buildings such as: leaky, normal, and tight (DOE, 2017). These three categories of building leakage do not account for the infiltration driving mechanisms and other building characteristics. Consequently, the assumed infiltration rates do not reflect the direct impacts of outdoor weather conditions (Ng L. C., Musser, Persily, & Emmerich, 2013). Therefore, the default settings for infiltration rates do not reflect the actual dynamic infiltration rates in the building energy simulations. For many applications this could be proven unacceptable ("unfit").

To more directly reflect the actual weather conditions, a study by the Pacific Northwest National Laboratory (PNNL) proposed a simplified approach to account for wind-driven infiltration rates into buildings (Gowri, Winiarski, & Jarnagin, 2009). The method uses an average wind speed coefficient for a square office building to calculate a base infiltration rate that is further varied with the average wind speed using modeling features within EnergyPlus. Even though this approach addresses wind-driven infiltration, it is highly simplified and also not accounting for the infiltration rates due to stack effect. Another study on airflow rate calculations indicates that the underestimation and overestimation due to surface averaged pressure coefficients are not negligible (Costola D. , Blocken, Ohba, & Hensen, 2010). Therefore, the simplified methods may not be sufficiently accurate to generate actual infiltration values in varying conditions and in configurations where the wind pressures show significant variation over the façade. The

assumption that a part of the façade can be represented by one single ELA and one single wind pressure coefficient (Cp) introduces a model reduction (in fact contributing to model form uncertainty) that will inevitably lead to a bias in the calibrated value of ELA.

Among the physical modeling methods, a research study developed a roadmap for performing full 3-D envelope simulations to calculate air leakage in buildings (Younes & Abi Shdid, 2013). This method realistically depicts the various cracks common in an envelope in terms of shape, location, and quantity, so it leads to a computationally demanding simulation model.

It is to be expected that different infiltration model assumptions will lead to large deviates in simulation outcomes. However, existing studies have not focused on comparing the influence of different infiltration rate calculation methods on the accuracy (or rather correctness) of building energy simulation results in a given usage scenario.

As an important contribution to quantify infiltration in buildings, the U.S. National Institute of Standards and Technology (NIST) developed multi-zone airflow network models and applied this to the DOE reference buildings (Ng, Musser, & Persily, 2012). Their approach makes it possible to calculate detailed annual whole building infiltration rates (Wang, 2013); (Ng, Emmerich, & Persily, 2014). Based on the NIST airflow simulation software CONTAM (CONTAM, 2015), the multi-zone airflow network models can use transient weather data to calculate air infiltration (through openings and leakages) at each time step of the simulated period. In this nodal-flow network, each branch in the network is defined through a flow-resistance equation, which requires knowledge about the ELA's of façade elements on a zone level as well as the flow resistance of internal

partitions and doors. If these values are known with accuracy, the model provides more or less realistic estimations for air infiltration since they consider the outdoor weather, internal zonal/system interactions, stack effect etc. (Ng, Musser, & Persily, 2012). The studies conducted by (Wang & Zhong, 2014), which aim to assess the energy savings realized by yearly air infiltration, use a coupled CONTAM-TRNSYS model to calculate energy savings in heating and cooling affected by air infiltration in the medium office DOE reference building model in 8 climate zones. The use of TRNSYS in their study is apparently justified by the availability of a data interface with CONTAM. (Ng, Emmerich, & Persily, 2014) use a similar approach with EnergyPlus. As shown in Figure 1.1 they used the yearly infiltration rate generated from CONTAM and input it into EnergyPlus to account for total energy performance of the building. However, this method needs predefined values of ELA's for CONTAM input.

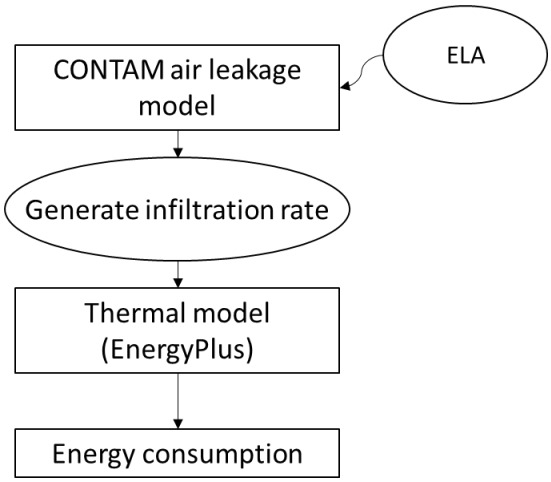


Figure 1.1 Framework of infiltration simulation in current methods using a nodal airflow model

Overall, there have been many different calculation methods of infiltration rates in both theoretical and practical studies. The approach shown in Figure 1.1 represents the

most common method is the yearly infiltration and transient airflow simulation method. The pre-simulated hourly infiltration rates of the whole building are exported and input in the energy model. It is important to understand the accuracy level of building performance analysis associated with different methods of infiltration rate calculations.

1.2.2 Direct measurement of air leakage

Several standards exist for the testing of infiltration and air leakages in building envelopes. Among these are conventionally blower door testing and tracer gas methods (Sherman & Grimsrud, 1980). It should also be stated here that other methods such as tracer smoke, sound transmission, and thermal imaging testing assist in locating leakages (McKenna & Munis, 1989). These testing methods are mostly carried out in accordance with ISO and ASTM standards. A more detailed discussion comparing the testing methods and their instrumentation is presented below.

1.2.2.1 Fan pressurization method

Air leakage in buildings can be determined by fan pressurization techniques. This method is synonymous to ‘blower door’ testing, which uses a large door or window-mounted fan to blow air into or suck air out of a building creating pressurization or depressurization to determine the airflow at different pressure differences (ΔP) across the building shell. The measurements are influenced by environmental conditions at the time of the test (Persily A. , 1982). The environmental factors influencing the measurement can be grouped into two categories: 1) the effects of air density differences on airflow measurements; and 2) the effects of wind pressures on airflow and inside-outside pressure differences. The air density differences are caused by inside-outside air temperature

differences at the time of the test, as well as differences between the air temperature at the time of the test. Differences in air density are also caused by variations in atmospheric pressure from the time of calibration to the time of the test and by changes in elevations (Sherman M. , 1998). Wind pressure variations over the surfaces of the building can cause uncertainties in the inside-outside pressure measurement (typically sensed at one location only).

1.2.2.2 Tracer gas method

Tracer gas methods provide an accurate measurement of the aggregated air flow through the many unknown gaps and cracks that appear in the construction of a building. Both the (ASTM, 2011) standard test method for determining air change in a single zone by means of tracer gas dilution and the (ISO-12569, 2017) thermal performance of buildings determination of air change in buildings tracer gas dilution method present similar testing procedures. This test method uses the measurement of tracer gas dilution to determine air exchange in a building zone. The measurement of tracer gas concentration, and often the volume rate of the tracer gas that is released into the zone, allows calculation of the volume rate of air that infiltrates or exfiltrates from the zone. From this, one can infer the air change rate from the following equation.

$$ACH = \frac{ln\, C_2 - ln\, C_1}{\Delta t} \quad (1.1)$$

Where

ACH: the change rate, [1/h];

C_1: gas concentration at start;

C_2: gas concentration at end;

Δt: time interval, [h].

To determine the average air change rate, one introduces a volume of tracer gas uniformly into the zone, ensures a uniform concentration, and then measures the tracer gas concentration at given times; then calculate the air change rate for that period as the difference between the logarithms of the initial and final tracer gas concentrations divided by the time interval as in Eq. (1.1).

Although these two direct measurement methods can provide a more accurate estimation of air leakage than physics-based modeling methods, some critical issues are always associated with them which in many cases stand as a barrier to using them. The implementation of direct testing requires a fully controlled environment, which is disruptive to occupants most of the time. Moreover, the instrumentation used for the testing is typically sophisticated which requires special site preparation and expertise, thus leading to an expensive and costly test. Finally, not every building under certain conditions can be tested, for instance it is very difficult to utilize blower door testing in large commercial building. Same with tracer gas, it is nearly impossible to reach the desired gas concentration on large spaces, and if one tried, it would be costly to fill the space with the gas at high enough concentrations. Therefore, alternative (i.e., indirect) methods are attractive to mitigate such issues and still arrive at a fit enough model, or a model that can compete with models based on direct identification of ELA.

1.2.3 Calibration methods

The most need for calibration occurs in energy auditing of buildings for the purpose of benchmarking, detecting faults or determining best retrofit interventions. In the latter case, the simulation is set up to improve energy efficiency by implementing the most optimal retrofit technologies. However, no current physics-based model can represent the infiltration flow in a real building well enough without reliance on actual observations. As discussed, this is a major downside of some simplified methods presented in the previous sections. One common way to mitigate that is by adding onsite observations and tuning model parameter values to reconcile predictions with observations. (Reddy T. A., 2006) presents a review on the various uses and benefits of calibration in the context of building energy simulation. Current calibration processes can be broken down into the following categories: general references (Kaplan, Caner, & Vincent, 1992); calibration based on iterative, manual, and pragmatic intervention (Kaplan, McFerran, Jansen, & Pratt, 1990a); calibration based on a suite of informative, graphical comparative displays (Bronson, Hinchey, Habrel, & O'Neal, 1992); calibration based on special tests and analytical procedures; and analytical or mathematical methods of calibration (Reddy T. A., 2006). It is crucial to understand that every model calibration process is influenced by sources of error and uncertainty including: model fidelity, data resolution and accuracy, improperly defined ranges of unknown input parameters and improper model assumptions.

As part of developing a calibration methodology for BPS, (Reddy, Maor, & Panjapornpon, 2007) presented heuristic and analytical procedures, which include a heuristically defined set of influential parameters. This is followed by performing a coarse grid search using Latin Hypercube (LHS) and Monte Carlo (MC) simulations to identify

“strong” parameters and a small set of solutions. They use the small solution set to predict changes in the building and building systems, and compute uncertainty in the calibrated model solutions. The objective is to tune the various inputs to the simulation program so that the predicted energy use matches the actual energy use. Once the calibrated model provides a good fit with the observed utility data, it is assumed that the program is able to predict building performance after operational and equipment changes accurately. This approach can be seen as the classical calibration approach. However, major retrofits of building components and use of new technologies and systems may pose additional uncertainties and render the model less fit for the intended purpose.

Some studies create a methodology that depends on constructing a preliminary model that relies only on the connectivity of the zones and on easily observable parameters such as zone volumes. Then they describe other parameters with uncertainty distributions, and take samples from the parameter distributions using MC methods to exercise the preliminary model, and analyze the results to determine what type of measurements will most reduce the uncertainties in the parameters that affect a specific quantity of interest (QoI) (Price, Chang, & Sohn, 2004). Due to a large number of parameters in any given building model, it is not practical to conduct exhaustive measurements. Also, due to the complexity and surrogate nature of many parameters, it may not be possible to perform experiments that can directly measure them. (Price, Chang, & Sohn, 2004) focus on determining which measurements should be done to reduce the uncertainty in the input parameters that have the greatest influence on the response.

In terms of implementation of traditional calibration, manual tuning of the model is the most common method in real practice. In this process, the modeler changes certain

parameter values in a heuristic and iterative manner to reconcile predictions with observations as best as possible. Most studies in this category focus on the procedures, guidance, and particularly graphical and analytical techniques that inform the model tuning process (Habrel, Bronson, & O'Neal, 1993c), which otherwise would be solely based on the modeler's knowledge and experience.

Instead of manually selecting calibration parameters and adjusting their values, this process can be automated. This is accomplished by advanced methods that rely on mathematical algorithms to infer parameter values based on best model agreement with observations. This approach is at the core of the approach used in this thesis. The standard statistical metric CVRMSE serves as the objective function that needs to be minimized as shown in Figure 1.2. It implies the search for the smallest possible integrated squared difference between model output and observations. This optimization approach formalizes calibration into a deterministic parameter estimation problem, and employs numerical algorithms to find the optimum, i.e., the parameter values that minimize CVRMSE. If multiple types of observations and outcomes need to be matched (e.g., energy consumption and zone temperatures), a compound CVRMSE based on chosen weights for both types of observation can be used. The choice of appropriate weights is not simple and deserves a separate treatment. For more details on BPS optimization, one can refer to the work of (Nguyen, Reiter, & Rigo, 2014). For now the choice of weights is based on a heuristic assessment of the relative magnitude of the two components of the objective function. The weights are typically chosen such that both have the same order of magnitude, i.e., equal weight.

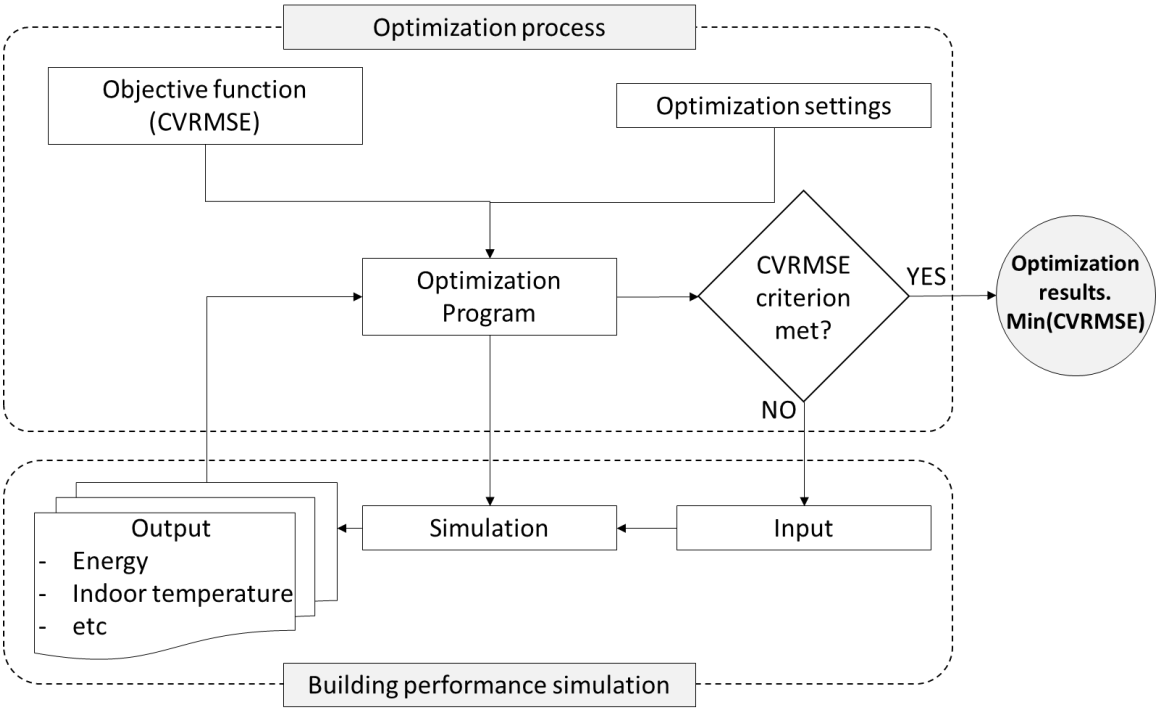


Figure 1.2 Simulation-based calibration process minimizing the calibration criterion

Many studies have investigated the optimization algorithm in Figure 1.2, such as (Ramos Ruiz, Fernández Bandera, Gómez-Acebo Temes, & Sánchez-Ostiz, 2016). They use a genetic algorithm to minimize a combined metric of R^2 and the coefficient of variation CVRMSE for dominant parameters. To select the most dominant calibration parameters, they apply the Morris method based on "relative deviation", whose value ranges are equally-possible discrete values that are consistent with supplied specifications and building documentations. (Djuric, Novakovic, & Frydenlund, 2008) use a simple heat balance model with built-in sequential quadratic programming (SQP) algorithm to minimize the root mean square error (RMSE) between predictions and observations. The entire set of parameters are calibrated since the model is relatively simple, and their upper and lower bounds are chosen based on on-site visit and authors' experience.

The high computational cost of detailed physical models is a burden that has been widely investigated. Such studies often use statistical surrogate models or meta-models to emulate the physical model within the optimization routine, which allows for exploration of a large parameter space and expedites the search for the optimum. A comprehensive model calibration framework for simultaneous multi-level building energy simulation is proposed in (Yang & Becerik-Gerber, 2015). In this work they implicitly adopt a normal multivariate linear regression emulator and quasi multi-objective optimization with a weighted objective function and linear programming algorithm. The study performs classification of model parameters to differentiate estimable and adjustable parameters, then selects important adjustable parameters for calibration using the Morris method. Additionally, (Robertson, Polly, & Jon, 2015) apply an optimization algorithm based on gradient calculation to calibrate a normal multivariate linear regression meta-model of a DOE-2.2 model against synthetic utility data of a residential building. This study uses the CVRMSE as the objective function and calculates sensitivity coefficients by Monte Carlo simulation to select a subset of six calibration parameters.

Uncertainty in calibration is another stream that heavily impacts the detailed building simulation models (Reddy T. A., 2006). In a collaborative research project (Reddy T. A., Maor, Jian, & Panjapornpon, 2006); (Reddy, Maor, & Panjapornpon, 2007), a two-step approach has been proposed to search within the parameters space for plausible solutions. The first step, a “bounded” coarse grid calibration involves a procedure that identifies both promising solutions of parameter values and parameters influential on the model discrepancy. The second step, a guided search calibration, looks for the final set of solutions through either manual or automated calibration methods. In particular, (Sun &

Reddy, 2006) proposed an analytical parameter estimation method for the guided search, which used a gradient-based nonlinear optimization technique to minimize a weighted value of the CVRMSE of both electricity energy use and demand residuals on monthly basis as the objective function. This method identifies calibration parameters based on the normalized sensitivity coefficients determined by the local sensitivity analysis approach. It also recognizes the mutual correlation among the calibration parameters to ensure they are mathematically identifiable.

A common issue in model calibration is overfitting, where one obtains incorrect and often implausible parameter values in fitting observations as they subsume model bias. Many studies attempt to address this issue in deterministic parameter estimation including the work of (Carroll & Hitchcock, 1993). They propose the regularization process by introducing a penalty term in the objective function to prevent overfitting. This takes the form of the sum of weighted square difference between a plausible solution and its corresponding default values. They reject those values deviating drastically from the default despite that they provide good agreement. Unfortunately, the calibration process in real practice is prone to overfitting due to the model's imperfection and limited observations. There are a few attempts to consider parameter uncertainties but they fail to translate those uncertainties into model predictions to inform building performance management practice. Lack of proper recognition of the main sources of discrepancy and inefficient use of data impairs the confidence of those methods in many real-life cases.

1.2.4 Validation method

Validation of BPS models in practice involves collection of observations and assessment of the model predictions' agreement with these observations, both of which will be briefly reviewed and discussed in this section.

1.2.4.1 Data gathering

It is well-known that the choosing of the correct inputs in our building performance model will largely determine the validity of the outcome. Therefore, the collection, recording and cleaning of descriptive data in practice has an intrinsic role in BPS, which often relies on practitioners' knowledge and experience. As a result, manual collection of extensive building information and data often leads to a model of sufficient validity, but in most cases it is costly in time and labor. This is probably because these methods often involve evidence-based parameter estimation at the level of building sub-systems, often with the help of local embedded sensor data. In addition, well-designed procedures regarding the retrieval and recording of information and data throughout the calibration process also contribute to success. However, most studies on calibration focus on compliance with calibration standards, like ASHRAE Guideline 14-2002 (ASHRAE, 2002), as the sole validation criterion. Lack of sufficient information and recorded observations for validation may impair the model's credibility and obscure the benefit of BPS calibration. Therefore, a systematic way to integrate the strengths of both approaches deserves further investigations on the gathering of information and recording of monitoring data.

Regarding the type of information and data, (Fabrizio & Monetti, 2015) provides a taxonomy in ascending order of collection effort: utility bills, as-built data, site visit or inspection, detailed audit, short-term monitoring, and long-term monitoring. (Raftery, Keane, & Costa, 2011) proposed a similar taxonomy with the focus on the hierarchy of reliability, which includes data-logged measurements, spots for short-term measurements, direct observation (site surveys), operator and personnel interviews, operation documents, commissioning documents, benchmark studies and best practice guides, standards, specifications and guidelines, and design stage information (e.g., the initial model).

It is important to mention that monitoring data in building operation can be classified at different scales, i.e., temporal, spatial, and categorical. The temporal scale concerns the coverage and resolution of monitoring data regarding its temporal variability, mostly due to weather and usage scenarios. The most common type of data in building performance management is the monthly utility bill, as it is readily available and reliable in most cases. The use of hourly or sub-hourly data, such as those from smart meters, building management systems (BMS), in-situ monitoring, etc. receives increasing attention (Heo & Zavala, 2012); (Djuric, Novakovic, & Frydenlund, 2008) (Yang & Becerik-Gerber, 2015). Short interval data is usually more informative than monthly data in model validation because of the embedded dynamic characteristics. However, the length of hourly or sub-hourly data used in common analysis is often limited to a few weeks of spot monitoring, or only aggregated at the whole-building level for smart meter data, as it is difficult for common methods to handle large amounts of data. This may affect the data's coverage of variations of weather conditions and expose the model to extrapolation risks. In addition, hourly or sub-hourly data contains relatively large variations because of

varying building usage. The fact that these variations are largely unknown, unless directly monitored, pose a significant difficulty. Its strong temporal correlation with energy outcomes makes it difficult to filter them out when identifying the performance of building fabric and energy supply systems. In general, performing a calibration on an in-situ experiment which is largely uncontrolled and at least only partly unmonitored, will put limits on the effectiveness of calibration of BPS models in practice.

It is worth mentioning that spatial and categorical are other scales that must be recognized in data collection. The spatial scale deals with the coverage and resolution of monitoring data with respect to its spatial variability, i.e., determined by spatial topology of the building and space functions. The categorical scale concerns the type of monitored state variables which are mostly directly related to the output of interest for building performance management. The most common type of output is power or energy use of electricity, gas, and/or other fuels. In contrast, room air temperature is less commonly monitored, as it is typically maintained by the HVAC system according to the thermostat setting, and therefore is often relatively constant and less informative (Mustafaraj, Marini, Costa, & Keane, 2014). (Roberti, Oberegger, & Gasparella, 2015) provide a case study that calibrates a model of a historical building to hourly indoor air and surface temperature. (Ramos Ruiz, Fernández Bandera, Gómez-Acebo Temes, & Sánchez-Ostiz, 2016) perform calibration of a building envelope by comparing it with observed interior temperature measurements.

All in all, model validation in real practice often builds upon its agreement with whole building monthly consumption data, which is often too aggregated and incomplete to reveal detailed dynamic characteristics. Monitoring data with high temporal and spatial

resolutions and/or belonging to multiple outcome categories, on the contrary, could be either too noisy, done under uncontrolled and unknown weather and usage conditions, or less relevant for the approximation of a specific QoI. Given the situation that there is always a lack of information about scenario and physical parameters (at least in uncontrolled experiments), it is important to treat model calibration as a process that delivers a stochastic outcome for the calibration parameters. This then requires a new approach to model validity, as embedded uncertainty is to some degree unavoidable. It then becomes necessary to test the resulting model from a risk-conscious decision-making perspective rather than testing accuracy per se. The latter is barely addressed in the literature. Therefore, a systematic method to assess data informativeness and calibration under uncertainty is worth further study as it will take model validation in building performance management to the next level. This thesis aims to cement a solid footing for this ambition.

1.2.4.2 Accuracy metrics

In common practice, a BPS model is deemed calibrated if its prediction is in agreement with observations, i.e., its goodness-of-fit reaches a certain threshold. This agreement is often quantified by the standard statistical metric CVRMSE:

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

Where

$\bar{y}$: the mean of all observation values,

y_i: the observation point, i=1,2,…n,

$\hat{y}_i$: the corresponding estimate of the observation.

This validation approach forms the calibration criterion established in ASHRAE Guideline 14-2002 (ASHRAE, 2002), that stipulates the tolerance limits of calibrated simulation models in terms of CVRMSE, which should be less than 15% with the use of monthly utility data and up to 30% with the use of hourly data for a model to be deemed valid.

However, calibration criteria based on goodness-of-fit of deterministic predictions are not very suitable to evaluate BPS models (Li, 2017). In reality, every BPS has uncertainty in model parameters and inbuilt uncertainties in systemic modeling assumptions. Outcomes should therefore be interpreted as stochastic outcomes which can be more informative in specific applications. To be more specific, the connection between the prediction uncertainty of a model and its accuracy under these goodness-of-fit metrics is valid for normal linear regression models rather than for BPS models because of the latter's disparate model assumptions (Reddy T. A., 2006). Similar observations and a more technical explanation can be found in the work of (Heo Y. , 2011). (Li, Augenbroe, & Muehleisen, 2017) develops a comprehensive uncertainty quantification of energy models of experimental setups carried out in a test cell for the purpose of blind model validation.

In summary, we argue that current standard deterministic calibration methods cannot evaluate the inherent probabilistic nature of predictions, and hence cannot inform these application studies that involve decision scenarios involving a conscious risk assessment. More generally speaking, deterministic models cannot be tested against a fitness criterion, as such testing requires that a model is tested against the probability that it produces the correct answer or decision in a simulation supported decision scenario. In this context, more informative and practical metrics to assess model accuracy, validity and

fitness and corresponding data informativeness are needed in the calibration of building performance models.

1.3 Goal and research questions

The review of research concerning calibration and validation of BPS models reveals that further study is needed to understand the role of prior information about the building and availability of monitored data in the calibration of the model. A primary goal is to improve the accuracy of the calibration parameters while avoiding the confounding effect of uncertainties across the calibration parameter set. In theory, this applies to all parameters of the building, but this thesis focuses on a subset of parameters, i.e., the crucially important façade parameters associated with infiltration, in particular on effective leakage area (ELA), which expresses the leakiness of a façade. The ELA needs to be diversified for different façades and façade elements in the BPS. In this study, the set of ELA parameters is seen as the anchor of a calibration framework that allows inspection of the "subsumption" effects on other parameters, as already introduced. This approach is based on the belief that in many buildings, air infiltration is a major unknown that is virtually impossible to observe directly without significant effort. The correct calibration of ELA is therefore important and taken as central focus, while other relevant parameters are introduced in the calibration to show their effect on the outcome of the overall calibration. An important practical goal of the thesis is that a process is developed that can serve as a guide to determine what monitoring plan one needs to adopt to calibrate ELA such that the resulting model is fit enough to perform the given task.

Therefore, the thesis targets the following research questions:

1. Will the effective leakage area (ELA) that is calibrated with limited "retail" data (i.e., customarily available monitor data) at different resolutions be close enough to one derived from direct measurements, which is considered the best obtainable estimate?
2. How can a new measure of fitness of a calibrated BPS be defined that can be correlated with model fidelity and monitoring data resolution? This question will be answered for one specific decision scenario dealing with energy benchmarking, requiring an answer to the question "is a model fit enough to be used in the benchmarking of an existing building against a predefined target value?".

To address these questions, we need to look at the calibration process differently than hitherto customary. We need to ask what we know about ELA values before calibration, and how certain we are about ELA values after calibration. This would normally invite a Bayesian calibration (BC) approach, but this is deliberately avoided in this thesis to keep the treatment closer to practice. Existing work on BC of BPS models is scarce, but recent progress is promising as reported in (Heo Y. , 2011) and (Chong, 2018). These studies reveal multiple problems that hamper efficient and reproducible application of BC of BPS models. To avoid these problems and reduce the computational footprint of the calibration methed, this dissertation uses a comparatively light-weight, home-grown replacement of BC. Accordingly, the work in this thesis is oriented to answer these research questions in the presence of unavoidable uncertainties in parameters, measurements and models.

1.4 Structure of the thesis

The thesis is outlined as follow:

- CHAPTER 1 presents background and motivations for the thesis. It starts with the question why we need calibration in the first place, then provides literature for modeling and current measuring techniques as well as calibration and validation methods.
- CHAPTER 2 provides theory for infiltration and the physics behind it. It also introduces fundamentals and crude uncertainty analyses.
- CHAPTER 3 provides an overview of the proposed framework and discusses theoretical benefits.
- CHAPTER 4 covers the uncertainty analysis of building model parameters related to ELA. It handles the process of uncertainty quantification in the context of wind pressure coefficient and wind speed, and it presents the experimental results of the direct ELA estimates.
- CHAPTER 5 provides details on the models preparation for the ELA calibration process. It also covers the methodological treatments on calibration in terms of sensitivity analysis as well as the process of quantifying uncertainty parameters of building performance models.
- CHAPTER 6 presents case study 1: Indirect ELA calibration in a commercial building.
- CHAPTER 7 presents case study 2: Indirect ELA calibration in a residential building.

- CHAPTER 8 provides an inclusive discussion about the results that are compiled from the two case studies.
- CHAPTER 9 introduces the concept of fitness level and applies it on a benchmarking pass/fail decision scenario.
- CHAPTER 10 presents a summary of the thesis with research conclusions and recommendations for future work.

CHAPTER 2. THEORY OF INFILTRATION AND UNCERTAINTY

Airflow control in buildings is a key component of building performance for reasons of air quality, comfort, and energy efficiency. There are several problems attributed to poor airtightness of buildings as indicated in the following diagram.

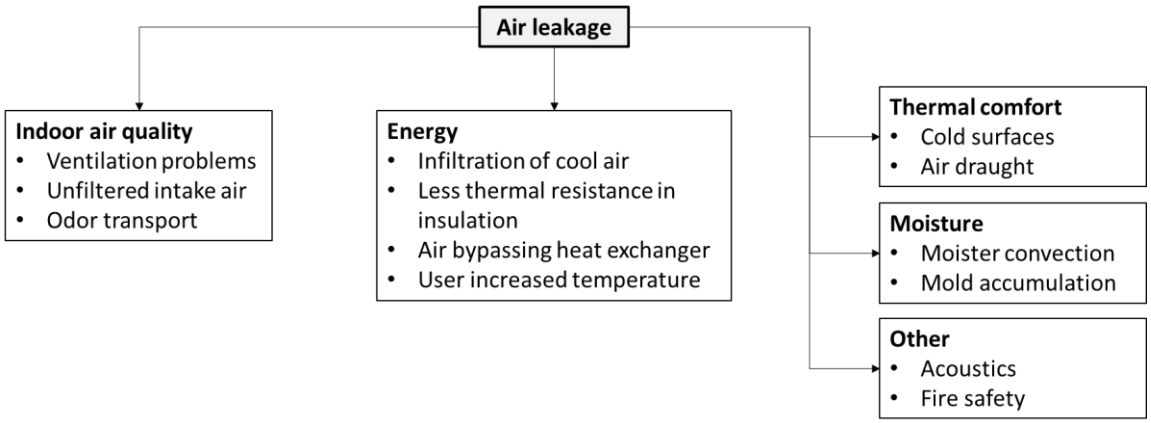


Figure 2.1 Consequences of poor building airtightness

The largest single energy end-use in buildings is space heating and cooling, which accounts for almost half the energy use in the building sector. The space conditioning load (i.e., the energy consumed by both space heating and cooling) is due to two main components: conduction and infiltration (Sherman M. , Air Infiltration in Building, 1980). Conduction causes direct heat loss through materials in the building envelope because of a temperature difference across it and expressed by the following formula:

$$E = U\ A\ \Delta T \tag{2.1}$$

Where

E: heat loss due to conduction [W],

U: conductivity of the envelope [W/m^2/K],

A: envelope area [m^2],

ΔT: inside-outside temperature difference [K].

Infiltration, on the other hand, is the mass transport of air through leaks in the building envelope and is caused by pressure differences across the envelope. The heat loss (in the heating season) due to infiltration is the product of the infiltration and the temperature difference and formulated as follow:

$$E = Q\ \rho\ C_p\ \Delta T \tag{2.2}$$

Where

E: heat loss due to infiltration [W],

Q: infiltration [m^3/s],

ρ: density of air [1.2 kg/m3],

C_p: heat capacity of air [1000 J/kg/K],

ΔT: inside-outside temperature difference [K].

Infiltration is transient as it is a function of weather, which is dynamic and impossible to model as a single factor in the heat balance. Most attempts are based on empirical formulas that express monthly infiltration amounts Q into a zone of a building or the building as a whole, usually expressed as the fraction of air change rate of the building volume per hour (ACH). Due to the complexity of wind and pressure patterns around the building, such attempts are merely a first order approximation, and cannot lead to accurate predictions of the energy consumption, unless the façade is tight enough to

safely assume that infiltration is always small, e.g., less than, say 0.1 ACH. This is however seldom the case, certainly over time as facades age.

2.1 Building simulation concerning airflow

The term *Building simulation* covers a wide range of scenarios and modeling activities related to buildings. In the context of this thesis we reduce the scope to the calculation of energy consumption and temperature field in response to given, time-variant, external conditions; involving several steps as follow:

1. **Specification of requested simulation outcomes:** the required model outcomes (or generally speaking Quantity of Interest (QoI)) that is determined by the purpose of the simulation. In this study we focus on a dynamic assessment of the energy consumption of the building as well as the indoor climate, specifically temperature.
2. **Specification of model inputs related to external conditions and usage:** this includes the outdoor climate and the operation of the building. For the outdoor climate, a modeler needs access to historical time series of (hourly) weather data is measured at a nearby meteorological station. If available, one can use an on-site weather station for higher granularity data. The usage scenario concerns the operation of the building as they relate to internal heat gains from people, equipment and lighting in the spaces.
3. **System modeling and simulation:** the process that involves the modeling of every component and defines all linkages, i.e. their physical aspects and controls. In most practical situations, system models are developed in a building modeling and simulation environment. The functionality and architecture of these modeling

environments vary significantly. In this thesis we use a small subset of models across the spectrum of fidelity, i.e., a reduced-order tool: EPC, in both monthly and hourly resolution, and a high fidelity tool: EnergyPlus (Crawley, Hand, Kummert, & Griffith, 2008).

2.1.1 Airflow modeling

As discussed in the previous section, the simulation model should define the "scenario" which specifies the weather and usage conditions the building is exposed to, and the physical model that captures the response of the building to these conditions. In practice, this distinction is commonly not preserved in the quantification of the airflow through the building, as airflow rates are often specified as pre-defined flows with a more or less fixed value throughout the simulation.

In general, airflow through a building is driven by a combination of wind, thermal buoyancy, and mechanically induced pressures. The corresponding physics are presented in the following section.

2.2 Modeling the airflow physics

To develop an effective model for predicting the infiltration in a given building, the basic physics of infiltration needs to be understood. In this section, the laws of airflow dynamics are applied to the special circumstances that drive air infiltration in buildings.

In order to evaluate the energy load of air infiltration, it is essential to determine the infiltration flow rate. This indicates the amount (volume) of air infiltrating into a building (and elsewhere exfiltrating from the building), based on which the zonal energy

loads due to this infiltration flow amounts can be determined. Infiltration is the airflow through the envelope in response to externally applied pressures (caused primarily by weather, i.e., wind pressure and temperature difference). They occur through leaks or cracks in the building shell. Several traditional methods assign a fixed value to the infiltration flow rate and do not take into account various dynamic factors that control the fluctuating infiltration flow rate, such as weather conditions, air temperature, wind speed, wind direction and several other dynamic factors. To study the actual flow, we need to separate the driving force (pressure difference) from the façade property that captures a resistance to that force. The resistance is inversely proportional to the leakiness, for which we customarily use ELA. It should be well understood that ELA is a "surrogate" parameter that is not observable in the true sense of the word or even directly measurable. It can be defined as "the equivalent amount of orifice area that would pass the same quantity of air as would pass collectively through a building envelope or component at a specified reference pressure difference" (AIVC, 1992).

As stated earlier, infiltration is caused by differential pressures across the envelope of a structure. These pressures are caused by the action of the weather on the structure. Weather induced pressures can be separated into two types: stack effect pressures as created by an indoor-outdoor temperature difference, and wind effect pressures as created by the dynamic forces exerted by the wind on a stationary object (Sherman & Grimsrud, 1980). In this section, calculations of both effects are presented separately based on the (Sherman & Grimsrud, 1980) model.

The pressures induced by the stack and wind effects cause flow through the leakage area of the building envelope. Using ELA is the property that relates driving pressure to resulting flow leads to:

$$Q = ELA\sqrt{\frac{2}{\rho}\Delta P} \tag{2.3}$$

Where

Q : air flow [m^3/s],

ELA : effective leakage area [m^2],

ρ : density of air [$1.2kg/m^3$],

ΔP : the applied pressure [Pa].

Although every part of the envelope can be given an effective leakage area, in a real situation it will be practically impossible to identify all of the facades and facade parts and roofs individually. Hence, the number of leakage variables is typically limited, for instance by considering each face of the building to have a single ELA. This assumes that the ELA of a part of the envelope is an aggregate value of the individual ELA's of its sub-components, which for the sake of energy load calculations is a plausible assumption. In this thesis the LBL model for the stack and wind effects calculations is adopted. The required formulas are given below; the full derivations and more information can be found in (Sherman & Grimsrud, 1980).

2.2.1 Stack effect

The stack effect pressure is caused by the existence of bodies of air whose difference in temperatures cause different densities. Based on the condition of hydrostatic equilibrium, the change in pressure with respect to height is proportional to the density. Consequently, In the case of a building, as the inside and outside bodies of air will have different temperatures, there will be a differential surface pressure that changes with height, this differential pressure is given by this formula:

$$\Delta P_s = \rho g H \frac{\Delta T}{T} \tag{2.4}$$

Where

ΔP_s: internal pressure shift [Pa],

g : acceleration of gravity [9.8 m/s^2],

H: height of the ceiling above grade [m],

ΔT : inside-outside temperature difference [ºK],

T: inside temperature [ºK].

Normally the internal pressure shift ΔP_s has relatively little impact compared to wind pressures on the overall infiltration, however in a case where the difference between inside and outside temperature is large or in the case of tall buildings, the stack pressure can have a considerable impact on infiltration. In those cases we take the stack effect pressure into consideration in the following formula:

$$Q_s = \frac{1}{3} \times ELA \times \sqrt{\frac{1}{\rho} \Delta P_s} \quad (2.5)$$

Where

Q_s : stack effect factor,

2.2.2 Wind effect

The flow of air around a building creates varying wind pressures on the building's skin. The pressure intensity, distribution, and nature over the building envelope vary by location. The wind pressure intensity depends on the air density, wind speed and building shape and urban surroundings characteristics. This can be expressed as a wind angle dependent normalized wind pressure coefficient: Cp. The general expression of the wind pressure intensity is then given by the following equation:

$$\Delta P_w = Cp \frac{1}{2} \rho v^2 \quad (2.6)$$

Where

ΔP_w: wind pressure [Pa],

Cp: wind pressure coefficient,

v: actual wind speed at a specified reference point [m/s].

This can be applied to any surface for which the Cp is known (as a function of wind direction). Conventionally, a reference wind speed, *v* can be defined to be the undisturbed wind speed at the height of the building. This convention for defining wind speed necessitates a method for converting wind speed obtained from a weather tower

measurement to this local reference, free-field wind speed. To do this, the following formula can be implemented:

$$v = v_0 \times \tau \left(\frac{H}{H_{met}}\right)^{\gamma} \tag{2.7}$$

Where

v_0 : wind speed at standard conditions [m/s],

H: height of the structure [m],

H_{met}: height of meteorological station [m],

τ, γ: constants that depend on terrain class; where terrain class characterizes the urban setting [-].

Terrain effects capture the fact that the vertical wind profile (see Figure 2.2) varies by geographic location when the surrounding terrain changes (i.e., surrounding shape of urban built form, trees, etc. as well as the geography will affect the free-field wind profile.) The standard undisturbed wind speed is defined at 10m height in rural terrain as shown in Figure 2.2.

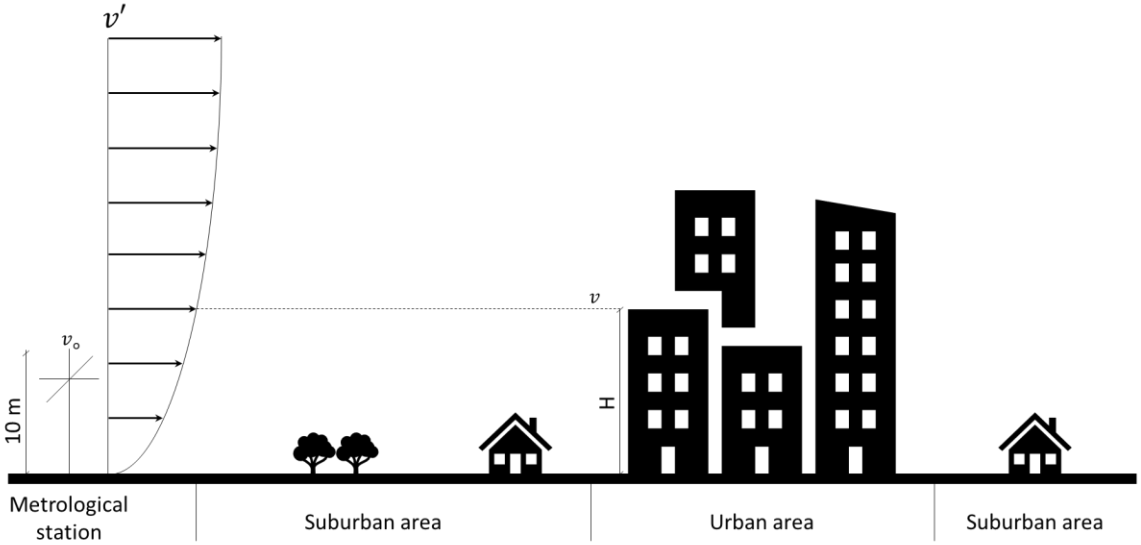


Figure 2.2 Impact of wind speed by terrain

To calculate the wind speed at one site from measured data at another site, we first use the above formula to calculate the standard wind speed for the measurement site; then the standard wind speed is used to calculate the wind speed at the desired site. The following equations are useful in calculating the actual wind speed:

$$\acute{v} = v_0 \times \acute{\tau}\left(\frac{\acute{H}}{H_{met}}\right)^{\acute{\gamma}} \tag{2.8}$$

$$v = \acute{v}\left[\frac{\tau\left(\frac{H}{H_{met}}\right)^{\gamma}}{\acute{\tau}\left(\frac{\acute{H}}{H_{met}}\right)^{\acute{\gamma}}}\right] \tag{2.9}$$

In these expressions, the primed quantities are for the wind measurement site. Values for the two parameters dependent on terrain class are shown in the table below. Note that this approach is based on the assumption that there is no strong correlation with wind direction. For some sites the upwind terrain varies with direction, such as, for

instance, in sites close to large bodies of open water. If we ignore the effect of wind direction we can use Eq. (2.3) in the calculation of infiltration induced by wind effect, if we can find a local approximation of wind pressure as function of wind speed. Normally, this is not possible or physically plausible, which mandates the calculation of Cp for all relevant parts of the skin, and for all wind directions.

The quantities necessary for calculating the terrain factor are in Table 2.1 (Sherman & Grimsrud, 1980):

Table 2.1 Terrain parameters for different classes

Class	γ	τ	Description
I	0.10	1.30	Ocean or other body of water with least 5km of unrestricted expanse.
II	0.15	1.00	Flat terrain with some isolated obstacles (e.g. building or trees well separated from each other).
III	0.20	0.85	Rural areas with low buildings, trees, etc.
IV	0.25	0.67	Urban, industrial or forest areas.
V	0.35	0.47	Center of large city.

2.3 Crude uncertainty analysis

A building energy model mimics the actual behavior of the building, mostly to assist decision-makers in achieving some objective. However, this model is not capable to correctly predict the real conditions of the building due to the following reasons. First, whether the model is built based on general assumption or even deep energy audits, the model may not accurately correspond to the actual building because the (partly assumed) input parameters do not provide complete and certain information. Second, the model cannot exactly capture the system behavior because of inherent simplifications of the physics. Hence, the model may not predict actual building behavior even with the best

possible values of the model parameters. Due to these model imperfections, the building performance model often yields predictions that unavoidably deviate from actual observed values.

Uncertainty analysis helps to quantify the effect of lack of knowledge that affects the fidelity of the model. The analysis explicitly captures the effects of incomplete knowledge on a QoI (Augenbroe, Heo, & Choudhary, 2011). Therefore, the quality of uncertainty quantification (UQ) strongly influences the rigor of an uncertainty analysis (UA). UQ encompasses the identification of uncertainty sources that possibly impact the outcomes. The quantification of uncertainty is done in the form of probability density functions of the values that parameters can have based on a lack of more precise information. The following sections discuss sources of uncertainty in the context of performance analysis; later sections will present uncertainty sources and quantification in the identified sources. Usually the quantified uncertainty is propagated through the simulation of the energy model with the use of various sampling methods, which results in probability distributions of model outcomes. The probabilistic outcomes can be translated into single values (e.g., by aggregation or statistical representation) if needed according to the objectives of the case in hand. Typically the mean and standard deviation of an approximated normal distribution are used for this purpose.

2.3.1 Sources of uncertainty

As stated earlier, the nature of uncertainties and how one deals with them depends on the given scenario and application. Uncertainty arises from a lack of knowledge about the behavior of the building or a specific system and the appropriate parameter value to use

which is normally assumed to have a fixed value in standard deterministic simulation. In general use of BPS to analyze some aspect of building performance, the uncertainties come from three main sources: physical parameter uncertainty (P), scenario uncertainty (S), and model inadequacy, also referred to as the earlier introduced model form uncertainty (MFU). These types of uncertainty are briefly discussed in the following paragraphs; more details can be found in the work of (de Wit, 2001), (Sun Y. , 2014), and (Wang Q. , 2016).

The parameter uncertainty in the energy model refers to that uncertainty associated with building physical and operational components. The model parameters that fall under this type of uncertainty specify thermal properties of materials, internal gains, properties of the HVAC system, and their operation and control settings (de Wit, 2001). In existing buildings, these parameters often diverge from their original specifications in the design documentation or from industry template values. This is because the nominal conditions used for performance testing cannot capture dynamic and stochastic building operation conditions. In addition, systems degrade over their life cycle, which gradually increases the magnitude of uncertainty in the system performance. Moreover, design and specification documents often lack a full description of system properties but provide information about system types. Therefore, we need to investigate the fundamental factors that cause uncertainty in model parameters in order to quantify parameter uncertainty in the models.

Scenario uncertainty resides in the external environment (i.e., weather conditions) and the use of the building (e.g., building usage, occupancy, and operation schedules). The actual weather conditions around the building (e.g., local ambient temperature, cloud cover, local wind speed) differ from the TMY/AMY weather data usually adopted in standard simulations. For instance, the local wind speed can be significantly different

compared to wind speed captured from a meteorological station that is located at a distance from the actual site. When use is made of a standard year, such as TMY (Hall, Prairie, Anderson, & Boes, 1978), the weather data is averaged by a statistically representative selection from 30-year weather data. In addition, actual in-use operation scenarios fluctuate from average fixed schedules typically assumed used in models. Although scenarios are inherently uncertain, modelers ignore deviations from "average" scenarios and assume that usage scenarios are fully known. In some cases this assumption is reasonable, for instance when using monthly average weather data and schedules in the performance measures that are calculated for a year and not dependent on day to day variations. However, for the purpose of this thesis the temporal scale of model calibration is typically much shorter, and short term variations are significant. For instance, the wind speed, which is highly fluctuating, is a crucial factor in ELA measurements as will be discussed in CHAPTER 4.

Building energy models are not capable to perfectly represent the reality, which in itself produces another type of uncertainty, i.e., model form uncertainty or MFU (Sun Y. , 2014). Inadequacy in a model varies depending on the choice of specific energy models; a higher resolution model is known to represent the reality more accurately than a lower resolution model. Yet, all energy models approximate the phenomena of physical heat transfer that occur in a building by abstracting their complexity into simplified models. In this process, some physical phenomena are ignored if they are regarded insignificant with respect to their effects on system energy performance. But the simplifications obviously lead to an inherent MFU that should be used in predictive simulations since the importance of MFU on predicted outcomes is significant. If that is the case, it could turn out that the model should be rejected for the selected purpose. In calibration however, no attempt is

made to explicitly quantify MFU. The reason is that calibration poses a different challenge. If a model has large MFU the calibration will be impacted in the sense that the calibration parameters will subsume all or part of (unknown) MFU. Stated differently, the calibration parameters will try to make up for the model behavior that is missed by the model but is obviously present in the observations. This phenomenon of subsumption is intrinsically present in any calibration exercise. If enough calibration parameters are used, one could hope that the calibration parameters actually subsume all of the MFU and thereby make the model more correct than it ought to be. This reasoning drives much of the thinking behind this thesis as one of the major goals is to compare the calibration of low fidelity (large MFU) with high fidelity (small MFU) models. The underlying inquiry is whether the low fidelity model can, after calibration, be as correct as the high fidelity model. This is a fundamental question that cannot be answered in generality but only in specific cases. The thesis focuses on outlining the approach to generate relevant answers. Following above reasoning it should be well understood that the calibrated parameters in a low fidelity model can be far off from the true values of these parameters. They "correct" the model for a selected QoI, but their calibrated values are incorrect. Through the comparison between the calibrated low and high fidelity models it is expected that this effect can be quantified.

2.3.2 ELA quantification

This section provides fundamentals about ELA quantification on the basis of given building conditions and usage/weather scenarios. According to the airflow physics described earlier, it is found that wind pressure coefficient (Cp) and local wind speed (v)

are the parameters that play an essential role in identifying the ELA. Therefore, we need to have a thorough understanding of both Cp and *v*.

2.3.2.1 Wind pressure coefficient

An averaged and simplified Cp can be obtained from several sources and databases (Costola, Blocken, & Hensen, 2009). However, actual wind pressure coefficients can vary largely over the surface of a building. For instant Figure 2.3 shows an example of local pressure coefficient distributions on the surface of a typical tall rectangular building for varying wind angles.

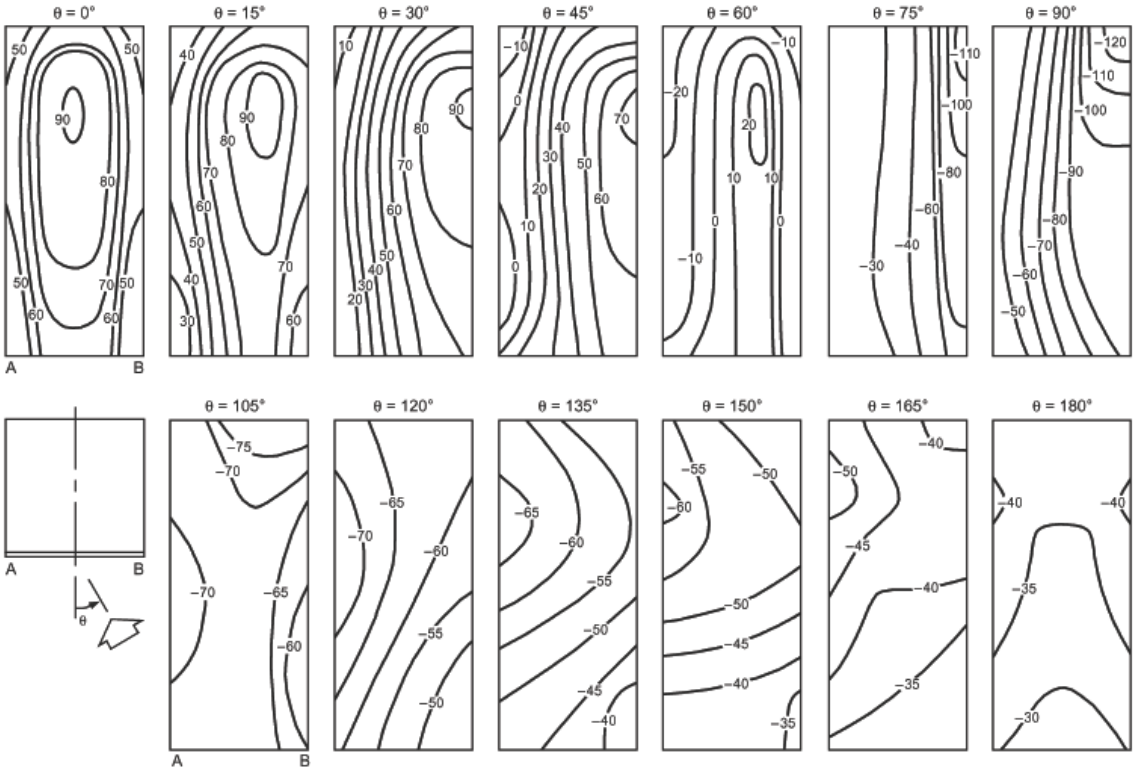


Figure 2.3 Cp contour on a tall rectangular building at varying wind angles (ASHRAE., 2009)

For accurate computation of the pressure fields on the facades, one will need a wind tunnel experiment or a high fidelity flow simulation. As the pressure field will be very

sensitive to global wind direction (measured in the undisturbed situation) we need to repeat the Cp calculation for every possible wind direction. This makes wind pressure computations for infiltration airflow rate calculations overly complicated and computationally expensive. In order to correct the pressure coefficients for variations in wind direction, (Swami & Chandra, 1988) suggest a formula to calculate Cp values for a rectangular free-standing building where wind angle is a parameter in the formula, as given by Eq. (2.10) along with example plot for different wall ratios in Figure 2.4:

$$Cp = 0.6ln\left[1.248 - 0.703sin\left(\frac{\alpha}{2}\right) - 1.175sin^2(\alpha) + 0.131sin^2(2\alpha G) + 0.769cos\left(\frac{\alpha}{2}\right) + 0.07G^2sin^2\left(\frac{\alpha}{2}\right) + 0.717cos^2\left(\frac{\alpha}{2}\right)\right] \quad (2.10)$$

Where

α: wind incidence angle,

G: natural logarithm for the ratio of width (W) of the wall under consideration to the width of the adjacent wall (L).

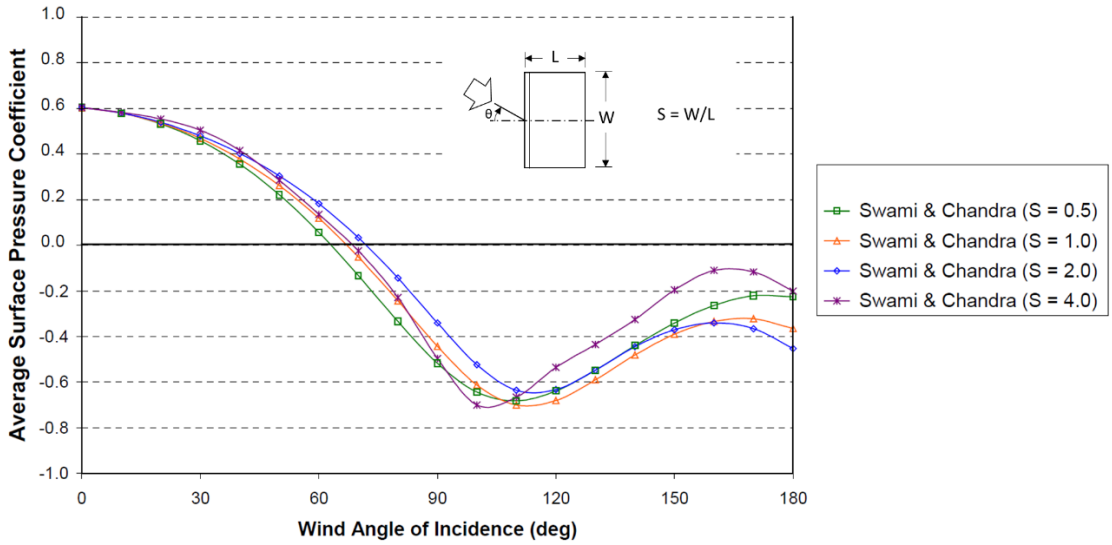


Figure 2.4 Surface pressure coefficient as a function of wind incident angle in the Swami and Chandra formula for different side ratios (Deru & Burns, 2003)

It should be noted that the above model is only valid for cases dominated by free field wind pressure. It would potentially lead to inaccurate results in situations that introduce turbulence to the flow, e.g., high terrain roughness or local shielding, irregular shaped buildings, etc.

In those situations, i.e., where the free field assumption does not apply, we can use alternative methods such as:

- *Cp generator (CpGen)*: a parametric model developed based on experimental data that calculates Cp values based on wind direction and building dimensions and obstacles taking the surrounding terrain into account (Knoll, Phaff, & de Gids, 1995). The Cp generator requires information about geometries and terrain roughness for adjacent obstacles and distant obstacles, respectively. As outcomes, it obtains surface-averaged coefficient values (Cp) for each wall at each wind direction. It can also approximate Cp for grid points on a façade, but they are generally less accurate.

- *CFD simulation*: general purpose high fidelity flow-temperature simulations that have been used to study airflow around buildings with focus on wind pressure on the building façade. However, despite the vast increase in the application of CFD to study wind flow around buildings, it is not common to use it as a source of custom Cp data for building simulation. The main reasons are the required level of expertise and the high cost of these simulations, both in terms of computational resources and user time, when compared to the building energy simulation itself (Costola, Blocken, & Hensen, 2009).
- *Direct wind pressure measurement*: it consists of a wind pressure sensor or transducer to be located on the surfaces of the building. As pressure is applied to the diaphragm of the sensor, a charge is generated on the sensitive surface of the sensor. The high-impedance electrostatic charge is then conditioned externally to a readout or recording device. This method is the most accurate compared to the other methods above; however, it requires high-end devices and special installation which is cumbersome in some cases. Also, the measurement must be long enough to collect enough data for all wind directions. In some cases this is a bottleneck. It should be noted that this adds to the control or rather observability of the experiment and as such reduces uncertainty in the calibration process considerably.

2.3.2.2 Local wind speed

As described in previous sections, the local wind speed can affect the infiltration significantly. This is because the wind data in weather files are usually measured at a meteorological station at a given height, and the approaching wind velocity v for height H of the building is modified from the measured meteorological wind velocity by Eq. (2.9). The exponent indexes γ and τ are regarded to depend on ground roughness, and are taken

to be standard values as in Table 2.1. However, these standard values are based on predominant mechanical turbulence (strong wind). Hence, these indexes should be modified further when the contribution of convective turbulence becomes significant. This implies that the use of constant values, e.g., 0.20 and 0.85 for γ τ respectively, contains an approximation error in calculating the velocity of approaching wind when the effect of stratification becomes strong due to unstable atmospheric conditions (Lim, Ooka, & Kikumoto, 2015). This error contributes to the uncertainty in the calculation of local velocity and hence wind pressures and resulting airflow rate *Q*. Wind velocity has a significant impact on the formulation of wind pressure coefficient. As concluded in the previous section, it is difficult to perform an accurate evaluation of Cp because of the various influencing parameters, including surrounding elements, building configuration, details of the building surface, and the characteristics of the approaching wind. It is clear that the uncertainty in the approaching wind velocity profile causes subsequent uncertainty in the wind pressures on the building surface.

2.4 Propagation of uncertainty

On the basis of the parameter uncertainties identified in the previous section, the uncertainty in the model output is calculated by propagation of the parameter uncertainties through the model. For lack of explicit information on the parameter distributions, a combination of normal and uniform distributions is assumed for all parameters in accordance with the GURA-W UQ repository (Sun Y. , 2014). The parameter ranges established from this generic source are interpreted as central 95% confidence intervals. Where necessary, the normal distributions are truncated to avoid physically infeasible values. The uncertainty in scenario and usage scenarios is not taken from the “vanilla” UQ

repository but based on local on-site observations, with uncertainty to represent unknown variability due to lack of full observability of extrapolation outside the strictly observed period.

Once the sources of uncertainty associated with model parameters (P and S) are quantified, they are propagated through a simulation engine in Monte Carlo (MC) fashion with an appropriate sampling technique (e.g., Latin hypercube sampling) (Wyss & Jorgensen, 1998). The MC method draws random values from uncertainty distributions. This process often requires a large number of samples to ensure convergence to the true probability density of a particular QoI. Hence, when a high-fidelity simulation model is deployed, this method is often not efficient due to high computational burden. LHS alleviates the computational burden by efficiently capturing the variability of the distributions of uncertain parameters. LHS partitions a probability density function into segments by the same magnitude of probability and draws a sample once from each of the segments. As a result, this method ensures the reliability of probabilistic outcomes with much smaller sample size.

CHAPTER 3. APPROACH AND METHODOLOGY

This chapter lays out the research methodology and explains its effectiveness toward the specific questions under study. The previous chapter focuses on identifying sources of model discrepancy, i.e., the gap between predictions and observations and the role that calibration parameters (i.e., the model parameters for which "best estimates" based on measurements are derived) play in this. These results are used to arrive at novel interpretations of the value of calibration for different model fidelities and different measurement data sets. The inherent "value" of the resulting model will be based on quantification of the fitness of the calibrated model, albeit that the scope of this thesis limits this to one particular fitness measure in one specific case.

The first part of the work is conducted along two main routes that both lead to an estimation of the primary parameter (vector) under study, ELA. The diagram below shows the two distinct ways to measure ELA, i.e., via direct measurement and indirect calibration from simple, routinely available measurements.

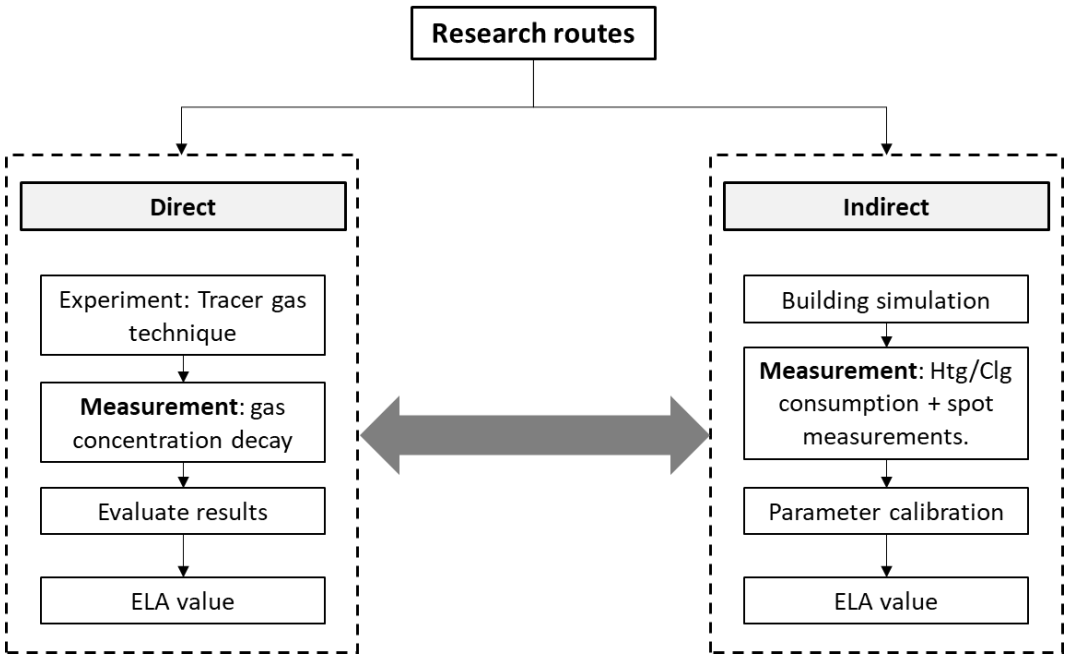


Figure 3.1 General outline of the core research topic: ELA estimation

From the testing techniques presented in the first chapter, the tracer gas method has been chosen for the direct calibration approach in the left part of Figure 3.1. The obvious reason for this choice is its practicality and relative simplicity. One measures the decay of a released gas concentration at a zonal level which is assumed to be the direct result of infiltration. On the other hand, indirect calibration basically depends on the measurement of the indirect effects of infiltration (e.g., on energy and temperature) and uses simulation tools and calibration techniques to estimate the "best estimate" of ELA.

3.1 Methodology framework

As depicted in Figure 3.2, the first step in the process of this work is to conduct sensitivity analysis of all candidate building parameters for a building under study. This step informs the choice of the set of calibration parameters and other "regular" parameters that are fixed, albeit, as explained below, that the set of regular parameters will be assumed

to be variable/uncertain to some extent. The next step is to use the model with a selected calibration parameter set in a calibration process that finds best estimates of the calibration parameters by making the difference between model outcomes and real data from the conducted direct or indirect experiments as small as possible. This step may be repeated multiple times with different sets of calibration parameters to verify how sensitive the outcomes are to the choice of the calibration set. In all steps, the ELA parameter(s) are always in the calibration parameter set. Every calibration trial with different parameter sets will lead to a new value (distribution) of the ELA parameters. If some of the regular parameters are assumed uncertain, the calibration will be repeated many times for different samples from these uncertain parameters (sampled in a way that they are represented according to their occurrence probability). For each chosen calibration set, the resulting values of an ELA parameter will then take the form of a probability distribution which represents the uncertainty in the calibrated value of ELA as a result of the (assumed) uncertainty in the regular parameters. It should be remembered that the result is only as good as the uncertainty estimates of the parameters allow.

All resultant ELA distributions will be compared and analyzed to reveal the effectiveness of the different calibration sets and the sensitivity of the calibrated parameters to other partially unknown parameters.

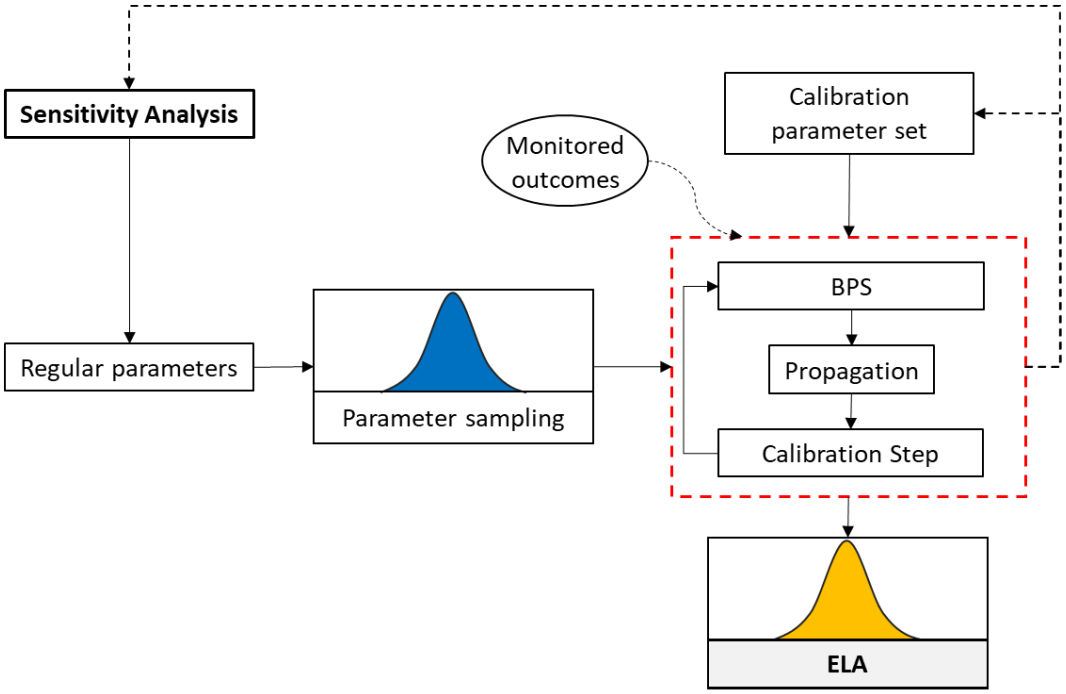


Figure 3.2 Proposed calibration process

The following sub-sections discuss every part of this process separately.

3.1.1 Sensitivity analysis

A sensitivity analysis (SA) of all uncertain model parameters is conducted by representing their possible values with probabilistic distributions and test their influence on a chosen outcome (QoI) of the simulation. SA reveals which parts of the total output uncertainty can be attributed to uncertainties in individual inputs or groups of inputs, and it ranks the importance of input parameters based on their influence on the uncertainties of model outcomes. Commonly used probability distributions include uniform distribution, triangular distribution, normal distribution, and log-normal distribution. The classical interpretation of probability considers the sample points of a specific parameter as observable realizations of uncertain events, and assumes that the frequency records of them allows the inference of probability distributions. The other view involves a quite different

interpretation. A subjective interpretation considers the probability distributions as a reflection of the decision-maker's subjective preferences following a 'rational preference' set of axioms, or the degrees of belief about these variables without necessarily referencing to frequency observations of them. These competing interpretations of a standard probabilistic setting are quite different, but they imply rather similar practical implementation features. For the targeted SA, it is sufficient to define plausible ranges of parameters and assume their distribution (e.g., uniform or normal) and rank their importance through well-established methods. The resulting ranking and interaction between parameters is used to determine the calibration set and regular set, where the boundary between the two sets is not fixed but variable.

3.1.2 Data collection for the calibration target

As introduced in section 1.2.4.1, well-designed procedures regarding the use of information and data throughout the calibration process have a significant impact on the success of the outcome. Most studies on calibration use compliance with calibration standards, like ASHRAE Guideline 14-2002 (ASHRAE, 2002), as validation criterion. Regardless of whether the calibrated model passes the test, there is usually insufficient proof to test the validity of calibrated parameter values. Clearly the quality and volume of the monitored outcomes as well as fidelity of the model play a role, but it is very hard to make their roles explicit. The most common type of monitored data in building performance management is the monthly utility bill, as it is readily available and reliable in most cases. The use of additional hourly or sub-meter data, building management systems (BMS), in-situ monitoring, etc. is discussed in (Heo & Zavala, 2012). This type of data is obviously more informative than monthly data for model validation because of the

embedded dynamic characteristics. Hence, increasing the volume and type of monitored outcomes helps to overcome uncertainties associated with the experiment setup, such that one can appropriately attribute the discrepancy between assumed and observed parameters.

In terms of data variety, it is always desirable for a data set to include simultaneous measurements of different types of variables, such as submetered energy use, lighting and plug loads, and indoor temperatures. When different types of variables can be simultaneously monitored over a relatively long period, model verification can be performed more comprehensively. One objective of this thesis is to measure the fidelity of a model over different resolutions and granularities of input data. Therefore, different setups of data monitoring are applied in the case studies. The monitored outcomes include:

- Energy: hourly electricity consumption.
- Temperature: indoor zonal temperatures.
- Meteorological conditions: on-site dry bulb temperature, relative humidity, global irradiance, and wind speed and direction.

3.1.3 Model calibration technique

Calibration is the essential part in the framework of this thesis. The most common automated model tuning approach formalizes calibration into a deterministic parameter estimation problem, minimizing the discrepancy between measurements and model outcomes as objective. This employs numerical algorithms to find the parameter values (for all parameters in a chosen calibration set) that minimize this objective function.

As implied in section 1.2.3, the issue with any calibration is that some parameters may subsume the effect of other parameters and inherent model discrepancies (MFU) as well as wrong scenario data and observation data errors. Jointly, they increase the uncertainty in the resulting model and more specifically make parameter estimates less reliable. The subsumption of these exogenous disturbances in the calibration parameters is directly intertwined with the interaction between all calibration parameters. The proposed approach to monitor and, if possible, quantify this effect is depicted in Figure 3.3.

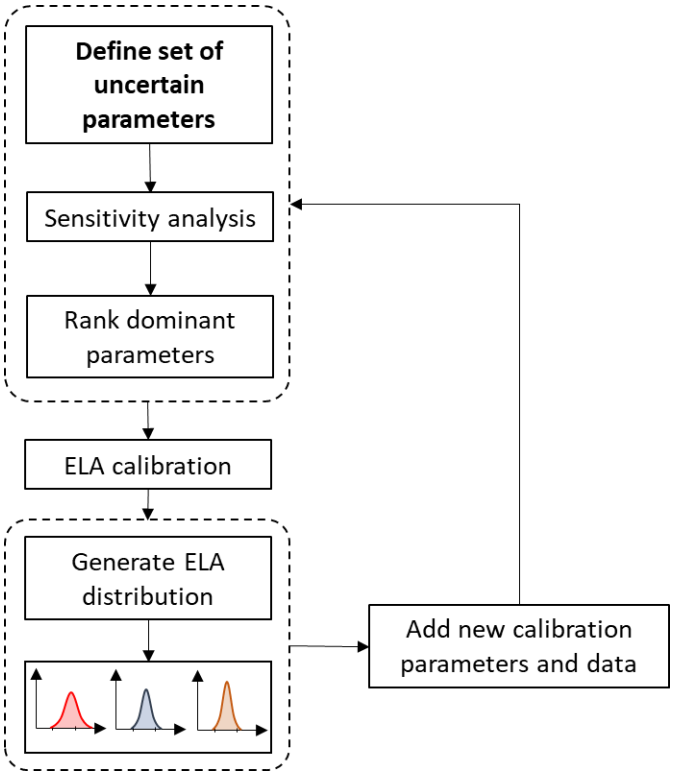


Figure 3.3 Iterative calibration steps to determine calibration parameter sets via sensitivity analysis

The process starts with defining all uncertain parameters (including ELA) followed by a sensitivity analysis to rank the most dominant parameters at this stage. The outcome of this step is to separate parameters into a calibration set and a regular set. Since our target

is to determine the value of ELA, we start the calibration with a small set containing only ELA as calibration parameter(s). Out of that calibration we can generate a distribution for the ELA vector by taking sample values for all uncertain parameters and repeating the calibration for every sample. The reduction of the calibration set is a repetitive process where every time additional calibration parameters are added. For every set a new ELA distribution is found in the calibration step. It could be expected that all ELA distributions are not the same due to the interaction between the calibration parameters. At this stage we apply a sensitivity analysis technique to more objectively account for parameter interaction effect. Particularly, the method used for this is the Morris method (Morris, 1991) which ranks the effect of uncertain parameters on ELA distributions. We stated earlier that the Morris method has been recognized as a suitable screening technique for building energy models (de Wit, 2001); (Moon, 2005). The method is computationally efficient as it is able to rank the sensitivity of many uncertain parameters using relatively small samples. Moreover, the method does not assume the relationship between parameters and model outcomes as linear, and evaluates the effects of parameters on the model outcome over the whole parameter space by exploring multiple regions sampled from the parameter space (Heo Y. , 2011). Hence, the method can capture nonlinear effects of individual parameters and interaction effects among parameters. More technical details are described in section 5.3.1.

3.1.4 Model validation

A crucial question is whether the proposed approach leads to a more accurate ELA estimate. This can only be judged against a known true value. However, as explained above, there is no established way that would give the true value of ELA. Reference values

generated in previous research as will be discussed in section 4.1, inform only the range of possible ELA values. Therefore, this study relies on the best available method to establish a "truth value" for ELA, which is the direct experiment with tracer gas to measure infiltration for a given building zone directly and derive a best estimate of ELA. The test is done at zonal level so that we can derive ELA for each façade or part of a façade of the building. There are test factors that have to be considered throughout the test period. They reflect experimental assumptions and can be described as follows:

1. The tracer gas mixes perfectly.
2. The effective volume of the enclosure is identified.
3. The zone only exchanges air through the façade; i.e., the other enclosure parts and any HVAC opening are sealed perfectly.
4. The factors that influence air infiltration remain unchanged throughout the experiment period.

The following conditions apply to the conducted experiments:

- CO2 is the gas used for the test with high concentration of 500 ppm and neutral level of 3500 ppm.
- Meteorological parameters are recorded and include: wind speed and direction, temperature (indoors and outdoors), relative humidity.
- Pressure sensors measure the wind pressure on the tested facades (only used in one of two case studies)

It is clear that test assumptions do not apply perfectly, and some conditions cannot be monitored with great accuracy. These facts lead to uncertainties in the experiment that

compound into a range of possible values of the derived ELA. Wind pressure is one of the main driving forces of infiltration and is one of the main sources of uncertainty that affects the resulting ELA value. Therefore, even with in-situ direct measurement, we still expect a (possibly large) range of values in the generated ELA estimate.

As discussed in the previous chapter, the formula for wind pressure uses a reference wind speed as the basis for the wind pressure coefficient Cp. Cp for a given location on a façade as explained earlier can be obtained in different ways either by measurement, empirical laws or using software. Ideally, wind tunnel is the most accurate method to derive Cp, but because of its complexity it is nearly impossible to be done for every study. Alternatively, we can use Cp generator. As explained in section 2.3.2.1, it is a tool that calculates wind induced surface pressure coefficients in the urban context. It is originally developed from experimental (wind tunnel) data to determine Cp values for any wind direction, building dimensions and adjacent and distant obstacles in the urban context (Knoll, Phaff, & de Gids, 1995). For the purpose of this study, it is important to investigate the uncertainty in Cp thoroughly. Hence, more about uncertainty quantification of Cp will be discussed in CHAPTER 4.

As alluded above, it is important to account for any uncertainty in the tracer gas experiment. By modeling all factors as uncertainties, the derived ELA will be in the form of a distribution. In fact, we will find different distributions depending on the assumed uncertainty (width of distribution) of the Cp estimates that we use. The diagram below visualizes the resulting ELA values in response to different types of Cp fidelities.

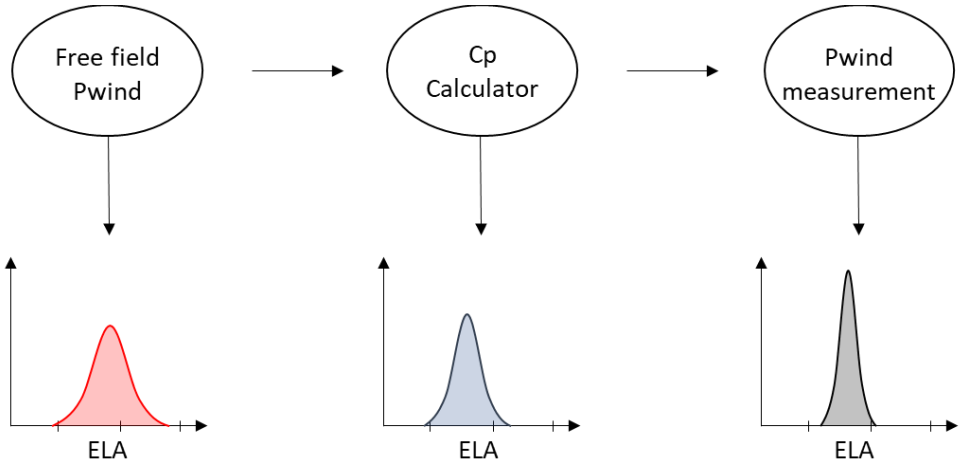


Figure 3.4 Visualization of the impact of different Cp calculation methods on ELA estimates

The figure depicts that with increasing accuracy of the experiment, the estimated ELA distribution should become "skinnier" thus increasing our confidence in the ELA estimate. One ultimate way to account for the effect of wind pressure is to measure it directly during the tracer gas experiment, e.g., with pressure transducers at the outside of the facade. Obviously the error in the measurement devices and sensors themselves and unaccounted disturbances during the experiment can never be avoided.

3.1.5 The role of Cp in the indirect calibration

When estimating ELA through an indirect experiment (i.e., calibration of the building energy model), we still need a reliable estimate of Cp to enable the calibration of ELA. This implies that Cp enters into the sensitivity analysis and hence into the calibration, either as member of the calibration parameter set or in the regular parameter set (with an uncertainty distribution that will vary with the method employed to derive Cp estimates). This constitutes a major contribution to the current understanding of calibration methods

under the combined uncertainty of model parameters and external conditions (in this case local wind pressure).

The significance of this work is to better understand how to improve the estimate of a specific calibration parameter to be as close as possible to the true value. In particular, the challenge is to verify whether the proposed iterative process of sensitivity analysis and increased number of calibration parameters leads to more confidence in the best ELA estimate where "best" estimate is defined as one that gives the smallest difference between model outcomes and measurements. This process will be repeated for different calibration cases, where each case deals with a certain simulation tool and a certain amount of measurement data. Based on the approach discussed above, every case results in a different ELA parameter distribution. Figure 3.5 provides a visualization of the expected result. Each distribution represents the best estimate obtained for a case as described above. It also shows the prior distribution and two additional dotted lines representing the results of direct calibration based on field experiments with different fidelity of Cp. The other two solid curves represent two cases with specific model fidelity and data resolution. The figure is only intended to prepare the reader for the next chapters that establish the estimates in two distinct case studies. Based on those, conclusions will be drawn with respect to type of model and outcome data that is needed to produce ELA estimates with a certain level of confidence. For this we use statistical measures to define the distance between best ELA estimate and experimentally measured ELA.

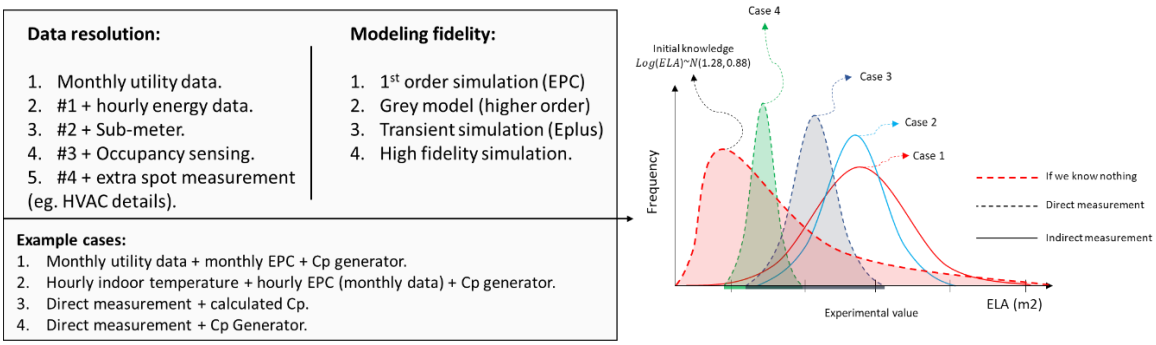


Figure 3.5 Illustration of expected ELA outcomes for different calibration cases

The comparison of the resulting ELA distributions is expected to lead to important findings about the relevance of direct versus indirect calibration and the role that data availability and model fidelity plays in the outcomes.

The process will be repeated for several model fidelities and different monitored outcomes to show how both will impact:

- Approximation of measured QoI (traditional measure of goodness of the calibrated model)
- Approximation of ELA (this represents a measure of goodness of the calibrated parameter values)

It should be noted that both measures are good indicators of the effectiveness of the calibrated model, but they nevertheless fall short of delivering a measure for a "one fits all" conformance test. For this an application tailored fitness measure is more adequate as elaborated in the next section.

3.2 Evaluation of a specific calibrated model

This research is also aiming to provide a reliability measure that helps to determine how much confidence one can have in using a calibrated model for a specific task, or in a specific simulation scenario. A novel fitness measure, for now loosely defined as the "measure of reliability" of a calibrated model would be the ideal outcome. If the calibrated model is proven reliable to be used for a task it is deemed fit for this task. By defining tasks of increasing complexity, a model's fitness level can be defined by the highest task it can perform. In this manner, the hope is that fitness level of a calibrated model can be predicted based on the resolution and quality of the data, number of calibration parameters and the fidelity of the underlying model. It is well recognized that such a generic assessment methodology is the holy grail of calibration research and some modesty is in order with respect to what this dissertation can achieve. In this study we will therefore only test our thinking on one particular setting. This setting is the benchmarking of existing buildings where a fitness measure is introduced that can effectively test the reliability of a calibrated model to verify its reliably in an existing building benchmarking application.

3.2.1 Comparing existing goodness of fit to proposed fitness levels

Section 1.2.4.2 discusses the statistical metric to evaluate the goodness of every calibration model. This type of validation is in fact the only criterion established in ASHRAE Guideline 14-2002 (ASHRAE, 2002) which provides a standard procedure for the whole-building calibrated simulation approach. First, a modeler should plan the calibration exercise by specifying a simulation software, the unit and interval of monitored data for calibration (i.e., monthly, hourly), and acceptable tolerances for model validation.

Second, the model requires some information to be collected onsite (e.g., building dimensions, construction specifications, system nameplates information, occupancy and operation schedules, and whole-building utility data). Third, based on collected information, a simulation model of the building can be constructed corresponding to other assumptions that are deemed to represent the actual building reasonably. Finally, one compares the simulation model outcomes to measured data, and refines the model until the discrepancy between predicted energy uses and measured energy use falls within acceptable tolerances. Under the assumptions of normal linear regression, one can translate these statistical metrics into uncertainty of certain model predictions, e.g. energy savings.

Within a calibration process, uncertainty analysis estimates the level of confidence that can be placed in our predictions. Without such a measure of "fitness", it is impossible to judge the accuracy of the prediction for making decisions. In other words, trusting a calibration model output associated with high uncertainty may lead to considerable risk to the validity of the model's results in a decision context. As has been stipulated before, the fidelity of the model is expected to play a dominant role in how fit the resulting calibrated model is. But it could be expected that for simple tasks a low fidelity model could do well enough, even if the values of the calibrated parameters are far removed from their truth values. The fidelity levels to be considered in this thesis (Figure 3.5) can be ranked within a spectrum from first order simulation (EPC) to high fidelity simulation (EnergyPlus with AirflowNetwork). Ideally, fitness levels could be determined for each specific case as depicted in Figure 3.7.

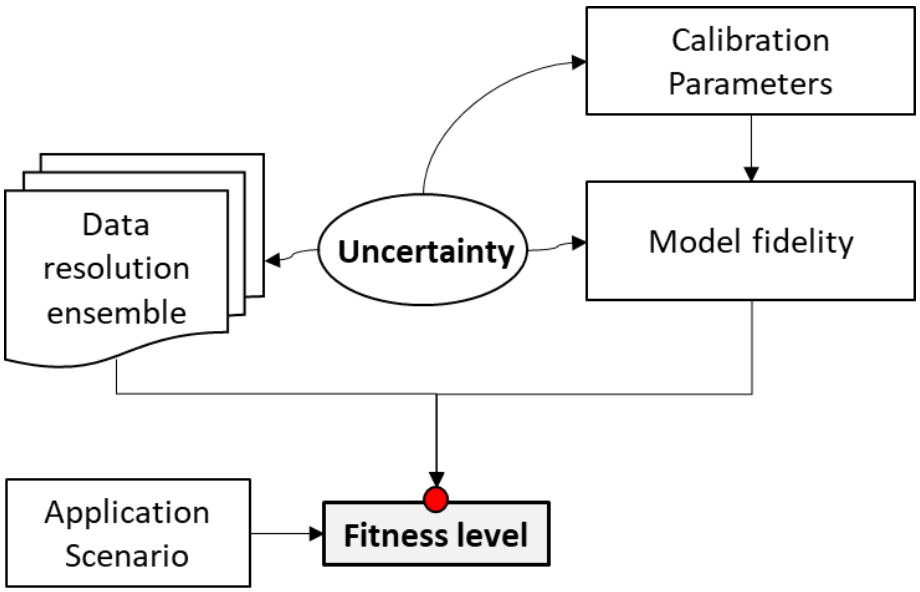


Figure 3.6 The case specific determination of fitness level

Fitness levels established this way could serve as a guide to determine what monitoring plan one needs to calibrate certain parameters (e.g., ELA), such that the resulting model is fit enough to perform a given task. To make this explicit, Figure 3.6 shows what type of result could be expected; each red dot represents the maximum fitness of a given case, i.e. a given combination of data resolution and model fidelity. The maximum value for each case is determined through iteration with different calibration and regular parameter sets. Once the maximum fitness for each case is established one can set a threshold (or required fitness level) for the calibrated model in a given scenario as shown in Figure 3.7, and find which case is fit enough to meet this requirement. This thinking requires that relevant decision scenarios and fitness measures have to be developed first. The vertical axis in the figure is for now undefined. It would have to be based on a set of application scenarios ordered along the axis according to increasing demand on model correctness. This poses a major challenge for the research community as a whole. The fitness level is expected to be higher for increased data resolution and model fidelity.

However, these levels are highly dominated by the choice of calibration parameters, which complicates the process.

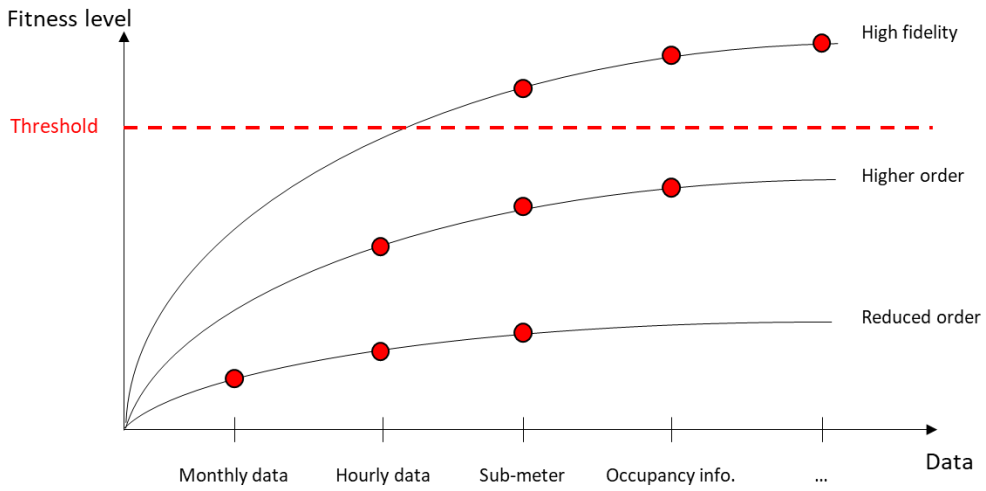


Figure 3.7 The expected result of the fitness determination

The projected outcome from the fitness levels is to determine which factors in the calibration have the most influence on the overall result so that one can make the judgment about what additional effort will make the calibrated model more effective. This objective is over-ambitious at this stage and a general treatment that would underpin Figure 3.7 is not attempted in this thesis.

3.3 Contribution to the field

The thesis develops a theoretical foundation and workable calibration process for a deeper understanding of building energy model calibration. It ties together model fidelity and data resolution in the presence of uncertainties and verifies the role of the choice of a calibration parameter set. It introduces a novel approach to the definition of fitness measures that could be shown to be more versatile and relevant than the present ASHRAE measures. Throughout this thesis, the approach is demonstrated and validated on the

calibration of ELA parameters, which are recognized as notoriously difficult to calibrate as their role is often subsumed in other parameters in models that pass the ASHRAE test. Revealing the latter effect through iteration with different parameter sets is a second important contribution of the thesis. The diagram below illustrates the computational framework with speculative use for the quantification of novel fitness measures.

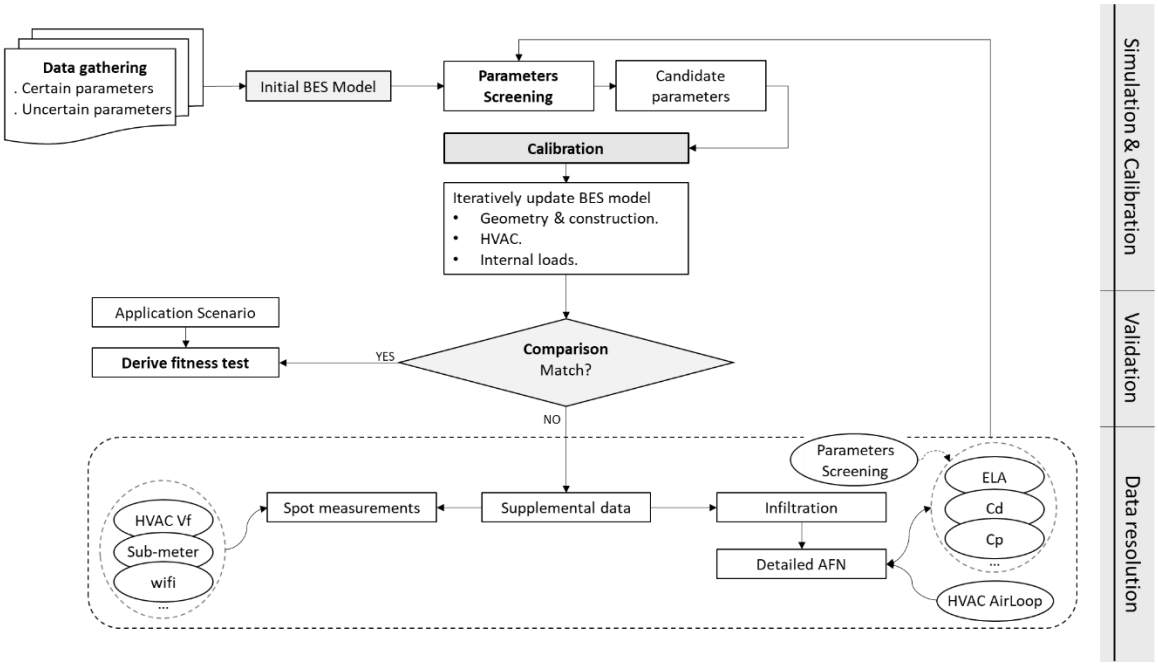


Figure 3.8 The overall iterative calibration process

The methodology described above is implemented on two case studies, one is a commercial building, and the other one is a residential building as described in CHAPTER 6 and CHAPTER 7.

CHAPTER 4. ELA ESTIMATES DERIVED FROM TRACER GAS EXPERIMENTS

This chapter details the derivation of ELA estimates from in-situ tracer gas experiments, using the physical models introduced in CHAPTER 2. The explicit quantification of all inputs in this direct calibration is conducted with focus on wind speed and Cp, which have been established as the major sources of uncertainty in the experiment. The underlying research issues are elaborated in the following sections.

4.1 Prior estimate of the uncertainty of ELA

As introduced earlier, ELA is not a material property that is directly measurable. It is a defined or "surrogate" property that can only be inferred from an experiment. It should be noted that the scope of ELA is in principle undefined and left over to the modeler's choice. For example, it could relate to a whole façade or the façade parts that border a given building zone, or even a subarea of a given façade, such as a window.

National standards and guidelines establish recommended ELA values that correspond to the normal and the best practice for different types of buildings and more specifically facades. One example is the ASHRAE fundamental (ASHRAE, 1997). Most of these values correspond to pressurization tests at specific conditions. The problem with the standardized values is that disparities have been reported between the specific measured data and the standard values, which indicates that actual ELA of existing buildings is often either much higher or lower than the values recommended by the standards.

Before attempting a calibration, it is important for this study to identify what we know about ELA beforehand as this will serve as a prior guess for the ELA values that are going to be estimated. In order for that, we need to obtain data on how leaky a building is in general, i.e., without much specific information. A recent study quantifies the uncertainty in ELA across a large spectrum of directly measured buildings in many locations and situations (Sun, et al., 2014). This is part of a larger effort that provides energy modelers with quantified uncertainty distributions for a variety of parameters and discrepancies in the internal modules of a building energy model.*

The uncertainty in ELA was quantified based on analyzing measured whole-building envelop airtightness data as summarized by (Emmerich & Persily, 1998). The majority of these data are previously reviewed by (Persily A. K., 1998) with some additional buildings added. The air leakage values are normalized by the area of the aboveground portion of the building envelope. The basic ELA formula in this situation is an enhancement to Eq. (4.1), which mainly focuses on the cases where a blower door test is used. This expression is given by:

$$Q = \frac{ELA}{10000}\sqrt{\frac{2}{\rho}\Delta P}\left(\frac{\Delta P_f}{\Delta P_r}\right)^n \tag{4.1}$$

Where

Q : air flow at ΔP_f [m^3/s],

* It focuses on realization uncertainty at the time that there is only a designed building. It has to be stipulated that this does not apply in the case of an existing building when on-site inspection can reveal many details that will help the building auditor to adjust an uninformative prior (as developed in above mentioned study) into a more precise adjusted prior. However, the distinction needs no further stressing as the ELA prior is in fact not used in this thesis and only presented as baseline to compare derived ELA estimates.

ΔP_f: applied pressure difference [Pa],

ΔP_r: reference pressure difference [Pa],

n: pressure exponent [-].

It should be realized that all tests are done by a standard blower door test at $\Delta P_f =$ 75 Pa. The outcome can be used to derive an ELA estimate at $\Delta P_r = 4$ Pa which is frequently used since it represents the average magnitude of the pressure difference under real life conditions. n is a factor that accounts for the shape and average length of the cracks. It is basically unknown, but studies have shown that a value of 0.65 is a good estimate with some variability between 0.6 and 0.75.

The ELA value was derived for each building in the dataset by (Emmerich & Persily, 1998) creating multiple samples of ELA for varying *n*. Assuming that every building comes from the following distribution:

$Log(ELA) \sim N(\mu, \sigma^2)$;

a distribution for ELA can be established. Figure 4.1 shows the result as a log-normal distribution with mean 1.282, and standard deviation 0.879 (Wang Q. , 2016).

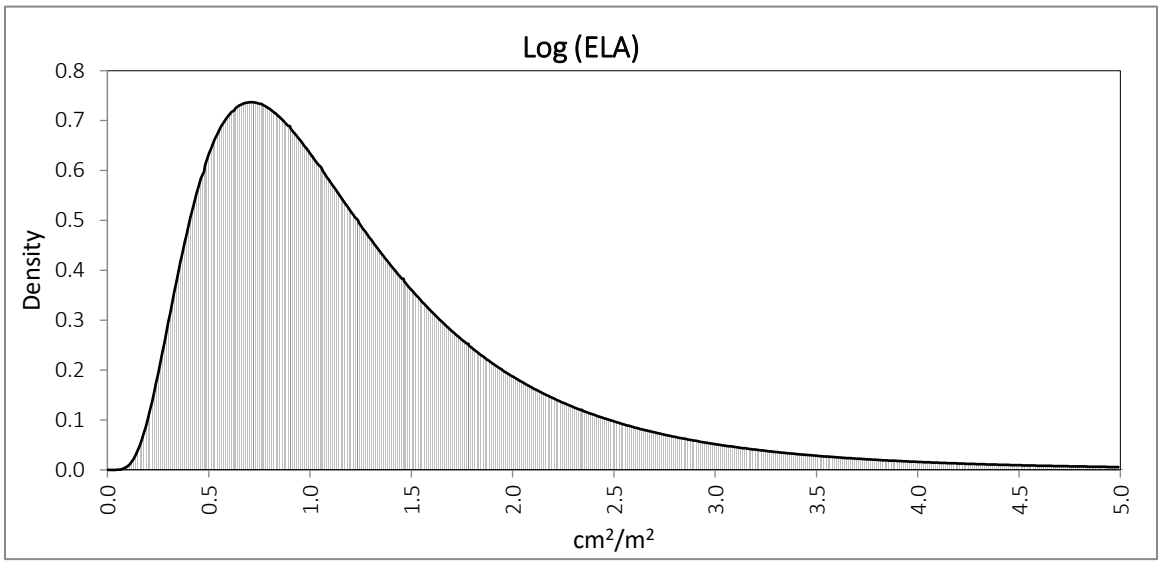


Figure 4.1 Predicted distribution of Log(ELA)

Note that this distribution is constructed from data for a large set of buildings of all kinds, including some old and very leaky buildings, hence the long tail with high ELA values. If one had no knowledge whatsoever about a building (new or existing), this distribution would be a useful prior albeit a rather uninformative one. Obviously as one has more information about the building such as age, façade technology, type of assembly method, contractor reputation, performance contract (if any), etc., one will use this information to make a more informative prior distribution. This is however not within the scope of this thesis as it is not relevant for the chosen method. Rather, the wide distribution in Figure 4.1 is used as a basic reference (or indeed as an ELA prior distribution in the Bayesian sense if no other information is available) for the values that will be obtained from all ensuing calibration efforts.

4.2 Wind pressure coefficient

As introduced in section 2.3.2.1, the wind pressure coefficient (Cp) is the key boundary condition that induces airflow between the external environment and building zones. Primary sources such as full-scale experiments or reduced-scale experiments in wind tunnels can provide custom and detailed Cp data for a specific building shape. Sometimes an urban scale CFD simulation can be a plausible and somewhat less expensive alternative (Costola, Blocken, & Hensen, 2009). However, both methods are time-consuming and expensive, which prohibits their use in routine building (design) assessments. Analytical formulas and databases are commonly used as alternative sources to obtain Cp data, which are straightforward to use and therefore found in most of the airflow modeling and building performance tools. A disadvantage of the latter sources is the limited amount of Cp data that is typically provided or derived, often based on simplifying assumptions. One significant assumption adopted by many of these sources is the validity of using surface-averaged Cp values instead of local Cp values with high spatial resolution.

From the sensitivity analysis conducted in our case studies in CHAPTER 6 and CHAPTER 7, it is shown that the effect of the uncertainty in wind pressure coefficients on the output of a building simulation model, relative to the effect of other uncertainties, is significant, which means that the contribution of Cp to the overall uncertainty in predicted outcomes is potentially significant. Hence, the aim of this section is to quantify the uncertainty in Cp more rigorously and generically, based on existing experimental data.

4.2.1 Experimental data

A widely accepted method to assess wind pressure coefficients of a building is a wind tunnel study. In a wind tunnel, a scale model of the building and its surroundings is immersed in a simulated boundary layer flow. Thus, Cp can be assessed from measurements of both surface pressures and wind velocities or dynamic pressures.

It must be noted that these wind tunnel values only approximate the desired full-scale values with some level of uncertainty. Indeed, due to scaling effects and simplifications in both the simulated boundary layer and the wind tunnel model, the acquired pressure coefficients only approximately capture the relation between wind velocities and surface pressures on the full scale building in an actual wind field (Sherman M. , 1998). For instance, buoyancy effects, which may significantly affect the boundary layer flow at low wind speeds, are absent in a typical wind tunnel flow.

Although potentially important, uncertainties resulting from these effects are not addressed in this chapter. We consider a wind tunnel study as a first and basically only step in the assessment of Cp. Therefore, given the specification of a wind tunnel experiment, we can identify the uncertainty in the measured pressure coefficients by comparing them with Cp values derived by existing tools and analytical methods rather than in a wind tunnel experiment. This follows the standard approach used in uncertainty quantification (UQ) when the discrepancy between a high fidelity and a lower fidelity model is characterized and quantified.

Accordingly, the ‘‘Tokyo Polytechnic University (TPU) wind pressure database’’ provides the experimental wind tunnel data used in this part of the intended uncertainty

quantification (Quan, Tamura, Matsui, Cao, & Yoshida, 2007). The database contains the results of test models in a boundary layer wind tunnel with a test section 2.2 m wide by 1.8 m high. The atmospheric boundary layer is simulated by turbulence-generating spires, roughness elements and a carpet on the upstream floor of the wind tunnel's test section. Different wind profiles were used to build the database. In most experiments, the mean velocity (with power-law exponent $\alpha = 0.20$) and the turbulence intensity profiles were in accordance with class III (rural area) in Table 2.1. The turbulence intensity at a height of 10 cm was about 0.25, and the test wind velocity at this height was about 7.4 m/s (Costola D. , Blocken, Ohba, & Hensen, 2010).

Our uncertainty analysis needs to consider the effect of various surrounding conditions to be in agreement with the variation of the wind tunnel experiment. Accordingly, the target model of the experiment was set at the center of a turntable of 200 cm diameter surrounded by similar buildings models with different variations and surrounding densities (i.e., C_A (0.1, 0.15, 0.20, 0.25, 0.30, 0.40, 0.50, 0.60)). The formulation of these densities is following Eq. (1.1):

$$C_A = \frac{area\ occupied\ by\ buildings}{area\ of\ site} = \frac{bd}{BD} \tag{4.2}$$

Where, b and d are the breadth and depth of the buildings. B and D are the average distances between corresponding points on adjacent buildings in two coordinate directions, as shown in Figure 4.2.

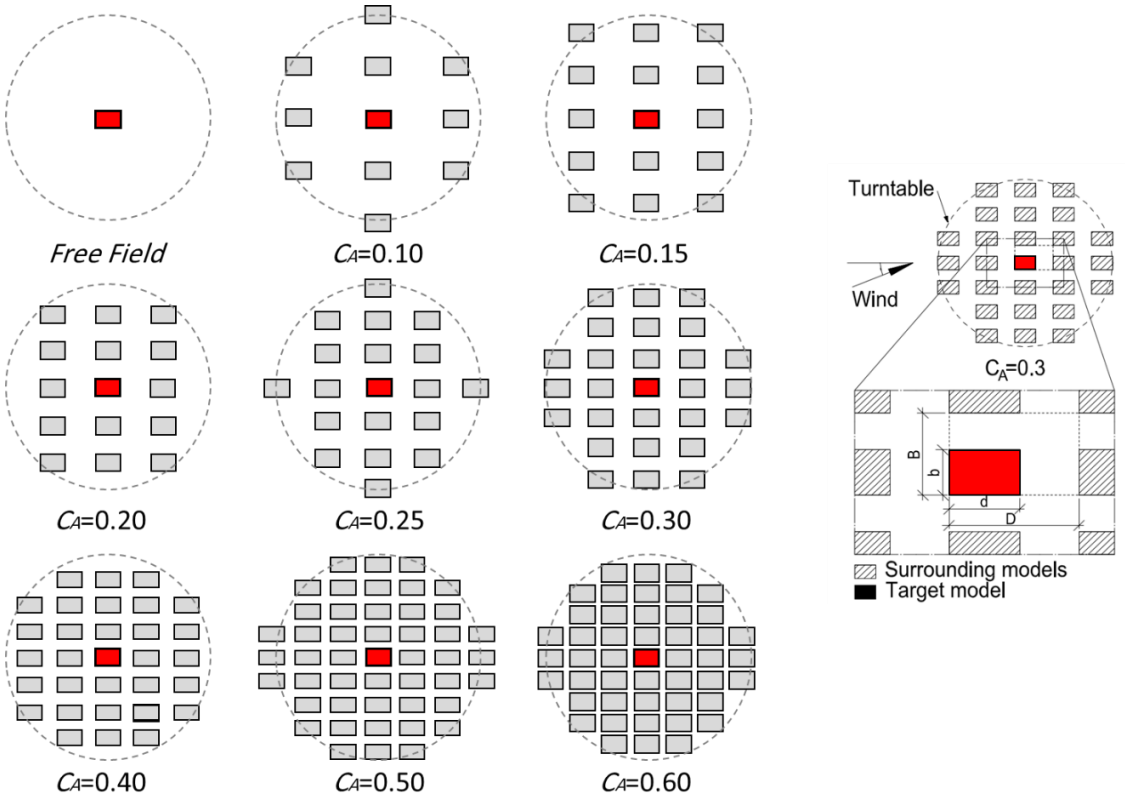


Figure 4.2 Variations of surrounding densities for Cp measurements

From all surrounding densities presented in the database, we used *Free Field* as well as C_A (0.10, 0.15, 0.30, 0.50) to quantify uncertainty against our Cp modeling techniques. The outcomes provided in the published database are Cp values at every measured tap on the surfaces of the model. We select 12 taps that are evenly distributed over each surface and used for comparison with Cp values generated from other models. It must be noted that the number of wind directions tested is not consistent for all cases (e.g., free field has only 7 angles, and C_A 0.10 has 19 angles), which could somewhat bias the final outcome of the uncertainty quantification.

The test models in this database include three kinds of roof types (i.e., flat, gable, and hip) with different variations of shape and size. In our study we select the sizes that

closely match our case studies, which are i) flat-roofed model with dimensions: Hight = 6cm, Breadth=16, Depth=24; and ii) gable-roofed with dimensions: Hight=12cm, Breadth=16, Depth=24, roof slop=26.7º. All these data are used to quantify the uncertainty associated with the empirical models as described in the following section.

4.2.2 Calculation models for Cp

To obtain a first impression of the uncertainties in Cp derived from existing data, *Swami&Chandra* and *CpGen* from section 2.3.2.1 are used to estimate the pressure difference coefficients for wind angles between 0º and 180º as shown in Figure 4.3 and Figure 4.4.

As introduced earlier, the model by Swami&Chandra is developed to calculate façade averaged pressure coefficients. CpGen modeling has some superior capability over Swami&Chandra, as it can simulate Cp in situations with added geometric details. CpGen is indeed based on regressions over a wider range of geometric parameters generated by a wide range of variants in wind tunnel tests. It remains an open question though whether the Cp outcomes of CpGen deliver (on average) a better result than Swami&Chandra. CpGen produces Cp results under different variations (i.e., free field or with obstacles). Figure 4.3 shows that the spread of outputs from the different models is not very considerable under the free field condition; however, it becomes more significant when obstacles are introduced into the model as presented in Figure 4.4. In both models, the degree of consensus between the two alternatives strongly depends on the wind angle.

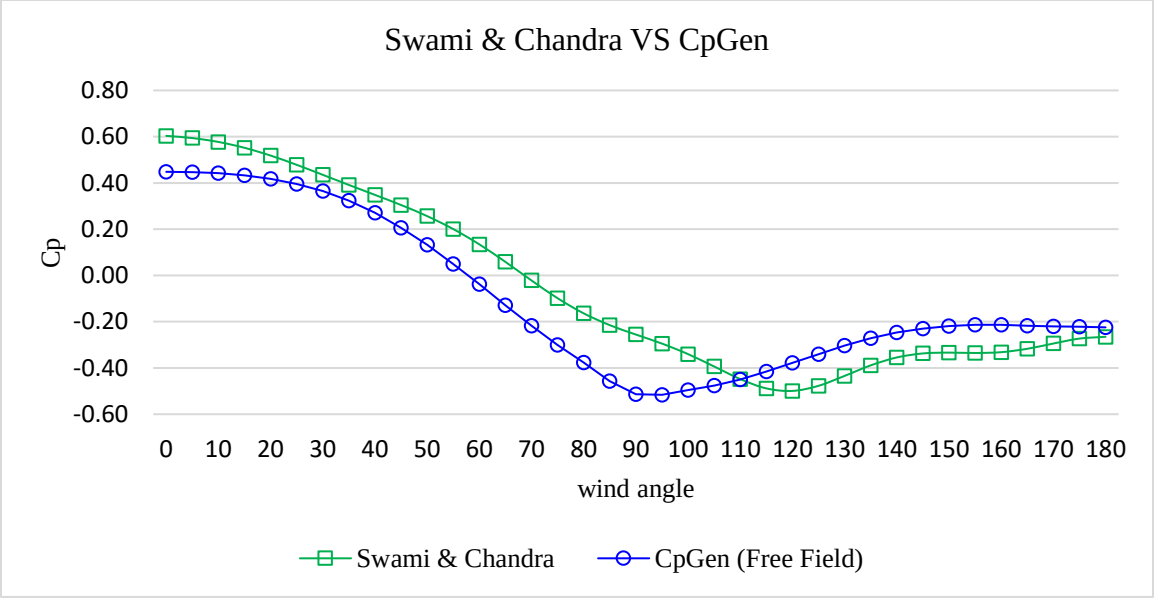


Figure 4.3 Comparison of Cp calculations under Free Field condition

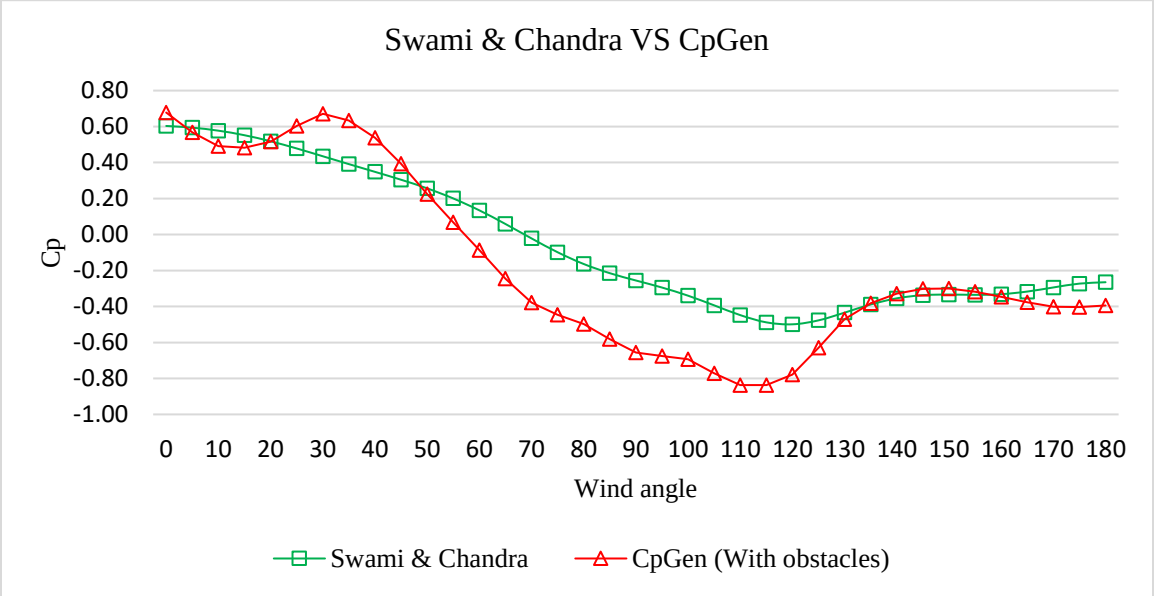


Figure 4.4 Comparison of Cp calculations with obstacles condition at a density level of C_A = 0.10

If we would consider adopting the Cp points in the model outcomes as a measure of the uncertainty, it is important to contemplate which factors contribute to that scatter

and, more importantly, which do not. Accordingly, the model outcomes depend on additional choices by the analyst:

- Several models require characterization of the wind velocity profile;
- The surroundings of the building under study have to be classified in a "shielding" class that fits the given case.

Although the above comparison seems to supports the assertion that significant uncertainty exists in wind pressure coefficients, evidenced by existing data, it does not provide a proper basis to assess this uncertainty in a particular Quantity of Interest (QoI) as predicted by a building simulation. More trusted data is required to provide reliable inputs to the simulation and to assess the impact of Cp uncertainty as well as other sources of uncertainty. Following standard UQ methodology, the Cp uncertainty is quantified by comparing "real" and modeled Cp values. The next section discusses the procedure in which the TPU database is used to quantify the uncertainties in Cp.

4.2.3 Quantification of Cp uncertainty

Figure 4.5 and Figure 4.6 give an item-wise presentation of the Cp results with respect to the available data and wind directions of the database. As the purpose of the wind tunnel tests was primarily to obtain empirical reference material, only the Cp results are shown here. The figures show the upper bound and lower bound for both models (Swami&Chandra and CpGen), within the 5 and 95 percentile, i.e., with 90 % confidence that a value falls inside the plotted range. As mentioned earlier, for this study we use five different urban variations to represent surrounding effects where each variation falls in a C_A class as previously explained. Our case studies fall within density value of C_A = 0.10.

The data presented below are for the north face of the model and surrounding density C_A = 0.10. Tables and figures for all other cases are presented in APPENDIX A.

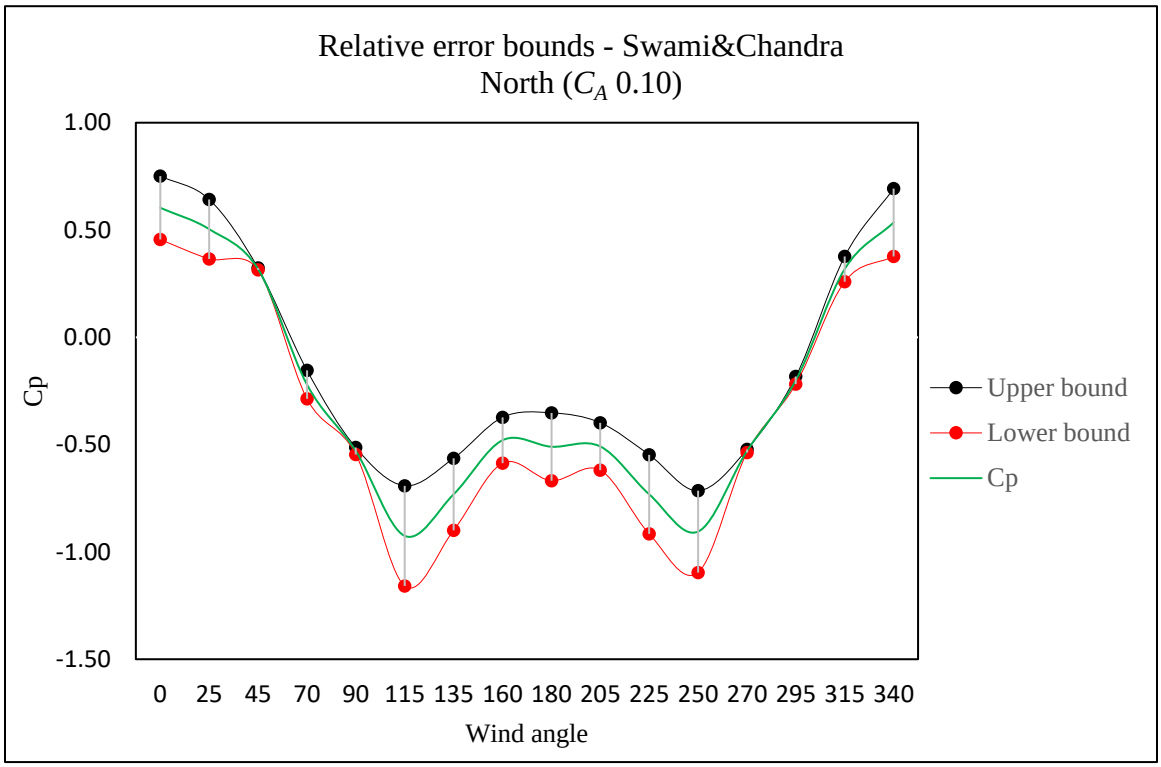


Figure 4.5 Results of Cp from Swami&Chandra model with 90% confidence interval under C_A=0.10

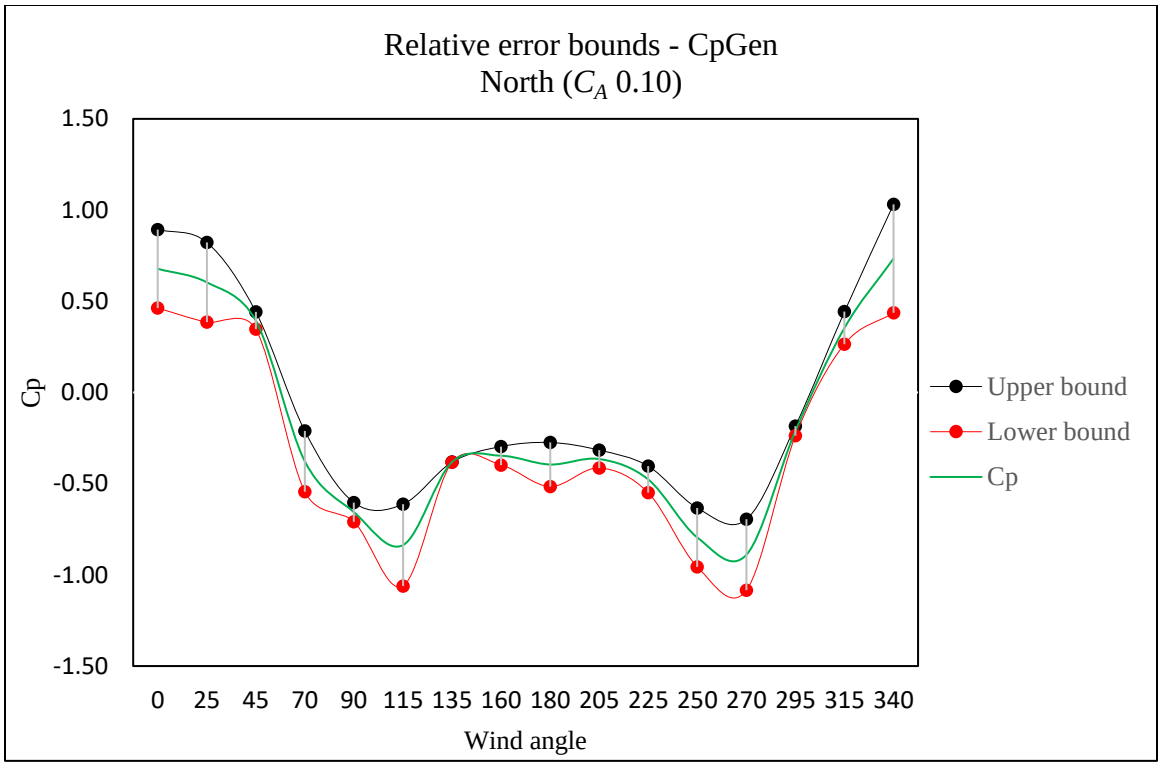


Figure 4.6 Results of Cp from CpGen model at 90% confidence interval under C_A=0.10

The results show that uncertainty in Cp models dominated by wind directions and surrounding density. In the free field analysis, both models demonstrated good agreement with the wind tunnel data, which results in relative standard deviations of 0.08 and 0.11 for Swami&Chandra and CpGen, respectively. With the higher surrounding density, the deviation in the models becomes more significant as can be seen in the above figures; the relative standard deviations are 0.25 and 0.29 for Swami&Chandra and CpGen, respectively. Although the variation is higher in CpGen, the Cp results lie within a quasi-consistent narrower range, and the uncertainty is distributed more evenly when compared with Swami&Chandra. We see this only happen on the lower densities (e.g., *CA*=0.10 and *CA*=0.15), and the reason for that is obvious in the case of Swami&Chandra, it is due to

the embedded empirical formulation that does not capture the characteristics of surrounding densities. Although CpGen uses empirical formulae regressed on wind tunnel experiments with a large number of urban characteristics, the Cp outcomes for each specific wind tunnel experiment are not available and can therefore not be used in the comparison. As a result, it is not surprising that the discrepancy between the specific experiments in our reference database and empirically-based CpGen results can be large. Table 4.1 shows uncertainty quantification in terms of standard error (standard deviation divided by the absolute sample mean) for both models and under different variations of surrounding densities.

Table 4.1 Standard error (SE) of Cp uncertainty under various surrounding densities

Surrounding density	**Uncertainty quantification (SE)**	
	Swami&Chandra	**CpGen**
Free Field	0.08	0.11
C_A 0.10	0.25	0.29
C_A 0.15	0.30	0.34
C_A 0.30	0.45	0.46
C_A 0.50	0.46	0.48

In conclusion, the uncertainty in Cp for two competing low fidelity models is assessed on the basis of existing real ("high fidelity") data. The analysis demonstrates that there is considerable uncertainty in resulting median Cp values compared to reality. This uncertainty is strongly dependent on wind direction and correlated with surrounding density. The immediate conclusion is that airflow predictions with current models that use Cp require uncertainty analysis (UA) to quantify the uncertainty in airflows based on the uncertainty in Cp. The above shows that a comprehensive wind tunnel database is an

adequate source to perform the UQ on Cp. The next section presents uncertainty in wind speed, which is another source of uncertainty that needs to be considered in the UA. The corollary for calibration is that uncertainty in Cp and local wind speed (whether measured or model generated) are direct inputs into the calibration of ELA.

4.3 Local wind speed

Much work has been dedicated to the mathematical representation of wind speed as being essential for BPS and ELA in particular. Typically, BPS follows the approach presented in section 2.2.2 to compute the local wind speeds at a local terrain. As discussed in the same section, Eq. (2.9) expresses the method that computes local wind speed at a certain height from the measured wind speed (H_{met}) at the meteorological station with the use of wind reduction factors. A simple empirical way to derive the local wind speed is to use a wind reduction factor that is expressed as a function of measurement height (H) and terrain class constants (γ,τ), as provided in Table 2.1. This model is also adopted by ASHRAE Fundamentals (ASHRAE., 2009).

The variability of wind reduction parameters creates major discrepancy between the output from the ASHRAE model and the actual measured local wind speed. Figure 4.7 gives an impression of the discrepancy between observed wind speeds (red and green) and converted wind speed (black) using the ASHRAE model. The hourly data presented here are for the first three days of January 2018; Eq. (2.9) is used to convert the wind velocity at 10 m height at Atlanta Hartsfield airport to a local wind speed. The observed local wind speeds are collected from two weather stations located on the campus of Georgia Institute of Technology (GaTech police station and Bobby Dodd stadium weather stations). The

purpose of using two observed data points is to confirm the resolution of the measured local wind speed within the same area, and the dotted lines in Figure 4.7 confirm that both observed data are in close agreement. Also, the results show that the converted wind speed does not represent the values that are measured on-site very well, which confirms that the simplicity of the terrain parameters in the wind reduction formula erodes the accuracy of the calculated local wind speed.

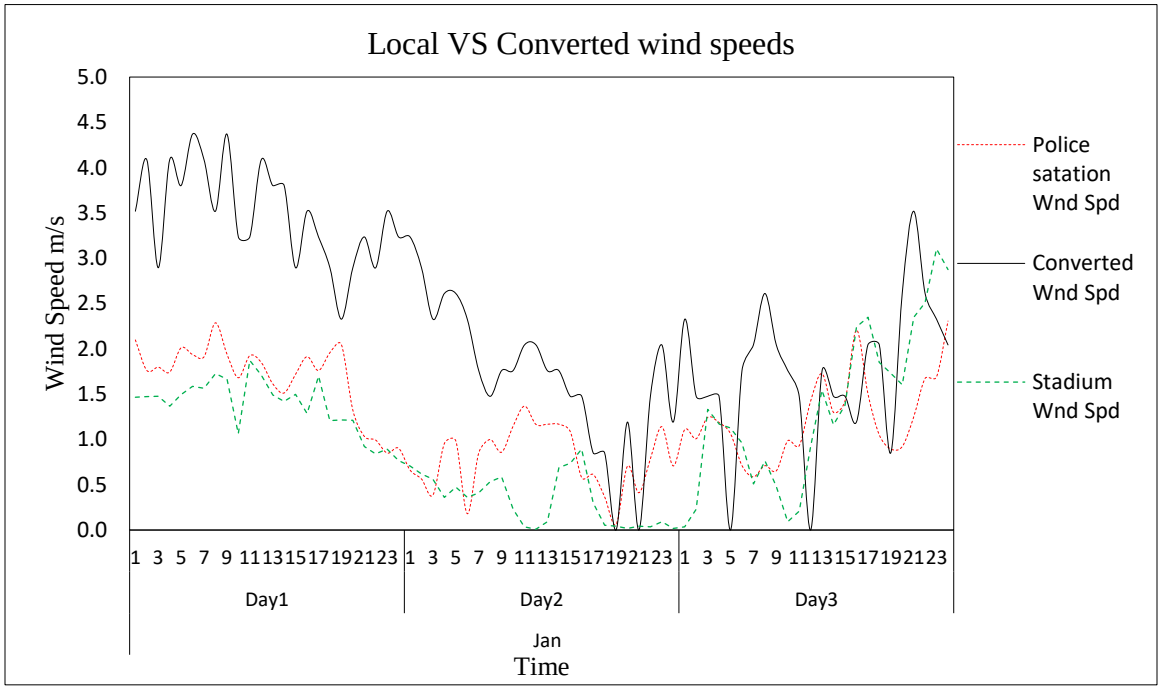


Figure 4.7 Pairwise comparison between measured local wind speeds and ASHRAE model wind speed

Hence, simulation of local wind speed has been the subject of several publications that aim to provide techniques for simulating wind speeds for dispersed sites. Many of these studies share a common drawback of not taking the high temporal correlation, or autocorrelation, into account. Autocorrelation occurs when the error term observations in a regression are correlated, which means that observations of the error term are not

independent of each other (Blanchard & Desrochers, 1984). This is a common issue in time series analysis and can be found clearly in wind speed data.

Therefore, this section describes the steps that perform an uncertainty quantification of local wind speed. Based on actual data, the model takes account of the autocorrelation and allows a time series to be generated, which preserves all the main characteristics of these data.

4.3.1 Model preparation

In order to account for the effect of temporal correlation, the model was built using a method that is common for time series analysis. It allows dependent variables to be studied and takes account of the nature of that dependence. It also allows the user to single out, from an entire class of models (ARIMA), one that would best represent the original data.

An Autoregressive Integrated Moving Average (ARIMA) is a class of models that explains a given dataset based on its own past values, and can be written as:

$$Y_t = a + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} \epsilon_t + \emptyset 1 \epsilon_{t-1} + \emptyset 2 \epsilon_{t-2} + .. + \emptyset q \epsilon_{t-q} \qquad (4.3)$$

Where

Y_t: value of regressed phenomenon at current time *t*;

Y_{t-p}: value of Y at lag p of time *t*;

β_p: autoregressive coefficient of lag p;

ϵ_{t-q}: errors of the autoregressive models of the respective lags.

Normally, an ARIMA model is characterized by three terms: *p,d,q*. where *p* is the order of the AR term, q is the order of MA term, and d is the number of differencing required to make a time series stationary. In regard of stationarity, the correlation between the state at two different times is defined only by the amount of time between them, which is known as stationarity. This property enables modelers to significantly reduce the complexity of the model; however, numerous time series cannot be adequately modeled as stationary. Such as the case for meteorological data, which tend to have significantly different statistical properties throughout the day and year. The statistical characterization can be made once the residuals, ϵ_t, of the model have been determined. If the residuals are assumed normally distributed, then the mean and variance of ϵ_t can be estimated. Consequently, with the regression parameters and the corresponding uncertainty defined, the model of ARIMA process is complete. More statistical analysis regarding weather data time series can be found in (Blanchard & Desrochers, 1984) and (Lee, Sun, Hu, Augenbroe, & Paredis, 2012).

4.3.2 Generating the statistical model

The framework to quantify the uncertainty of local wind speed is shown in Figure 4.8. The process consists of four stages: 1) obtaining wind speed data, 2) process the dataset, 3) compute the statistical difference for uncertainty quantification, 4) use the data in the simulation tool. The following paragraphs will discuss every stage in more detail.

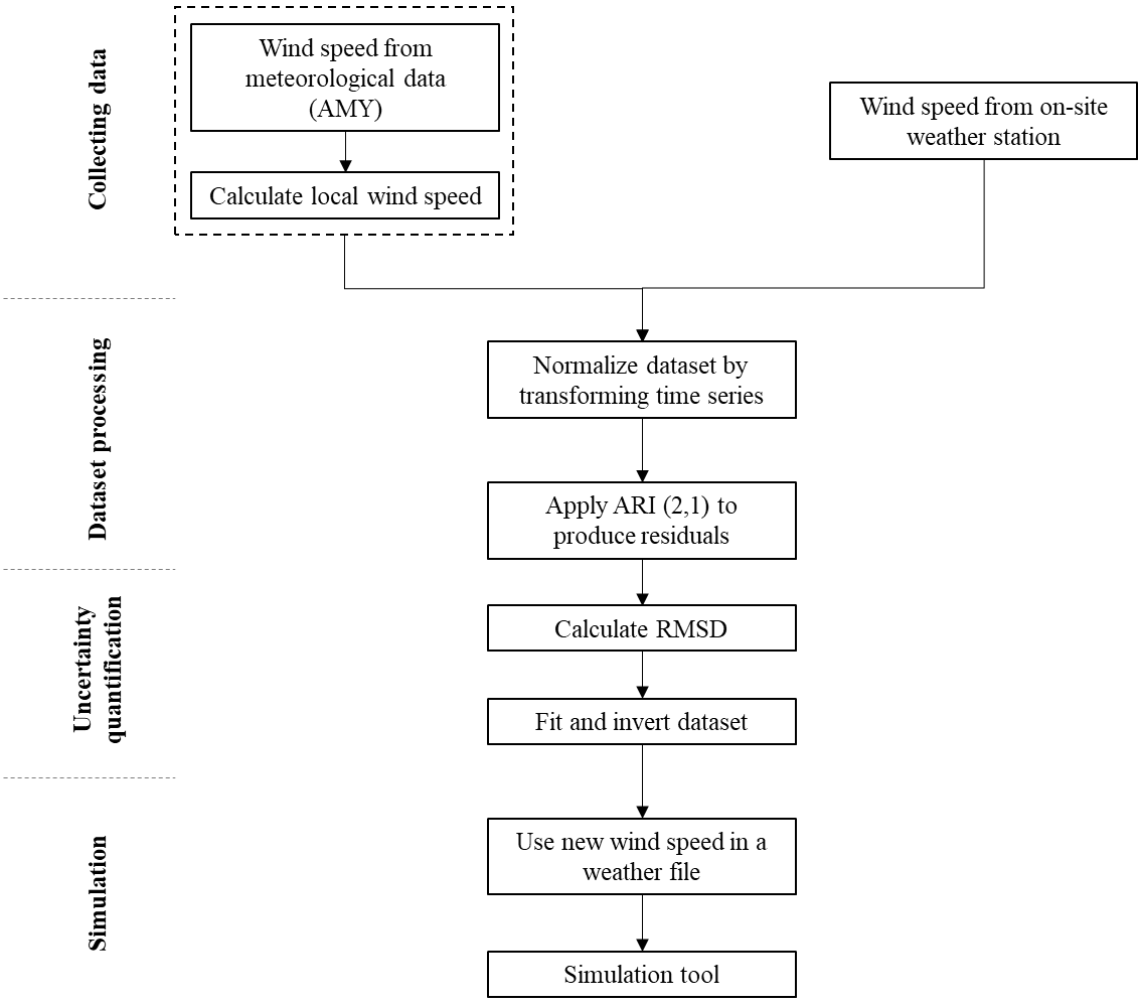


Figure 4.8 Framework of local wind speed uncertainty quantification

The first stage is to gather the required meteorological data. As our case studies located in Atlanta, the weather file used in the first run of simulations is the Actual Meteorological Year (AMY) for Hartsfield airport of 2018. Then we apply Eq. (2.9) to calculate the local wind speed for every hour in the time series. From Table 2.1 we used class I to represent the airport terrain conditions, and class III to characterize the terrain condition of Georgia Tech. On the observation side, an on-campus weather data was collected to represent the actual data of local wind speed. These data will then be processed

to quantify the discrepancy between calculated local wind speeds and the measured ones on location.

The next stage consists of the determination of the *p*, *d*, and *q* of Eq. (4.3) in addition to the preliminary estimation of the ARMA parameters of the model that will be used in the next step.

The first task is determining whether the selected dataset contains any special behavior. For this we need to perform an autocorrelation function to check for the seasonality of the data. The analysis confirms that our datasets are nonstationary, which requires a differencing process to overcome this problem. The implementation revealed that the model reached stationarity with the first order of differencing; this determines the value of *d* to be equal to 1.

Besides, we need to determine the set of lags in the term *p* that defines which previous hours the current state is regressed upon. While this step may appear trivial, it is quite important to construct the model properly. Because only a limited amount of data is available at each hour, only a limited number of lags can be included. In addition, using too many prior coefficients as predictors can lead to over-fitting, resulting in poor predictive capability. As a result, the number of lags in the term *p* is chosen to be 2 based on the partial autocorrelation analysis, which also suggests that our time series can perform well with autoregressive model only; this means our final model will take the form of ARIMA (2,1,0) or simply ARI (2,1).

Before we proceed with autoregression analysis, we need to ensure that the residuals of the linear process will follow a Gaussian form. That means the measured values

themselves should be Gaussian as well. For this, a Box-Cox Transformation is used to 'normalize' the data. In this process, sampled data are inverted through an approximate Cumulative Distribution Function (CDF), and then through the inverse standard normal CDF to obtain a set of samples that more closely approximate samples from a normal distribution. Once the data have been normalized for each hour, ARI (2,1) is performed to generate the required residuals.

Once all above is satisfied, we can start with the process of autoregression. The ARI (2,1) model was obtained by minimizing the sum of the squares of the differences between the actual speeds and those given by the model. If the fitted model is adequate, then the autocorrelations of the residuals should be uncorrelated and normally distributed with mean 0 and variance $\sqrt{n}$. If the assumption of normality is justified, then 95% of the residuals should be within the $\pm 2\sqrt{n}$ boundaries. The model is finally accepted if these residuals are found to be uncorrelated.

The next step is to define the residuals and then capture the means and covariance of each hour using standard statistical methods. The generated residuals from both datasets (i.e., AMY and on-campus wind speed data) are then compared against each other based upon Root Mean Square Deviation (RMSD), which results in a relative standard deviation. This outcome represents the uncertainty quantification value that we are aiming for. Accordingly, the quantified uncertainty for the local wind speed from our process is equal to 0.42. As Figure 4.9 presents, this number shows high uncertainty in the context of calculated to actual local wind speed, and that confirms the fact that wind speed is a very unstable phenomenon and cannot be completely described by a probability distribution.

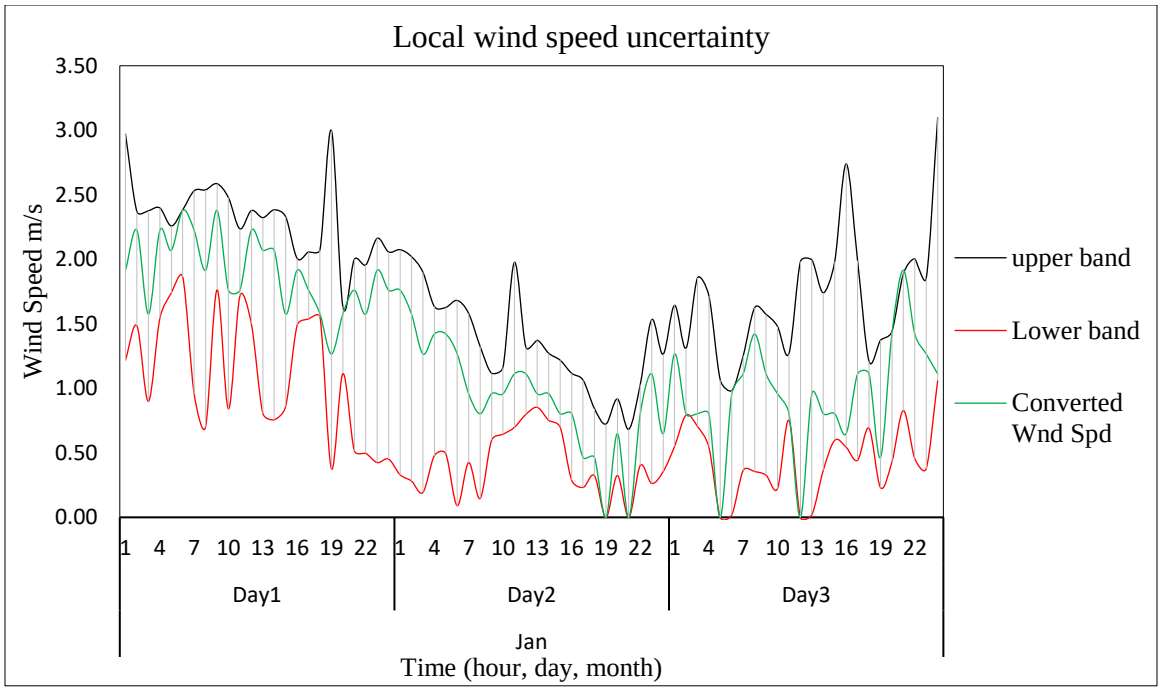


Figure 4.9 Uncertainty bands reflected in the ASHRAE model local wind speed for the first three days of January

The final step in the process is the generation of the weather files that contain the newly fitted (i.e., subtraction of residuals from the original data) and inverted local wind speeds. The new wind speeds will serve as variables that change with respect to the provided uncertainty range. Those uncertainties in wind speeds are determined by multiplying the residuals by the uncertainty value; after that we fit the data by subtracting the new residuals from the original data to produce the final converted wind speed. This is done by specifying a surrogate parameter in the simulation tool that allows us to apply the defined uncertainty range to generate new wind speed values (i.e., time series samples) every time we run the simulation. In other words, every simulation will use a new weather file with new precalculated local wind speed values. The process of generating the wind speeds with uncertainty can take place inside the simulation tool within the process of sensitivity analysis and calibration.

4.4 Propagation of uncertainty

The notion of uncertainty propagation and the essential need for it was briefly introduced in section 2.4. In this section we propose the uncertainty propagation through a building energy model to assess the resulting uncertainty in the yearly consumed electricity in kWh/m^2/year. This exercise is done to give an illustration of the effect of different sources of uncertainty on a resulting model outcome. An evaluation of this uncertainty on its own merits may give an intuitive idea of its significance and the relevance to account for it. To fully appreciate the effect, we analyze the impact of uncertainty information on, or rather its contribution to, the prediction of a QoI with a building energy simulation model.

The uncertainty in the model output, i.e., yearly energy consumption, is investigated at three different levels. Firstly, the uncertainty in the building performance resulting from only Cp. Secondly, the effect of the uncertainty from local wind speed only. Note that in these two stages, all other parameters in the model are fixed at deterministic values. Finally, the uncertainty in all model parameters (chosen in accordance with the "vanilla" UQ repository introduced in section 5.4) is propagated through an entire year simulation with the building energy model. Although the specific building is immaterial for the illustration, it should be noted that the building in the second case study in CHAPTER 7 is used, specifically its EnergyPlus model used there.

4.4.1 Propagation of wind pressure coefficient

As presented earlier in this chapter, the uncertainty in Cp is quantified in terms of marginal distributions for different wind directions. Figure 4.10 shows the propagation

results based on 500 random samples from the joint distribution over Cp values. This is a sufficient number to obtain an accurate estimate of mean and standard deviation of the QoI. All parameters other than the Cp were kept at fixed base values. For most of these parameters, the base value was set to an estimate of their mean value.

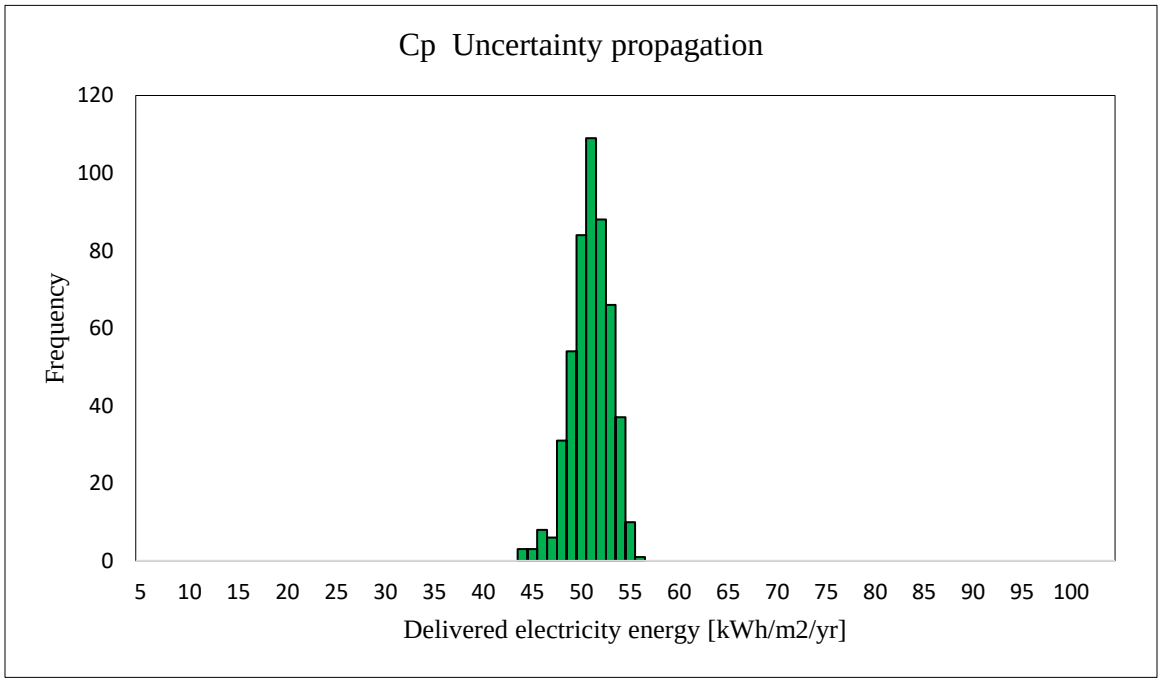


Figure 4.10 Frequency distribution of the delivered energy with respect to Cp uncertainty propagation

The result of Cp propagation indicates that the uncertainty resulting from only Cp is moderately significant. It should be noted that the ELA value in the example building was on average 0.32 cm^2/m^2, which introduces some leakiness. This explains the relative sensitivity of the energy consumption to the variation in Cp, i.e., between 45 and 55. Moreover, the delivered electricity in this propagation process is sensitive because of the way this building was operated, which mostly depends on cooling. Hence, the effect of air infiltration on cooling is huge, and the major part of this is cooling. This explanation needs to be handled with care as it is applicable as well for the following two subsections.

4.4.2 Propagation of local wind speed

The propagated uncertainty of only local wind speed leads to a QoI distribution as presented in Figure 4.11. similar to Cp, this is based on 500 random samples from the joint distribution over wind speed values. Again all parameters other than the wind speed were kept fixed at base values.

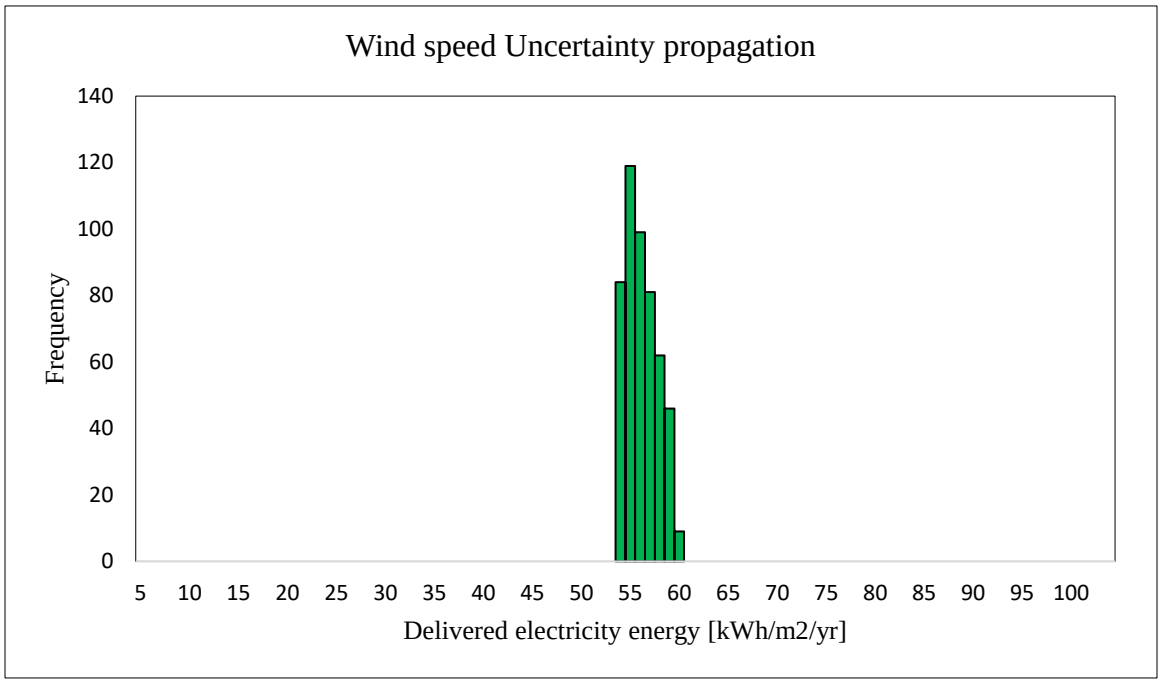


Figure 4.11 Frequency distribution of the delivered energy with respect to wind speed uncertainty propagation

Here the results of local wind speed propagation show a variation on the outcome between 55 and 60. Although this range is relatively smaller than what obtained from the Cp, the impact of wind speed is moderately significant.

4.4.3 Propagation of all parameters

For all parameters, besides those addressed above, we interpret the ranges of uncertainty quantification of different uncertainty sources in our models that are identified in Table 5.1 as central 95% confidence intervals. Where necessary, the normal distributions are truncated to avoid physically infeasible values. Figure 4.12 shows the results of the propagation of the uncertainty in all parameters. The outcome distribution is based on 500 random samples.

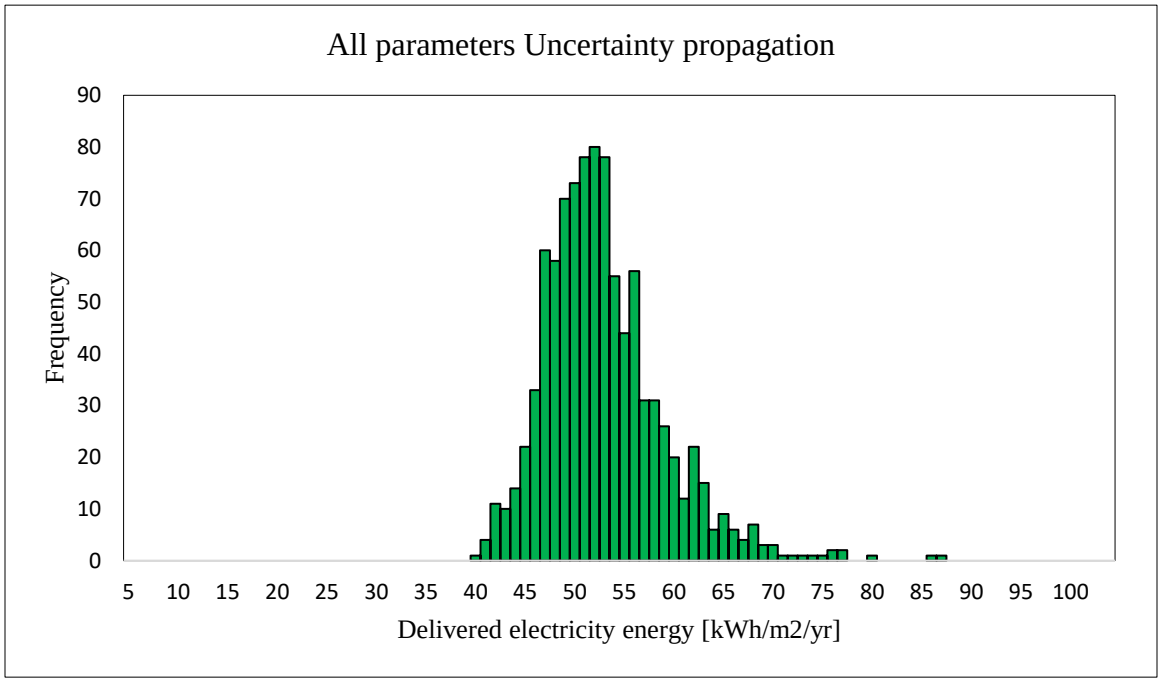


Figure 4.12 Frequency distribution of the delivered energy with respect to all parameters uncertainty propagation

It is clear from here that the uncertainty in the outcome is quite pronounced, as indeed confirmed by the variation in outcome, which is between 40 and around 80. It should be kept in mind though that the vanilla (i.e., non case specific) UQ repository has in general wide distributions. This is in particular true for the very wide ELA distribution

(the uninformative prior as discussed previously). It is no surprise that the combination of all uncertainties leads to unrealistically wide distributions.

4.5 Experimental case studies

It is important to establish an ELA benchmark based on in-situ measurement, which we will also refer to as "*direct* ELA calibration" as distinctive from the *indirect* calibration based on a QoI that is readily available, i.e., without the need for an in-situ experiment. The tracer gas technique is used for the experiment on the two case studies (as will be discussed in CHAPTER 6 and CHAPTER 7). The first case study is a commercial building with six thermal zones. The second case study is a two-story detached house and divided into eight thermal zones that are served by two central air conditioning systems. The resulting direct ELA is conducted with the conditions described in section 3.1.4. The following subsections provide details on the ELA experiment process.

4.5.1 ELA determination based on in-situ measurement for case study 1

Only one zone ("event area") in this case study has been subjected to the experiment, which leads to the determination of the ELA of the front façade of the building, referred to as "front ELA" (see Figure 6.7 zone 5). The experimental results reveal that this space has an average ACH value of 0.68 1/h during the conducted experiment. Now we can conduct the direct ELA calibration using Eq. (2.3), but it should be remembered that this calculation is affected by sources of uncertainty characterized by Cp and wind speed. This is due to the fact that we cannot know the precise wind speed and direction, nor the correct Cp values for different parts of the designated façade. Therefore, deriving a best guess for the experimentally determined ELA requires an uncertainty analysis, which leads

to an estimate in the form of a distribution of estimated ELA values. As mentioned in section 4.2, we can test the ELA with two different types of Cp models or measurements; in this case we use the empirical calculations based on Swami&Chandra Eq. (2.10). The direct calibration provides a distribution of where the actual ELA could be, and the outcome ranges between 0.22 and 1.00 cm^2/m^2; Figure 4.13 shows the Monte-carlo simulation results of the experimental ELA against the prior knowledge that we discussed in section 4.1.

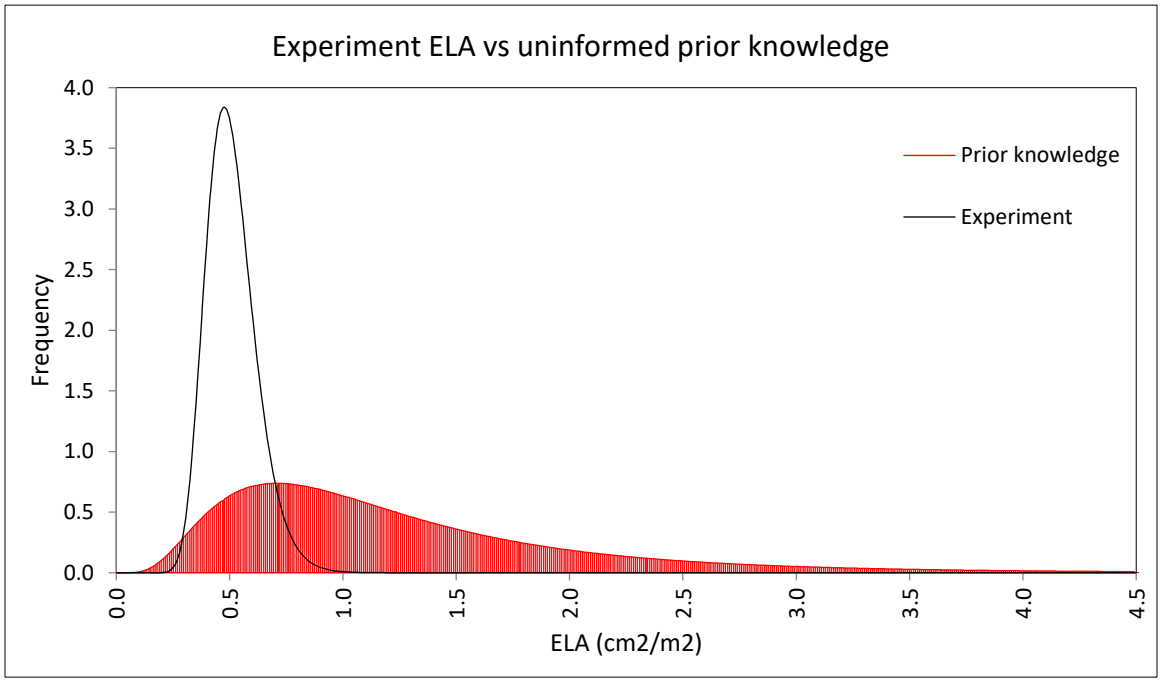


Figure 4.13 The result of experimentally determined ELA (black) against the uninformative prior (red)

The results indicate that what we know about ELA if no information is available, is a very wide distribution that would be nonsensical to use in any type of building energy analysis. This underscores the need for some level of ELA calibration (direct or indirect) before proceeding. It should be noted however that the building energy auditor will have expert knowledge and some onsite information that will allow him or her to reduce the

prior uncertainty significantly. In such a case, it becomes a matter of judgment whether ELA calibration should be performed.

4.5.2 ELA determination based on in-situ measurement for case study 2

We carried out a tracer gas test on three zones in the building, which represents the front, back, and right facades. Using the data collected in the conducted experiment it is found that these spaces have average ACH values of 0.38 1/h, 0.41 1/h, and 0.23 1/h, respectively. The right façade is represented by a single room with only one small window, which explains the small ACH compared to the others. We generate the ELA value following the same fashion that we used in the previous case study. However, for this building we have installed both an on-site weather station and a pressure sensor during the experiment. This provides higher resolution and more precise data to calibrate the "direct" ELA value. Hence, we show the results of ELA for two conditions i.e., "with" knowing the wind pressure value given by the pressure sensor, and "without". Without pressure sensor, the outcomes range between 0.12 and 0.75 cm^2/m^2, 0.14 and 0.66 cm^2/m^2, and 0.04 and 0.44 cm^2/m^2, for the front, back, and right facades, respectively. On the other hand, the direct calibrations with the pressure sensor show significantly narrower distributions that range between 0.26 and 0.43 cm^2/m^2, 0.30 and 0.49 cm^2/m^2, and 0.08 and 0.31 cm^2/m2 for the front, back, and right facades, respectively. Figure 4.14 shows the results for the "back ELA" of both experiments (i.e., with Cp only and with wind pressure sensor) against the prior knowledge that we discussed in section 4.1).

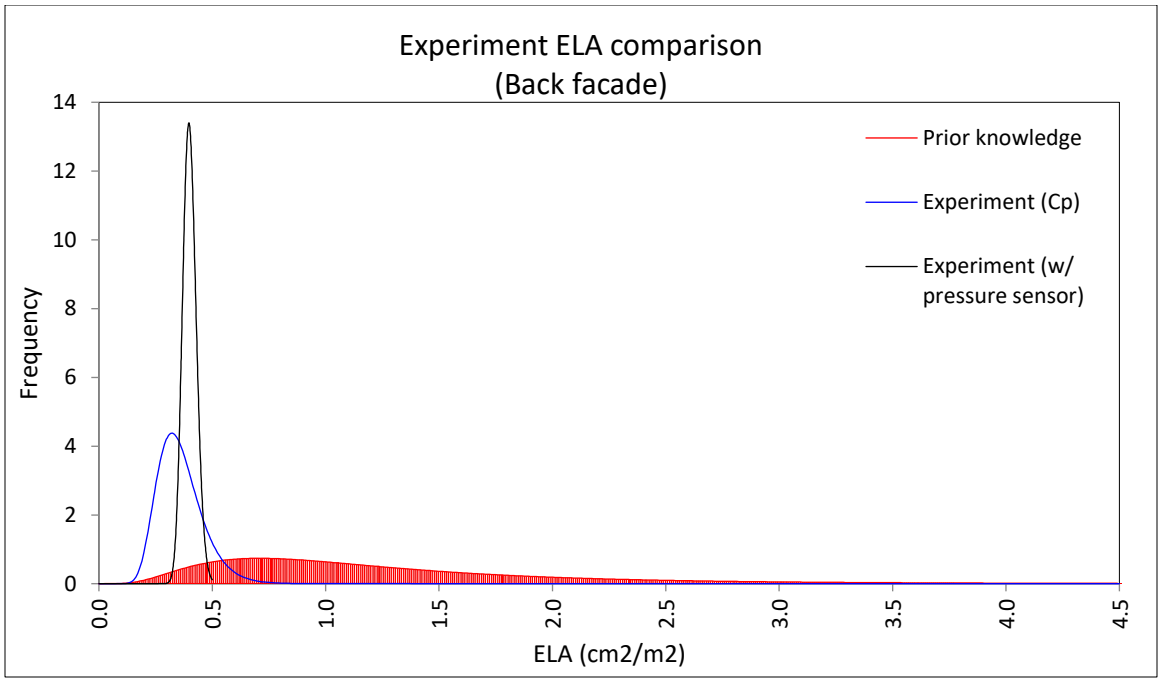


Figure 4.14 "Directly" calibrated ELA against uninformative prior (red); with the use of a pressure sensor (black), and without pressure sensor (blue)

Similar to the findings in the previous case study, the results indicate that the prior knowledge about ELA is very uninformed compared to the ELA found from on-site measurements. Summarizing the work presented, the most important recognition is that there is still some uncertainty in the determined ELA, which explains the distribution derived from the experiment. The underlying uncertainties are in the Cp, wind pressure as well as the experiment devices. It is not surprising that ELA estimates become significantly more informative when more information regarding wind pressure and local wind speed is available.

CHAPTER 5. MODEL PREPARATION FOR INDIRECT CALIBRATION

Irrespective of the model being used, there are several essential steps that require consideration prior to and during the model build phase. These steps impact on model performance and include the choice of model fidelity, modeling approach, and specific steps in the calibration. The following sections discuss the preparation we depend upon in building our calibration models.

5.1 Calibration models and fidelity

5.1.1 Reduced order models

The Energy Performance Calculator (EPC) tool is used as the prime representative of the class of low resolution, first-order simulation tools. It is a simple energy estimation tool developed by the building technology group at the school of Architecture at Georgia Tech according to ISO-13790 (ISO-13790, 2008) with some modifications. It represents the lowest fidelity level considered in this thesis. The tool has an Excel interface and requires basic inputs such as building geometry, construction type, system used, schedule and appliance, and equipment specification. The tool results show the thermal energy need for heating and cooling, and delivered energy (electricity and gas) of the building throughout the year, either at monthly or hourly intervals. The first is done with simple heat transfer and zonal heat balance calculations. Delivered energy is calculated based on simple conversion factors for the HVAC system to meet the thermal heating and cooling demand. Direct electricity use for other energy consumers in the building is added to the

delivered energy by standard calculations. ISO-13790 was developed for normative benchmarking purposes and used mainly in Europe. Through the removal of some of the original normative assumptions, the tool can also be used as a low-order building simulation tool. This has good use in the early design stages and for simple evaluation of new and existing buildings. For comparative analysis the tool has proven to be accurate enough for most applications (Kim & Augenbroe, 2013). It should be mentioned that there is an updated standard (ISO-15014), which adds some refinements to the calculation. This is immaterial for this study as EPC is primarily used as a (random) exemplar of a family of low-order simulation tools.

Based on ISO-13790 and after removing the normative assumptions, we follow the standard process of consulting design specifications, construction documents and operation manuals to build the energy model of the given building. The main parameters of the energy model can be broadly summarized within the following groups: a) building envelope properties (e.g., thermal transmittance, emissivity, solar absorptance, heat capacity), b) internal loads (appliances, lighting, occupants), c) properties of the HVAC systems, d) properties describing ventilation and infiltration, and e) external environment (weather data). This is followed by an onsite audit and interviews with the building manager, which leads to adjusted values assigned to the model parameters such that the energy model represents the actual building as used and operated.

The current version of EPC accounts for the airflow rate in the zonal heat balance using the air change rate as an input. However, it must be noted that customizations have been added to the standard EPC to account for the necessary airflow calculations based on ELA values following the fundamentals described in section 2.2. This modification

substitutes for the simple infiltration calculations that are traditionally done in ISO-13790 and adds a term to the zonal heat balance that is characterized by the amount of air with the outside temperature and is driven by the wind pressure difference. In this procedure, the ELA is the infiltration input instead of the air change rate, and the subsequent airflow calculation follows Eq. (2.3) and Eq. (2.5) for wind effect and stack effect, respectively.

5.1.2 Higher order models

In energy performance assessment with simulation, the goal is to predict energy use in a way that reflects what is expected from a real system "as closely as needed and/or possible". Although this goal is vague, it is to be expected that some use cases of building energy simulation require an extended level of detail and model refinements to achieve it. In such cases, the resulting model exceeds the simplified low resolution approach and can be denoted as a higher fidelity model. For instance, a low-resolution model describes the characteristics of major components, whereas the higher fidelity simulation model requires a detailed level of model parameters that describe the physical behavior of individual parts. Thus, this type of model allows the modeler to specify in detail the parameters that influence the building energy behavior based on information about thermal properties of the envelope, control set points, occupant loads, primary and secondary HVAC system properties, and hourly weather data for the location of the building. Accordingly, we assume in this thesis that EnergyPlus represents all aforementioned characteristics and can be used as our high fidelity model. In addition, it is essential for this thesis to implement the modeling methods that can best estimate the ELA value. The following sections elaborate on the modeling approach to achieve this.

5.1.2.1 EnergyPlus-AirflowNetwork

As the focus of this thesis is to accurately predict the ELA values, different methods to account for the airflow rate in EnergyPlus need to be inspected. There are different simplifications (assuming infiltration rates similar to what we discussed for EPC), but the only correct way to account for the airflow rate in the model with the use of ELA at each surface is through the Airflow Network feature. The Airflow Network model provides the ability to simulate multizone airflows driven by wind and also by forced air distribution systems and can specify the properties of a surface "linkage" through which air flows. This linkage is always associated with a heat transfer surface (wall, roof, floor, or a ceiling) or subsurface (door, glass door, or window) with both faces exposed to air. The associated leakage component for this surface can be a crack (or surface ELA) in an exterior or interior heat transfer surface or subsurface. For our purpose, the ELA is used to define the air leakage at each outer surface of the zonal geometric model, and the relationship between pressure and airflow follows the fundamentals discussed in section 2.2.

5.1.2.2 Wind speed manipulation

As previously described, wind speed is an essential part in the characterization of ELA calculation; therefore, it is important to include it correctly as a transient variable in the energy model. Unfortunately though, there is no direct way in EnergyPlus to account for the wind speed's uncertainty. As a result, it requires specific manipulation to tackle this issue through EnergyPlus's built-in objects. Currently, the "Site Weather Station" and "Height Variation" objects in EnergyPlus are used to specify the measurement conditions for the climatic data in the weather file. These conditions deal with the height above ground

of the air temperature sensor, the height above ground of the wind speed sensor, as well as coefficients that describe the wind speed profile due to the terrain surrounding the weather station (see Eq. (5.1)). Therefore, these weather objects are useful and can be manipulated to customize the weather data input during the simulation process.

$$V_z = V_{met}\left(\frac{\delta_{met}}{z_{met}}\right)^{a_{met}}\left(\frac{z}{\delta}\right)^{a} \tag{5.1}$$

Where

V_{met}: the wind velocity that is extracted from the weather file [m/s],

δ: the boundary layer thickness for the given terrain type,

a: terrain-dependent coefficient,

z: the height of the standard wind speed measurement at the meteorological station as well as the model site [m].

Introducing this formulation is essential to understand how local wind speed is manipulated in our model (note that Eq. (5.1) is equivalent to Eq. (2.9)). As discussed in section 4.2.3, the uncertainty quantification is performed based on a time series process with actual wind speed data, and the generated wind speed values from the time series are used to produce weather file samples for the simulation process. Accordingly, we can adapt Eq. (5.1) to introduce a "surrogate" term that can assign an uncertainty range to the wind speeds that are extracted from the weather file. To do this unobtrusively within EnergyPlus, we use the boundary layer thickness δ_{met} from Eq. (5.1) as our surrogate term where the V_{met} term still recalls the actual wind speed values from the weather file. The remaining

terms in the equation need to be normalized by assigning values that equal to *one* for z_{met} and a_{met}, and *zero* for a to disable their dependency. This is expressed in Eq. (5.2).

$$V_z = V_{met}\left(\frac{\delta_{met}}{1}\right)^1\left(\frac{z}{\delta}\right)^0 \tag{5.2}$$

$$V_z = V_{met}\,\delta_{met} \tag{5.3}$$

The consequent result will be V_{met} multiplied by the surrogate term δ_{met} depending on the uncertainty quantification in section 4.3.2. The final expression follows Eq. (5.3). It should be noted that this is a method that captures the specified uncertainty in the wind speed without resorting to rewriting a specific part of the EnergyPlus code, necessitating an intrusive intervention in the simulation software.

5.2 Modeling approach

The structure of the calibration exercise follows the diagram in Figure 5.1. The procedure of the proposed calibration methodology relies on three distinct parts. The first part is the choice of the underlying model to be calibrated. In this thesis, a distinction is made between four model resolutions, i.e., EPC-monthly, EPC-hourly, EnergyPlus-(monthly data), and EnergyPlus-(hourly data). It should be noted that the last two (EnergyPlus) are equal in fidelity but different in resolution of outcome and hence the formulation of the calibration criterion. The first two are different in resolution (EPC-monthly is based on 12 averaged monthly calculations, whereas EPC-hourly performs 8760 hourly calculations), as well as resolution of the outcome calibration. The determination

of these models depends, among others on the availability and granularity of the quantity of interest (QoI-i) used for the calibration. One interesting point to make here is that if a QoI is only available at monthly granularity, the use of an hourly simulation would require the simulated outcomes to be aggregated per month before substitution in the calibration criterion. Intuitively, this approach raises doubts on the effectiveness of the use of a high granularity (hourly) model. The intent of this chapter is to perform the necessary studies to confirm or contradict this intuition.

Every model in the calibration process is conjoint with a calibration criterion (CC) that is determined by the measured QoI's. In the case studies only two types of QoI are considered: QoI-1 is the energy consumption, either available at monthly or hourly resolution, and always as whole building aggregated electricity consumption; QoI-2 is the indoor temperature, available at hourly level, but not for the full calibration period, and only in selected zones. Based on the availability of QoI-1 and sometimes QoI-2, two calibration criteria can be constructed: i) CC-1 which uses only QoI-1; ii) CC-2 which uses QoI-1 and QoI-2 together. The resolution of CC-1 and CC-2 (monthly or hourly) can be chosen independently, depending on the resolution of model outcomes and measurements.

The incremental nature of the process is driven by the characterization of the calibration set (CS-i), which is the set of calibration parameters that enters the calibration process steps for a particular model and a specific calibration criterion. The choice of the calibration set is informed by a sensitivity analysis (SA) process that is performed at every calibration step. In the current calibration process, the number of calibration sets is limited to four sets (i.e., CS-1, CS-2, CS-3, and CS-4). This is executed in both case studies.

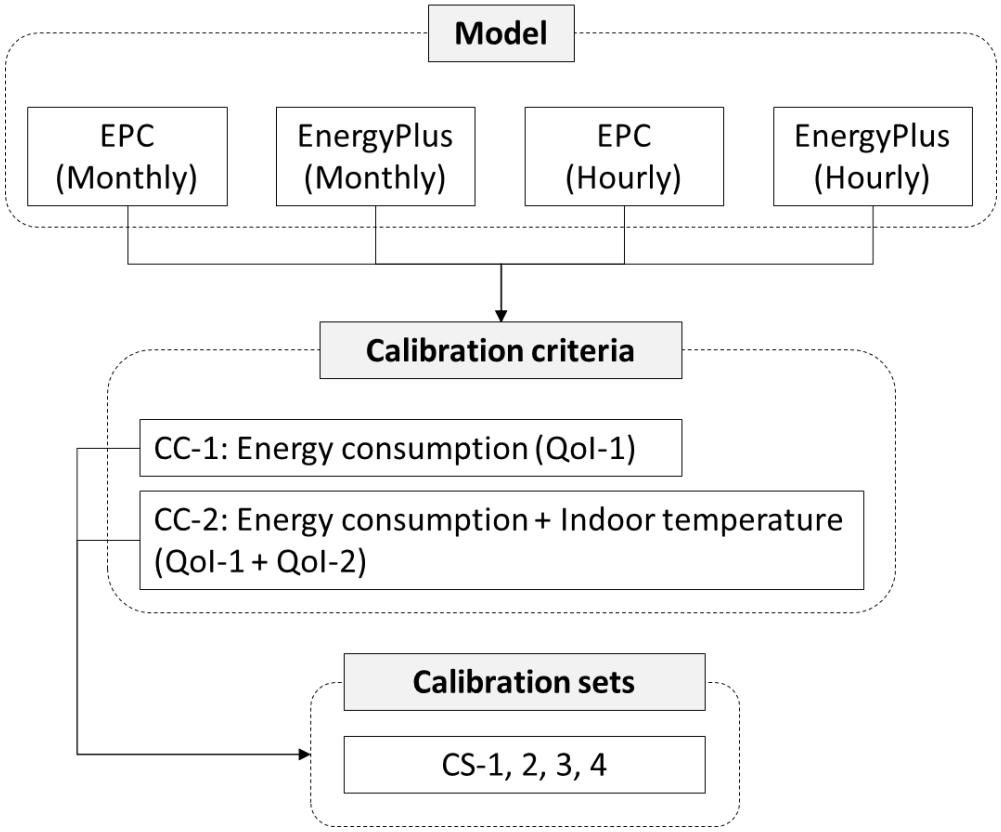


Figure 5.1 The structure and variants of the calibration process

Every specific calibration step is characterized by the *model*, a *calibration criterion*, and a *calibration set*. For example, EPC-hourly, CC-1, and CS-3. More details are provided in CHAPTER 6 and CHAPTER 7 which show the application of the structure and variants in the two case studies.

5.3 Methodological treatment on calibration

5.3.1 The use of sensitivity analysis

Sensitivity analysis (SA) can help in understanding the relative influence of input parameters on the output (Saltelli, Tarantola, Campolongo, & Ratto, 2004). In the field of building energy models, combining simulation tools and sensitivity analysis can be useful

as it helps to rank the input parameters (or family of parameters) and then to select the most appropriate ones to be considered, depending on the objective of the modeling. The Morris method (Morris, 1991), (Campolongo, Cariboni, & Saltelli, 2007) is used to perform the parameter screening. It is a well-known method that utilizes the distributions of parameter uncertainties to determine the relative ranking of all uncertain parameters. It is deployed for every step in our calibration exercise, where every step uses a larger calibration set, and consequently a smaller uncertain parameter set. The Morris method produces a ranking across the uncertain parameter set and an indication of potential interaction between individual parameters. Besides, the standard deviation of the elementary effects (EE) is a relevant indicator of non-linearity in input parameters of the model or interactions with other input variables involved in the model.

For the normative model calibration we use SimLab 2.2 (a software designed for Monte Carlo-based uncertainty and sensitivity analysis) to execute the Morris method (SimLab, 2008). For the transient model, the sensitivity analysis is programmed and executed in python 3.0 based on this work (Saltelli, et al., 2008). The total electricity consumption (monthly or hourly) is used as the QoI in the analysis while the process of SA is repeated iteratively prior to every consecutive calibration step (i.e., with changing calibration parameter sets) as described in Figure 3.2; more details regarding SA attributes and outcomes will be discussed in CHAPTER 6 and CHAPTER 7 for every case study.

5.3.2 Calibration treatment

Once the parameters have been chosen and the model has run for a set of samples of possible parameter values selected from their respective distributions, the calibration is

repeatedly used for every sample to generate distributions for the selected parameters. The goal of the calibration process is to generate distributions of best estimates of the true ELA. To measure the reliability of an ELA distribution to its true value, CVRMSE is used to evaluate the calibrated model by checking whether there is an acceptable agreement between the simulated and measured QoI. As discussed earlier, this agreement is only a proxy to the reliability of ELA. CC-1 is a straightforward calibration criterion needing no special consideration. CC-2 considers QoI-1 and QoI-2 altogether which necessitates a specification of how the optimization should satisfy both targets: minimize the difference in energy consumption as well as indoor temperature. This multi-objective optimization problem is simplified in a single combined discrepancy using a weighting factor as defined in the following formula:

$$J(W_E, W_T) = W_E \sum_{i=1}^{N} (\Delta E_i)^2 + W_T \sum_{i=1}^{N} (\Delta T_i)^2 \quad (5.4)$$

Where

W_E: weighting factors for energy consumption.

W_T: weighting factors for temperature.

ΔE_i: the difference between a reference energy consumption value and the actual value at time, *i*.

ΔT_i: the difference between a reference indoor temperature value and the actual value at time, *i*.

On the left hand side of Eq. (5.4), W_E and W_T are weighting factors for QoI-1 and QoI-2, respectively, and the objective function turns into a single objective once these weighing factors are set.

It should be noted that the advocated approach comprises deterministic optimizations repeated for all samples of the uncertain parameters. Although there is also prior information about the uncertainty in the calibration set, this information is not used in this approach. It could be if one were to employ Bayesian calibration (BC). This comes at a heavy price (mainly computation expense) while there is a little demonstratable payoff. Indeed, the posterior distributions of the calibration parameters could be heavily influenced by the assumed priors. If this is undesirable, the only recourse is to assume uninformative priors which de facto will deliver the same results as advocated in our method. So, rather than using BC, a less cumbersome method is used, which can be referred to as "poor-man's Bayesian calibration". It is composed of a series of independent deterministic calibrations without use of prior distributions for the calibration parameters. Each single deterministic calibration uses sampled values from the parameter probability distributions. The results of all deterministic calibrations, together with the occurrence probability of the sampled parameter values used in each step can be assembled in a probability distribution of every calibration parameter that is considered.

As described in the methodology chapter of this thesis, the process of calibration is repeated multiple times until the "smallest" value of a discrepancy metric, i.e., the difference between measured and calculated QoI is achieved. For every energy model (with varying resolution and fidelity), a fixed number of consecutive calibration steps is performed, where each step is preceded by a sensitivity analysis (SA) step. Each

consecutive calibration step is conducted with a larger set of calibration parameters. Parameters that are removed now become part of the normal set of parameters with specified uncertainty ranges. We limit the treatment to four distinct steps where each step corresponds to a different size of the Calibration Set (CS-i). It is important to mention that every next SA step is conducted after the previous calibration step is finished, and the resulting calibration parameters' values can now be used in the next step of the process. This way the new SA reflects the better judgment of the uncertainty of a parameter based on the results of the previous calibration step. This will, by consequence, deliver a more reliable sensitivity ranking.

5.4 UQ of model parameters of energy models of existing buildings

A major source of uncertainty is the fact any model parameter is only "known" to some extent; hence its value needs to be approximated by a probability distribution. This would be less relevant if a parameter is deemed a calibration parameter, but in many cases the modeler may be reluctant to include every parameter that is not known precisely in the calibration set. This section focuses on quantifying parameter uncertainty in a general context of building energy modeling, with focus on existing buildings. Uncertainty in model parameters depends on the level of model granularity. For instance, transient simulation models can describe the physical behavior of an individual component through its model parameters at a detailed level, while those in normative models capture the characteristics of systems only at an aggregate level. Accordingly, uncertainty associated with different levels of model parameters should be separately investigated. One can refer to (Heo Y. , 2011) who investigates uncertainty in normative model parameters by quantifying uncertainty in several macro-level parameters and their relationships. Also

many studies have extensively investigated quantification of uncertainty in detailed simulation models (de Wit, 2001); (Macdonald, 2002); (Moon, 2005), with the caveat that the first two focus solely on new (as designed) building specifications whereas the latter focuses on realized buildings. The distinction between the two is very significant when it comes to conducting a UQ of the parameters in either model. The reasons for this are easy to understand. In the design case, the primary objective of the UQ is to define "realization uncertainty", i.e., how the design specifications will be realized in the future real building. Such UQ thus deals with tolerances in material properties, deviations from the specs, natural variability (also referred to as batch uncertainty) across components, workmanship, aging, and similar effects. For existing buildings there is fundamentally no realization uncertainty as the building is a real "brick and mortar" specimen. In theory, the physical parameters of the realized building are fully observable and any uncertainty can thus be eliminated. This is only true in theory though, because the amount of work to determine material compositions and properties, such U-Values, etc. would be clearly prohibitive. On top of this, certain properties, such as ELA, emissivity, convective heat transfer coefficients and others, even U-value, are not directly observable as we have argued earlier. For the latter, one would indeed always rely on calibration rather than direct observation. In summary, it can be stated that the UQ of model parameters will be different for design and existing buildings. For scenario uncertainty there is also a significant difference, as the design case solely works from the designed use by the client organization whereas for the existing case, actual onsite use can be monitored to some extent, thus typically reducing uncertainty in the scenario parameters. For the third category, model form uncertainty (MFU), there is in principle no distinction between the designed and existing building case.

The next section details the approach to uncertainty in model parameters at the start of the calibration process of existing buildings.

5.4.1 Baseline parameters of building model

The uncertainty quantification of model input parameters such as thermophysical properties, systems, and internal heat gains is relatively straightforward. Once sufficient data are collected, the uncertainty in these input parameters can be characterized by probability distributions (e.g., normal distribution) with standard statistical methods (Bedford & Cooke, 2001). This applies to model and scenario parameters, although the quantification of the latter could become difficult if the parameters are time series or if correlations are apparent among parameters. One example is the weather conditions usually specified as an external scenario with 8760 hourly values in the case of hourly simulations for a year. In some cases when there is no direct observation for the full simulation period, the weather variable can be characterized with an autoregressive process to generate stochastic values from the measured weather data. This process is discussed for the quantification of wind speed in section 4.3.

The characteristics of construction materials used in the building fabric are captured by the physical parameters of the building energy model. Uncertainty in thermal properties of materials largely arises from variations in manufacturing and specific realizations related to workmanship (Wang, Augenbroe, & Sun, 2014). The latter study shows the need for extensive on-site observations to quantify uncertainties in the model parameters of the building as realized and as operated. Within the scope of this thesis, this is not attempted. Rather, the uncertainty in model parameter uncertainties is assumed to be similar to the

realization uncertainty assumed in the design case. Note that this assumes that we basically treat the uncertainty in model parameters in the calibration as prior information, without adding detailed posterior information from the actual building; although were available, we color the prior uncertainty distributions based on some readily on-site observable information. With this in mind, we are able to use uncertainties as for example quantified in (Macdonald, 2002) and (de Wit, 2001) and other resources to quantify thermal properties of materials as well as other uncertainty sources. More recently, most current known ranges were brought together in an Uncertainty Quantification repository (Lee, Sun, Augenbroe, & Paredis, 2013). For instance, the standard deviation of uncertainties in solar absorptance and emissivity is 5% and 2%, respectively. These uncertainties were selected ("colored") based on the material types of our case studies; however, it must be recognized that the uncertainty range quantified as standard deviation differs depending on the type of materials. The mentioned Uncertainty Quantification Repository (UQ Repository) provides energy modelers with quantified uncertainty distributions for a variety of parameters and discrepancies in the internal modules of a building energy model. The repository was developed to serve the Georgia Tech Uncertainty and Risk Analysis Workbench (GURA-W), details of which are described in (Lee B. , 2014) and (Lee, Sun, Augenbroe, & Paredis, 2013). Table 5.1 shows the quantified uncertainty sources and their references.

The HVAC system is parametrized by the thermal efficiency of heating/cooling generation, more specifically the coefficient of performance (COP). The specification of the system is typically documented in the manufacturer's manual, which presents the system efficiency under full-load standard testing conditions. However, the system's

efficiency can fluctuate over time depending on its actual operating conditions. The quantification of HVAC system parameter uncertainty is highly dependent on the granularity of the HVAC system model in the building energy model. Based on (Chong & Lam, Uncertainty Analysis and Parameter Estimation of HVAC Systems In Building Energy Models, 2015), the uncertainty quantification of the system's COP is 5% of its capacity, and 25% of the fan efficiency (Wang Q. , 2016).

Occupants, lights, and plug loads in a building are all forms of internal heat gains. Heat gains from occupants depend on the number of people and their metabolic rates in spaces. Even if the number of people in the space temporarily fluctuates, schedule of building's occupancy is considered as the fixed profile that captures the average occupancy pattern. Occupants' metabolic rates depend on individuals' activity level. Referring to GURA-W UQ repository, the uncertainty of metabolic rate is quantified between 70 and 130.

In an effort to quantify plug loads, a study by (Dunn & Knight, 2005) was conducted in form of a survey of 30 buildings in UK. The survey suggests that plug loads can range between 124 W and 229 W per person in a building with 158 W per person as an average with a standard deviation of 21%. Based on these values, one can calculate the range of possible plug loads from the number of occupants for a particular building under investigation.

Table 5.1 Uncertainty quantification of different uncertainty sources in our models

Parameter	Uncertainty quantification	Reference
1. HVAC		
Cooling COP	Relative normal (1, 5%)	(Chong & Lam, Uncertainty Analysis and Parameter Estimation of HVAC Systems In Building Energy Models, 2015)
Heating COP	Relative normal (1, 5%)	(Chong & Lam, Uncertainty Analysis and Parameter Estimation of HVAC Systems In Building Energy Models, 2015)
Fan efficiency	Relative normal (1, 25%)	(Wang Q. , 2016)
2. Material		
Wall conductivity	Relative normal (1, 4%)	
Roof conductivity	Relative normal (1, 4%)	
Wall solar absorptance	Relative normal (1, 5%)	
Roof solar absorptance	Relative normal (1, 5%)	(Sun Y. , 2014), (Lee B. , 2014), (Wang Q. , 2016), (Macdonald, 2002)
Wall emissivity	Relative normal (1, 2%)	
Roof emissivity	Relative normal (1, 2%)	
3. fenestration		
Glazing transmittance	Relative normal (1, 7%)	
Glazing conductivity	Relative normal (1, 7%)	
4. Internal heat gain		
People	Relative uniform (70, 130)	(Sun Y. , 2014)
Lighting	Relative normal (1, 21%)	(Dunn & Knight, 2005)
Plug loads	Relative normal (0.75, 15%)	(Hejab & Parsloe, 1992)

Table 5.1 (continued)

5. *weather*		
Cp	Relative normal (1, 30%)	CHAPTER 4
Local wind speed	Relative normal (1, 42%)	
6. *effective leakage area*		
ELA	LogNormal (1.28, 0.88)	

As emphasized before, in a calibration process, the scenario parameters need to be chosen with care, using all available posterior information. This mainly applies to the usage of the building (occupancy, HVAC control setpoints, lighting/appliance) and weather, usually requiring the translation of nearest weather data to the local site. This is highly dependent on the data collection during the period used for the calibration. Uncontrolled calibration experiments have by definition large uncertainty in scenario parameters which results in poor calibration results, although this goes unnoticed in many studies when no uncertrainty analysis is conducted.

5.5 Measurement of the discrepancy between ELA distributions

At a certain point in the analysis, it is essential to measure the goodness of the estimated ELA against the direct ELA estimation for each case. This step is important to help understand and determine what level of information was able to achieve the best estimate when compared to the real ELA experiment. In statistics, there are a variety of methods that one can use to measure the distance and divergence between two distributions. In this section we propose one effective method that is widely utilized to quantify the similarity between two probability distributions, which is called Kullback–Leibler divergence (Kullback & Leibler, 1951).

The Kullback–Leibler divergence (KL) (a.k.a., relative entropy) is a divergence measure and widely used for measuring the fit of two distributions. It is important to note that although the KL is often used to see how "different" two distributions are, it is not a metric. Importantly, it is neither symmetric nor does it obey the triangle inequality. However, the KL is always non-negative and equals *zero* only if the two distributions are identical, which by definition means that the larger KL value is the larger distance between the two distributions. Hence, for probability distributions *p* and *q* over the same domain, the KL is defined as (Bishop, 2006):

$$KL(p \parallel q) = log\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2} \tag{5.5}$$

Where KL is the Kullback–Leibler divergence of distributions *p* and *q*, and $KL(p \parallel q) \geq 0$. The two case studies described in the following chapters will deploy the methods laid out above.

CHAPTER 6. CASE STUDY 1: INDIRECT ELA CALIBRATION IN A COMMERCIAL BUILDING

This chapter shows the implementation of the proposed methodology consisting of a stepwise calibration process on a real commercial building. The main intention of the case study is to validate the proposed approach where validation centers on the obtainable accuracy of ELA estimates for the given building.

6.1 Building description

The calibration is conducted for a multipurpose building located in Atlanta, GA called the gathering spot (TGS). The building is one floor with a unique double-height ceiling and floor area of 1950 m^2. The building has six zones in total with a mixture of HVAC systems as shown in Figure 6.1 and specification in Table 6.1.

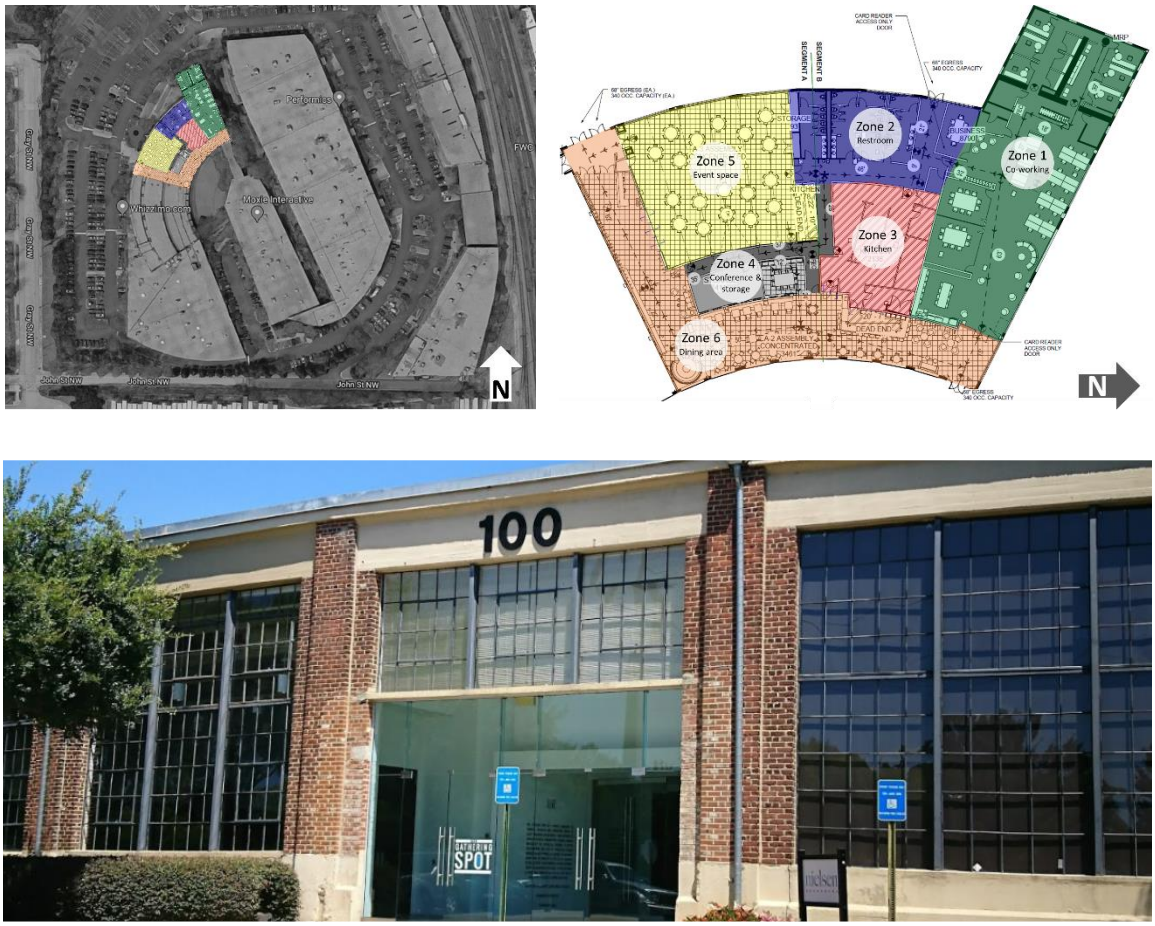


Figure 6.1 Zones distribution and overview of The Gathering Spot (TGS)

Table 6.1 Case study characteristics

Zone	Area (m^2)	HVAC Units
Zone 1: Co-working	565	FCU-1, FCU-2, FCU-10, FCU-15, FCU-16, FCU-17, FCU-18, FCU-19, FCU-20
Zone 2: Restroom	218	FCU-12, FCU-13
Zone 3: Kitchen	200	AHU-3
Zone 4: Conference & Storage	187	FCU-11, FCU-21
Zone 5: Event space	337	AHU-1, AHU-2, FCU-3
Zone 6: Dining area	446	FCU-4, FCU-5, FCU-6, FCU-7, FCU-8, FCU-9, FCU-14

The setpoint temperature of the building is 21℃ for heating and 24℃ for cooling. The primary energy source for all systems is electricity. In addition, the building is equipped with indoor temperature sensors in every zone.

The energy consumption data that is available in this building is electrical energy consumption in kWh. These data are collected for the entire year of 2018. Indoor temperature sensors were installed to collect data from the beginning of January 2018 until the end of the year at a resolution of 5 minutes. The indoor temperature sensor data have been aggregated to average hourly resolution to be used in the calibration criterion as QoI-2.

6.2 Calibration of the low-resolution model (EPC-monthly) with monthly consumption data

As the first step, uncertainties in model parameters are quantified by reviewing published literature and industry standards. CHAPTER 4 and CHAPTER 5 describe the process of quantifying the uncertainties in energy models in detail, and Table 5.1 is used to summarize the uncertainties around the initial values assigned to model input parameters. The uncertainty information is essential for the consecutive step in the calibration process: in particular for parameter screening and iterative calibration under uncertainty. The latter refers to calibration when some model parameters are viewed as uncertain. The following sections present the technical workflow and implementation of the calibration process.

6.2.1 Sensitivity analysis

6.2.1.1 General remarks on the presentation of results

The results of the SA and elementary effects (EE) analysis are presented as two instructive graphs. One is a horizontal bar chart to rank the sampled parameters based on their importance, which is characterized by each input variable *i* absolute average (μ_i^*). The other representation of SA outcome is the traditional Morris 2D plot with an (x,y) point for each input variable: the x-axis represents the absolute average (μ_i^*), and the y-axis represents the standard deviation of the EE (σ_i). This information can be supplemented by the ratio (σ_i/μ_i^*) as an indicator of linearity (or non-linearity), as defined below.

First, we can say that the input factor *i* has a monotonic effect on the response function if all estimates of EE_i have the same sign, increasing or decreasing, depending on the sign of the elementary effects. In this case, μ_i^* is equal to the absolute value of μ_i. If σ_i/μ_i^* is smaller than 0.10, then the elementary effects are almost constant, and the input variable *i* has an almost linear effect on the model. A true linear response corresponds to $\sigma_i/\mu_i^* = 0$.

If the ratio σ_i/μ_i^* is smaller than 0.5, most elementary effects have the same sign, and the model response can be considered as monotonic with respect to the input factor *i*. This fact justifies using the ratio σ/μ^* as an indicator for almost linear (if $\sigma_i/\mu_i^* < 0.1$) or monotonic influences (if $\sigma_i/\mu_i^* < 0.5$).

For highly scattered elementary effects ($\sigma_i/\mu_i^* > 1$), a non-monotonic behavior is clearly established and, in this case, (σ/μ^*) stays in the interval greater than 1. The absolute average μ^* is more influenced by the standard deviation σ.

Accordingly, by plotting three straight lines of slopes $\sigma/\mu^* = 0.1$, 0.5 and 1, respectively, we can identify the elementary effects in a graphical scatter plot. Those factors are 1) almost linear (below the line $\sigma_i/\mu_i^* = 0.1$), 2) monotonic ($0.5 > \sigma_i/\mu_i^* > 0.1$), 3) almost monotonic ($1 > \sigma_i/\mu_i^* > 0.5$), and 4) the factors that are non-linear or have interactions with other factors ($\sigma_i/\mu_i^* > 1$). Defining these four zones also provides a means of inspecting the results of the sensitivity analysis in case the results contradict what is comprehended from the physical point of view.

6.2.1.2 SA implementation

We apply the Morris method SA to identify the parameters that have the most impact on the variation of monthly energy consumption. In this case, monthly energy consumption is chosen as the principle Quantity of Interest (QoI) because it is the QoI that is monitored and used as the data in the ensuing calibration step. The dependency of the delivered energy on the chosen variables can be shown through the resulting distribution of the QoI as a function of the possible variation (variance) of input variables, as described above.

As stated earlier, the first step for the SA process is to estimate uncertainties in model parameters specific to the building case. The identification of the base values is based on design documents, the generic UQ repository and, where necessary, building audits augmented by expert knowledge. A 20-level design of 23 parameters with a sample

size of 480 is applied to the building energy models. Table 6.2 provides the initial list of the uncertain parameters in the EPC model and their uncertainty ranges.

Table 6.2 Parameter uncertainty in model testing and acronym definition

Acronym	Definition	Uncertainty range	Comment
ELA_Back	ELA of the E side (cm^2/m^2)	LogNorm(1.28,0.88)	
ELA_Front	ELA of the W side (cm^2/m^2)	LogNorm(1.28,0.88)	
ELA_Right	ELA of the N side (cm^2/m^2)	LogNorm(1.28,0.88)	
Cp	Wind pressure coefficient	RelativeNorm(1, 0.25)	The uncertainty is relative to every Cp value
Wind Speed	Local wind speed (m/s)	RelativeNorm(1, 0.42)	The uncertainty is relative to every hour wind speed value
Clg COP	Cooling COP	Norm(3.75, 0.18)	
Htg COP	Heating COP	Norm(1.75, 0.09)	
fan power	HVAC fan power (W/l/s)	Norm(1, 0.25)	
Clg SP wd	Cooling setpoint in the weekdays (ºC)	Tring(20, 24, 26)	
Htg SP wd	Heating setpoint in the weekdays (ºC)	Tring(18, 22, 24)	
Roof U	Roof U-value ($W/m^2 \cdot K$)	Norm(0.24, 0.02)	
Wall U	Wall U-value ($W/m^2 \cdot K$)	Norm(0.35, 0.02)	
Window U	Window U-value ($W/m^2 \cdot K$)	Norm(3, 0.6)	
Solar Trans	Window's solar transmittance	Norm(0.75, 0.05)	
Occupancy	Occupancy internal gain (m^2/person)	Uniform(8, 16)	
Lighting	Lighting internal gain (W/m^2)	Norm(12, 3)	
Appliances	Plug loads internal gain (W/m^2)	Norm(16, 2.5)	
Sch Occ wd	Occupants schedule in the weekdays	Tring(0.7, 0.9, 1)	
Sch Occ we	Occupants schedule in the weekends	Tring(0.1, 0.15, 0.25)	
Sch Lgt wd	Lighting schedule in the weekdays	Tring(0.85, 0.9, 1)	
Sch Lgt we	Lighting schedule in the weekends	Tring(0.15, 0.2, 0.25)	
Sch App wd	Appliances schedule in the weekdays	Tring(0.85, 0.9, 1)	
Sch App we	Appliances schedule in the weekends	Tring(0.1, 0.15, 0.20)	

As it is always included in the calibration sets from CS-1 to CS-4, ELA will not enter the SA process (excpt in the first SA process), which helps to reveal other relatively

important parameters to be included in the consecutive calibration steps with larger calibration sets, as confirmed in the SA results for these steps.

At the beginning, we initiate the process with parameters screening of all input variables in the sample pool, which also includes the ELA parameters. Although ELA parameters are always included in the first calibration set CS-1, this step is important to understand the characterization and influence of these parameters as a whole on the energy model. Therefore, from Figure 6.2 we notice that ELA-Front (No. 4 in the graphs) has high level of importance on the SA rank, and also indicates a nonlinear effect or interaction with other variables because it is situated in the fourth region in the interaction graph with $\sigma/\mu^* > 1$. The remnants SA results as well as the calibration process are presented in the next section.

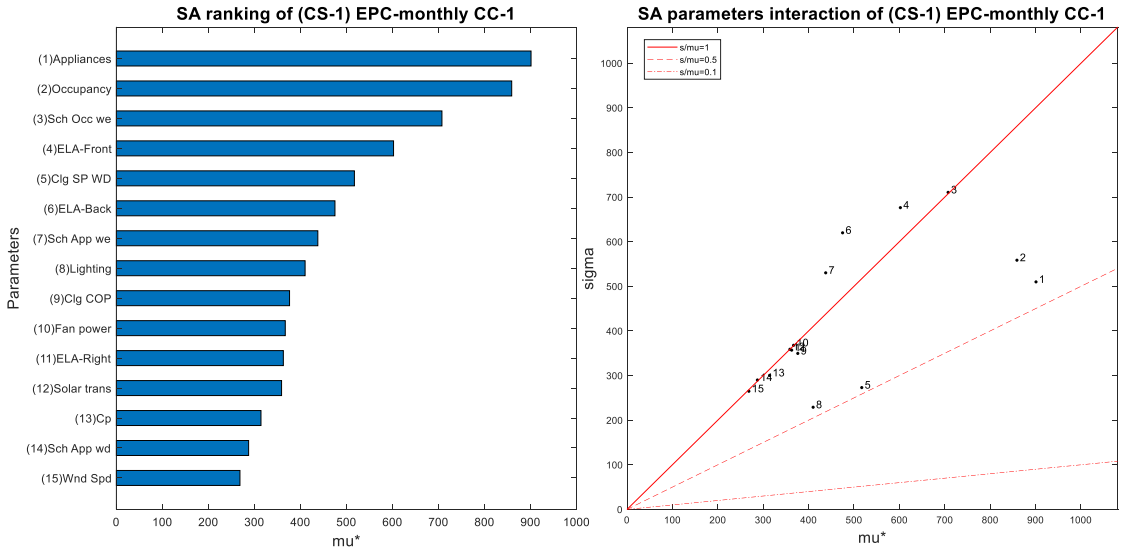


Figure 6.2 SA results of dominant parameters in CC-1, for calibration set: CS-1

6.2.2 Calibration process

After finding the most dominant parameters by conducting the SA, we select the top ones that have the most impact on the QoI, which is also expected to have the largest influence on the results of the calibration. As we move on with the same procedure, more parameters are selected to be added to the prior calibration set leading to a larger CS and a likewise smaller uncertain parameter set. Table 6.3 shows the parameter sets that have been selected in the four steps in this fashion.

Table 6.3 Calibration parameters sets of EPC/CC-1 model used in the calibration steps

CS-1	CS-2	CS-3	CS-4
Front ELA	Front ELA	Front ELA	Front ELA
Back ELA	Back ELA	Back ELA	Back ELA
Right ELA	Right ELA	Right ELA	Right ELA
	Appliances	Appliances	Appliances
	Occupancy	Occupancy	Occupancy
	Appliances Schedules	Appliances schedules	Appliances Schedules
	Cooling setpoints	Cooling setpoints	Cooling setpoints
		Cp	Cp
		Fan power	Fan Power
		Heating setpoints	Heating setpoints
		Lighting	Lighting
			Wind speed
			Occupancy schedules
			Cooling COP
			Heating COP
			Window U-value

In the first calibration we select ELA as the lone calibration parameter(s). Note that in this case the granularity of separately identifiable façade parts leads to three ELA parameters. This step is important to construct an idea of what would be the value of ELA without being influenced by concurrent calibration of other parameters. It is expected that

the resulting ELA will have a (potentially wide) distribution because of the influence of the uncertainties in the regular parameter set. To show the influence of the uncertain model parameters, the deterministic calibration is run multiple times, where each sample has a value for each regular but uncertain parameter, sampled in accordance with their probability distribution. Combining the deterministic calibration results of all samples delivers a distribution of calibrated ELA values. The calibration process is automated to run 50 samples resulting in 50 values for ELA in the EPC-monthly model. The façade configuration of the building leads to distinguishing three dominant parts for which a different ELA should be assigned or calibrated, for simplicity denoted here by “front”, “back”, and “right” ELA. As an example, the resultant of the entire calibration exercise of CC-1 for “front ELA” is shown in Figure 6.3.

Focusing on the outcome of CS-1, one notices a rather wide and uninformative distribution, which was expected because the EPC model has low resolution, the measured data is low resolution, and no other parameters are being calibrated except the ELA. The most probable values of ELA results established from this first step are used as input in the SA that drives the selection of parameters in the next calibration step; the SA result of the top 15 parameters in the second step is shown in Figure 6.4. It leads to the construction of the next calibration set, CS-2 as listed in Table 6.3. This process is continued until the final CS-4 is selected containing all or most of the parameters in the pool, which is almost deterministically calibrated with only a few remaining uncertainties involved, mostly stemming from the partly unknown scenario.

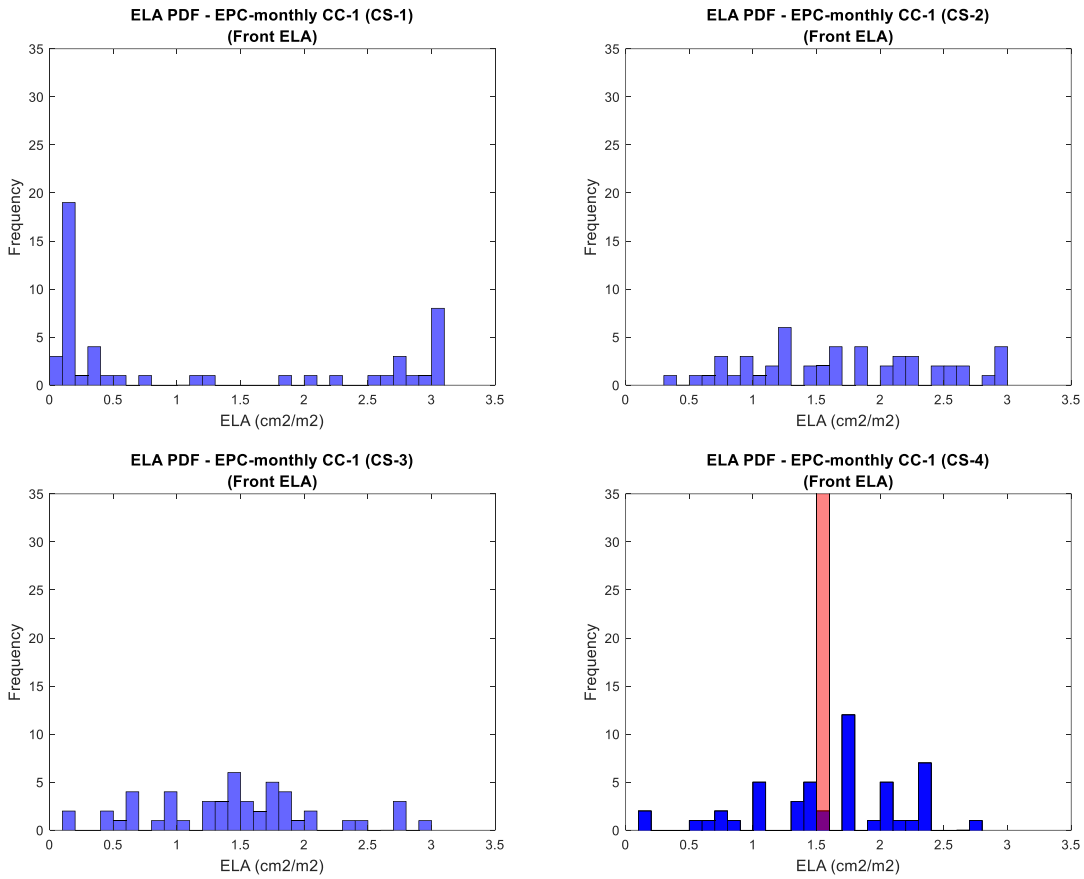


Figure 6.3 Calibration results of ELA (CS-1 through CS-4) for CC-1

Figure 6.3 also shows the ELA results of the next steps of the procedure (with CS-2, 3, 4). Recall that these results are based on QoI-1 (monthly energy use) and CC-1. Each step results in a different ELA distribution. ELA values of CS-2 through CS-4 prove to be still as wide as CS-1. The possible reason for this is the additional input variables still have nonlinear effect or interaction with the ELA, or the MFU is high, which puts severe limits on the ability to arrive at a more precise ELA estimate. This can be extracted from the SA results from Figure 6.2 to Figure 6.6 as all dominant parameters seem still to have high influence on the model with μ_i* ranging between 500 kWh/yr to 800 kWh/yr. However, CS-4 shows a rather constrained distribution compared to the other distributions. Wind speed in this calibration set is the most dominant parameter as shown in Figure 6.6, which

could indicate that the combination of Cp from the previous set and wind speed in this final calibration results in a more pronounced ELA estimation. This implies in this particular case that if ELA is a calibration parameter then uncertainty of Cp and wind speed is not crucial. Finally, the red bar in CS-4 is the deterministic calibration value for ELA when all parameters are included in the calibration. The purpose of this step is to construct a static point of where ELA can be located if we assume no uncertainty in the model and assume (falsely) that all unknown parameters can be calibrated based on the available data. This result can however be considered unrealistic as will be further discussed in section 6.5.

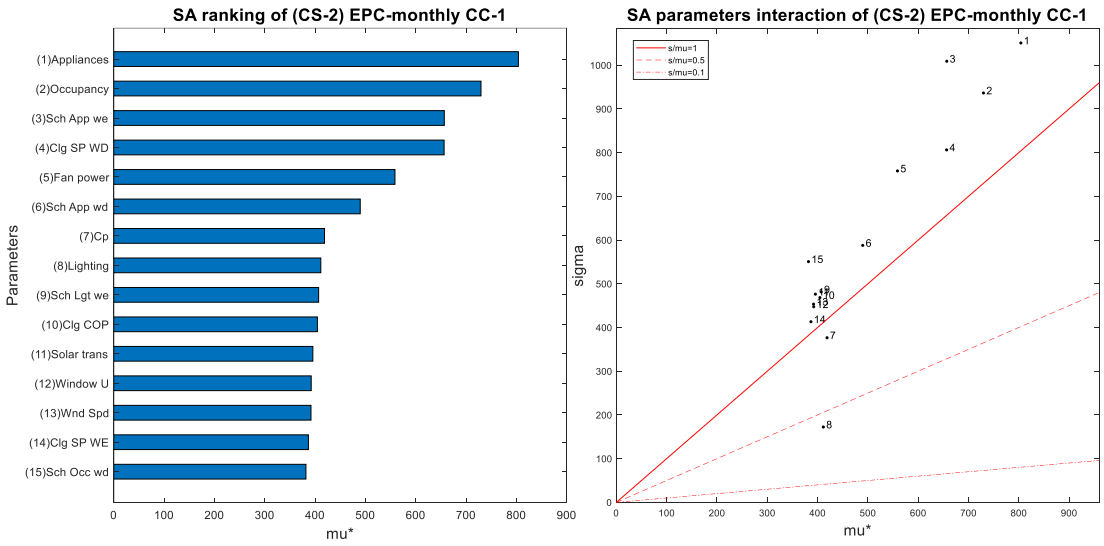


Figure 6.4 Ranking of dominant parameters for CC-1, with calibration set CS-2

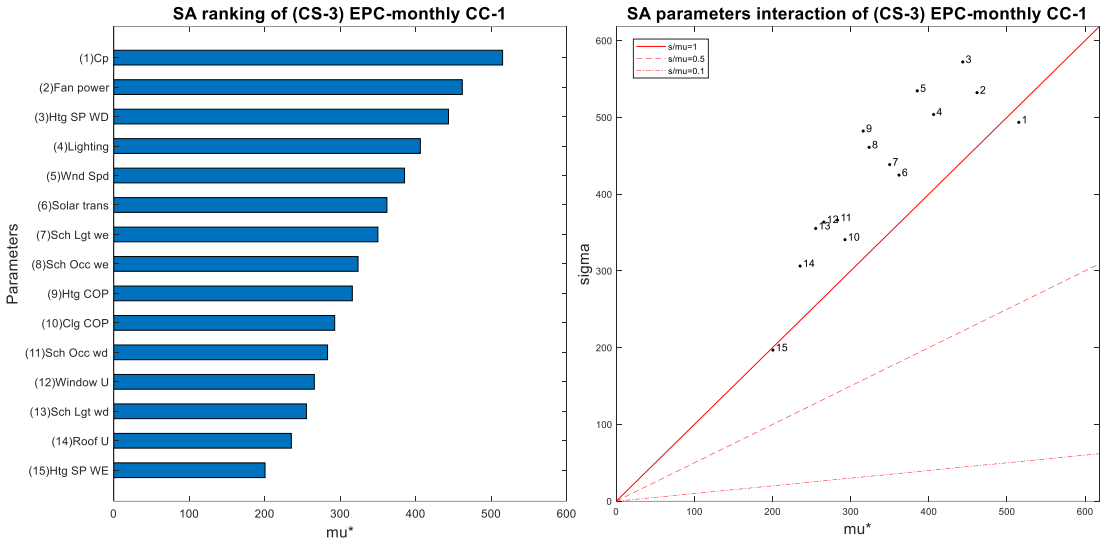


Figure 6.5 Ranking of dominant parameters for CC-1, with calibration set CS-3

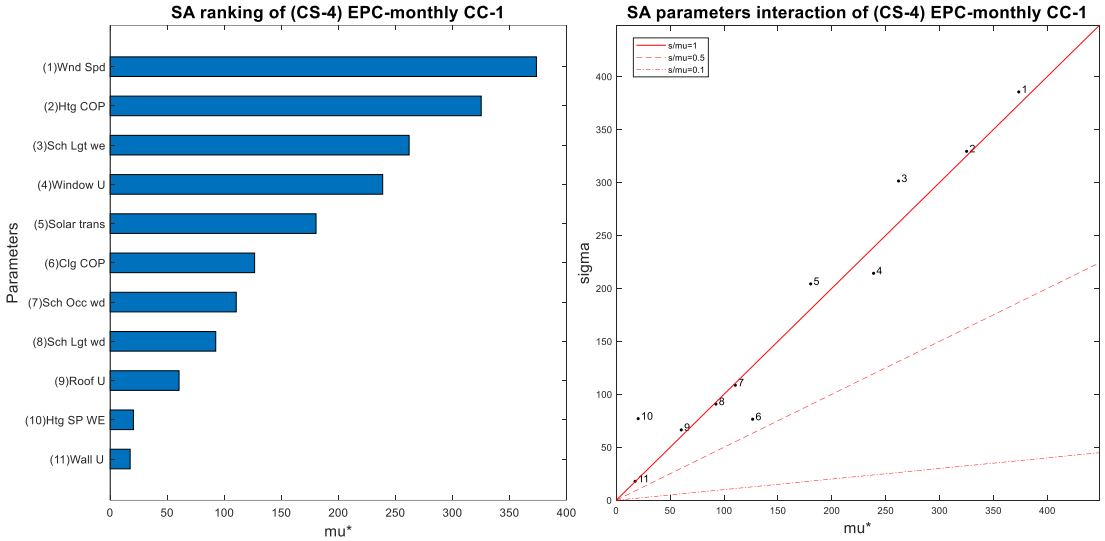


Figure 6.6 Ranking of dominant parameters for CC-1, with calibration set CS-4

6.2.3 Adding QoI-2 in the calibration

The calibration process is conducted following the same fashion but with more granular data to test the potential improvement if additional measurement data is added to the pool. In this case we include indoor temperature in zone 5 (see Figure 6.7) for an entire year. With the additional outcome the calibration criterion is changed to include QoI-1 and

QoI-2. This leads to CC-2 as introduced earlier, with the reflection that the resolution of both outcomes can be different, in this case monthly for energy and hourly for temperatures. There are different ways to define the calibration target, i.e., minimizing the discrepancy in multiple outcomes, as in this case 1) the discrepancy in monthly energy, and 2) the discrepancy in hourly temperatures in a selected zone. In this case we perform the calibration for the aggregated discrepancy by a weighting factor of W_E=1 and W_T=15.6 for both outcomes from Eq. (5.4) following the explanation in 5.3.2. This way a single objective is minimized, and essentially the same calibration process from the previous section is repeated but this time with EPC-hourly.

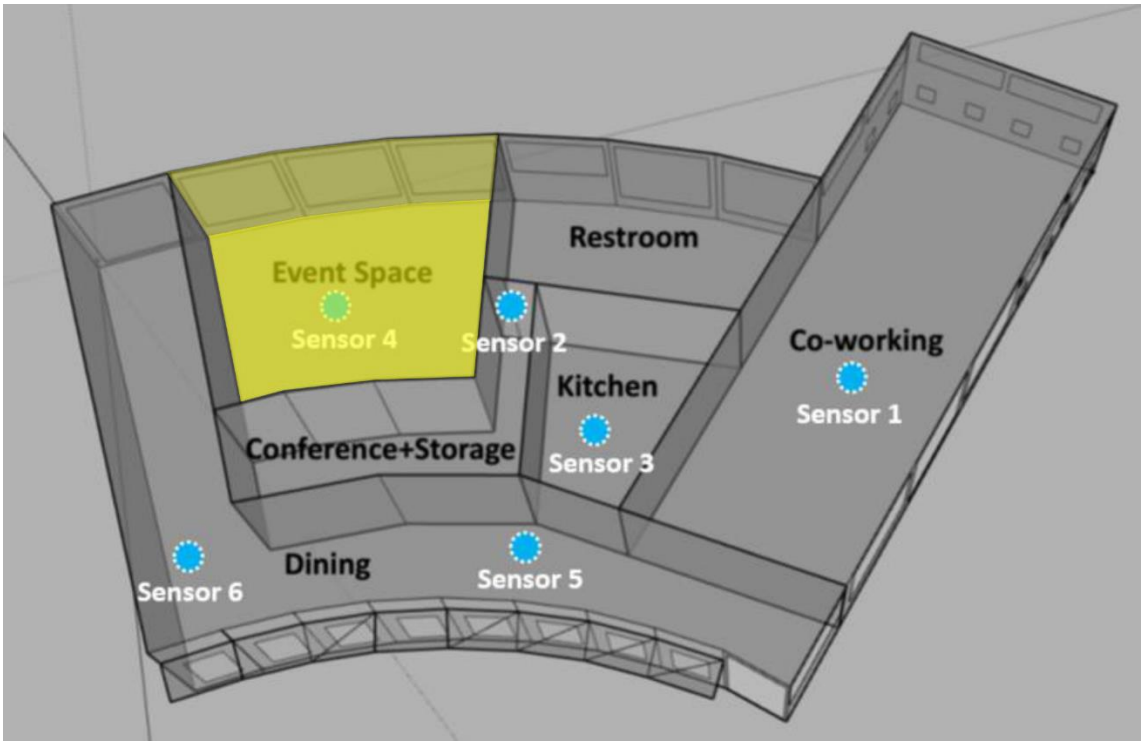


Figure 6.7 3D layout of temperature measurement locations. The tested zone is highlighted yellow (zone 5)

The SA process of the combined criteria results in new sets of calibration parameters as presented in Table 6.4.

Table 6.4 Calibration parameters sets of EPC/CC-2 used in the calibration steps

CS-1	CS-2	CS-3	CS-4
Front ELA	Front ELA	Front ELA	Front ELA
Back ELA	Back ELA	Back ELA	Back ELA
Right ELA	Right ELA	Right ELA	Right ELA
	Cooling setpoints	Cooling setpoints	Cooling setpoints
	Appliances	Appliances	Appliances
	Heating setpoints	Heating setpoints	Heating setpoints
	Appliances schedules	Appliances schedules	Appliances schedules
	Lighting	Lighting	Lighting
		Fan Power	Fan Power
		Solar transmittance	Solar transmittance
		Window U-value	Window U-value
		Cp	Cp
			Cooling COP
			Wind Speed
			Occupancy
			Occupancy schedules
			Wall U-value

Similar to the previous process in CC-1, we begin with a parameter screening to understand the characterization and influence of these parameters as a whole on the energy model. Therefore, from Figure 6.8 we notice that ELA-Front (No. 4) has high level of importance on the SA rank with a nonlinear effect or interaction with other variables (i.e., $\sigma/\mu^* > 1$). This finding is very similar to what we established from the first SA result in CC-1. Hence, this suggests that with the implementation of the weighting factor, the ELA still has the same magnitude of importance on the model.

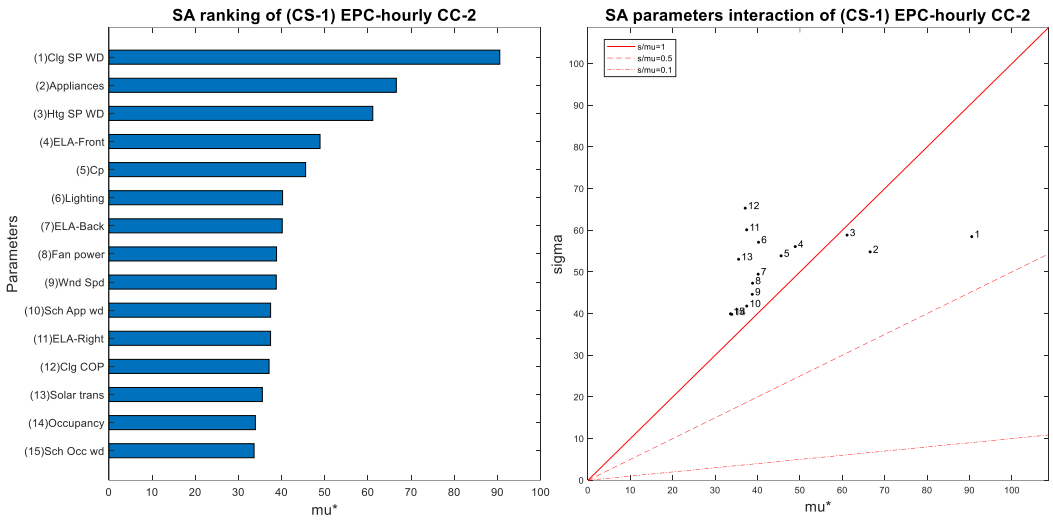


Figure 6.8 Ranking of dominant parameters for CC-2, with calibration set CS-1

Consequently, Figure 6.9 shows the results of the entire calibration exercise of CC-2 for front ELA.

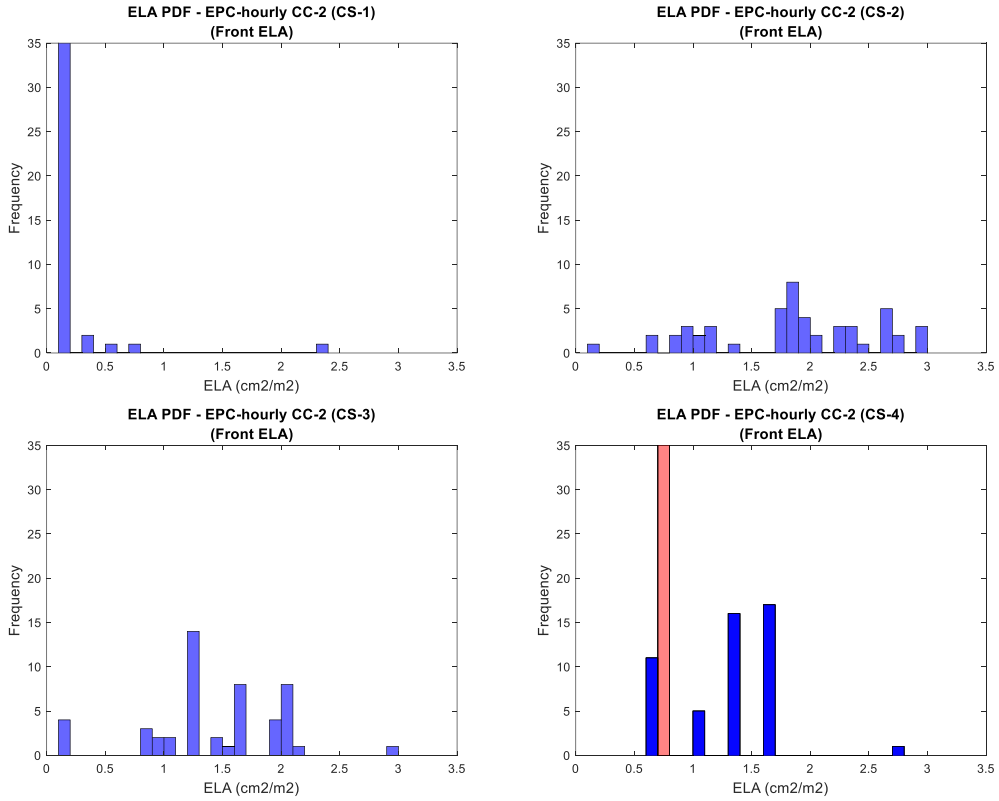


Figure 6.9 Calibration results of ELA (CS-1 through CS-4) for CC-1

The calibration exercise results in more constrained distributions compared to that we establish with CC-1. Although the ELA distribution with CS-1 delivers a pronounced distribution compared to other ELA estimates, this distribution is hardly credible since the systemic error in the tool my force the ELA to the lowest value admissible, i.e. the lower constraint that we assigned (i.e., 0.05). This is an example of how a small calibration set can lead to large bias in the estimate of the parameters, even when an uncertainty range is assumed in non calibration parameters. The SA can provide an indication of this. It can be seen from Figure 6.8 that the first three important parameters are located in the region between 0.5 and 1 (i.e., $1 > \sigma/\mu^* > 0.5$), which indicates an almost monotonic effect with the model, whereas ELA (No. 4) is in the interaction region. Therefore, ELA at this point seems to have a great impact on the model, which makes it always pick low values to achieve better optimization, apparently making up for large MFU inherent in the model.

Further, ELA distributions found with CS-2 and CS-3 are still uninformative with the caveat that CS-3 is forming a peak around 1.25 cm2/m2. The impact of the additional parameters can be explained by the mean value (μ^*) of the fan power (No.1), solar transmittance (No.2), window U-value (No.3), and Cp (No.4), which are generally larger than the other parameters, see Figure 6.11.

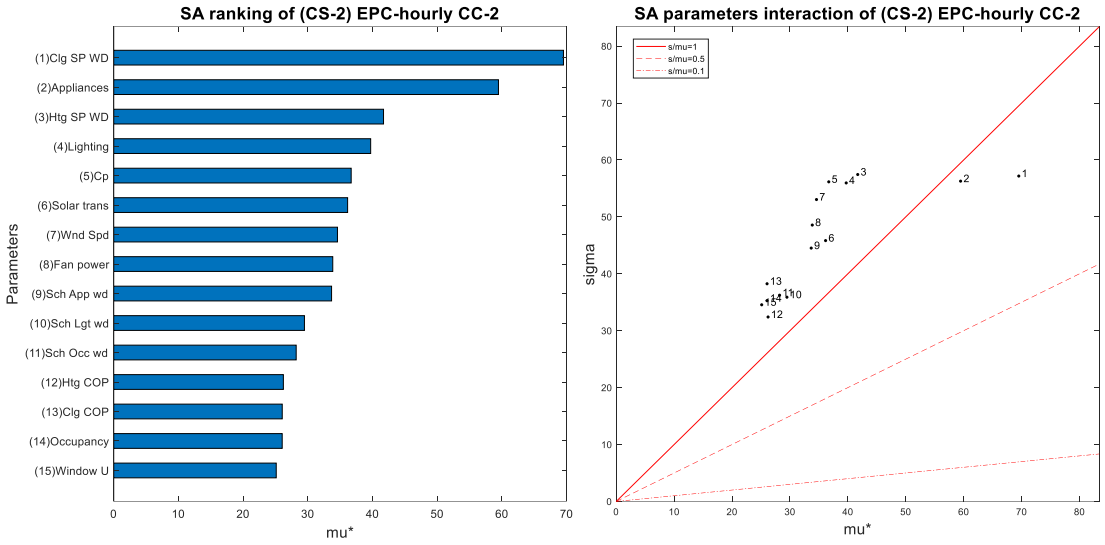


Figure 6.10 Ranking of dominant parameters for CC-2, with calibration set CS-2

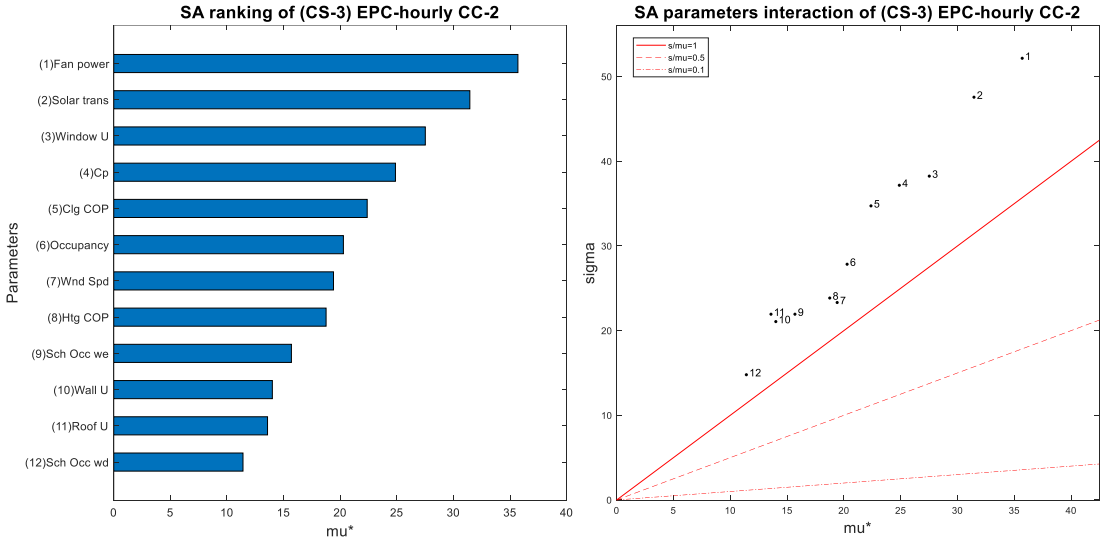


Figure 6.11 Ranking of dominant parameters for CC-2, with calibration set CS-3

The ELA distribution found with CS-4 is slightly less uninformative because it has less shorter tails compared to the others. We can expect that the addition of wind speed in the calibration set affects the model significantly as can be extracted from Figure 6.12.

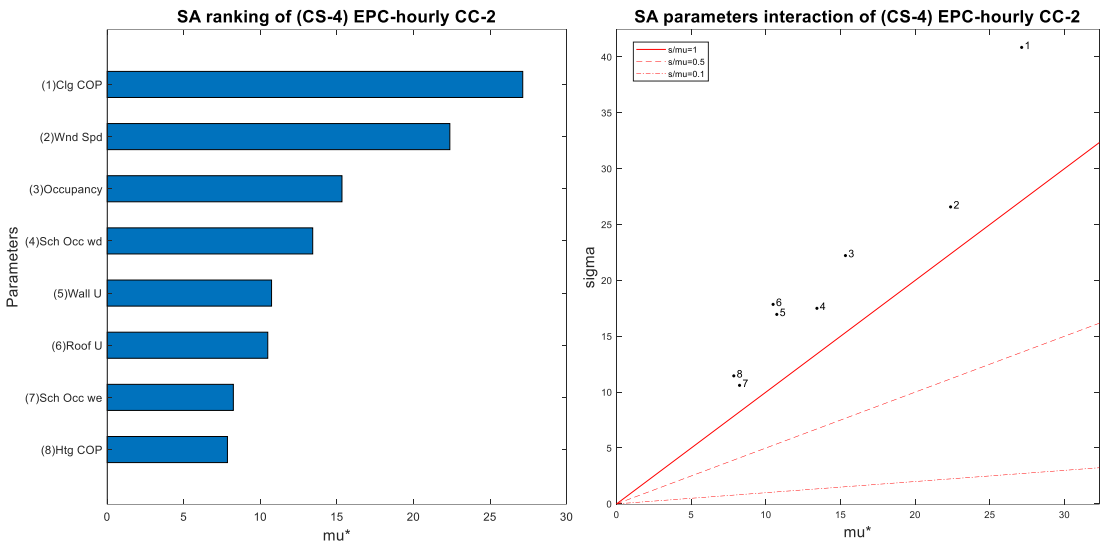


Figure 6.12 Ranking of dominant parameters for CC-2, with calibration set CS-4

However, all distributions are uninformative about the true ELA value, and if we look at all distributions in one picture, we come to the conclusion that the ELA distributions are only slightly better than the uninformative prior we established earlier. By comparing results for EPC-monthly with EPC-hourly, the results indicate that the higher resolution model provides a "better" estimation of ELA although the improvements are marginal and both models do not lead to credible ELA estimates. We will discuss the reflection of these results with the goodness of the energy model in CHAPTER 8.

6.3 Calibration of EnergyPlus model with monthly consumption data

Quantification of uncertainties in model parameters depends on the choice of a building energy model as even "similar" parameters can have a different realization depending on the granularity and related simplifications of each model. For instance, a low-resolution model describes the characteristics of major components whereas the higher fidelity simulation model requires a detailed level of model parameters that describe the

physical behavior of individual parts. Owing to the different granularity level, parameter uncertainty in a higher order dynamic simulation model (in our case EnergyPlus) can differ from that in the normative model. Table 5.1 is still an adequate summary of the uncertainties around the initial values assigned to the EnergyPlus model parameters, and the ranges reflect the correct associations with this model fidelity.

6.3.1 Sensitivity analysis

With the Morris method, 54 independent samples are generated to obtain the elementary effects of individual parameters. 1100 simulations are run to support the appropriate selection of the calibration parameters among the total 58 uncertain parameters as listed in Table 6.5.

Table 6.5 Parameter uncertainty in model testing and acronym definition

Acronym	Definition	Uncertainty range	Comment
ELA-Back	ELA of the E side (cm^2/m^2)	LogNorm(1.28,0.88)	
ELA-Front	ELA of the W side (cm^2/m^2)	LogNorm(1.28,0.88)	
ELA-Right	ELA of the N (cm^2/m^2)	LogNorm(1.28,0.88)	
Cp-Front, Back, Right (21 points)	Wind pressure coefficient	RelativeNorm(1, 0.25)	The uncertainty is relative to every Cp value
Wnd-Spd	Local wind speed (m/s)	RelativeNorm(1, 0.42)	The uncertainty is relative to every hour wind speed value
Clg-COP	Cooling COP	Norm(3.75, 0.02)	
Htg-Coil-Eff	Heating COP	Norm(1.75, 0.09)	
Fan-Eff	HVAC fan efficiency	Tring(0.2, 0.6, 0.9)	
Clg-sp-WD	Cooling setpoint in the weekdays (ºC)	Tring(20, 24, 26)	
Htg-sp-WD	Heating setpoint in the weekends (ºC)	Tring(18, 22, 24)	
Clg-sp-WE	Cooling setpoint in the weekdays (ºC)	Tring(22, 26, 28)	
Htg-sp-WE	Heating setpoint in the weekends (ºC)	Tring(16, 18, 22)	

Table 6.5 (continued)

Roof-cond	Roof U-value (W/m^2·K)	Norm(0.24, 0.02)
Wall-cond	Wall U-value (W/m^2·K)	Norm(0.35, 0.02)
Wndw-cond	Window U-value (W/m^2·K)	Norm(3, 0.6)
Wndw-Trans	Window's solar transmittance	Norm(0.75, 0.05)
Occ-1, 2, …, 6	Occupancy internal gain (m^2/person)	Uniform(8, 16)
Lights-1, 2, …, 6	Lighting internal gain (W/m^2)	Norm(12, 3)
EE-1, 2, …, 6	Plug loads internal gain (W/m^2)	Norm(16, 2.5)

Figure 6.13 shows the ranking of uncertain parameters in the EnergyPlus model for consecutive choices of the first calibration parameter set (CS-1). They are ordered by their relative importance in the selected QoI which is limited to aggregated monthly energy consumption, corresponding to the sparse availability of measured data in this case study. The ranking is based on the relative change in output means μ_i^*. The right side of Figure 6.13 shows the interaction wedge among the sampled parameters.

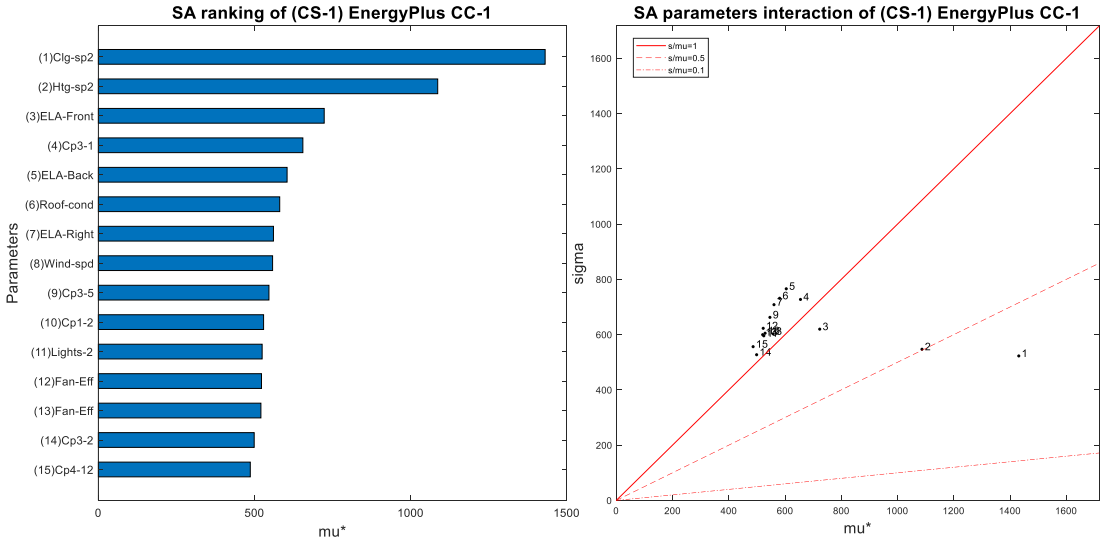


Figure 6.13 SA results of dominant parameters for CC-1, with calibration set CS-1

When comparing this SA result with its peer obtained for EPC (see Figure 6.2), we notice that the SA for both EPC and EnergyPlus shows that ELA has a substantial effect

on QoI-1 among all input variables. This interpretation must be handled with care, as the variance-based SA shows the effect of uncertainty in a parameter on the total resulting uncertainty in a chosen QoI. The reason that ELA is dominant can be largely attributed to the fact that the initial assessment of uncertainty in the ELA parameter is large. Refinement of the uncertainty ranges in parameters (as more information is used in their estimates) may lead to significant changes in the ranking order. This however does not change our procedure as the primary objective is the calibration of ELA which is therefore always the only calibration parameter in CS-1, regardless of the SA outcome of the initial step.

6.3.2 Calibration process

Table 6.6 shows the four parameter sets that are selected to conduct the EnergyPlus calibration.

Table 6.6 Calibration parameter sets of EnergyPlus/CC-1

CS-1	**CS-2**	**CS-3**	**CS-4**
Front ELA	Front ELA	Front ELA	Front ELA
Back ELA	Back ELA	Back ELA	Back ELA
Right ELA	Right ELA	Right ELA	Right ELA
	Coil heating efficiency	Coil heating efficiency	Coil heating efficiency
	Occupancy	Occupancy	Occupancy
	Roof conductivity	Roof conductivity	Roof conductivity
	Lights	Lights	Lights
		Cp	Cp
		Appliances	Appliances
			Wind speed
			Window transmittance
			Cooling COP
			Cooling setpoint
			Heating setpoint

Similar to EPC calibrations in the previous sections, CS-1 only contains ELA for the three major façade parts as the lone calibration parameters. This is required to construct an idea of what would be the value of ELA without being influenced by concurrent calibration of other parameters, but being influenced by lack of precise information about the values of the non-calibration parameters.

Following the established procedure, the highest probable ELA value from CS-1 is used as new inputs for the SA of the full calibration set leading to a new ranking as shown in Figure 6.15. From that ranking we construct a new set of calibration parameters, i.e., CS-2 in Table 6.6. This will be continued in a similar fashion for CS-3 and CS-4. All calibrations are done with only QoI-1 (monthly energy). Accordingly, the calibration is automated to run 50 samples from the calibration parameter set resulting in 50 values for ELA. The results of this calibration exercise for "front ELA" are shown in Figure 6.14.

ELA results of CS-1 and CS-2 are still as wide as those we established with EPC-monthly ranging between 0.05 to 3.25 cm^2/m^2. It can be seen from Figure 6.15 that though the μ* value of the input variables is varying with the position, but the μ* of the heating efficiency (No. 1) and the occupancy (No. 2) are generally larger than the other parameters. Hence, these two parameters are affecting the model output in a significant way with less impact on the ELA.

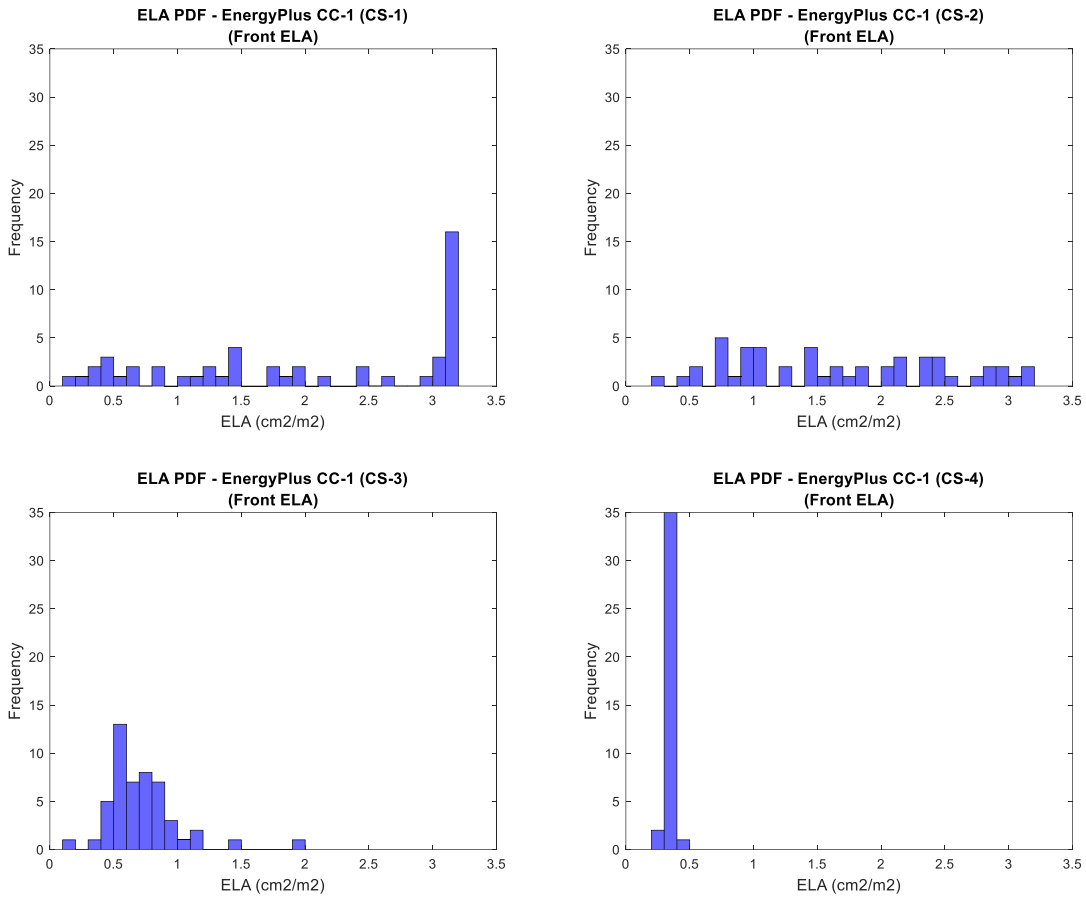


Figure 6.14 Calibration results of ELA (CS-1 through CS-4) for CC-1

On the other hand, CS-3 and CS-4 are more informative than the distributions that were developed from the EPC calibrations. In particular, ELA calibrated with EnergyPlus has a higher probability of being in the range 0.3-1.0 cm^2/m^2 in CS-3, and around 0.3 cm^2/m^2 in CS-4. This indicates that the higher fidelity of EnergyPlus may lead to a modest improvement in the resulting ELA estimate, assuming uncertainties in all non-calibration parameters to be identical. This can be extracted from Figure 6.16 and Figure 6.17; among all of the input parameters, the Cp parameters and wind speed have the highest impact on the energy consumption (QoI-1), whereas the other parameters have minimum impact on the model output. Therefore, their influence on energy consumption variations can be neglected. Moreover, it can be observed that the corresponding σ values of Cp and wind

speed are large, indicating that these two parameters have the greatest interaction with the other parameters when all parameters are affecting the model output. In other words, energy consumption shows a nonlinear behavior with respect to these two parameters.

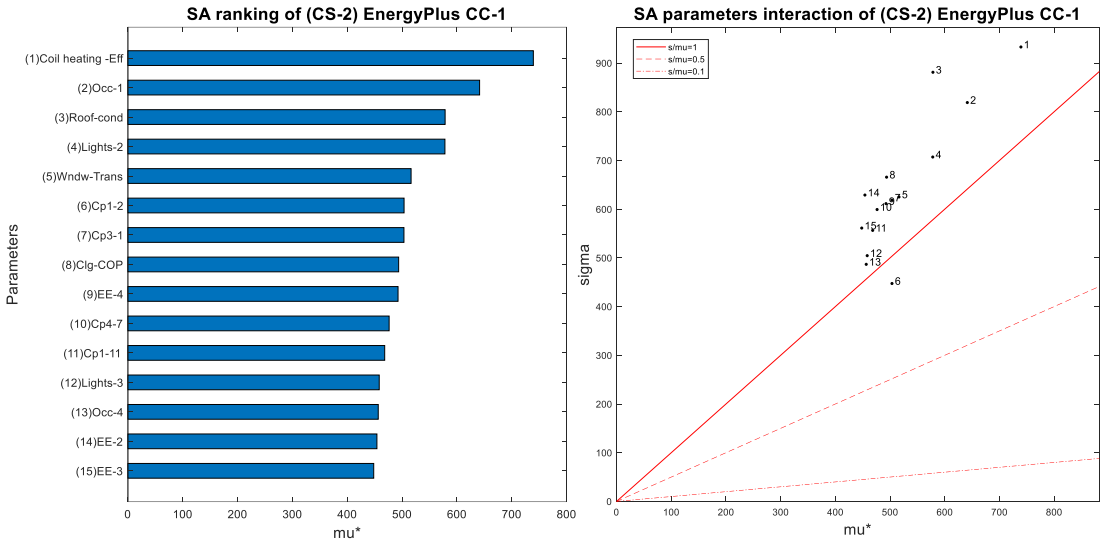


Figure 6.15 SA results of dominant parameters for CC-1, with calibration set CS-2

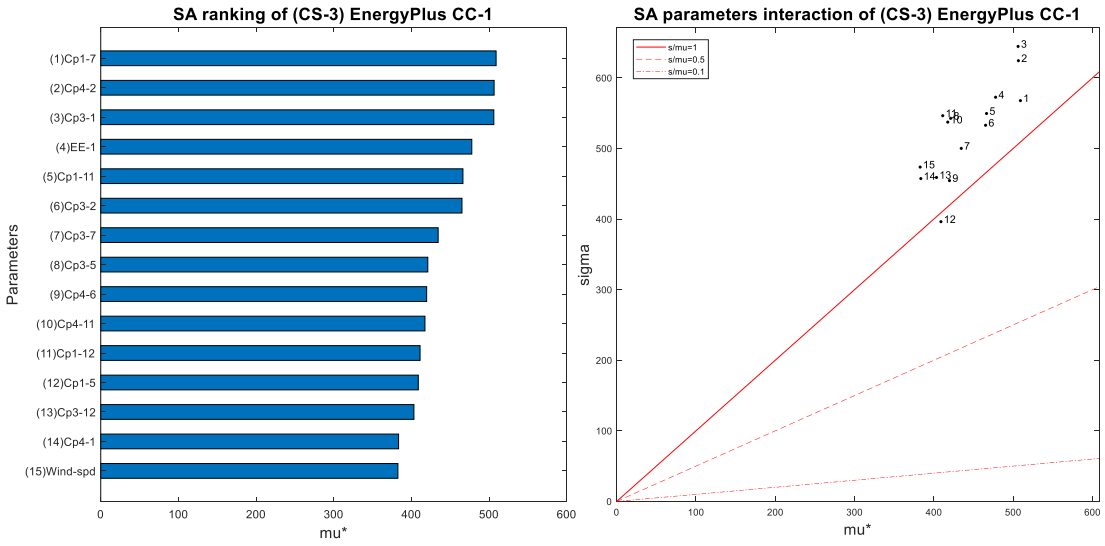


Figure 6.16 SA results of dominant parameters for CC-1, with calibration set CS-3

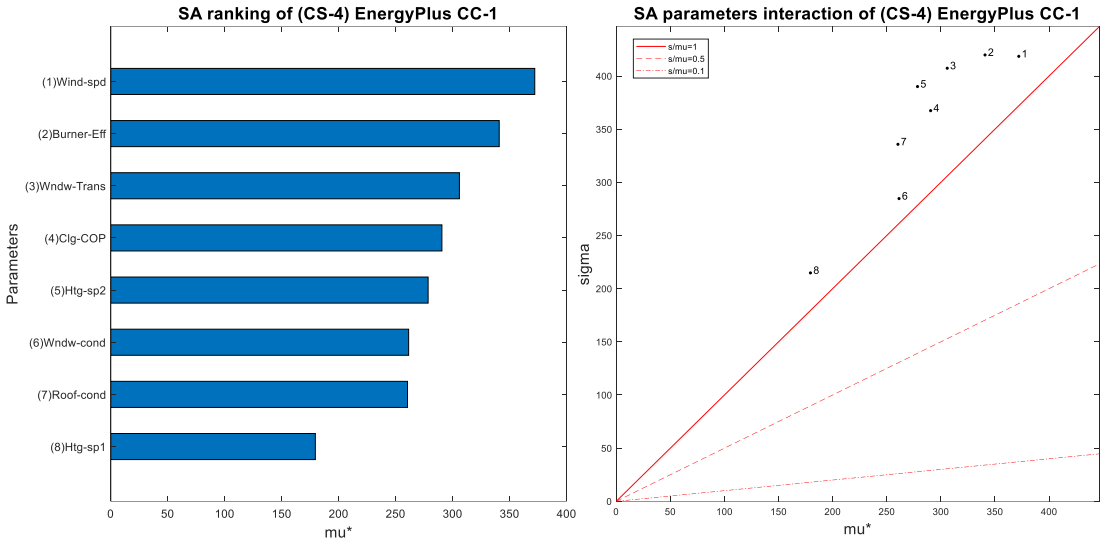


Figure 6.17 SA results of dominant parameters for CC-1, with calibration set CS-4

6.3.3 Adding temperature information in the calibration

The final step is the inclusion of QoI-2 (indoor temperature) of zone 5 into the calibration target. As explained before, the calibration is conducted with the target to minimize the discrepancy in both measured energy consumption and indoor temperature. In this case we perform the calibration for the aggregated discrepancy in a similar fashion to EPC-hourly using weighting factors of W_E=1 and W_T=15.6 for both outcomes from Eq. (5.4). This way a single objective is minimized, and essentially the same calibration process from the previous section can be used. The SA for the combined calibration process results in new sets of calibration parameters as presented in Table 6.7.

Table 6.7 Calibration parameters sets of EnergyPlus/CC-2

CS-1	CS-2	CS-3	CS-4
Front ELA	Front ELA	Front ELA	Front ELA
Back ELA	Back ELA	Back ELA	Back ELA
Right ELA	Right ELA	Right ELA	Right ELA
	Occupancy	Occupancy	Occupancy

Table 6.7 (continued)

	Coil heating efficiency	Coil heating efficiency	Coil heating efficiency
	Roof conductivity	Roof conductivity	Roof conductivity
	Cp	Cp	Cp
		Appliances	Appliances
		Lighting	Lighting
		Wall conductivity	Wall conductivity
			Window conductivity
			Window transmittance
			Wind speed

Figure 6.18 shows overlay distributions of the entire calibration exercise of CC-2 and CC-1 for the front ELA.

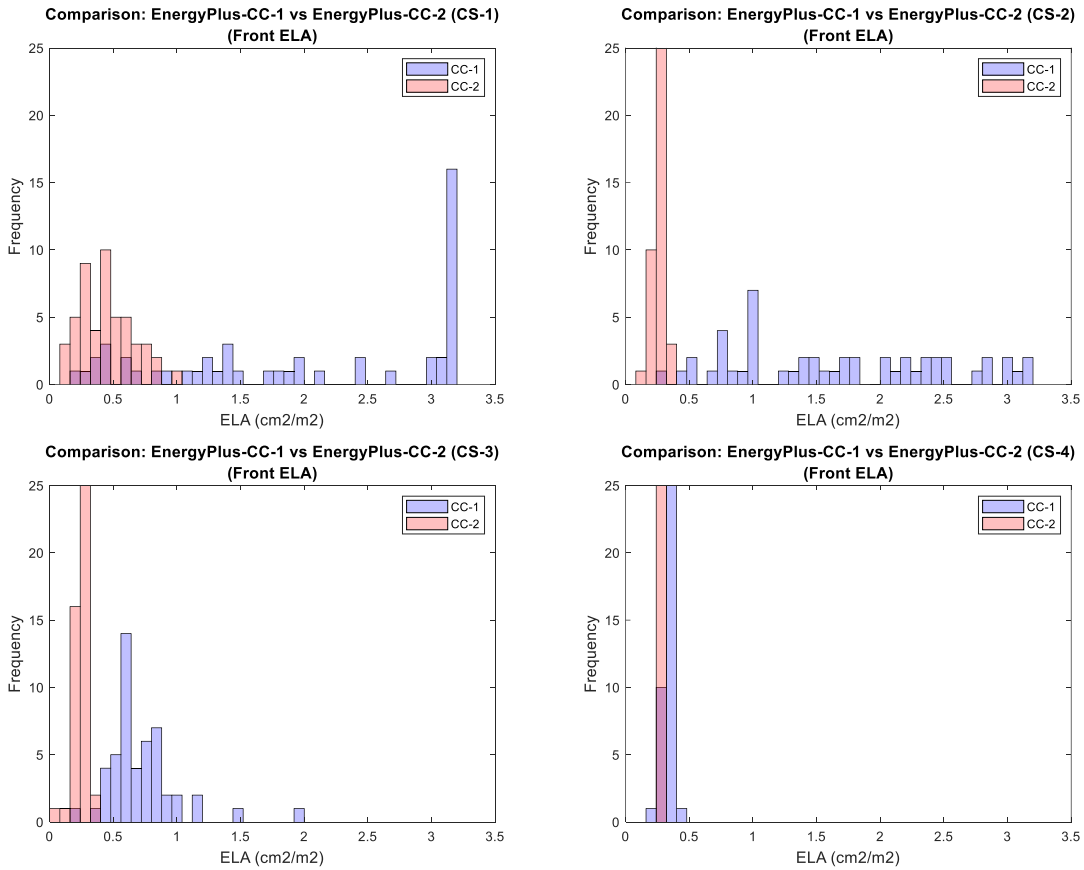


Figure 6.18 Calibration results of ELA (CS-1 through CS-4) for CC-1 and CC-2

ELA distributions of this calibration exercise are more informative and significantly constrained when compared with the results of CC-1 in EnergyPlus (blue bars). The results of SA in Figure 6.19 indicate that ELA-front (No. 4) has high μ* but at the same time located in the region between 0.5 and 1 ($1 > \sigma/\mu^* > 0.5$), which means that ELA at this point has less chance to have interaction with other parameters; therefore, more conclusive predictions. We also notice that the additional parameters in CS-2 and CS-3 have led to similar distributions of ELA, with the most probability being around 0.25 cm^2/m^2. This can be explained by the significant impact of the Cp (No. 4, 6, etc.) on the ELA estimation as it becomes a calibration parameter starting from CS-2 as shown in Figure 6.20. Besides, the corresponding standard deviation (σ) values of Cp are also large, indicating that the Cp have great interaction with the other parameters when all parameters are affecting the model output.

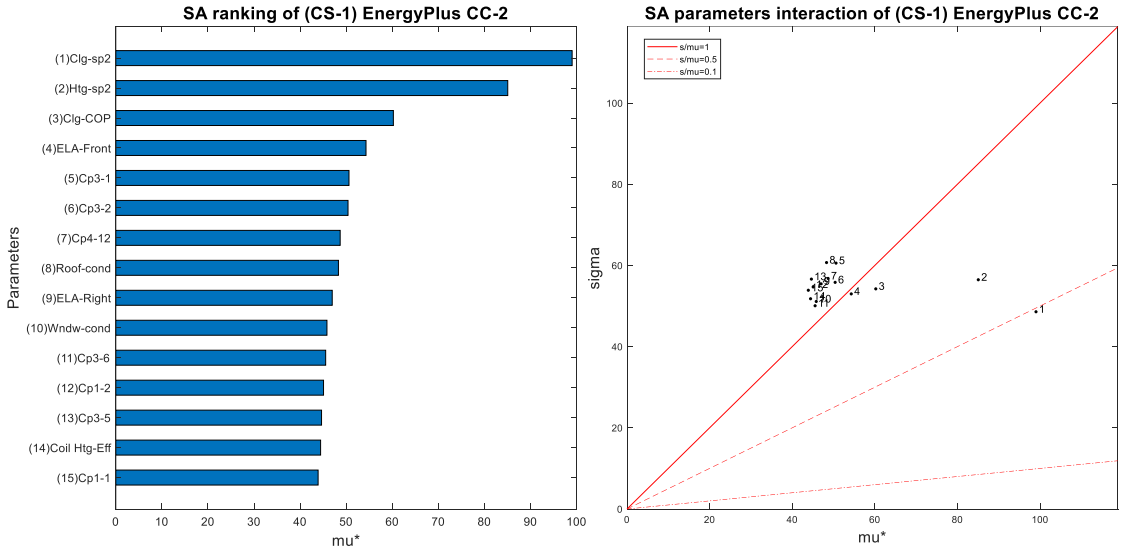


Figure 6.19 SA results of dominant parameters for CC-2, with calibration set CS-1

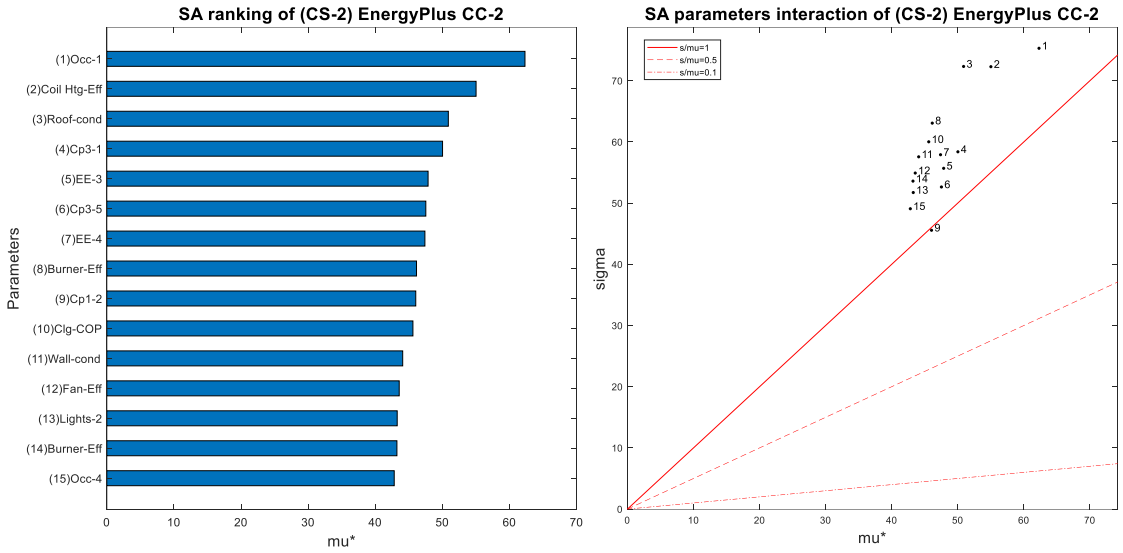


Figure 6.20 SA results of dominant parameters for CC-2, with calibration set CS-1

On the other hand, with CS-4 the result for CC-2 is not much different from that obtained for CC-1. In particular, ELA calibrated with EnergyPlus for CC-2 has a higher probability to be at 0.25 cm^2/m^2. This indicates that the higher fidelity of EnergyPlus may lead to a modest improvement in the resulting ELA estimate, assuming all uncertainties in non-calibration parameters to be identical. The SA results of CS-3 and CS-4 are shown in Figure 6.21 and Figure 6.22. The prudent conclusion at this point is that given the low resolution of the measurements and the lack of precise estimates of the non-calibration parameters, the increased fidelity of the simulation model leads to only modest improvements in general. But for the full calibration set (CS-4) the results for both CC-1 and CC-2 are remarkably narrow and close, indicating that some confidence in the resulting ELA estimate is warranted.

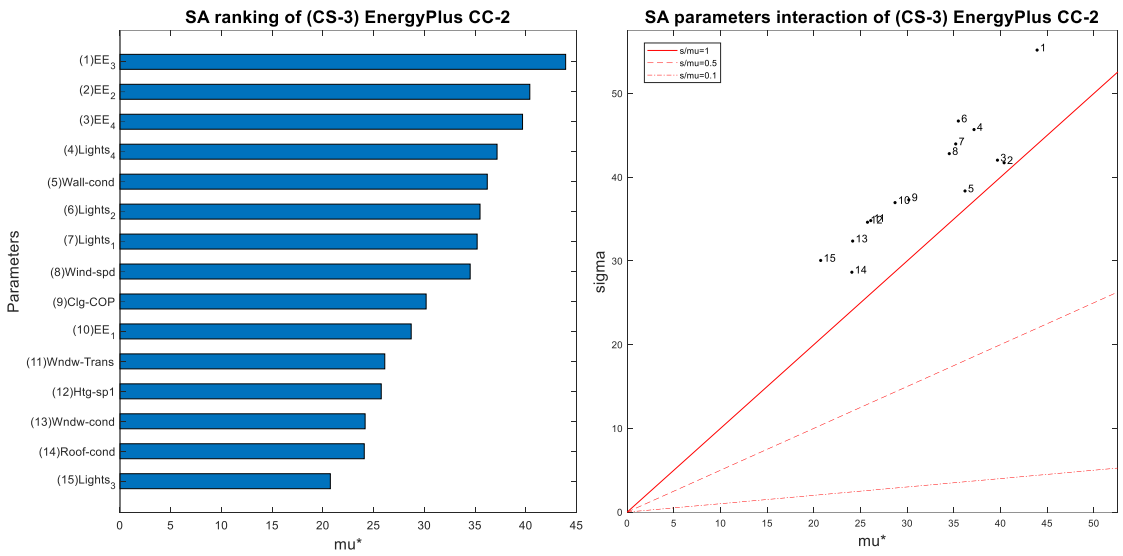


Figure 6.21 SA results of dominant parameters for CC-2, with calibration set CS-3

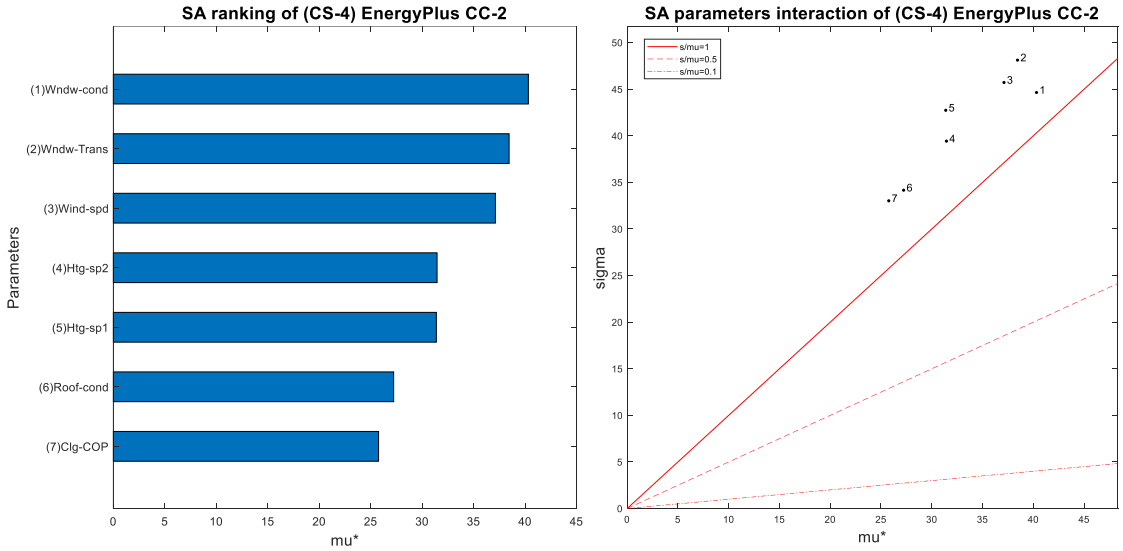


Figure 6.22 SA results of dominant parameters for CC-2, with calibration set CS-4

6.4 Evaluation and results analysis

Comparing the calibrated EPC models with the calibrated EnergyPlus models in terms of calibration metrics (i.e., CVRMSE), leads to the following observations. Figure 6.23 and Figure 6.24 use boxplots to show the variation of the statistical measures over the

two models (EPC and EnergyPlus), for the two calibration criteria, i.e., CC-1 (uses QoI-1 only) and CC-2 (QoI-1+QoI-2) and for the consecutive selection of calibration sets (CS-i).

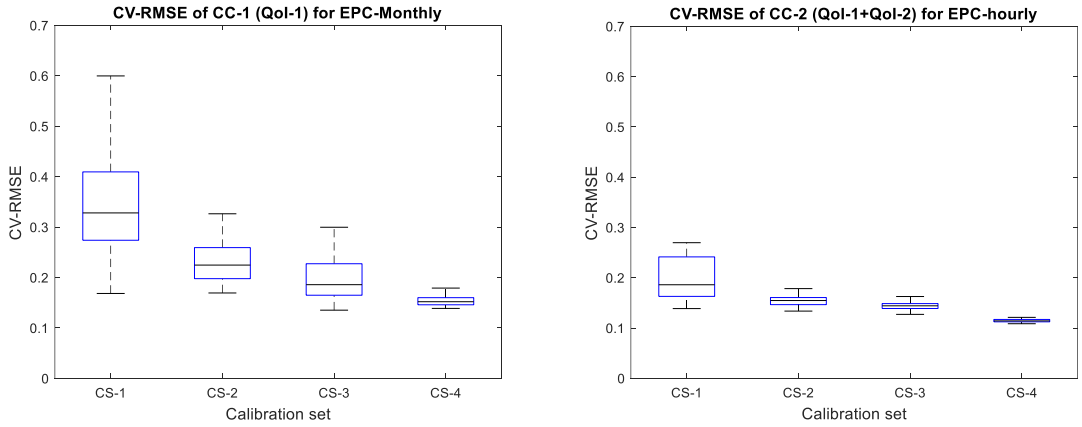


Figure 6.23 The variation of CVRMSE on energy prediction among the calibration sets for EPC using CC-1 (left) and CC-2 (right)

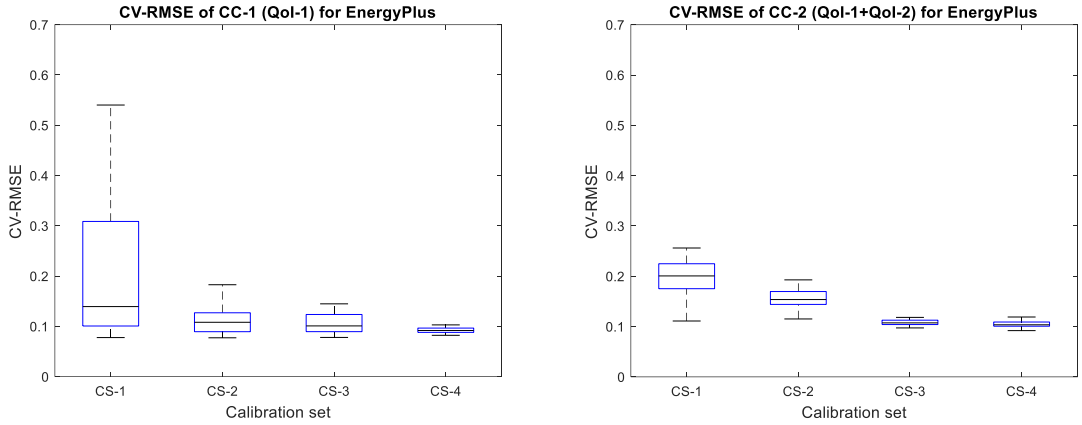


Figure 6.24 The variation of CVRMSE on energy prediction among the calibration sets for EnergyPlus using CC-1 (left) and CC-2 (right)

It becomes evident that the variation in the prediction error of CS-1 is relatively larger than all other calibration sets, which is true for both EPC and EnergyPlus. For example, the CVRMSE from CS-1 in the EPC-monthly/CC-1 model may be as small as 0.18 or as large as 0.60 across the four cases. On the other hand, CS-2, CS-3 and CS-4 give smaller variations in CVRMS, with medians much smaller than with CS-1. This means that

these calibration sets are more stable since the variations among different models are smaller. More discussion about this is provided in CHAPTER 8.

Besides, we cannot rely only on CVRMSE to evaluate the accuracy of the calibrated models. In fact, it is important to inspect whether the observed outcomes are in close similarity with the predicted outcomes. Hence, Figure 6.25 to Figure 6.28 show pairwise comparisons between EPC and EnergyPlus energy data as well as the observed and predicted data within the model itself; every figure presents four graphs for every calibration set (CS-i).

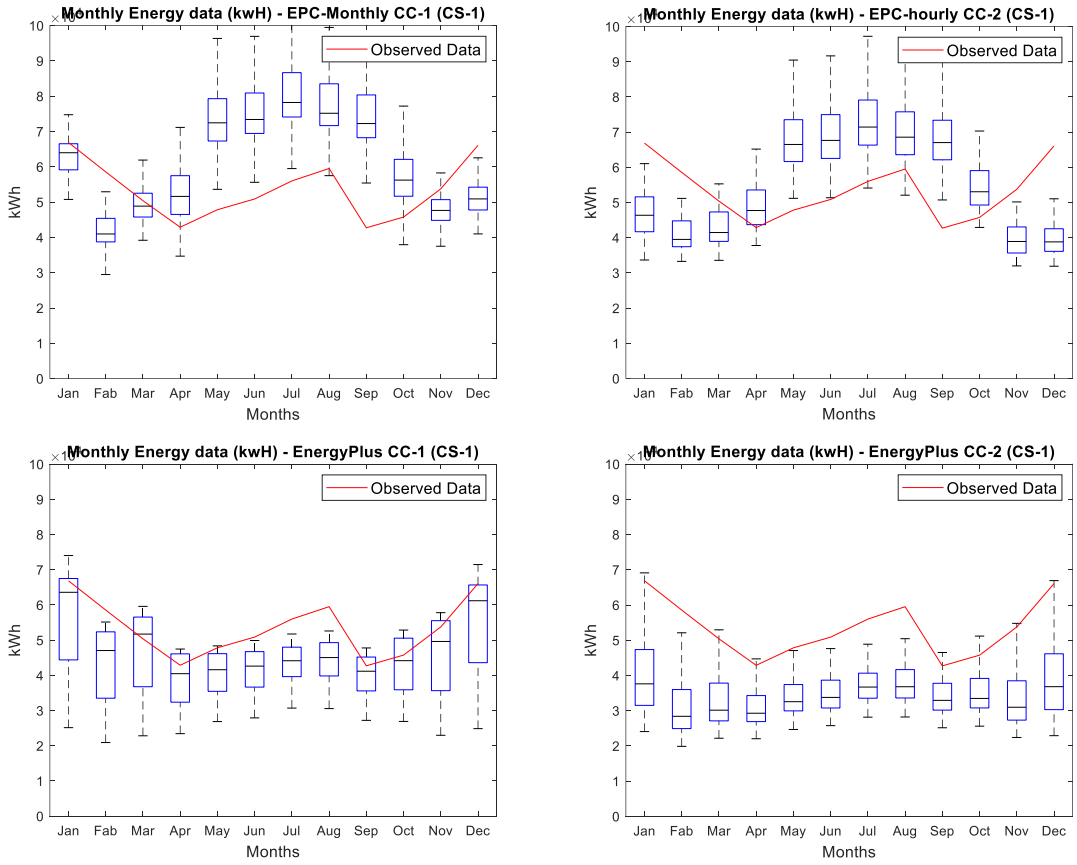


Figure 6.25 Energy consumption with CS-1: EPC (top), EnergyPlus (bottom). Red (observed data), box (prediction)

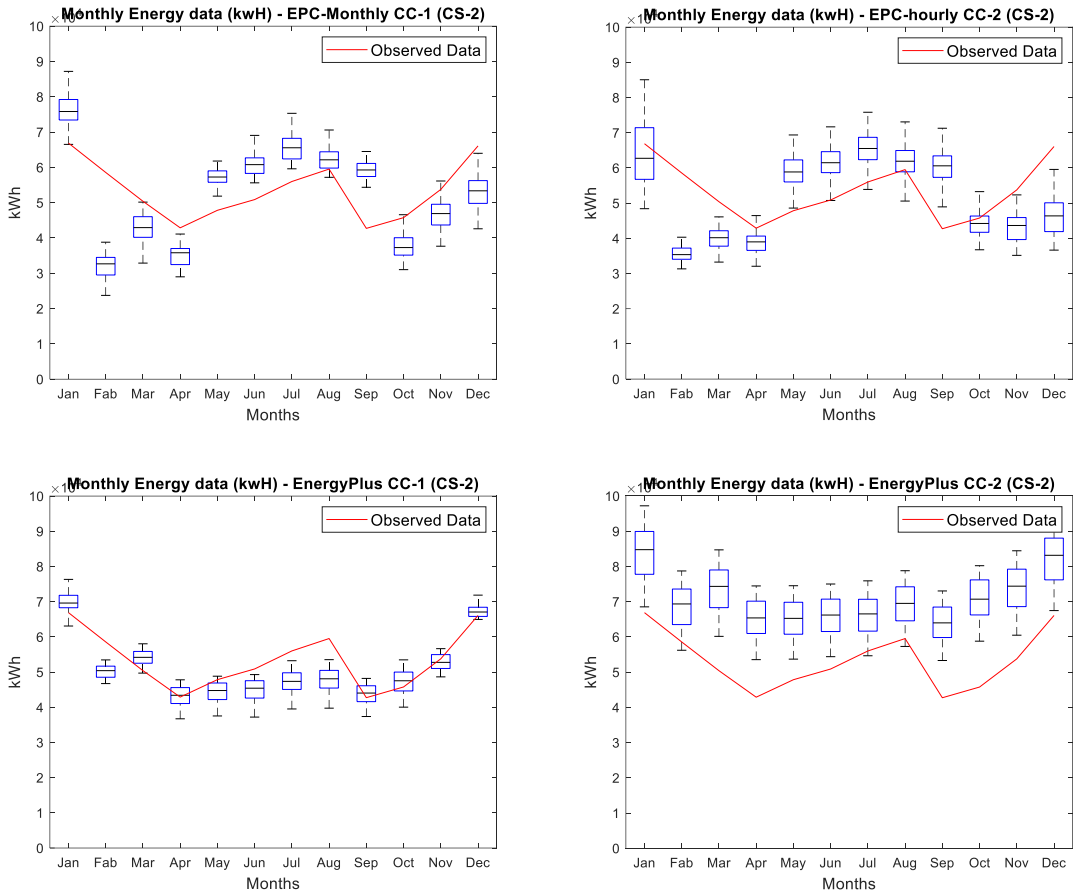


Figure 6.26 Energy consumption with CS-2: EPC (top), EnergyPlus (bottom). Red (observed data), box (prediction)

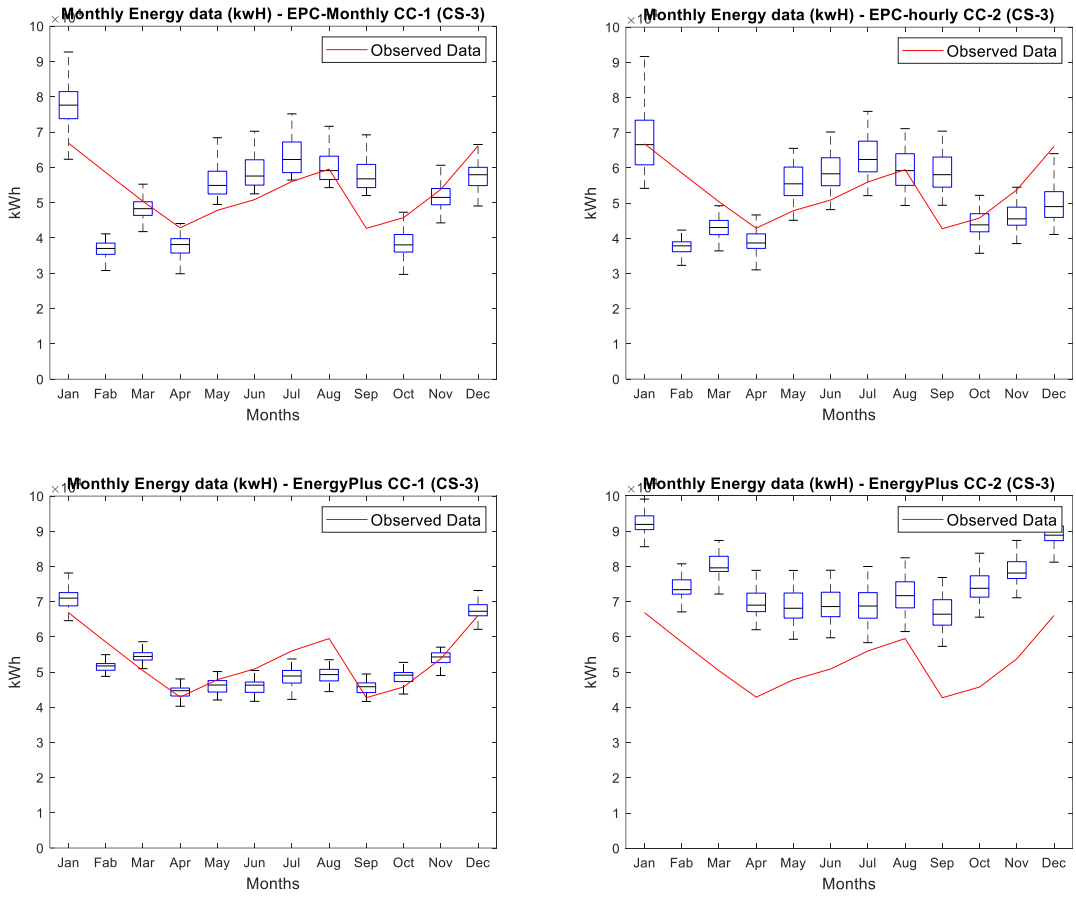


Figure 6.27 Energy consumption with CS-3: EPC (top), EnergyPlus (bottom). Red (observed data), box (prediction)

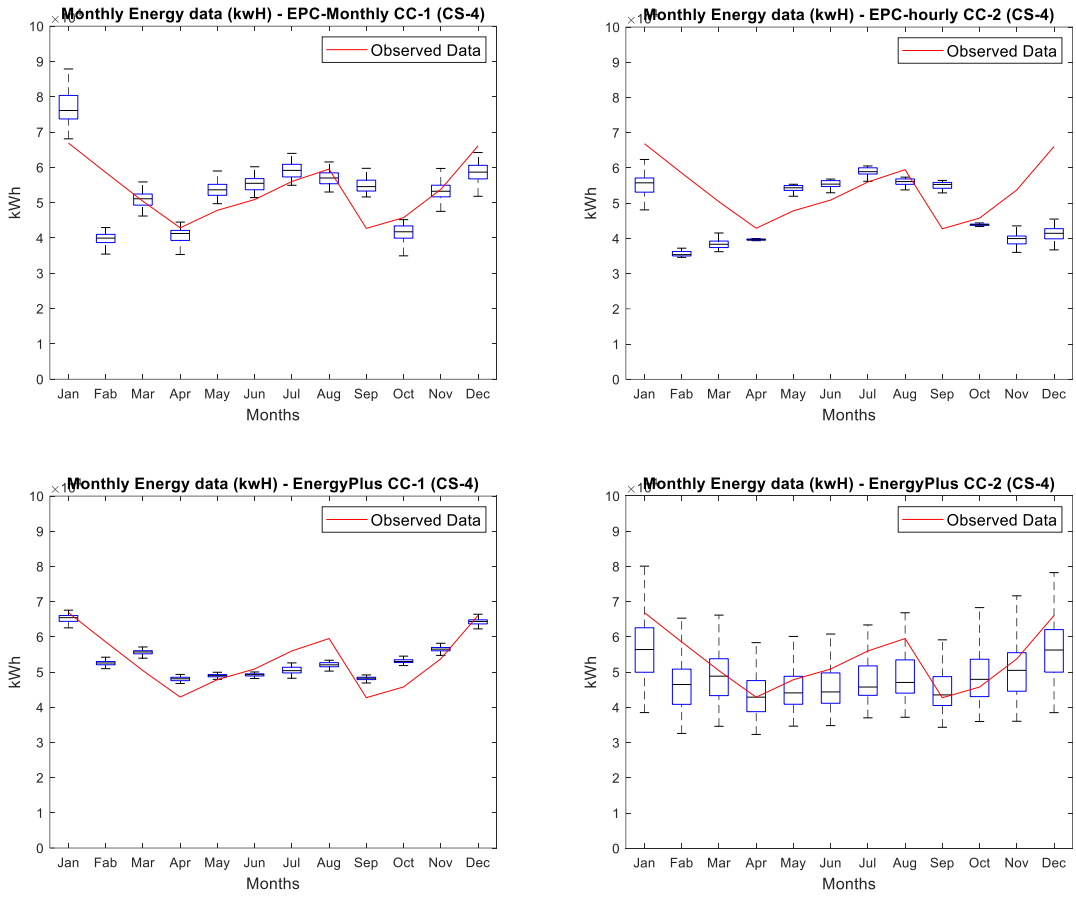


Figure 6.28 Energy consumption with CS-4: EPC (top), EnergyPlus (bottom). Red (observed data), box (prediction)

The figures above present a comparison between the energy consumption outcomes of the calibration criterion 1 (QoI-1) and the calibration criterion 2 (QoI-1 + QoI-2). Although the CVRMSE with CC-2 is seen to be better than with CC-1, the energy consumption predicted by the model is in contrast with the better CVRMSE values obtained with CC-2. Figure 6.29 presents CVRMSE values of CC-2 for QoI-1 only; these results indicate that the error in energy prediction becomes higher when we add QoI-2 to the model. This designates that the overall CVRMSE performs better with QoI-2 than with QoI-1, and based on the weighting factor that we assigned for the combined calibration,

we achieved a smaller CVRMSE value without doing much in improving the energy consumption prediction.

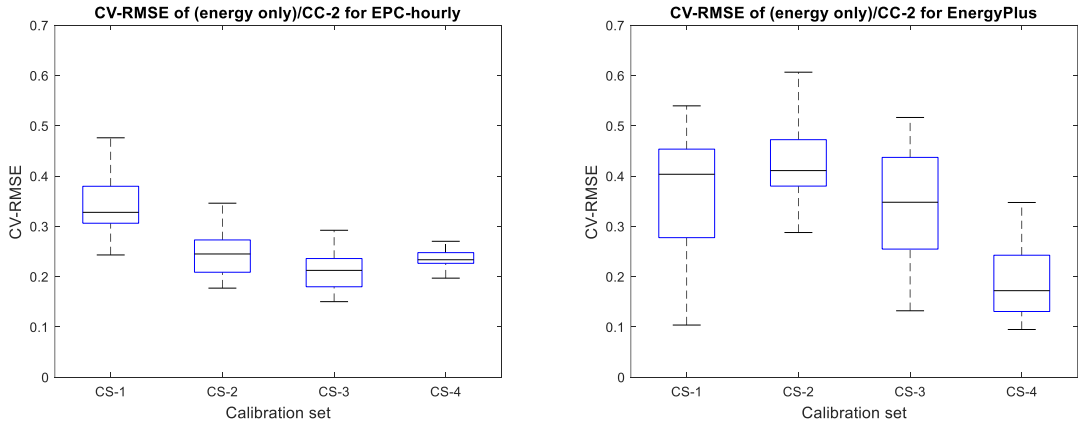


Figure 6.29 The variation of CVRMSE on energy prediction on CC-2 among the calibration sets for EPC-hourly (left) and EnergyPlus (right)

Finally, the above figures (6.25 through 6.28) show that EPC models have better agreement with the observation than EnergyPlus. This suggests that the low resolution information and data used in the EnergyPlus model and its calibration are potentially insufficient to support accurate predictions.

6.5 ELA determination based on in-situ measurement

Figure 6.30 compares the ELA estimates derived in this chapter, using both EPC and EnergyPlus models against the in-situ instrumentally derived ELA. Only the best cases for each model and best performing calibration set and calibration criterion are shown in the form of distribution fit run on the histograms of the original results.

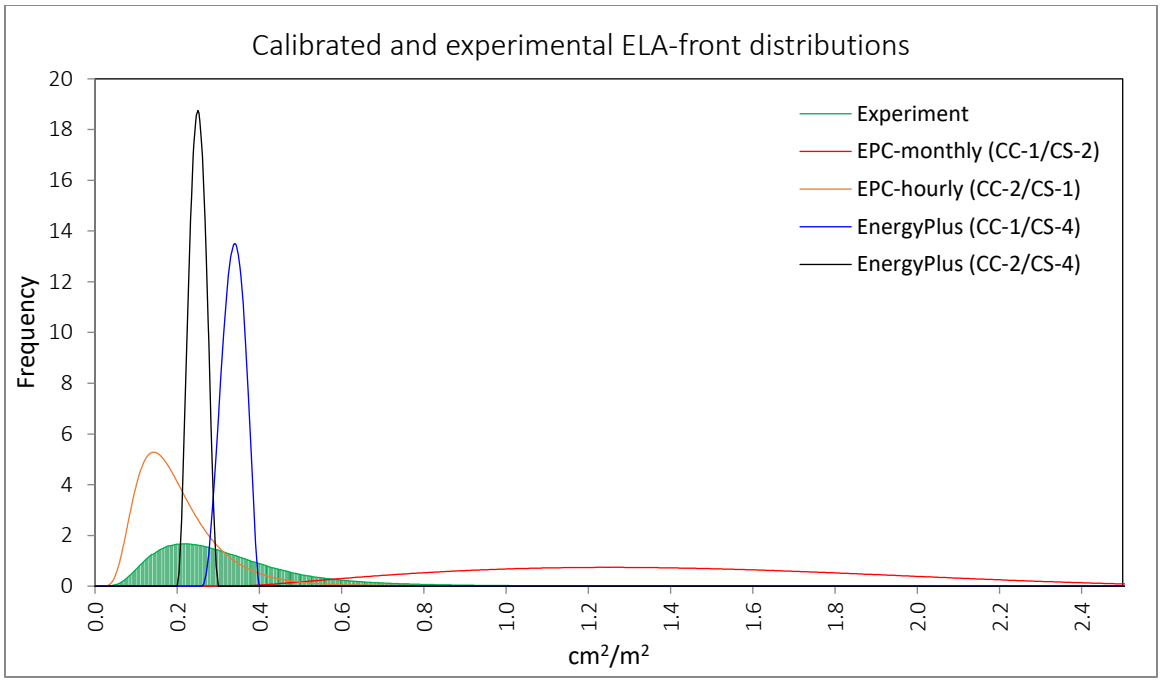


Figure 6.30 Best results of ELA calibrations with EPC and EnergyPlus in comparison with experimentally determined ELA (green)

The shaded green distribution in Figure 6.30 illustrates the direct ELA calibration, which is assumed to be our best representation for the true ELA. The direct calibration of ELA provides a reasonable estimate, although the distribution is unexpectedly wide (the 95% confidence interval (CI) is between 0.33 and 0.76). Assuming that for engineering purposes an 80% confidence interval is acceptable, it would situate the ELA engineering estimate between 0.38 and 0.66, which is much larger than most papers on experimentally derived ELA seem to indicate. This points to the fact that the reported confidence in experimentally determined ELA based on tracer gas are misplaced based on our findings. However, as indicated earlier, the uncertainty in wind pressures during the experiment can be significantly reduced if dynamic wind pressures are measured. The effect of this will be seen in case study 2 (CHAPTER 7).

In terms of ELA estimation through our calibration process, Table 6.8 provides the characterization of the ELA outcomes of each calibration set in the calibration process. As described in CHAPTER 5, we use Kullback-Leibler divergence (KL) to measure the divergence of the estimated ELA distribution from the one we established experimentally. Besides, we define the goodness of the estimated ELA by measuring its probability (P) within the 80% CI engineering range that we introduced in the previous paragraph; in this case the highest probability value is more desirable. As an additional aspect to support our decision toward the best ELA evaluation, we must take into consideration the span of the estimated ELA distribution; in this case we look at the standard error (SE), where we assume that the narrower distribution leads to a more stable model with respect to the ELA estimation. Thus, with the combination of these three criteria, we can eventually determine which ELA distribution represents the best results among all others. All best estimations are highlighted green in Table 6.8.

Table 6.8 Measurements of the distance and divergence between estimated and direct ELA

Model	**Measurements of Direct ELA at CI= 80%**											
	CS-1			CS-2			CS-3			CS-4		
	P	KL	SE	P	KL	SE	P	KL	SE	P	KL	SE
EPC-monthly/CC-1	0.05	1.99	0.08	0.03	1.89	0.07	0.03	1.93	0.07	0.02	2.02	0.07
EPC-hourly/CC-2	0.62	0.36	0.02	0.02	2.17	0.07	0.02	1.98	0.07	0	4.22	0.03
EnergyPlus/CC-1	0.02	2.22	0.07	0.02	2.02	0.07	0.50	0.41	0.04	1	0.31	0.01
EnergyPlus/CC-2	0.57	0.42	0.03	1	0.20	0.01	1	0.20	0.01	1	1.40	0.001

Accordingly, none of the EPC models resulted in a reasonable estimation that can efficiently represent the true ELA. Both EPC-monthly and EPC-hourly (red and orange distributions, respectively) are so wide that they hardly improve on the non-informative prior. However, in the case of EPC-hourly we notice that even though it is wide, the

distribution is much closer to the direct ELA estimation, which of course leads to better estimation compared to that with EPC-monthly. This is a 0.63 probability as opposed to 0.03 in the EPC-monthly. It is important to note that CS-1 is considered the best ELA distribution among the others in EPC-hourly, but this does not necessarily mean that CS-1 as an energy model is the best. In fact, the tendency of the ELA estimations to pick values from the lower bound in our assigned constraints makes this estimation more closer to the direct ELA. This is an aberration of the particular model rather than a structural benefit. Therefore, this as a whole leads to the inevitable conclusion that the ELA calibration in these models, i.e., with a data resolution of only monthly aggregated energy consumption leads to rather useless results. One should remember that the given building has precisely ONE single but unknown ELA value, and it is clear that this combination of data and model does not let us make a reasonable estimate of that value.

On the other hand, both EnergyPlus models perform much better in terms of ELA estimation combined with the fact that EnergyPlus with the energy-temperature criterion (CC-2) has a narrower distribution and is more in line with the mean value of the green distribution. Both distributions are situated within the 80% CI of the direct ELA with probabilities equal to 1. We notice that the KL value in EnergyPlus/CC-2 is higher than CC-1. This does not mean that CC-1 has led to a better distribution than CC-2, but because CC-2 gives a narrower distribution (SE = 0.001), the divergence by definition becomes larger than for CC-1 with a slightly wider distribution (SE = 0.01). Hence, the KL value in this case becomes larger. More discussion of the relation between ELA estimation and the energy results is provided in CHAPTER 8.

CHAPTER 7. CASE STUDY 2: INDIRECT ELA CALIBRATION IN A RESIDENTIAL BUILDING

7.1 Building description

The second case study is a typical residential house located in the Atlanta area (see Figure 7.1). The house has two stories with footprint area of 118 m^2 (total building area = 236 m^2). The structure of the building is mainly wood studs and plywood, and the roof is gabled. The first and second floors are separately airconditioned with two central airconditioning systems. The energy model for this case study is constructed based on personal audit and resources of typical house construction information.

Figure 7.1 Zones distribution and overview of the residential case study

In this case study we conduct the calibration process in the same fashion as the previous case study. However, for this building outcomes are collected at higher granularity, i.e., both hourly aggregate electricity consumption and hourly indoor temperature in addition to aggregated monthly energy consumption. The building is equipped with an on-site weather station to measure dry-bulb temperature, global solar radiation, local wind speed, and wind direction. The period for these measurements and associated calibration is three months, starting from August to the end of October 2019.

Accordingly, this building has more calibration cases than the previous case study, and the following sections will follow this structure:

- **Section 7.2:** calibration criterion 1 (CC-1) with EPC-monthly and QoI-1-M (monthly energy consumption).
- **Section 7.3:** calibration criterion 1 (CC-1) with EPC-hourly and QoI-1-H (hourly energy consumption).
- **Section 7.3.2:** combined EPC calibration with the addition of QoI-2 (hourly indoor temperature); (QoI-1-H + QoI-2). This is calibration criterion 2 (CC-2) with EPC-hourly.
- **Section 7.4:** calibration criterion 1 (CC-1) with EnergyPlus-(monthly data) and QoI-1-M.
- **Section 7.5:** calibration criterion 1 (CC-1) with EnergyPlus-(hourly data) and QoI-1-H.
- **Section 7.5.2:** combined EnergyPlus calibration with the addition of QoI-2; (QoI-1-H + QoI-2). This is calibration criterion 2 (CC-2) with EnergyPlus-(hourly data).
- **Section 7.6:** evaluation and results analysis of energy models.
- **Section 7.7:** comparisons of the ELA estimates and the experimentally measured ELA distributions.

7.2 Calibration of the low fidelity and low-resolution model (EPC-monthly) with monthly consumption data

The process starts with estimating the uncertainties in model parameters. Base values are assigned to model parameters based on building audits, design documents, operation manuals, and ASHRAE standards to make the model as close to existing building conditions as possible while respecting the fact that many parameters are only known approximately. Based on collective expert knowledge from the literature study, uncertainties are quantified in relevant model parameters, starting from Table 5.1. The SA process provides the parameter screening that helps to determine the dominant uncertain parameters for every step in our calibration process.

7.2.1 Sensitivity analysis

The Morris method provides the elementary effects of every individual parameter. At this stage of the process, monthly energy consumption is chosen as the primary QoI because it is monitored and used as the data in the ensuing calibration step. For every energy model (with varying order and fidelity) a fixed number of consecutive calibration steps is performed, where each step is preceded by a SA step. The procedure is exactly the same as previously explained in case study 1.

A 20-level design of 23 parameters with a sample size of 480 is applied to the building energy models. Table 7.1 provides the list of the uncertain parameters in the EPC model and their uncertainty ranges, while Figure 7.2 ranks the role of each parameter in the resulting energy consumption distribution based on a change in variable means. This is

visualized through the customary relative ranking graph, which shows the relative importance of variables on the outcome.

Table 7.1 Parameter uncertainty in model testing and acronym definition

Acronym	Definition	Uncertainty range	Comment
ELA_Back	ELA of the NW side (cm^2/m^2)	LogNorm(1.28,0.88)	
ELA_Front	ELA of the SE side (cm^2/m^2)	LogNorm(1.28,0.88)	
ELA_Right	ELA of the SW side (cm^2/m^2)	LogNorm(1.28,0.88)	
Cp	Wind pressure coefficient	RelativeNorm(1, 0.25)	The uncertainty is relative to every Cp value
Wind Speed	Local wind speed (m/s)	RelativeNorm(1, 0.42)	The uncertainty is relative to every hour wind speed value
Clg COP	Cooling COP	Norm(3, 0.15)	
Htg COP	Heating COP	Norm(1.5, 0.08)	
fan power	HVAC fan power (W/l/s)	Norm(1.5, 0.38)	
Clg SP wd	Cooling setpoint in the weekdays (ºC)	Tring(22, 25, 27)	
Htg SP wd	Heating setpoint in the weekdays (ºC)	Tring(17, 20, 23)	
Roof U	Roof U-value ($W/m^2 \cdot K$)	Norm(0.25, 0.02)	
Wall U	Wall U-value ($W/m^2 \cdot K$)	Norm(0.30, 0.02)	
Window U	Window U-value ($W/m^2 \cdot K$)	Norm(3, 0.6)	
Solar Trans	Window's solar transmittance	Norm(0.75, 0.05)	
Occupancy	Occupancy internal gain (m^2/person)	Uniform(80, 120)	
Lighting	Lighting internal gain (W/m^2)	Norm(8, 1.6)	
Appliances	Plug loads internal gain (W/m^2)	Norm(12, 2.4)	
Sch Occ wd	Occupants schedule in the weekdays	Tring(0.7, 0.9, 1)	
Sch Occ we	Occupants schedule in the weekends	Tring(.90, 0.95, 1)	
Sch Lgt wd	Lighting schedule in the weekdays	Tring(0.70, 0.80, .85)	
Sch Lgt we	Lighting schedule in the weekends	Tring(0. 70, 0.80, .85)	
Sch App wd	Appliances schedule in the weekdays	Tring(0.70, 0.80, .85)	
Sch App we	Appliances schedule in the weekends	Tring(0. 70, 0.80, .85)	

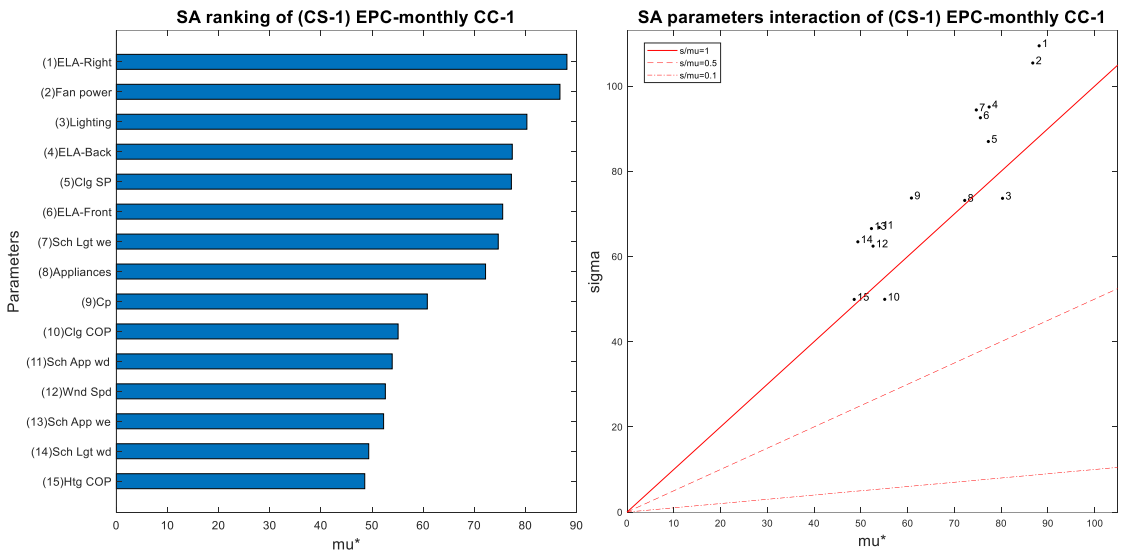


Figure 7.2 SA results of dominant parameters for CC-1, with calibration set CS-1

The first step in the calibration process is conducting a parameter screening of all input variables in the parameter pool, which includes the ELA parameters as well. As described earlier, this initial step is important to understand the characterization and influence of these parameters as a whole on the energy model predictions. Therefore, from Figure 7.2 we notice that ELA parameters (No. 1, 4, and 6 in the graphs) have a high level of importance in the SA rank, and also indicate a nonlinear effect or interaction with other variables because of their location in the fourth region in the interaction graph (i.e., $\sigma/\mu^* > 1$). This can be primarily attributed to the fact that the initial assessment of uncertainty in the ELA parameter is large. Refinement of the uncertainty ranges in parameters (as more information is used in their uncertainty estimates) may lead to significant changes in the ranking order. This however does not change our procedure as the primary objective is the calibration of ELA, which is therefore always calibration parameters in CS-1 through CS-4, regardless of the SA outcome of the initial step. The remainder of SA results as well as the calibration process are presented in the next section.

7.2.2 Calibration process

After finding the most dominant parameters by conducting the SA, we select the top ones that show substantial impact on the QoI and are therefore expected to have a large influence on the results of the calibration. As we move on with the same procedure, parameters are selected to be added to the prior calibration set leading to a larger CS and a likewise smaller uncertain parameter set. Table 7.2 shows all calibration sets that have been selected in the four steps.

Table 7.2 Calibration parameters sets of EPC-monthly/CC-1 used in the calibration steps

CS-1	**CS-2**	**CS-3**	**CS-4**
Front ELA	Front ELA	Front ELA	Front ELA
Back ELA	Back ELA	Back ELA	Back ELA
Right ELA	Right ELA	Right ELA	Right ELA
	Lighting	Lighting	Lighting
	Lighting schedules	Lighting schedules	Lighting schedules
	Appliances	Appliances	Appliances
		Solar transmittance	Solar transmittance
		Fan power	Fan power
		Cooling setpoints	Cooling setpoints
		Wall U-value	Wall U-value
			Cooling COP
			Cp
			Roof U-Value
			Wind Speed

This is very similar to the approach explained in case study 1. The façade configuration of the building leads to distinguishing three dominant parts for which a different ELA should be assigned, for simplicity denoted here by front, back, and right ELA (see Figure 7.1). The resultant of CS-1 through CS-4 for “back ELA” is shown in

Figure 7.3. We chose to discuss this particular façade because it gives the most illustrative results about how the process moves from one step to the next.

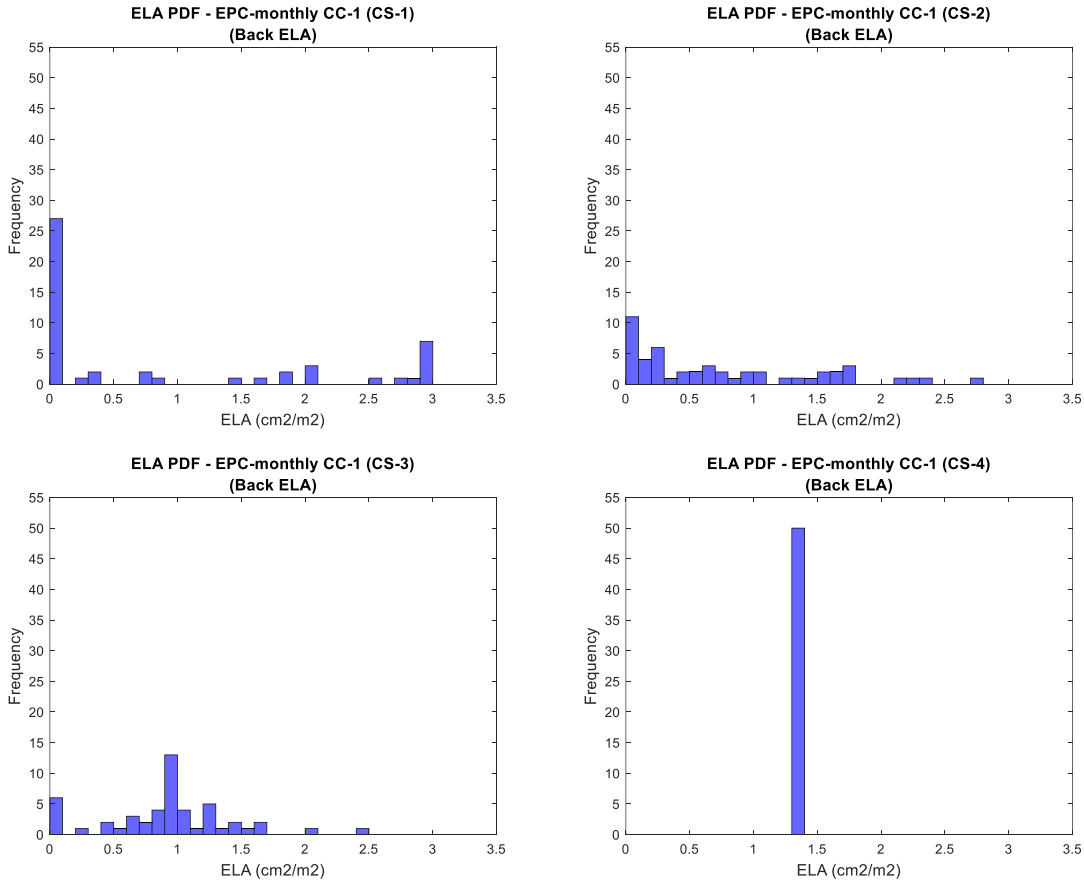


Figure 7.3 Calibration results of ELA (CS-1 through CS-4) for EPC-monthly/CC-1

Similar to the conclusion from the first case study, it turns out that the ELA distribution in CS-1 is very wide. This is not unexpected as the EPC model is low resolution, the measured data is low resolution, and no other parameters are being calibrated except the ELA. The next step is to run another SA, which results in a new ranking as shown in Figure 7.4. Then, the same calibration is conducted with the single discrepancy (CC-1) in the objective. From Figure 7.3 we notice that for CS-2 the ELA distribution is still as wide as with CS-1, which means that the additional parameters in this calibration set have minimal impact on the prediction of ELA. This can be seen from the

SA results in Figure 7.4 where the highest μ* values are found for the lighting, lighting schedules, and appliances (No. 1, 2, and 3, respectively). Another group of parameters with slightly lower influence includes solar transmittance, fan power, and cooling setpoints (No. 4, 5, and 6, respectively), while the remaining parameters are identified as negligible parameters. For this step all parameters exhibit a σ/μ* ratio >1, except for the energy consumption by lighting and appliances, which suggests that most parameters exhibit either non-linear behavior or interaction effects with other parameters or both.

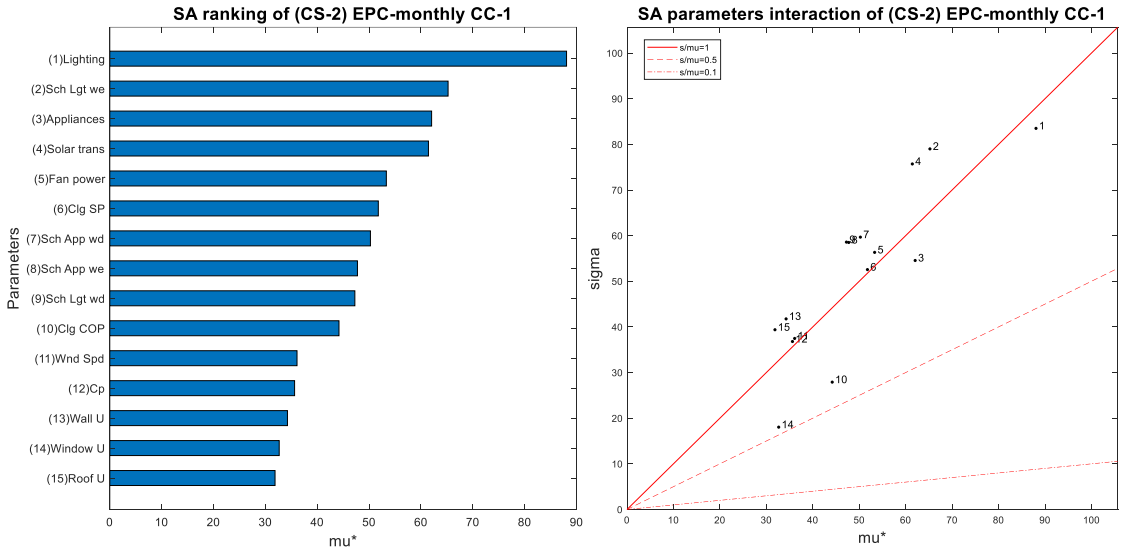


Figure 7.4 SA results of dominant parameters for CC-1, with calibration set CS-2

On the other hand, one can observe that ELA in CS-3 forms a clear peak around 0.9 cm^2/m^2 even though it is still a rather wide distribution ranging between 0.1 cm^2/m^2 to around 2 cm^2/m^2. This suggests that the new parameters have considerable interaction with ELA. Considering their impact on the model outcome, the most important factors on this step are fan power, cooling setpoints, and solar transmittance. These are followed by wall U-value, cooling COP, and by wind speed. It also can be seen from Figure 7.5 that the first

three important parameters have a ratio of $\sigma_i/\mu_i^* > 1$, which indicates a non-linear or interaction effect with other parameters within the model.

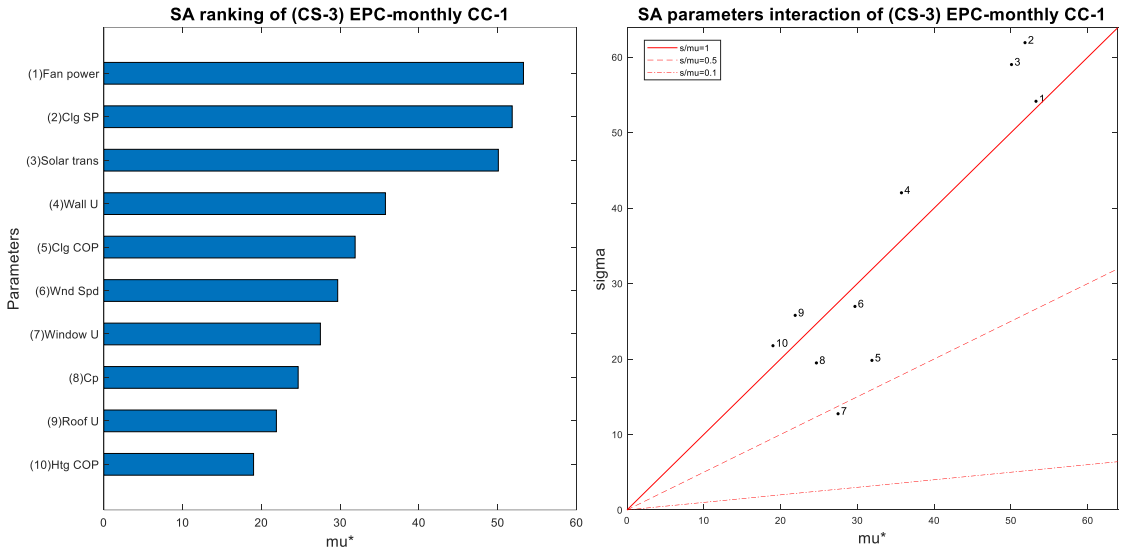


Figure 7.5 SA results of dominant parameters for CC-1, with calibration set CS-3

Finally, CS-4 distribution is more determined with a higher probability of being at 1.35 cm^2/m^2. This indicates that the more information we add to the low resolution model may lead to a modest improvement in the resulting ELA, assuming all uncertainties in non-calibration parameters to be identical. This can be explained in Figure 7.6; among all of the input parameters, Cp (No. 1) has the highest impact on the energy consumption (QoI-1), whereas the other parameters have minimum impact on the model output. Therefore, their influence on energy consumption variations can be neglected. Moreover, it can be observed that the corresponding σ value of Cp is almost monotonic, indicating that Cp mostly does not has interaction with the other parameters, which helps in constructing a more constrained ELA prediction. It should be noted that our confidence in the end result is low because we expect that the large MFU in the EPC-monthly is subsumed in the

resulting calibration parameter values. This is further elaborated in the overall assessment in the next chapter.

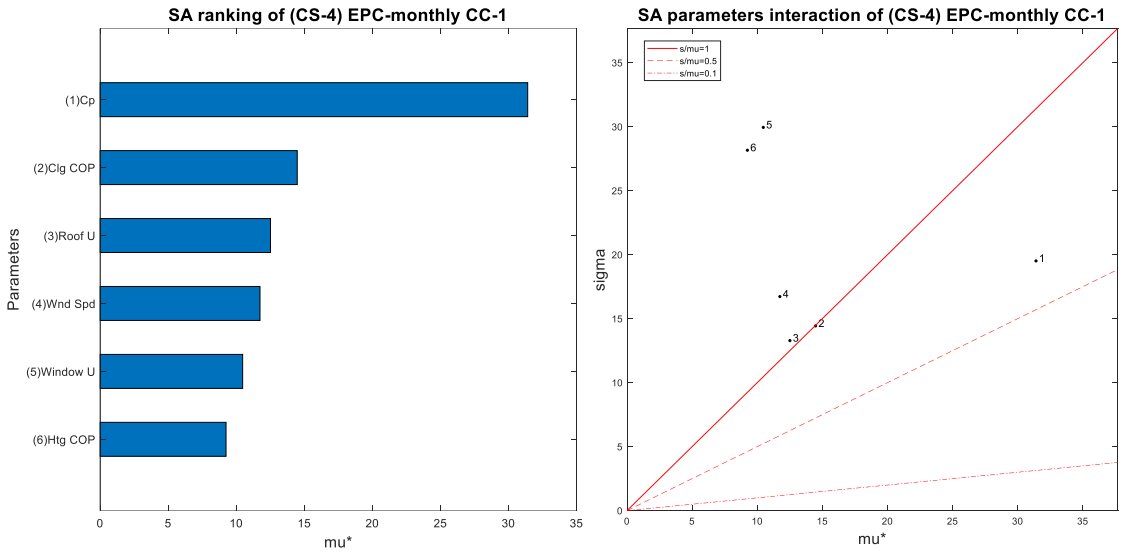


Figure 7.6 SA results of dominant parameters for CC-1, with calibration set CS-4

7.3 Calibration of the low-fidelity model (EPC-hourly) with hourly consumption data

It is acknowledged that with higher resolution data, the calibration is expected to perform better in terms of parameters estimation. Hence, the outcomes of this case study are extended to hourly electricity (QoI-1-H) for the period between August 1st to end-October. In this EPC-hourly model we use the same parameter inputs that are used in the EPC-monthly model, and subject to the calibration procedure in the same fashion as before, keeping in mind that this case uses calibration criterion 1 (CC-1) with hourly energy consumption (QoI-1-H) instead of monthly aggregates. Next, we present SA for parameter screening to determine dominant uncertain parameters for every step in our calibration process.

7.3.1 Calibration process

Table 7.3 shows all parameter sets that have been selected for the EPC-hourly model.

Table 7.3 Calibration parameters sets of EPC-hourly/CC-1 used in the calibration steps

CS-1	CS-2	CS-3	CS-4
Front ELA	Front ELA	Front ELA	Front ELA
Back ELA	Back ELA	Back ELA	Back ELA
Right ELA	Right ELA	Right ELA	Right ELA
	Lighting	Lighting	Lighting
	Cooling setpoints	Cooling setpoints	Cooling setpoints
	Appliances	Appliances	Appliances
	Lighting schedules	Lighting schedules	Lighting schedules
		Solar transmittance	Solar transmittance
		Appliances schedules	Appliances schedules
		Fan Power	Fan Power
		Cp	Cp
			Cooling COP
			Wall U-value
			Wind speed
			Roof U-Value

After running the calibration with the same procedure we followed earlier, the results of all calibrations for “back ELA” are shown in Figure 7.7.

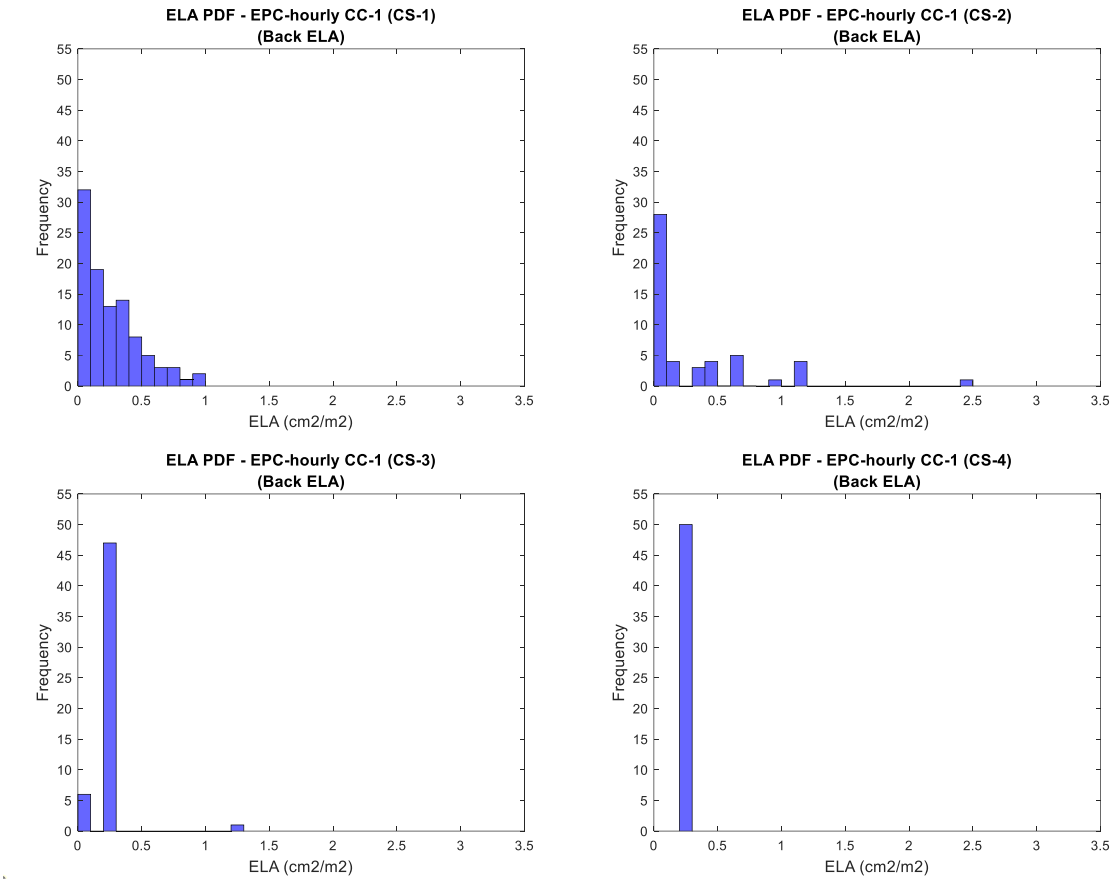


Figure 7.7 Calibration results of ELA (CS-1 through CS-4) for EPC-hourly/CC-1

This shows that although the ELA distributions in CS-1 and CS-2 are rather wide, it reveals a significant improvement compared to the first calibration set of the EPC-monthly calibration. This can be explained from the SA results in Figure 7.8 where we notice that besides its considerable importance, back ELA (No. 4) indicates an almost linear effect with other parameters on the energy consumption with a ratio of $0.5 < \sigma/\mu^* < 1$. A similar observation is found in the second SA results as shown in Figure 7.9, where lighting, cooling setpoints, and appliances are the most dominant variables, but each parameter has a different σ/μ^* ratio from low to high effect with the other parameters. This supports the conclusion that higher granularity in the observed outcomes leads to better parameter prediction, but this cannot be generalized without more cases.

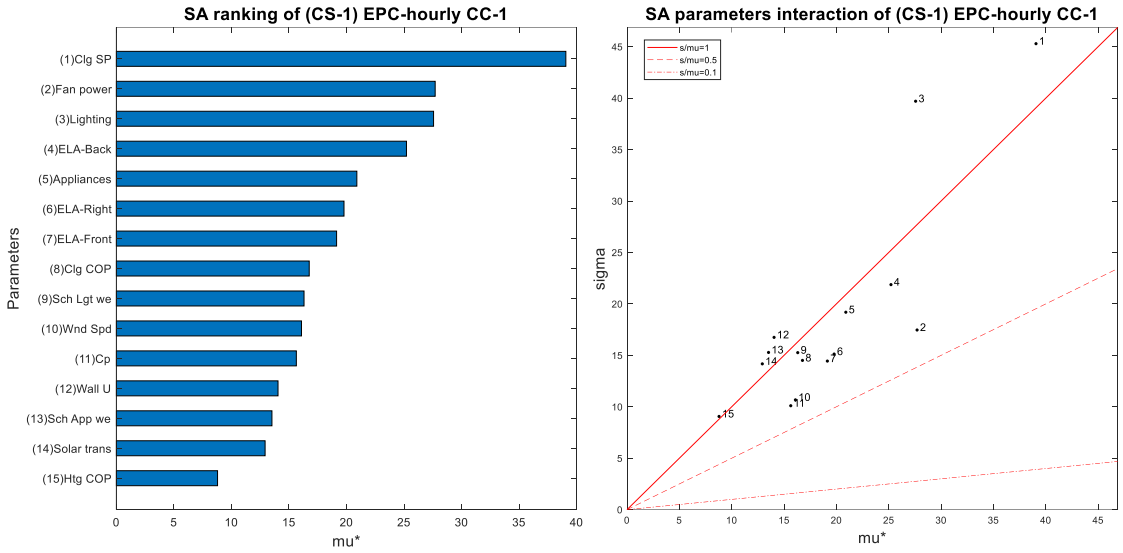


Figure 7.8 SA results of dominant parameters for CC-1, with calibration set CS-1

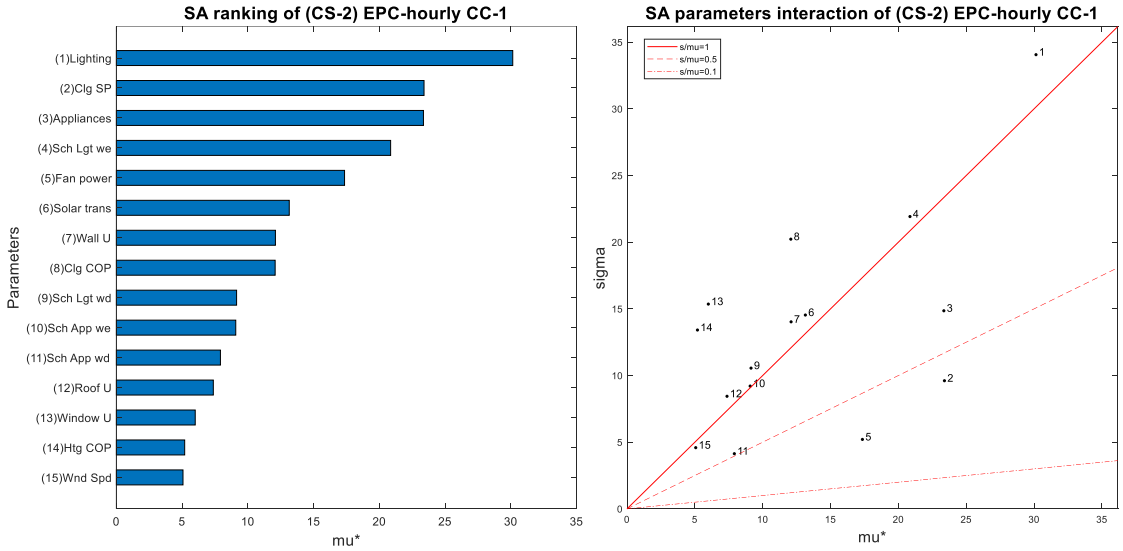


Figure 7.9 SA results of dominant parameters for CC-1, with calibration set CS-2

The combination of the parameters in CS-3 seems to have a stronger influence on the variability of ELA as shown in the interaction part of the SA in Figure 7.10. This is shown in the plot of (μ_i*, σ_i) in large values of mean and standard deviation of solar transmittance, appliances schedules, and fan power. Cp on the other hand shows to be a

very important parameter with high μ* value, but its overall σ/μ* ratio is between 0.1 and 0.5, which indicates an almost linear effect with the model outcome.

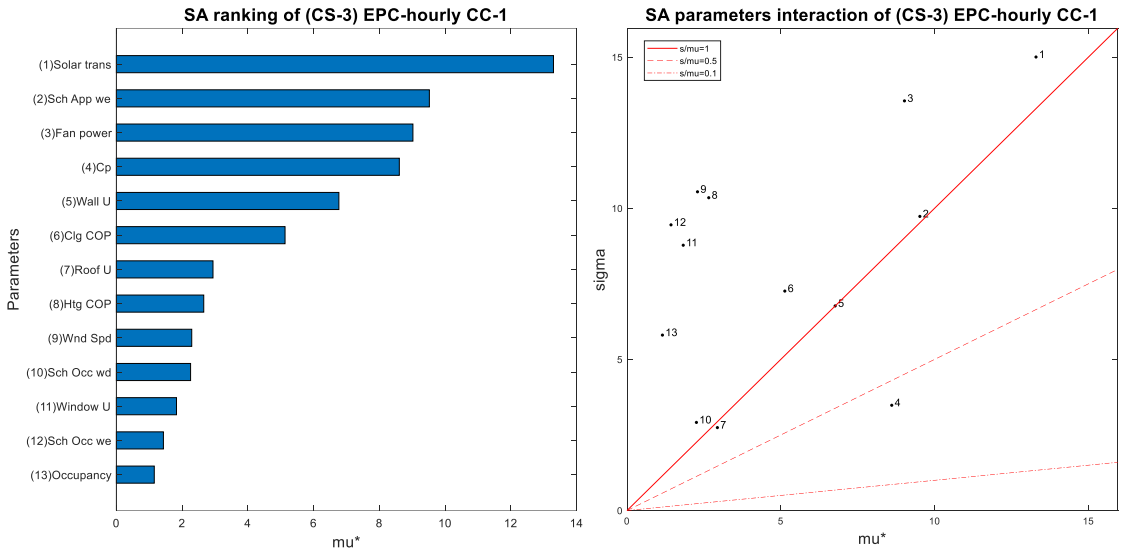


Figure 7.10 SA results of dominant parameters for CC-1, with calibration set CS-3

Therefore, as CS-3 achieves a highly constrained ELA distribution, CS-4 as a consequence becomes as determined and very similar distribution as with CS-3. This also can be observed by the low μ_i* values of the parameters, which indicates that the remaining parameters may have minimal improvement than CS-3. The SA outcome of CS-4 is shown in Figure 7.11.

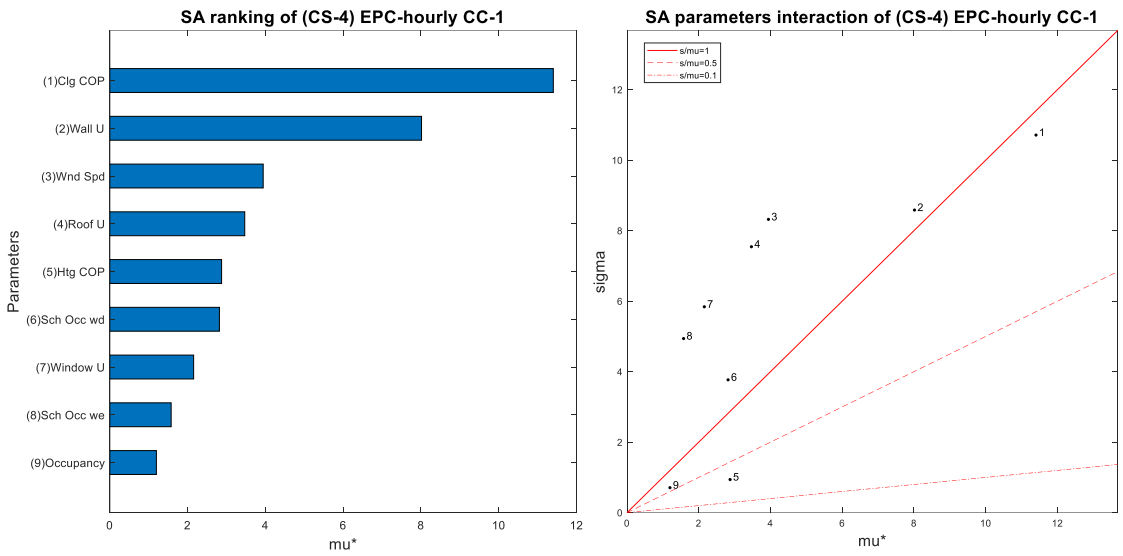


Figure 7.11 SA results of dominant parameters for CC-1, with calibration set CS-4

7.3.2 Adding temperature measurements in the calibration

Similar to the first case study, the calibration process is repeated with higher resolution outcomes to test the potential improvement if additional outcomes are added to the pool. The indoor temperature in zone 4 is measured for three months (August till the end of October). As stated earlier, the indoor temperature is an additional QoI that can be predicted with the energy model simulation, and thus used in the calibration target. In this case we perform the calibration for the aggregated discrepancy in a similar fashion to EPC-hourly using weighting factors of W_E=1 and W_T=2.8 for both outcomes from Eq. (5.4). This way a single objective is constructed to be minimized, and essentially the same calibration process from the previous section can be used. The SA for the combined calibration process results in new sets of calibration parameters as presented in Table 7.4.

Table 7.4 Calibration parameters sets of EPC-hourly/CC-2 used in the calibration steps

CS-1	CS-2	CS-3	CS-4
Front ELA	Front ELA	Front ELA	Front ELA
Back ELA	Back ELA	Back ELA	Back ELA
Right ELA	Right ELA	Right ELA	Right ELA
	Lighting	Lighting	Lighting
	Cooling setpoints	Cooling setpoints	Cooling setpoints
	Lighting schedule	Lighting schedule	Lighting schedule
	Appliances	Appliances	Appliances
		Fan power	Fan power
		Cp	Cp
		Wall U-value	Wall U-value
			Solar transmittance
			Wind speed
			Cooling COP
			Heating setpoint
			Window U-value

The SA process in this section is similar to the previous process in CC-1. At the beginning we conduct a parameter screening of all uncertain variables to understand the influence of these parameters as a whole on the energy model. Therefore, from Figure 7.12 we notice that ELA-Back (No. 3) has a high level of importance on the SA rank, and also indicates a nonlinear effect or interaction with other variables because according to its location in the interaction graph with a ratio of $\sigma/\mu^* > 1$. Hence, this suggests the implementation of the current weighting factors has little influence on the dominance of ELA on the combined calibration criterion. It can be expected that overweighting the temperature discrepancy (by choosing a larger W_T) may change that. This is not inspected further here.

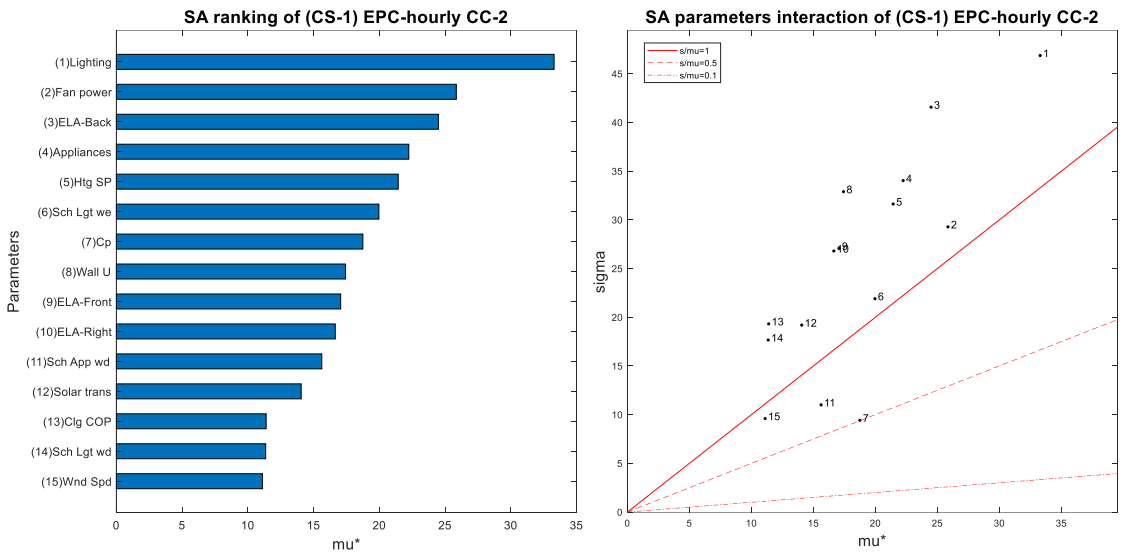


Figure 7.12 SA results of dominant parameters for CC-2, with calibration set CS-1

The sample pool in Table 7.1 is used in the SA process, and the ELA result established from the first step is used as input in the SA that drives the selection of parameters in the next calibration step; the SA result of CS-2 is shown in Figure 7.13.

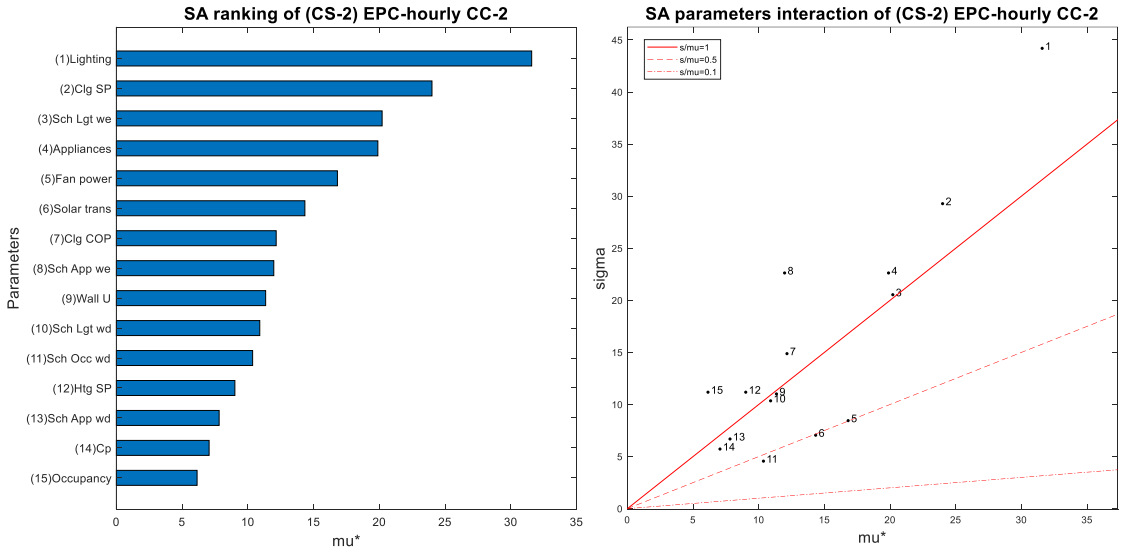


Figure 7.13 SA results of dominant parameters for CC-2, with calibration set CS-2

The calibration process is run with the same calibration sets that are identified in Table 7.4, with the target to minimize the weighted discrepancy in both measured energy

consumption and indoor temperature. Figure 7.14 shows overlaid distributions that compare all back-ELA results between EPC-hourly CC-1 and CC-2. It should be recalled at this point that CC-2 is the combined calibration with QoI-1-H (hourly energy consumption) and QoI-2 (hourly indoor temperature).

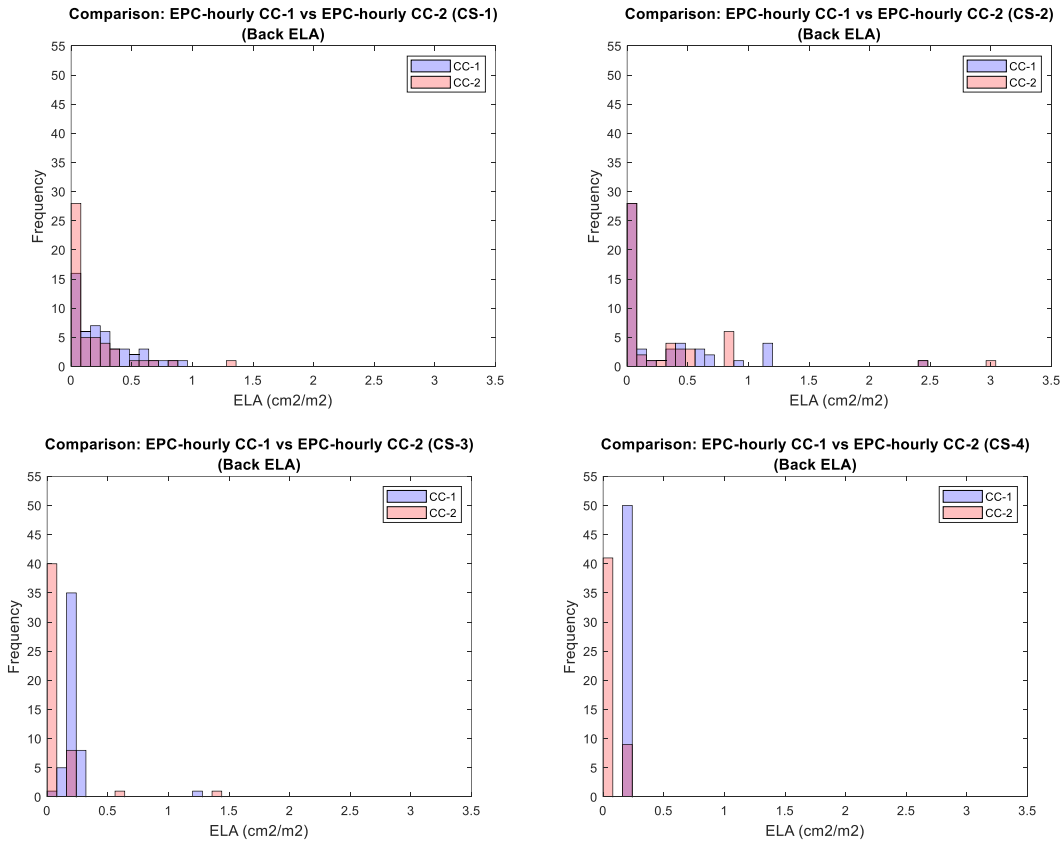


Figure 7.14 Calibration results of ELA (CS-1 through CS-4) for CC-1 and CC-2

The ELA results of CC-2 are not very much different from the results established for CC-1, especially with CS-1 and CS-2. However, it can be noticed that there is a persistent tendency in every calibration result that forces ELA values to the lower edge of our calibration constraint which is set at 0.05 cm^2/m^2. Therefore, the prudent conclusion at this point is that given the lack of precise estimates of the non-calibration parameters, the increased fidelity of the simulation model leads to only modest improvements.

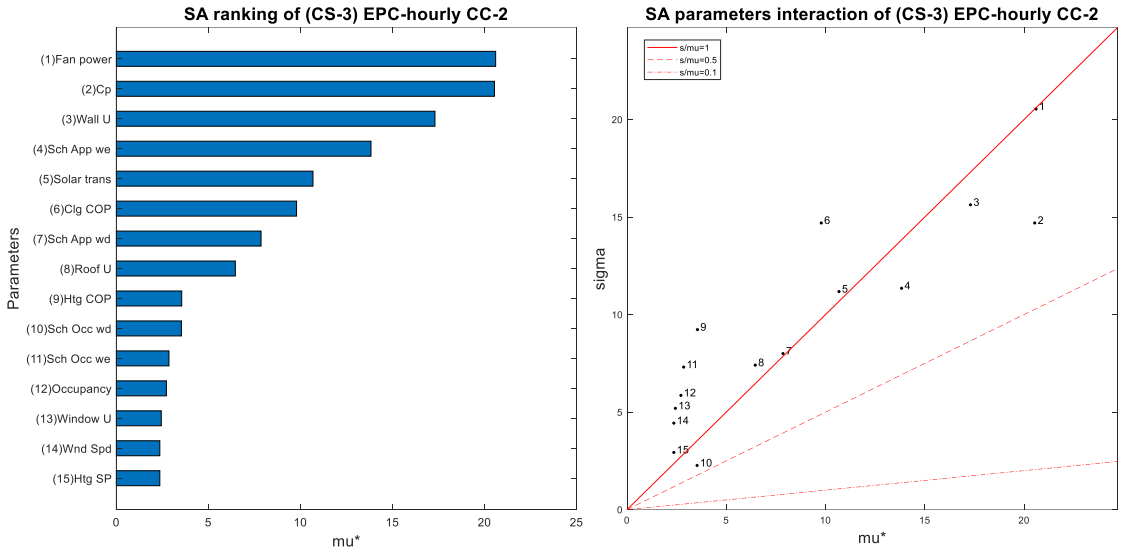


Figure 7.15 SA results of dominant parameters for CC-2, with calibration set CS-3

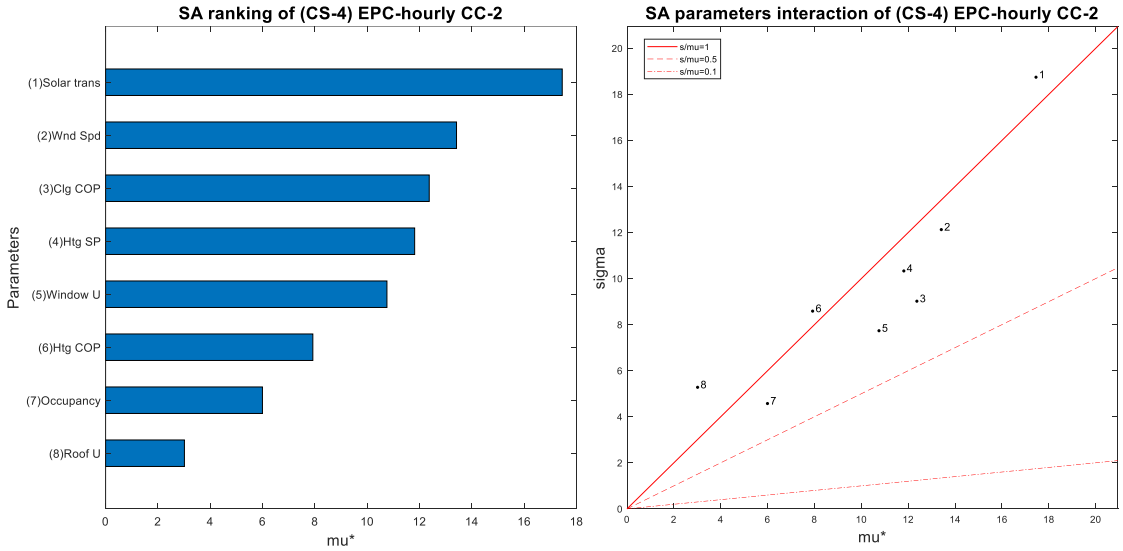


Figure 7.16 SA results of dominant parameters for CC-2, with calibration set CS-4

7.4 Calibration of EnergyPlus model with monthly consumption data

The selected higher fidelity model in this study is EnergyPlus, with base values and uncertainties assigned to model parameters through the same method we established for EPC, but this time with more details as required by the tool. Table 5.1 summarizes the uncertainties around the initial values assigned to model parameters. Cp is based on the

Swami&Chandra model with the uncertainty quantified in CHAPTER 4. SA is again applied for parameter screening to determine the dominant uncertain parameters for every step in the calibration process.

7.4.1 Sensitivity analysis

With the Morris method, A 20-level design of 48 independent samples are generated to obtain the elementary effects of individual parameters. 980 simulations are run to support the appropriate selection of the calibration parameters among the total 48 uncertain parameters as listed in Table 7.5.

Table 7.5 Parameter uncertainty in model testing and acronym definition

Acronym	Definition	Uncertainty range	Comment
ELA_Back	ELA of the NW side (cm^2/m^2)	LogNorm(1.28,0.88)	
ELA_Front	ELA of the SE side (cm^2/m^2)	LogNorm(1.28,0.88)	
ELA_Right	ELA of the SW side (cm^2/m^2)	LogNorm(1.28,0.88)	
Cp_ (21 points)	Wind pressure coefficient	RelativeNorm(1, 0.25)	The uncertainty is relative to every Cp value
Wnd_Spd	Local wind speed (m/s)	RelativeNorm(1, 0.42)	The uncertainty is relative to every hour wind speed value
Clg_COP	Cooling COP	Norm(3, 0.15)	
Htg_Coil_Eff	Heating COP	Norm(1. 5, 0.08)	
Fan_Eff	HVAC fan efficiency	Tring(0.2, 0.6, 0.9)	
Clg_sp_WD	Cooling setpoint in the weekdays (ºC)	Tring(22, 25, 27)	
Htg_sp_WD	Heating setpoint in the weekends (ºC)	Tring(17, 20, 23)	
Clg_sp_WE	Cooling setpoint in the weekdays (ºC)	Tring(22, 25, 27)	
Htg_sp_WE	Heating setpoint in the weekends (ºC)	Tring(17, 20, 23)	
Roof_cond	Conductivity of the roof (W/m·K)	Norm(0.13, 0.01)	
Wall_cond	Conductivity of the wall (W/m·K)	Norm(0.18, 0.01)	
Wndw_cond	Conductivity of the window (W/m·K)	Norm(0.95, 0.2)	
Wndw_Trans	Window's solar transmittance	Norm(0.75, 0.05)	

Table 7.5 (continued)

Occ_1, 2, 3, 4	Occupancy internal gain (m^2/person)	Uniform(80, 120)
Lights_1, 2, 3, 4	Lighting internal gain (W/m^2)	Norm(8, 1.6)
EE_1, 2, 3, 4	Plug loads internal gain (W/m^2)	Norm(12, 2.4)

The results from 48 independent evaluations of Morris method, each with 20 trajectories, are summarized in Figure 7.17. The figure shows the ranking of uncertain parameters in the EnergyPlus model for the first calibration parameter set (CS-1). The parameters are ordered by their relative importance in the selected QoI-1 (aggregated monthly energy consumption), corresponding to the availability of measured data. Each point represents the mean of the 20 absolute mean values μ* (x-axis) and standard deviations σ (y-axis) of the elementary effects for each parameter.

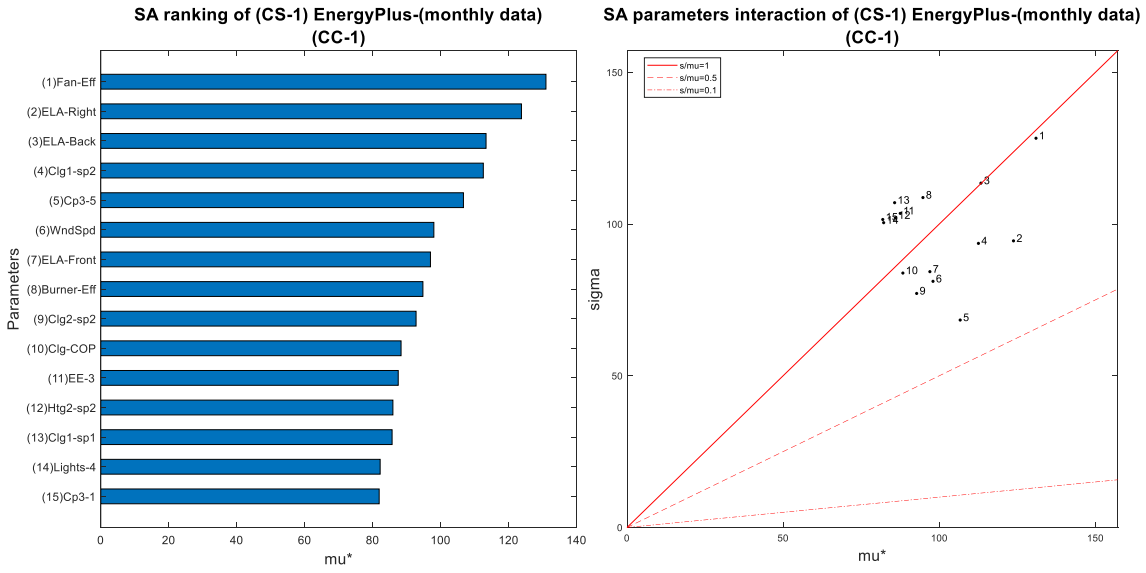


Figure 7.17 SA results of dominant parameters for CC-1, with calibration set CS-1

We notice from Figure 7.17 that ELA has a significant impact on the energy model. However, as we focus on the Back-ELA (No. 3), we notice its location is on the border between the third and the fourth regions in the Morris wedge plot. This indicates that back-ELA can be nonlinear or has an almost monotonic effect on the model. If the latter was

true, then we should expect a more rational estimation of ELA values. This is discussed in the next section.

7.4.2 Calibration process

Table 7.6 shows the four calibration sets that have been selected to conduct the EnergyPlus calibration based on the parameters screening.

Table 7.6 Calibration parameters sets of EnergyPlus-(monthly data)/CC-1 used in the calibration steps

CS-1	CS-2	CS-3	CS-4
Front ELA	Front ELA	Front ELA	Front ELA
Back ELA	Back ELA	Back ELA	Back ELA
Right ELA	Right ELA	Right ELA	Right ELA
	Fan efficiency	Fan efficiency	Fan efficiency
	Cooling setpoints	Cooling setpoints	Cooling setpoints
	Cp	Cp	Cp
		Lights	Lights
		Cooling COP	Cooling COP
		Window transmittance	Window transmittance
		Wind speed	Wind speed
			Wall conductivity
			Window conductivity

In the first calibration task we include all ELAs as lone parameters to construct an idea of what would be the estimation of ELA without being influenced by concurrent calibration of other parameters. The results of 50 samples for back ELA are shown in Figure 7.18.

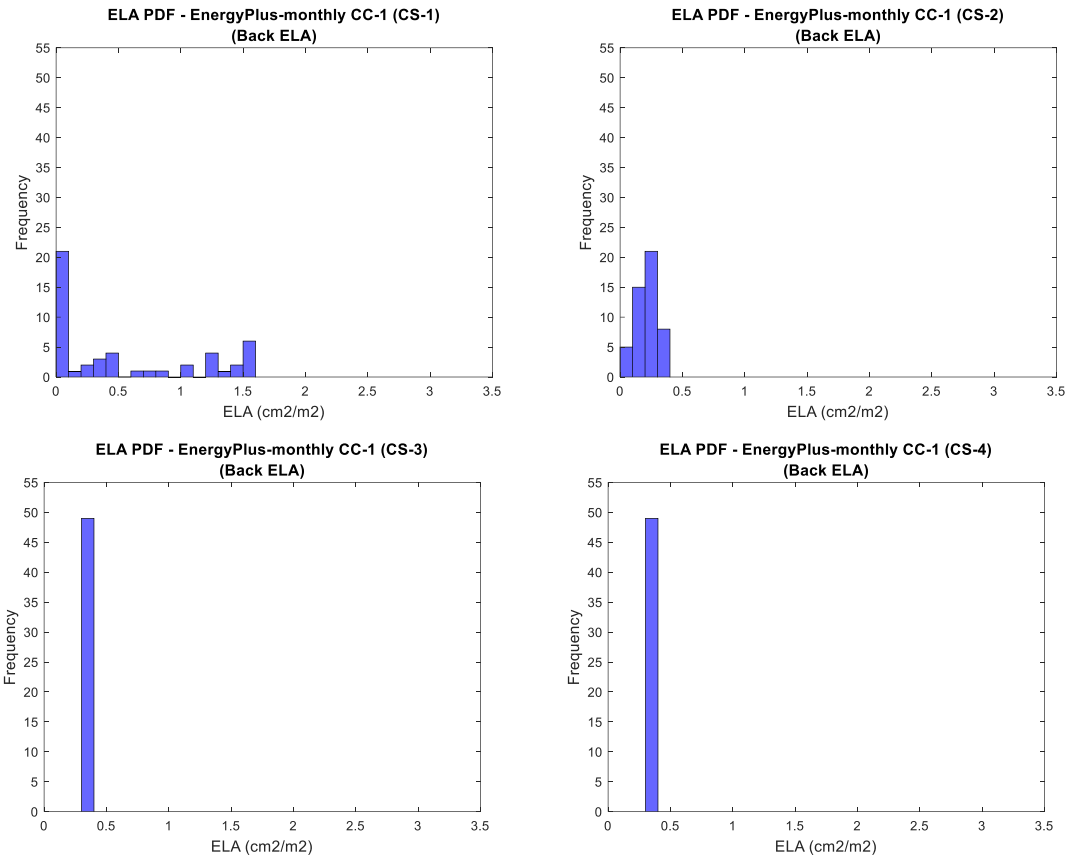


Figure 7.18 Calibration results of ELA (CS-1 through CS-4) for EnergyPlus-(monthly data)/CC-1

These results are interestingly different from those we established with the EPC models. Use of CS-1 leads to a wide distribution of ELA and is not very informative. This is because no other parameters are being calibrated except the ELA. Also, from the SA of CS-1 (Figure 7.17) we see that although the ELA parameter has a dominant impact in terms of output means μ*, its variance is also as high, which indicates a nonadditive relationship with other parameters.

However, the ELA distribution has a narrow shape with CS-2, i.e., within a range between 0.05 and 0.3 cm^2/m^2. This calibration set contains parameters that are related to the HVAC system, plug loads, as well as Cp (see Figure 7.19). Among these parameters,

fan efficiency and Cp have the highest variance (i.e., non-linearity with other parameters) compared to the other selected parameters, which suggests that the combination of parameters in CS-2 has a pronounced impact on the model output, which is one of the conditions for informative outcomes of calibration.

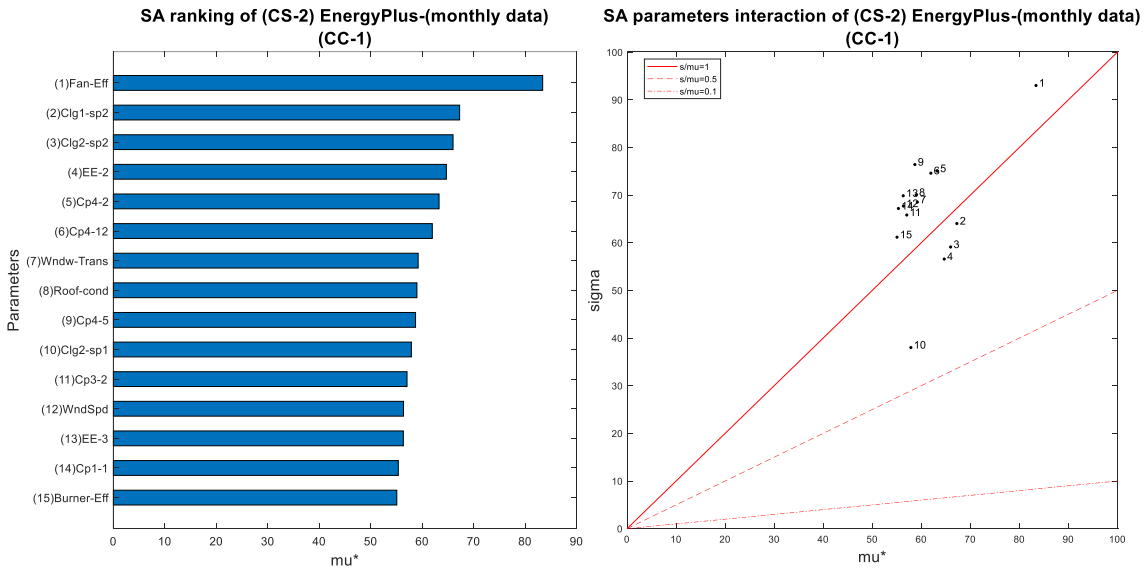


Figure 7.19 SA results of dominant parameters for CC-1, with calibration set CS-2

Subsequently, ELA in the calibration with CS-3 results in a very determined distribution at 0.3 cm^2/m^2 because, at this point, the model has all the influential parameters that affect the value of the ELA, as shown in Figure 7.20. Although all parameters in CS-3 have a nonlinear or interaction effect (or both) with other input variables, the means of their elementary effects are far less to have a distinct impact on the ELA estimation. Thus, the calibration of CS-3 produces one ELA value, which is also the same value that CS-4 produces.

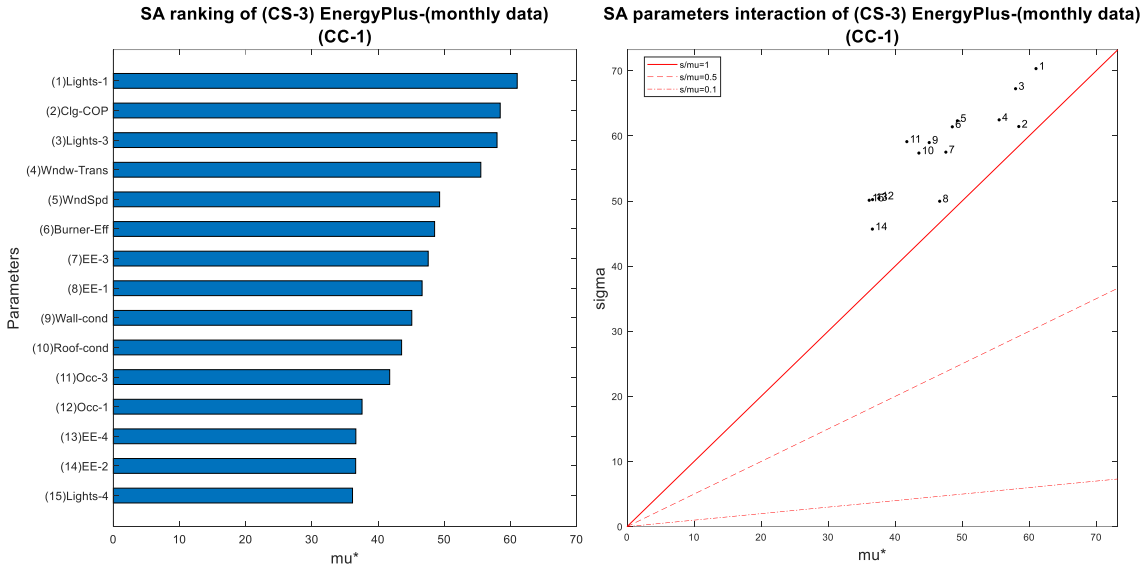


Figure 7.20 SA results of dominant parameters for CC-1, with calibration set CS-3

7.5 Calibration of EnergyPlus models with hourly outcomes

The EnergyPlus-(hourly outcomes) model is identical to the one we built for the calibration with monthly outcomes, so the same procedure as explained earlier is followed. Table 7.5 summarizes the uncertainties around the initial values assigned to the model parameters.

7.5.1 Calibration process

Table 7.7 shows the four parameters sets used in the EnergyPlus-(hourly data) model.

Table 7.7 Calibration parameters sets of EnergyPlus-(hourly data)/CC-1 used in the calibration steps

CS-1	CS-2	CS-3	CS-4
Front ELA	Front ELA	Front ELA	Front ELA
Back ELA	Back ELA	Back ELA	Back ELA
Right ELA	Right ELA	Right ELA	Right ELA
	Cooling setpoints	Cooling setpoints	Cooling setpoints
	Wind speed	Wind speed	Wind speed

Table 7.7 (continued)

Cp	Cp	Cp
Cooling COP	Cooling COP	Cooling COP
	Window conductivity	Window conductivity
	Roof conductivity	Roof conductivity
	Burner efficiency	Burner efficiency
	Electric equipment	Electric equipment
		Fan efficiency
		Occupancy
		Wall conductivity

Figure 7.21 shows the ranking of the top 15 uncertain parameters in the EnergyPlus-(hourly data) model for the first calibration parameter set (CS-1). The parameters are ordered by their relative importance in the selected QoI, which is hourly energy consumption, corresponding to the resolution of measured energy consumption. The result shows that μ^* and σ values of the sampled points are different, and there are certain fluctuations in their trend of variations. However, by focusing on the Morris wedge, we realize that the conjugate interaction of back-ELA (No. 3) is decreasing when it affects the model energy output as opposed to the SA results that we established for the monthly outcome calibration previously. This finding will affect our ELA estimation as described below.

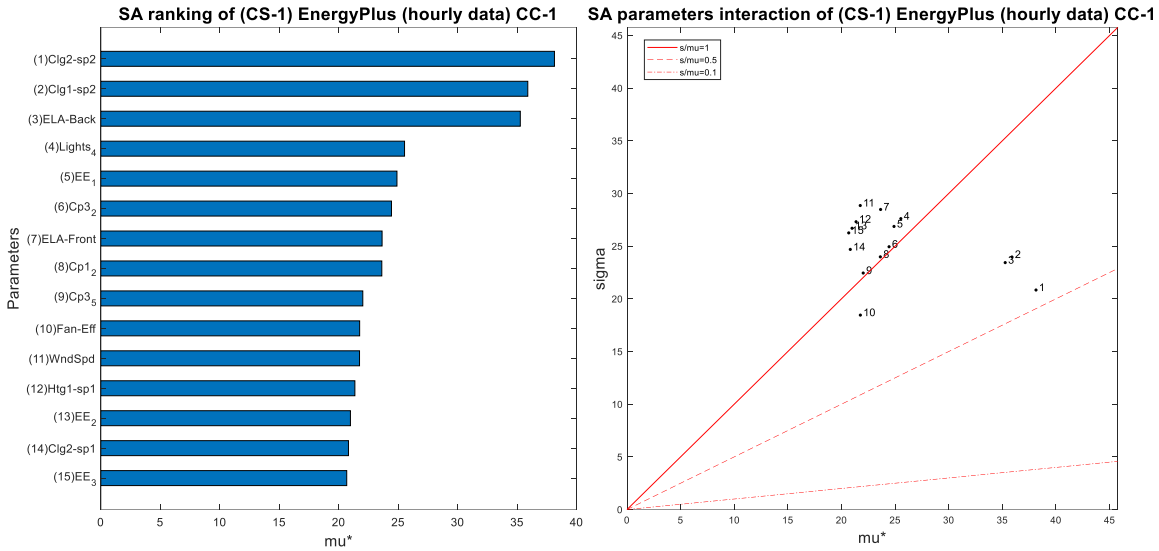


Figure 7.21 SA results of dominant parameters for CC-1, with calibration set CS-1

Consequently, we run the calibration steps with the same procedure we followed earlier, and the results of all calibrated cases for "back-ELA" are shown in Figure 7.22.

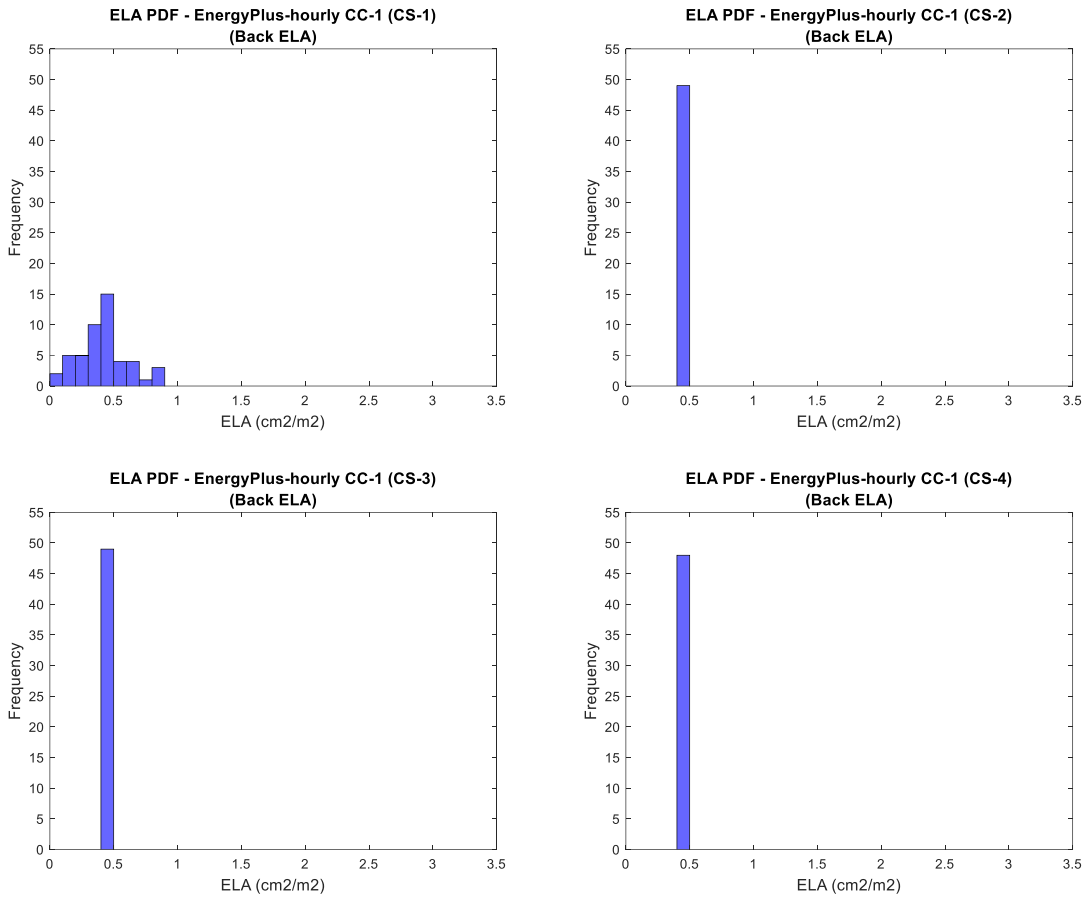


Figure 7.22 Calibration results of ELA (CS-1 through CS-4) for EnergyPlus-(hourly data)/CC-1

The first calibration set (CS-1) results in a narrower ELA distribution compared to the same calibration set that we established previously with EnergyPlus-(monthly outcomes). This is obviously due to the magnitude of the monotonic effect of ELA in the SA exercise (see Figure 7.21). It should be remembered that we perform the hourly calibration for the period from August to the end of October. Thus, both the data resolution and the period of data collection have a substantial role in characterizing the model's parameters, and possibly reduce its uncertainty. We also notice a clear peak in the ELA of CS-1 around 0.45 cm^2/m^2, which suggests a possible determination of ELA value in the later calibrations, which is found to be true as the CS-2 result reveals.

ELA in the calibration with CS-2 results in a more distinct distribution at 0.45 cm^2/m^2 because, at this point, the model has all the influential parameters that affect the value of the ELA estimation. It can be seen from Figure 7.23 that the SA outcome of CS-2 highlights cooling setpoints (No. 1 and 3), wind speed (No.2), and Cp (No. 4) to appear with a significant influence on the model outcome. The remaining parameters can be classified as less important, although not negligible for a number of them. Only cooling setpoints have a monotonic effect with σ/μ^* between 0.1 and 0.5. All other parameters show a non-linear influence and/or interactions with other parameters ($\sigma/\mu^* > 1$ for many of them). Therefore, with all parameters that help to form a determined ELA estimation in the model, we expect that every consecutive calibration will produce the same ELA distribution as no other influential parameters are involved, which indeed proves to be the case as evidenced in Figure 7.22. The SA results of the various calibration sets are presented below.

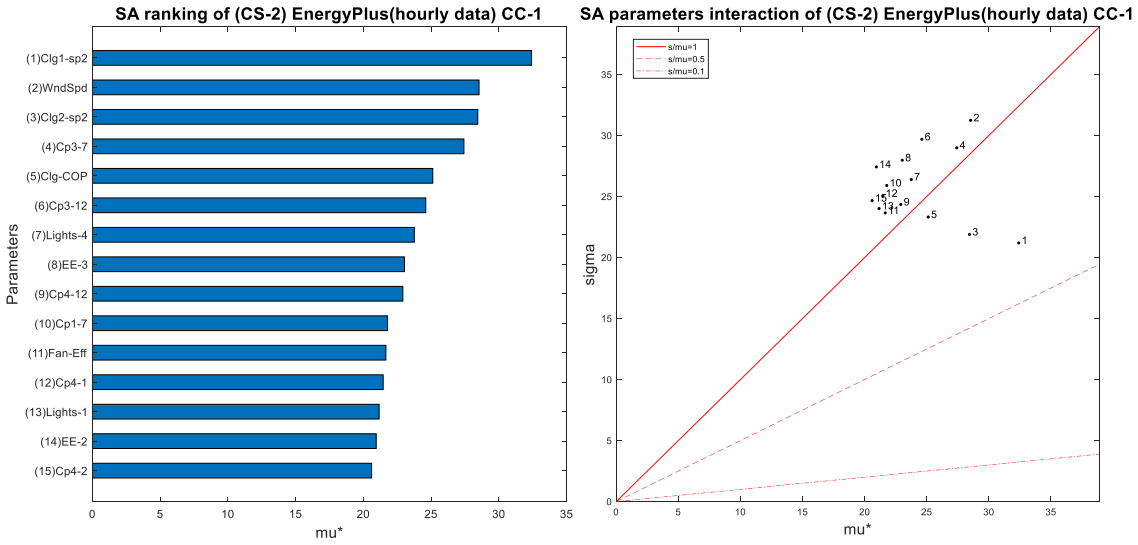


Figure 7.23 SA results of dominant parameters for CC-1, with calibration set CS-2

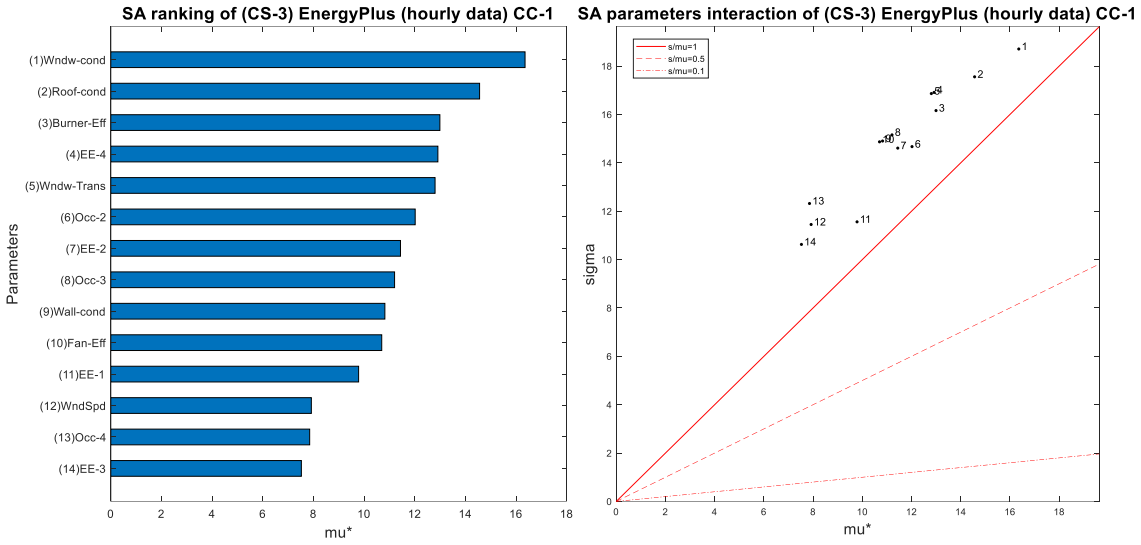


Figure 7.24 SA results of dominant parameters for CC-1, with calibration set CS-3

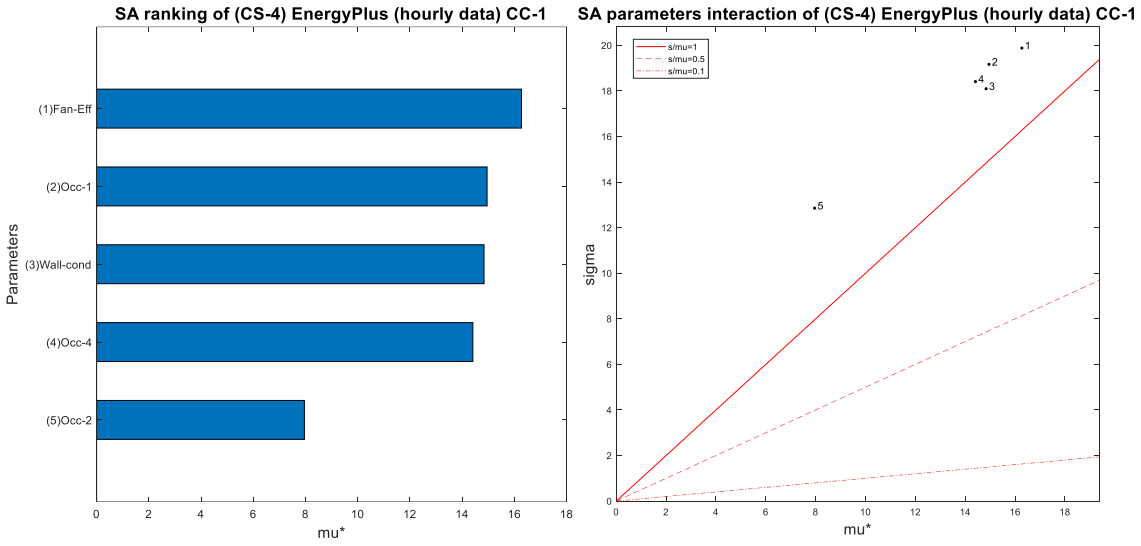


Figure 7.25 SA results of dominant parameters for CC-1, with calibration set CS-4

7.5.2 Adding temperature information in the calibration

Similar to the previous case, the calibration process is now repeated with higher resolution outcomes to test the potential improvement if additional outcomes are added to the calibration criterion. As in the EPC-hourly/CC-2, we use the indoor temperature in zone 4, which is measured for two months. In this case we perform the calibration for the

aggregated discrepancy in a similar fashion to EnergyPlus-(hourly data) using the same weighting factors of W_E=1 and W_T=2.8 for both outcomes respectively, from Eq. (5.4). The SA for the combined calibration process results in new sets of calibration parameters as presented in Table 7.8.

Table 7.8 Calibration parameters sets of EnergyPlus-(hourly data)/CC-2 used in the calibration steps

CS-1	CS-2	CS-3	CS-4
Front ELA	Front ELA	Front ELA	Front ELA
Back ELA	Back ELA	Back ELA	Back ELA
Right ELA	Right ELA	Right ELA	Right ELA
	Cooling setpoints	Cooling setpoints	Cooling setpoints
	Cp	Cp	Cp
	Fan efficiency	Fan efficiency	Fan efficiency
		Roof conductivity	Roof conductivity
		Lighting	Lighting
		Occupancy	Occupancy
		Wind speed	Wind speed
			Window transmittance
			Appliances
			Wall conductivity
			Window conductivity

The SA process in this section is similar to what was previously done in CC-1. Therefore, from Figure 7.26 we notice that back-ELA (No. 2) has high level of importance on the SA rank, and also indicates a nonlinear effect or interaction with other variables because it is located in the fourth region in the interaction graph (i.e., $\sigma/\mu^* > 1$).

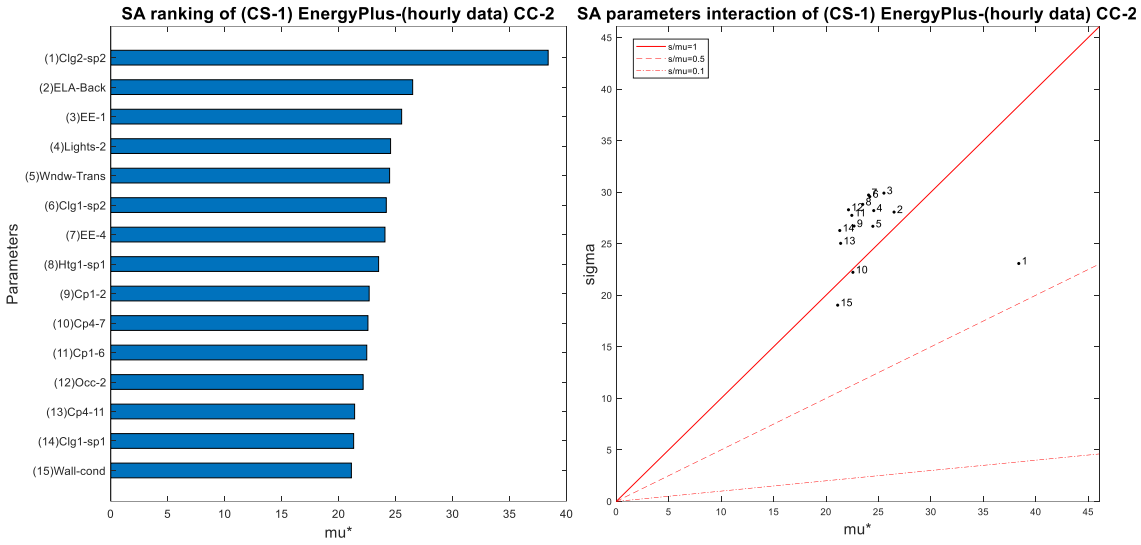


Figure 7.26 SA results of dominant parameters for CC-2, with calibration set CS-1

The calibration process is conducted with the same sets that are identified in Table 7.8, with the target to minimize the weighted discrepancy in both measured energy consumption and indoor temperature. Figure 7.27 shows overlaid distributions that compare all back-ELA results between EnergyPlus-(hourly data) CC-1 and CC-2.

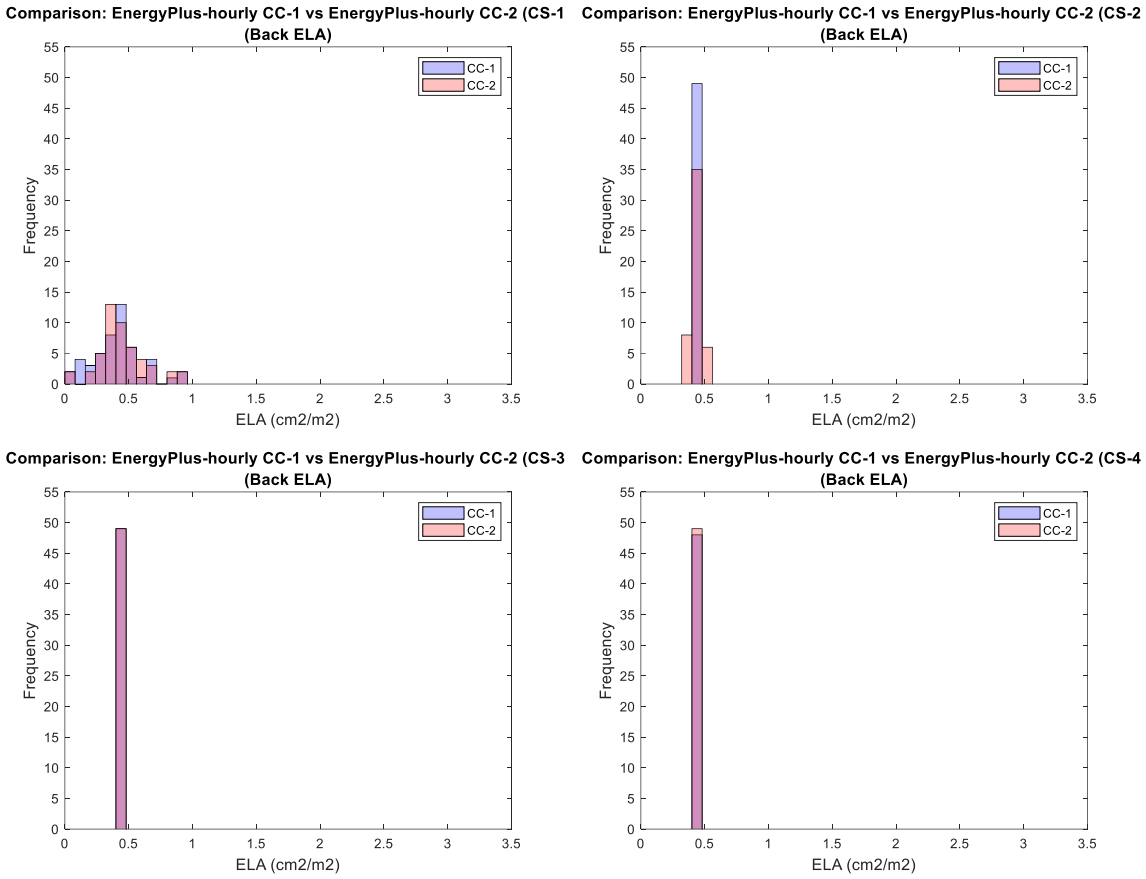


Figure 7.27 Calibration results of ELA (CS-1 through CS-4) for EnergyPlus-(hourly data)/CC-2

The ELA outcoms in this calibration exercise results in similar distributions obtained for CC-1. Even though with CS-2, ELA shows a little wider distribution, the peak value still best matches that for CC-1. We can say that the model at this point has all the influential parameters that affect the value of the ELA; in particular cooling setpoints and Cp, where the latter also has interaction with other parameters in the model as shown in Figure 7.28. Consequently, every consecutive calibration produces the same ELA distribution as no other influential parameters are involved. This finding suggests that feeding this model with additional information (i.e., indoor temperature) has minimal impact on the ELA estimation. The remaining SA outcomes are presented below.

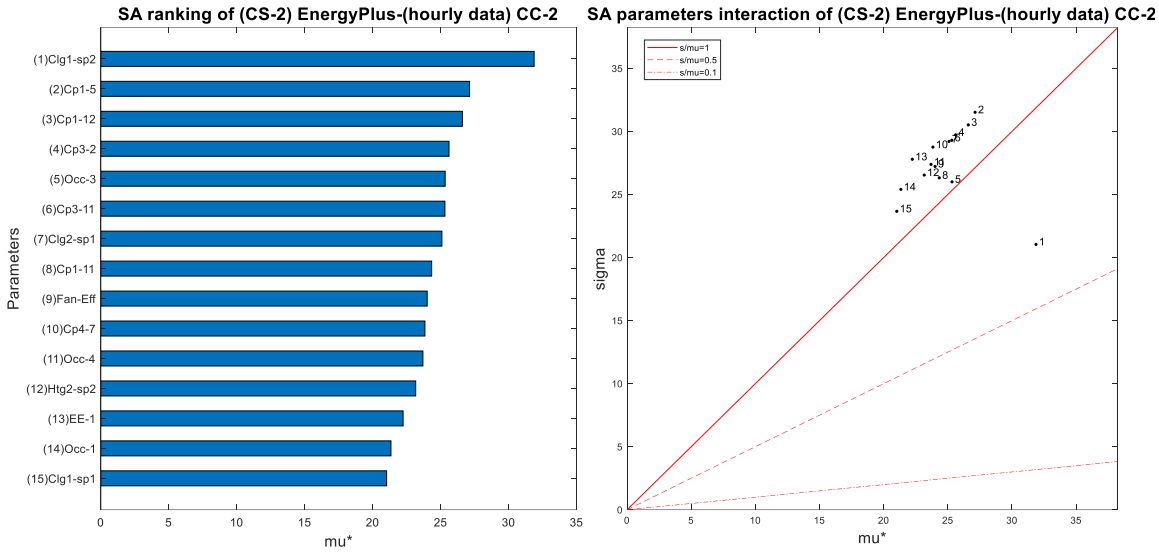


Figure 7.28 SA results of dominant parameters for CC-2, with calibration set CS-2

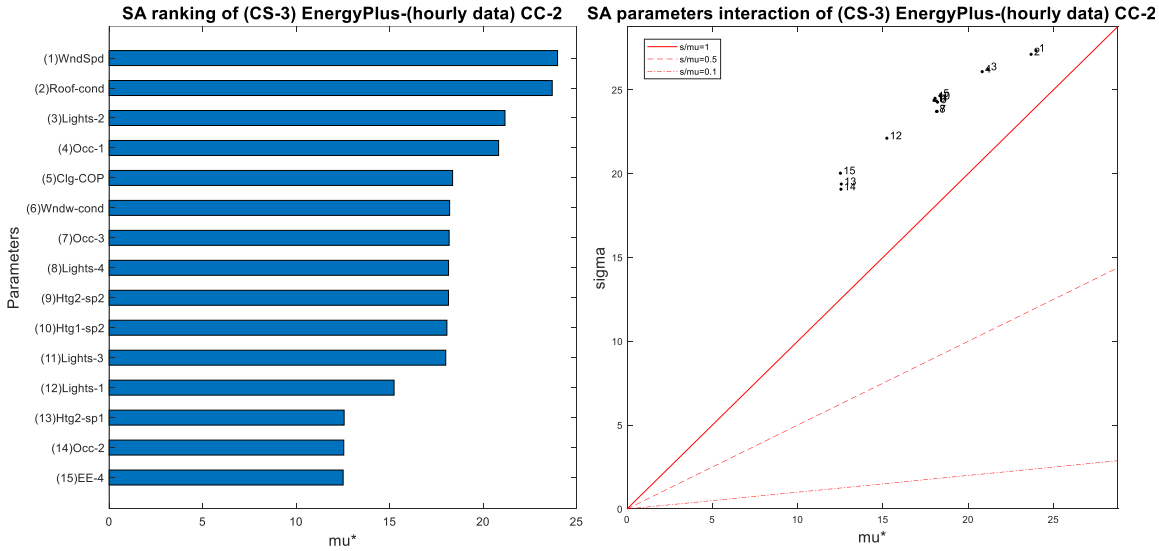


Figure 7.29 SA results of dominant parameters for CC-2, with calibration set CS-3

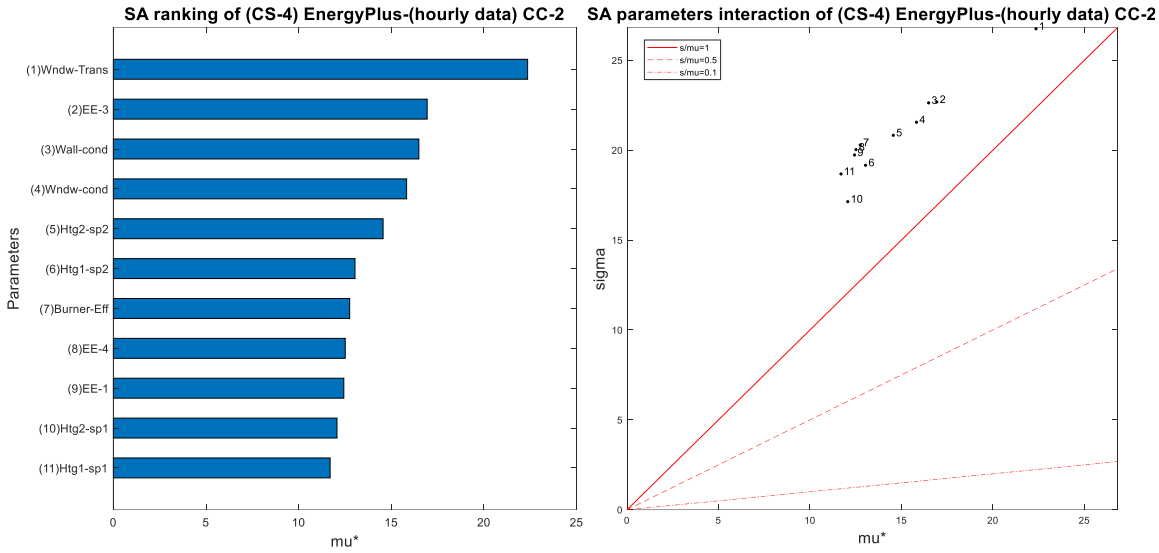


Figure 7.30 SA results of dominant parameters for CC-2, with calibration set CS-4

7.6 Evaluation and results analysis of energy models

This section presents a comparison between the CVRMSE results of EPC and EnergyPlus models with the different calibration criteria. The details are provided in the following subsections.

7.6.1 Evaluation and results analysis of monthly calibrations

We compare the calibrated EPC-monthly cases with the calibrated EnergyPlus-(monthly data) cases in terms of the performance metric CVRMSE. Figure 7.31 uses boxplots to show the variation of the statistical measures with the four calibration sets (CS-i) for one calibration criterion, i.e., CC-1 (uses QoI-1-M only). It is obvious that the variation in the prediction error with CS-1 is relatively larger than other cases for both EPC and EnergyPlus, with a median CVRMSE of 0.49 and 0.35 for EPC and EnergyPlus, respectively. The whole picture of the results shows a gradual decrease in CVRMSE values, with CS-4 resulting in the smallest median and variation in CVRMSE over the other

cases. This means not only that CS-4 overall performs better than the other calibration sets, but also that it is more stable since the variations among different cases are smaller. CS-2 in both models have median values that are greater than 15%, which means that both CS-1 and CS-2 in EPC and EnergyPlus models do not meet the ASHRAE guideline. The overall performance of CS-3 in both models is similar to CS-4 (below 15%), which proves to be within the acceptable range.

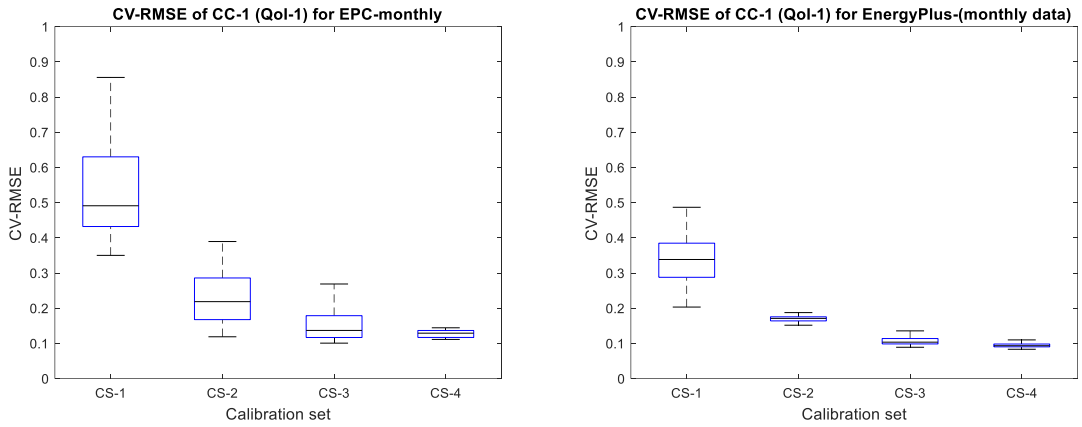


Figure 7.31 The variation of CVRMSE on monthly energy prediction among the calibration sets for EPC (left) and EnergyPlus (right)

Yet, we cannot rely only on CVRMSE to evaluate the accuracy of the calibrated models. Hence, it is important to inspect if the observed data is in close similarity with the predicted outcome. Figure 7.32 to Figure 7.35 show a pairwise comparison between EPC and EnergyPlus energy predictions with the calibrated models and observed outcomes.

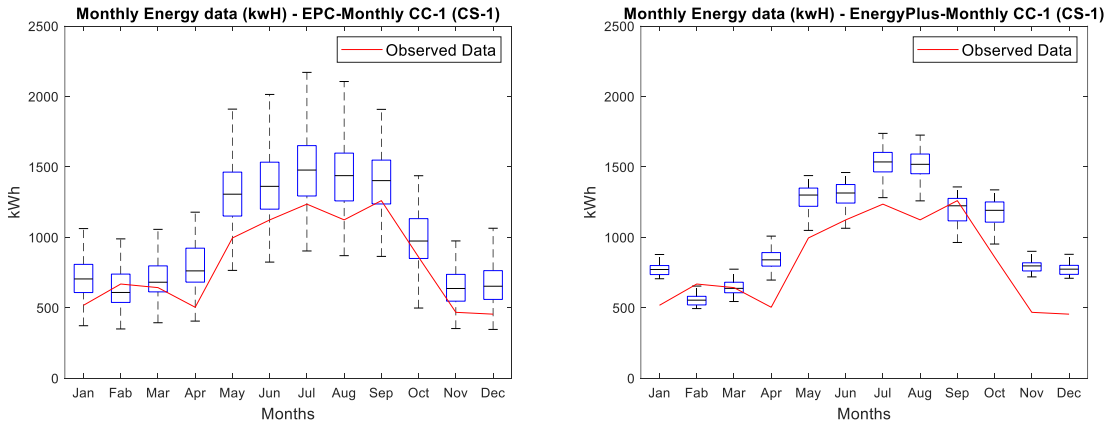


Figure 7.32 Energy consumption with CS-1: EPC (left), EnergyPlus (right). Red (observed data), box (prediction)

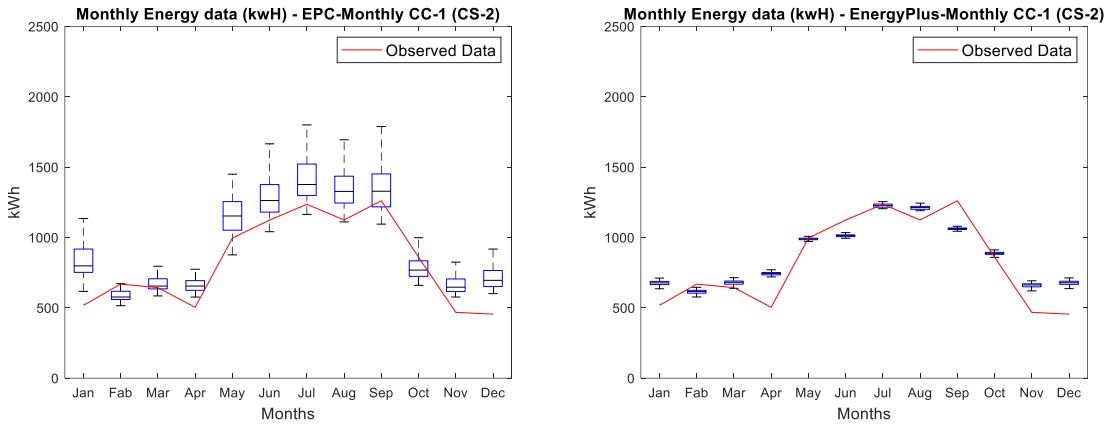


Figure 7.33 Energy consumption with CS-2: EPC (left), EnergyPlus (right). Red (observed data), box (prediction)

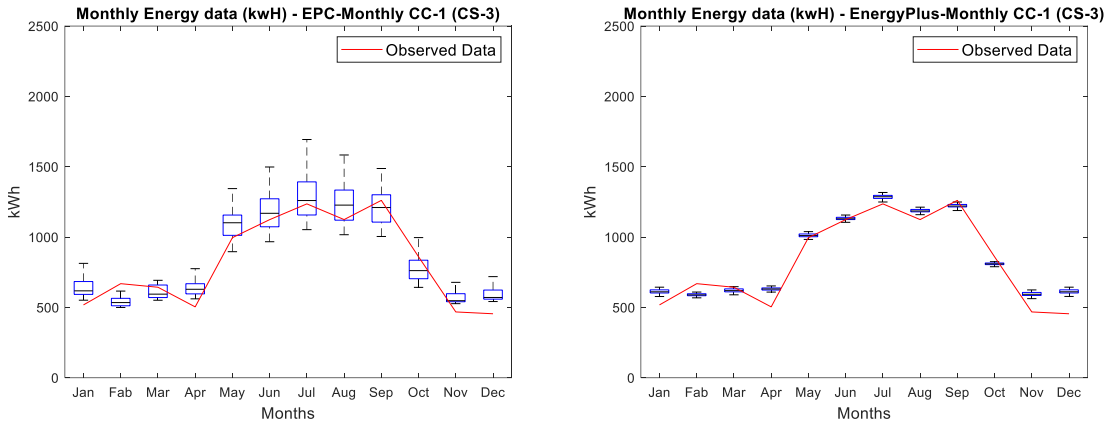


Figure 7.34 Energy consumption with CS-3: EPC (left), EnergyPlus (right). Red (observed data), box (prediction)

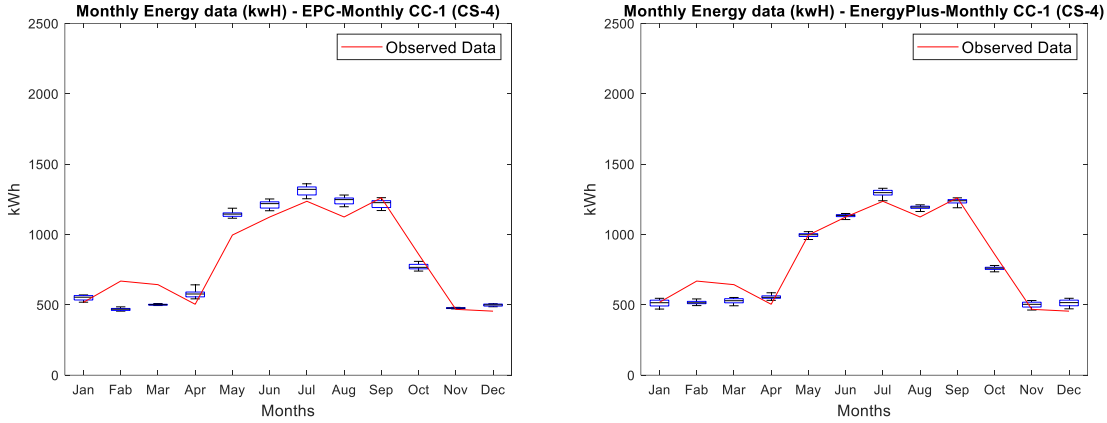


Figure 7.35 Energy consumption with CS-4: EPC (left), EnergyPlus (right). Red (observed data), box (prediction)

In general, models that are calibrated with fewer parameters seem to have considerable deviation compared to those with more calibration parameters. From these results, it can be concluded that the calibrated EPC models are as good as the calibrated EnergyPlus model in terms of model optimization, but it must be recognized that a model that passes can still have useless approximations of calibrated parameters, which is discussed in CHAPTER 8 with reflections on the realized ELA estimate. Moreover, the results indicate that EnergyPlus simulation cannot do better than a low resolution model if one has only low-resolution observations.

7.6.2 Evaluation and results analysis of hourly calibrations

With the more granular data and increased fidelity of the models, we provide a comparison between the CVRMSE results of EPC-hourly and EnergyPlus-(hourly data) with their different calibration criteria. Figure 7.36 and Figure 7.37 show the variation of the statistical measures over the calibrated models for two calibration criteria, i.e., CC-1 (uses QoI-1-H only) and CC-2 (QoI-1-H + QoI-2), along with the consecutive selection of calibration sets (CS-i). The variation in the prediction error with CS-1 and CS-2 is relatively larger than the other calibration cases which is seen clearly for both EPC and EnergyPlus. For example, the median CVRMSE with CS-1/CC-1 is 0.47 and 0.33 for EPC and EnergyPlus, respectively; and with CS-1/CC-2 it is 0.44 for EPC and 0.25 for EnergyPlus. In contrast, with CS-2, CS-3, and CS-4 we see smaller variations in CVRMSE for both models, with medians much smaller than with CS-1. This means that these calibration sets are more stable since the variations among different models are smaller. Another conclusion to draw from the CVRMSE values is that the calibrated outcomes for CC-1 are as good as for CC-2. More discussion about this is provided in CHAPTER 8.

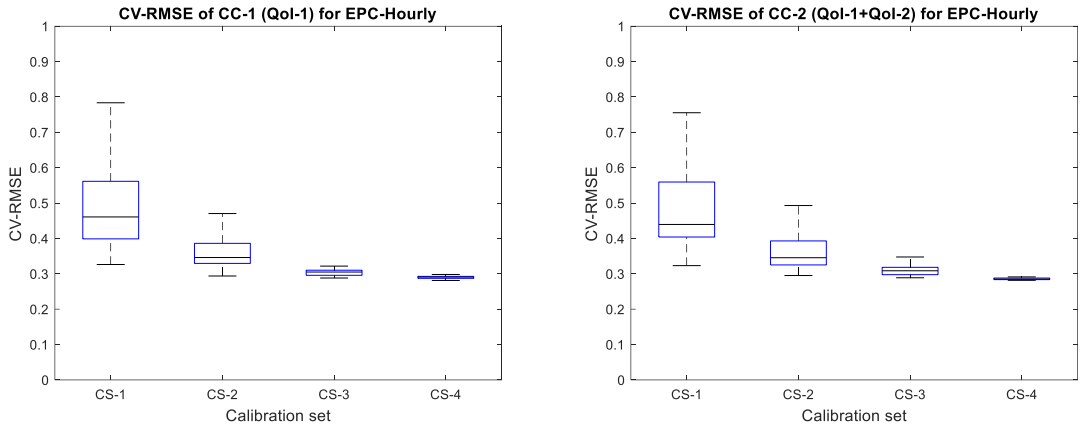


Figure 7.36 The variation of CVRMSE on energy prediction among the calibration sets for EPC. CC-1 (left), CC-2 (right)

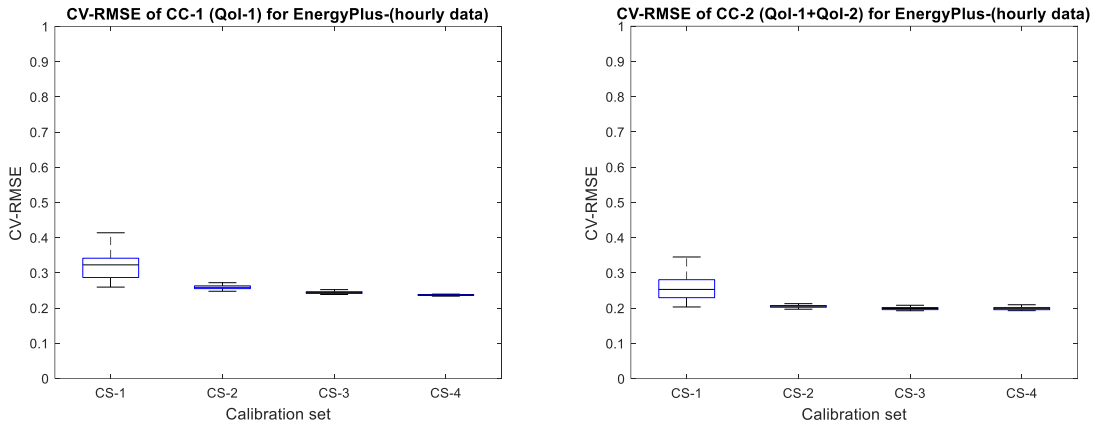


Figure 7.37 The variation of CVRMSE on energy prediction among the calibration sets for EnergyPlus using CC-1 (left) and CC-2 (right)

It is not enough to depend only on CVRMSE to evaluate the accuracy of the calibrated models, so it is important to check whether the observed data is in close similarity with the predicted outcome. Figure 7.38 to Figure 7.41 show pairwise comparison of the energy consumption predictions between EPC-hourly and EnergyPlus-(hourly data) for different calibration criteria; every figure presents four graphs for every calibration set (CS-i). The data presented are for the first two days of August only.

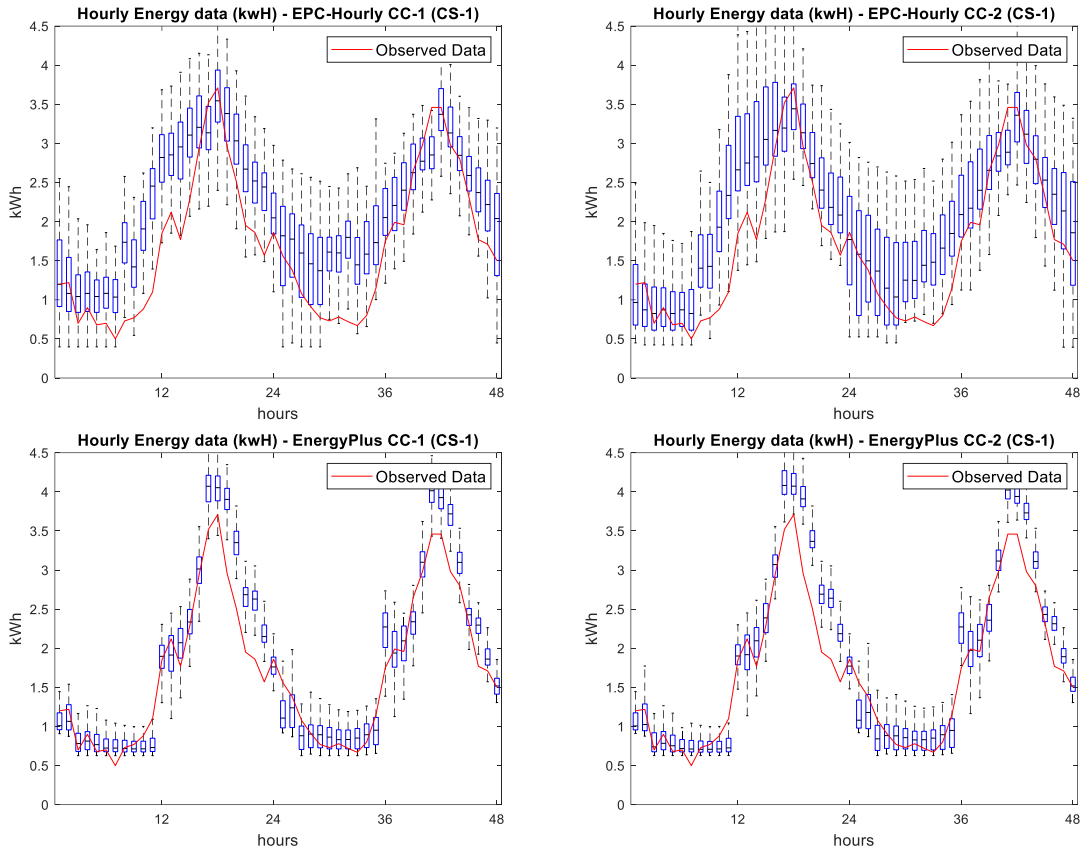


Figure 7.38 Energy consumption with CS-1: EPC (top), EnergyPlus (bottom). Red (observed data), box (prediction)

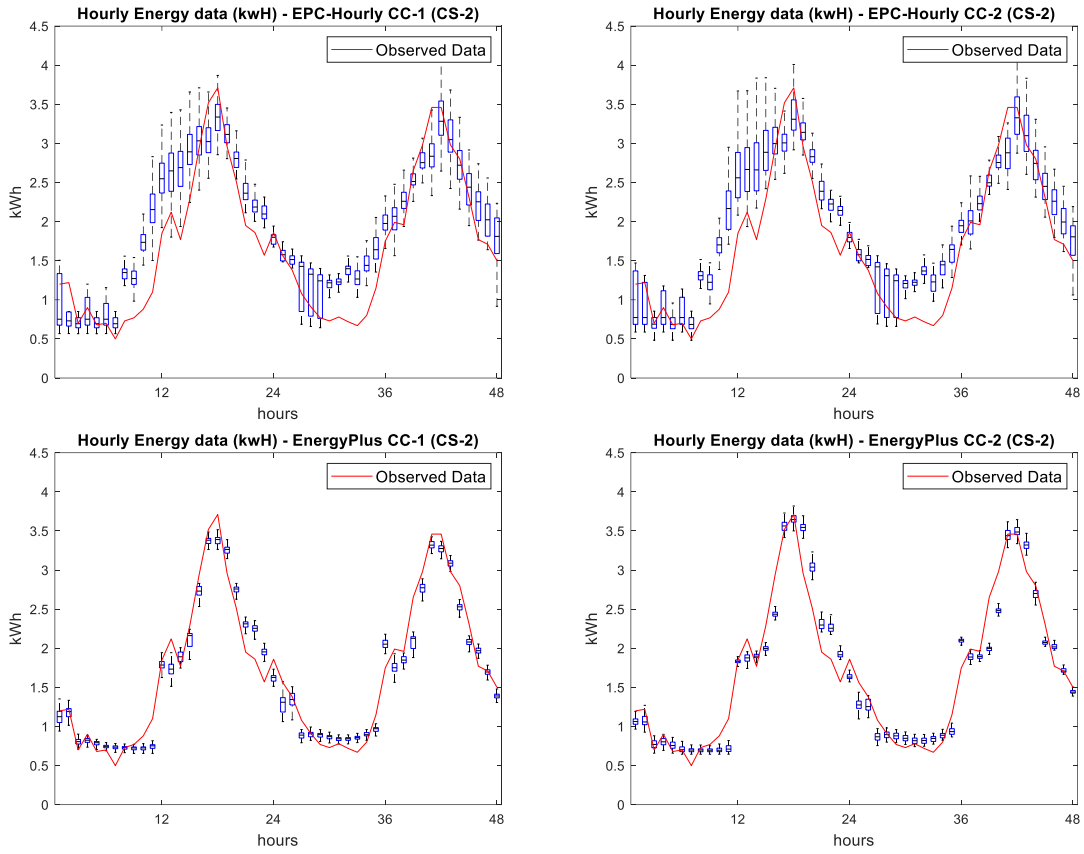


Figure 7.39 Energy consumption with CS-2: EPC (top), EnergyPlus (bottom). Red (observed data), box (prediction)

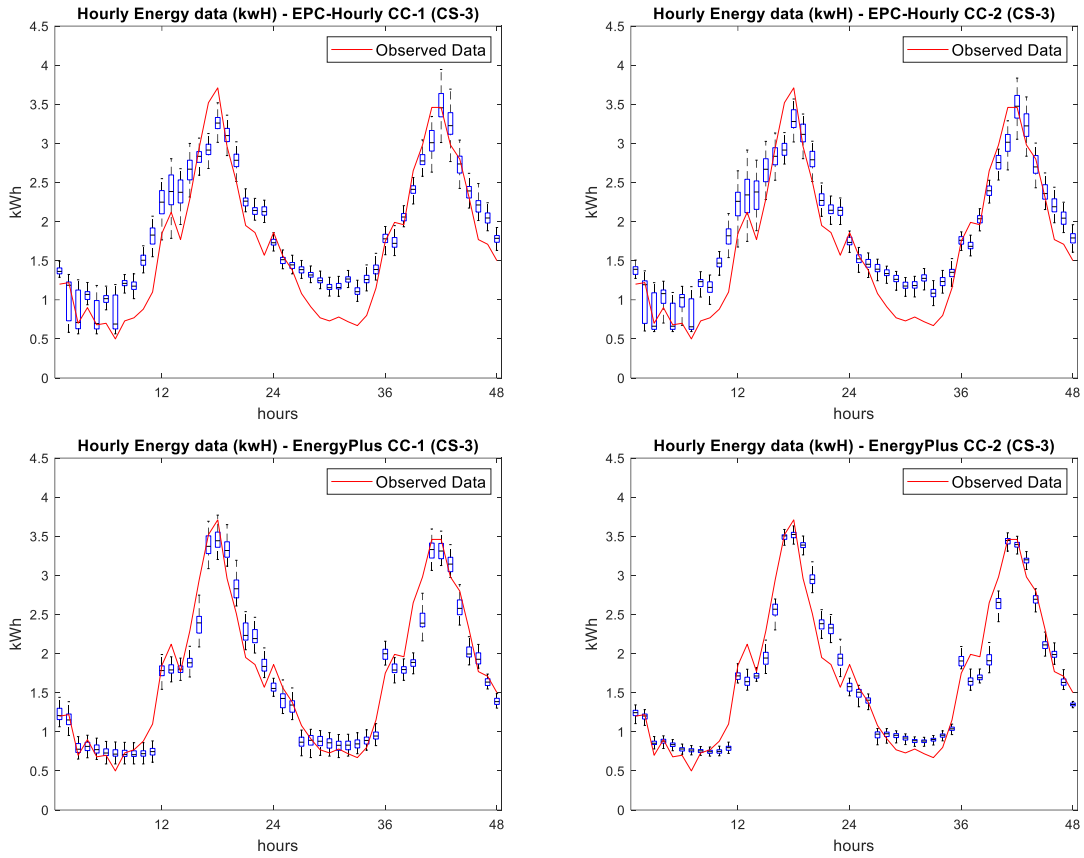


Figure 7.40 Energy consumption with CS-3: EPC (top), EnergyPlus (bottom). Red (observed data), box (prediction)

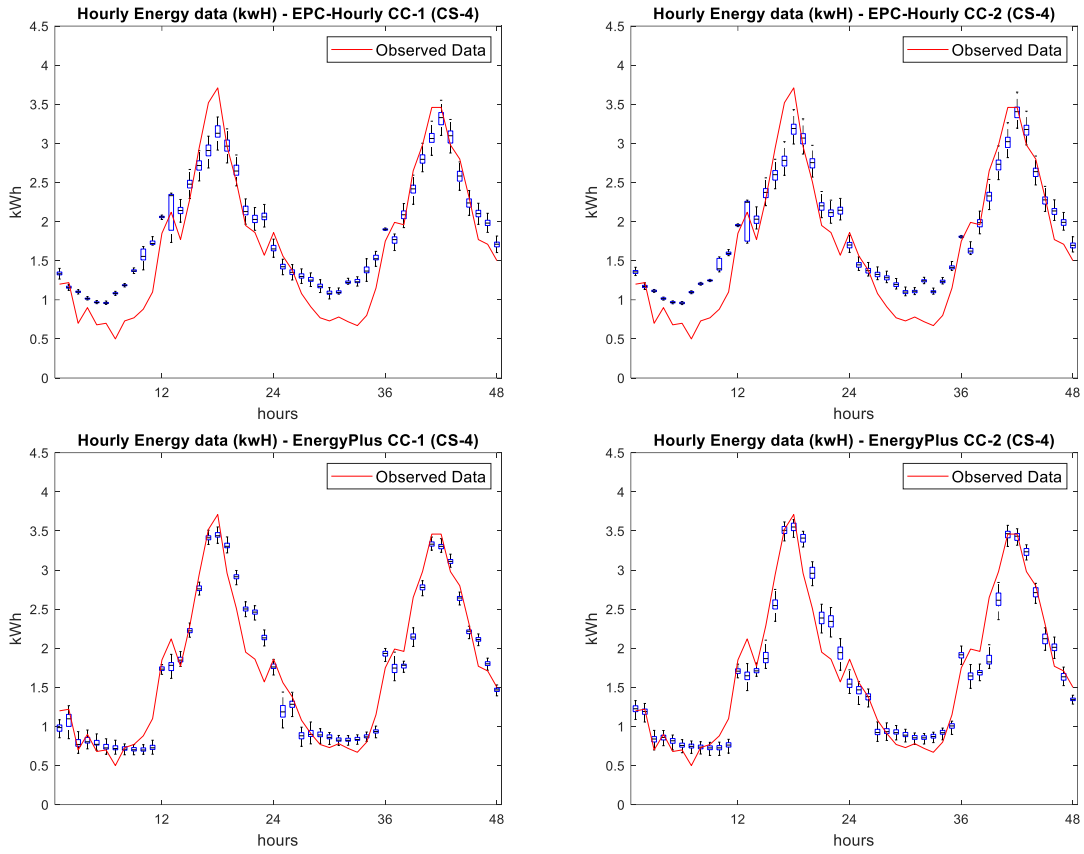


Figure 7.41 Energy consumption with CS-4: EPC (top), EnergyPlus (bottom). Red (observed data), box (prediction)

From the figures above we notice that the energy consumption results for CC-2 have not led to a significant improvement compared to CC-1, which means that the additional QoI (indoor temperature) has minimal impact on the model output. This finding suggests that our energy model with hourly energy data can perform well enough without feeding it with more information. However, this reasoning cannot be generalized without more cases.

7.7 ELA determination based on in-situ measurement

Figure 7.42 shows a comparison of the ELA calibrations that are generated with EPC and EnergyPlus models against the experimental ELA with measured wind pressure.

The distributions presented in the plot are the best models of each tool with the respective calibration set.

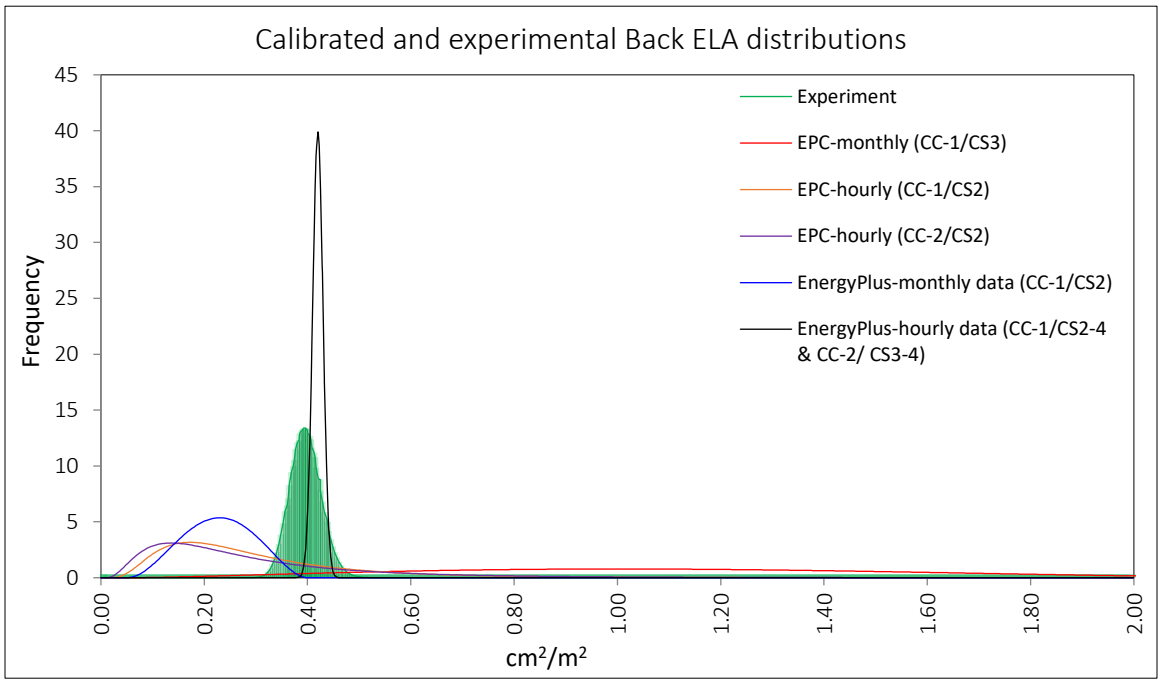


Figure 7.42 Best results of ELA calibrations with EPC and EnergyPlus in comparison with experimentally determined ELA (green)

The shaded green distribution in Figure 7.42 represents the direct ELA calibration with the use of the pressure sensor. Therefore, this is assumed to be our best prospect for the true ELA. The direct calibration of ELA provides an enhanced estimate where the 95% confidence interval is between 0.34 and 0.46 cm^2/m^2. Assuming that for engineering purposes an 80% confidence interval is acceptable, it would situate the ELA engineering estimate between 0.36 cm^2/m^2 and 0.44 cm^2/m^2 which seems a reasonable range based on the tracer gas experiment.

In terms of ELA estimation through our calibration process, Table 7.9 provides the characterization of the ELA of each calibration set outcomes in the calibration process, and the best estimations (for a given tool/data/CC combination) are highlighted in green.

Table 7.9 Measurements of the distance and divergence between estimated and direct ELA

Model	Measurements of Direct ELA at CI = 80%											
	CS-1			CS-2			CS-3			CS-4		
	P	KL	SE	P	KL	SE	P	KL	SE	P	KL	SE
EPC-monthly/CC-1	0.01	2.57	0.078	0.01	2.37	0.074	0.03	1.88	0.071	0	830	0.001
EPC-hourly/CC-1	0.08	1.43	0.017	0.07	1.30	0.025	0	7.22	0.005	0	23.06	0.001
EPC-hourly/CC-2	0.09	1.38	0.018	0.11	1.27	0.023	0.01	2.52	0.007	0	35.67	0.003
EnergyPlus-(monthly data)/CC-1	0.01	1.29	0.046	0.009	1.66	0.009	0	10.25	0.001	0	10.25	0.001
EnergyPlus-(hourly data)/CC-1	0.20	0.96	0.023	0.98	0.42	0.001	0.98	0.42	0.001	0.98	0.42	0.001
EnergyPlus-(hourly data)/CC-2	0.20	0.96	0.023	0.84	0.25	0.003	0.98	0.42	0.001	0.98	0.42	0.001

Consequently, none of the EPC-monthly cases results in a reasonable estimation that can represent the true ELA. CS-3 is considered to give the best ELA distribution (red distribution) for EPC-monthly given that its probability value is the highest (0.03) and KL is the lowest (1.88), but its standard error (SE) is high indicating that it has very wide distribution as clearly observed from the graph. It is worth paying attention to CS-4 outcomes; although this calibration set has the lowest SE, its probability is *zero*, and the KL value is the highest (830), indicating that it is extremely far from the direct ELA estimate, which makes it the least reliable case. This implies in this particular case that, in both monthly models (EPC and EnergyPlus), the high effect of Cp and wind speed makes ELA deviate from the true value due to their subsumption effect, which happens because they are in the same formula. On the other hand, all other EPC distributions are clustering behind the green distribution with slightly different variations. EPC-hourly/CC-1 and CC-

2 (orange, and purple distributions, respectively) are nearly similar, where both slightly intersect with the green distribution. However, this still could be a less informative estimation when compared with the experimental determination of ELA.

EnergyPlus models perform better than EPC in for both monthly and hourly outcomes. The blue distribution represents EnergyPlus-(monthly data). It is interesting to note that CS-2 is considered to give the best ELA distribution over the other cases due to its better combination of goodness probability, KL, and SE (0.009, 1.66, 0.009, respectively), but this does not necessarily mean that CS-2 as an energy model is the best. Furthermore, EnergyPlus-(hourly data) proved to be the best models for both CC-1 and CC-2 with identical distributions that are situated within the direct ELA distribution. Both models have cases with probability equals to 1 as seen in CS-2 to CS4 and CS-3 to CS-4 for EnergyPlus-(hourly data)/CC-1 and CC-2, respectively.

Considering the above, the primary conclusion is that all ELA distributions derived with monthly outcomes are broad and substantially diverge from the experimentally derived best estimate for ELA. In fact, they hardly improve on the non-informative prior. This of course leads to the inevitable conclusion that with a data resolution of only monthly aggregated energy consumption leads to rather useless results for ELA estimates. Models calibrated with hourly outcomes perform better; this is most pronounced for the EnergyPlus model calibrated on hourly outcomes, which shows better ELA prediction than all other models. This confirms that we can achieve better parameter estimation with higher resolution outcomes and a higher fidelity model. The other (disappointing) conclusion is that the low fidelity model with high (hourly) resolution, although able to closely match the measured energy outcomes, does not come close to the best estimate of ELA.

CHAPTER 8. DISCUSSION OF RESULTS

The previous chapters reveal that the calibration process reduces the mean bias errors between measured and energy model predictions used in two case studies. The conducted calibration resulted in ELA estimates based on different calibration sets (CS-i) and different criteria (CC-1 and CC-2). This chapter provides an inclusive discussion of the results presented in CHAPTER 6 and CHAPTER 7 and reflections on the resulting ELA approximation.

8.1 Discussion

We evaluate the validity of calibrated models in terms of agreements between predicted and monitored energy uses. Table 8.1 shows CVRMSE values of all calibrated models that are described in CHAPTER 6 and CHAPTER 7. The first column in the table indicates the type of model in the calibration process for both given case studies. The CVRMSE is provided in terms of mean, μ, and standard deviation, σ, with the four calibration sets (CS-i). The table also shows the best ELA results that were generated from every model's calibration exercise with its KL and probability value, P, as well as the CS-i with which the best estimate was achieved.

Table 8.1 overall view of the models' energy performance and the resulted ELA

Model	CVRMSE								Best ELA result		
	CS-1		CS-2		CS-3		CS-4				
	μ	σ	μ	σ	μ	σ	μ	σ	KL	P	CS
Case study 1											
EPC-monthly/CC-1	0.35	0.11	0.23	0.04	0.20	0.04	0.15	0.01	1.89	0.03	*2*
EPC-hourly/CC-2	0.20	0.04	0.15	0.01	0.14	0.01	0.11	0.00	0.36	0.62	*1*
EnergyPlus/CC-1	0.21	0.14	0.11	0.03	0.11	0.03	0.09	0.01	0.31	1	*4*
EnergyPlus/CC-2	0.19	0.03	0.16	0.02	0.11	0.01	0.10	0.01	1.40	1	*4*
Case study 2											
EPC-monthly/CC-1	0.54	0.15	0.24	0.10	0.16	0.06	0.14	0.04	1.88	0.03	*3*
EPC-hourly/CC-1	0.50	0.15	0.36	0.05	0.30	0.01	0.29	0.00	1.30	0.07	*2*
EPC-hourly/CC-2	0.50	0.15	0.36	0.05	0.31	0.02	0.29	0.00	1.27	0.11	*2*
EnergyPlus-(monthly data)/CC-1	0.34	0.07	0.17	0.01	0.11	0.01	0.09	0.01	1.66	0.009	*2*
EnergyPlus-(hourly data)/CC-1	0.32	0.04	0.26	0.01	0.26	0.08	0.24	0.00	0.42	0.98	*2 to 4*
EnergyPlus-(hourly data)/CC-2	0.26	0.03	0.21	0.00	0.20	0.00	0.20	0.00	0.42	0.98	*3 to 4*

The results show an overall consistent estimation of the calibration parameters among models with the same type of observations and calibration sets and time resolutions. The statistical measures indicate that the calibrated EPC models can predict the energy consumption as accurately as the calibrated EnergyPlus models. Also, the validation measures tell that the calibration process improves the accuracy of the model by gradually reducing the CVRMSE values with CC-1 and CC-2. It should be noted that in both case studies, all CVRMSE values with CS-1 of the calibrated models are higher than what is stipulated in ASHRAE Guideline 14 for a model to be deemed valid (i.e., 15% and 30% for the monthly and hourly data resolution, respectively) with the only exception for EnergyPlus-(hourly data)/CC-2.

A closer look at the accuracy of calibrated models across different cases shows that after calibration, i.e., refined quantification of uncertainties associated with important parameters and model bias, the best overall accuracy is achieved when all types of

observations (QoI-1 + QoI-2) are used regardless their time resolutions. However, calibration with QoI-2 (i.e., indoor temperature) tends to add only a minor improvement to the overall calibration process, which indicates that the benefit of temperature monitoring is not significant; probably because its relative small hour-to-hour variations over the entire period do not contribute too much information in addition to the energy consumption. This is particularly true in operated buildings where the temperature fluctuations are controlled within a narrow band. Only nights and weekends will then add relevant temperature information to the calibration. For this and other reasons, our case studies confirm that the energy has more reliable sensitivity to ELA than temperature. In situations when the calibrated model predicts temperatures very well (small CVRMSE), the (large) MFU gets always subsumed in the parameters making them less reliable. The effect of this interpretation can be seen more in the low fidelity EPC models.

The results in the above table indicate that models that are calibrated with fewer calibration parameters seem to have considerable deviation compared to those with more calibration parameters. In itself, this is not surprising as the calibration of smaller parameter sets suffers from more uncertainty in the non-calibration parameters. This does not always mean that the more calibration parameters the better model we achieve. In fact, feeding the model with a large number of calibration parameters, could lead to parameters subsumption that we described in CHAPTER 1. The proposed calibration process and the method of selecting calibration parameters from the pool based on a repetitive SA and deterministic calibration is set up to verify the subsumption effect. Surprisingly, the effect could not be established. Generally it is found that the model predictions get better with larger calibration sets and the ELA estimate converges to its best estimate. This may not be true

though for all other calibration parameters. The major advantage of the stepwise calibration is that one is able to verify how much precision one sacrifices if the calibration is conducted for a small set of parameters. This approach can be extended to improve the precision of the outcome by reducing the uncertainty in the other parameters. The SA results indicate on which parameters one should focus first if one were to take this approach.

We further compare the calibrated models in the context of ELA prediction. From this perspective we conclude that all EPC models provide less accurate ELA estimations comparing to what is achieved by EnergyPlus. In addition, the best ELA distributions in EPC models are either wide or far from the best estimate for ELA, which suggests that the MFU in these models is so significant that a reliable ELA estimate is more likely not reachable. This is also supported by the CVRMSE value for which the best ELA estimate was achieved. The first best ELA estimate in case study *one* is achieved at CS-2 with a mean CVRMSE value of 0.23, and CS-1 with a mean CVRMSE of 0.20; the similar finding is also realized in case study *two,* (see Table 8.1). However, it can be seen that the addition of QoI-2 in the calibration process leads to slightly better prediction, and this is verified in both case studies as presented in the ELA distributions. Accordingly, the CVRMSE values where the best ELA estimates were achieved with EPC models do not meet the ASHRAE guideline, which is in line with the expectation that ELA calibration results for these low-resolution models lead to less informative and hardly usable results.

On the other hand, the EnergyPlus models show more consistent outcomes regarding both CVRMSE and ELA estimates. This means that the ELA predictions become more reliable and definitely lead to a more informative result. We see this clearly in

EnergyPlus-(hourly data) in case study *two* where both ELA estimates and CVRMSE have reached constant outcome from CS-2 to CS-4 with a probability of 0.98 and KL 0.43.

Another important conclusion is about calibrations that ignore uncertainties in dominant parameters. Such situations arise when a modeler deterministically calibrates a model (e.g., CS-1, CS-2, or CS-3) while ignoring uncertainty in other parameters. In doing so he/she has to choose deterministic input values based on person judgment, which unavoidably leads to a modeler's bias. For instance, the outcomes for the two case studies show that the distribution in CVRSME is large for CS-1 but fast reducing for CS-2 and CS-3. This means that a deterministic calibration can hit a random point within the bands of CVRMSE outcomes. More importantly, the modeler's bias will result in a random outcome for the calibration parameters. Our results show that the band for ELA is large, especially for CS-1 and CS-2, and to some extent even for CS-3, which means that the parameter's estimation of a deterministic calibration can be at any point within the range of the parameter outcomes. As a consequence, it can be said that while ignorance of uncertainty, many current calibration methods with a limited number of calibration parameters lead to very unreliable parameter approximation. The only way to avoid this is to consider the largest set of parameters for calibration as well as parameter uncertainty. This however leads to a large computational effort for the optimization where in many cases the optimum is missed. Nevertheless, this is recommendable as the potential negative effect of large calibration set (i.e., subsumption) plays a smaller role than expected at the outset of this study. Of course, this will have to be confirmed by more case studies.

8.2 Concluding remarks

The objective of this chapter is to verify the first research question: will the effective leakage area (ELA) that is calibrated with measured, readily available outcomes at different resolutions be close enough to one derived from direct measurements, which is considered the best obtainable estimate? Several main conclusions can be drawn:

1. Based on the SA results we can conclude that scenario uncertainty and parameter uncertainty are major uncertainties affecting the prediction capability with current energy models. This conclusion is driven by the fact that ELA always appears to be a dominant parameter that has significant impact on the outcomes of energy models. The reasoning for this is that ELA is a surrogate parameter that is sensitive to other unknown sources and effects that enter into the ELA estimates e.g., Cp and wind speed. This interaction affects the determination of a precise ELA estimate.
2. The proposed calibration framework confirms that more calibration parameters can achieve better parameter estimation. However, this statement needs qualification in many cases (especially in monthly models) as some calibration parameters have only significant impact on the energy consumption but not much influence on the ELA calibration. Moreover, the "improvement" that is obtained in the ELA estimate is often not significant and in fact not a major improvement of the uninformative prior we established in section 4.1. Calibrations with hourly outcomes typically show better ELA estimation compared to monthly resolution. The high-fidelity models (i.e., EnergyPlus) prove to give a significantly closer estimate of the directly calibrated ELA than other models.

3. The considerable reduction in the CVRMSE with the higher resolution models is driven by the selection of calibration parameters. It is found that calibration with an additional outcome type (i.e., indoor temperature) can improve the overall CVRMSE, but that is highly dependent on the choice of the weighting factor and the setup of the hourly temperature simulation. Calibration with energy and indoor temperatures can in fact result in worse energy predictions compared to the best calibration with only energy data. It is not hard to see why this is always the case and is theoretically unavoidable.
4. Finally, in terms of data variety, it is always desirable for a data set to include simultaneous measurements of different types of variables, such as HVAC systems and separately submetered energy use, lights, and occupancy. When different types of variables can be simultaneously monitored over a relatively long period, model verification can be performed in a more comprehensive manner than we do on our buildings. Moreover, comparing the accuracy among different types of variables may also lead to clues about the sources of prediction discrepancy. Therefore, based on the patterns of energy prediction discrepancy over a calibration period, we suspect that uncertainty in some parameters, e.g., lights and plug loads, which are mostly quantified at the whole-building scale, had better be refined to the level of thermal zones.

In summary, the developed calibration framework is able to correct prior misguided assumptions about true ELA values, but results still depend on the prior estimates of other uncertainty in the energy model, uncertainty quantification, and the interaction with other parameters. One point to be emphasized is that prior estimates are set up based on collective

expert knowledge and can be made more precise when there is additional knowledge from prior uncertainty quantification studies. The resulting prior estimates can then be further refined through the calibration exercise.

CHAPTER 9. THE ROAD TOWARD A FITNESS MEASURE TO TEST A CALIBRATED MODEL

This chapter demonstrates the capability of the new framework to support uncertainty analysis for risk-conscious decision making. CHAPTER 2 presents the importance of recognizing the role of uncertainty in a calibration process, i.e., the level of confidence that can be placed in our predictions as revealed by an uncertainty analysis. Without a measure of model reliability, it is impossible to judge the accuracy of a prediction for making decisions with confidence. Any "calibrated" model generates outcomes for a selected QoI that are by definition uncertain (mostly propagated from uncertain model parameters). Depending on the level of uncertainty this may pose considerable risk to the validity of the model's use in a decision context.

As summarized earlier, one of the key limitations of current methods is the implementation of deterministic approaches that ignore the influence of uncertainty on the model. It is a common misconception that increasing the fidelity of the model plays a dominant role in improving the validity of the model. However, accuracy should not be the primary concern when the model is used to generate predictions that are used in a given decision context. Rather, the question becomes how "fit" a model is to support the decision making. It could be expected that for simple tasks a low fidelity model will do well enough, even if its accuracy is not very great. In a calibration context, this observation is significant. Indeed a model is typically declared accurate if it is able to closely approximate a measured QoI, even if the values of the calibrated parameters could be far removed from their true values. As one puts more effort in the calibration, one may be able to improve the

approximation of the true calibration values, but it cannot be verified as the true values are unknown and in many cases unobservable, even through dedicated experiments. For the above reasons it seems plausible to look for a new measure that determines whether a model can fulfill a given role adequately. Such fitness measure will help to determine when and through what type of calibration the resulting model is fit enough.

In this chapter we implement a method to determine how much confidence one should have in using a calibrated model for a specific task, or in a specific simulation scenario. If the calibrated model is proven sufficiently reliable for the task it is deemed fit for this task. By defining tasks of increasing complexity, a model's fitness level can be defined by the highest task it can perform. In this manner one can relate the determined fitness level of a calibrated model to the necessary resolution of outcomes for the calibration (calibration criterion), number of calibration parameters (calibration set), and the fidelity of the underlying model that is used. More methodological background of this method is described in section 3.2.

In the scope of this thesis, a full-fledged fitness framework and applicable measures could not be developed. Instead a case-specific exercise is shown as the first step towards that. In this exercise we test our thinking on one particular decision setting. This setting is the energy benchmarking of existing buildings, and a fitness measure is introduced that can effectively test the reliability of a calibrated model to perform the benchmarking at a requested level of reliability.

9.1 Proposed decision scenario: energy benchmarking of existing buildings

This section applies the fitness level concept to help determine the reliability of a model for energy assessment and thereby provide input to the decision to determine whether the building meets a specified energy benchmark or not. This scenario is for instance relevant when cities issue regulations that mandate that all existing buildings must meet certain energy consumption limits, or pay a penalty or additional tax. As the regulation focuses on benchmarking the energy efficiency of the building for standard use and given climate, consumption records of previous years cannot be used for that purpose. Instead it forces the use of an energy model that predicts energy consumption in a given (mandated, hypothetical) usage scenarios and climate. The first step is obviously to develop an energy model that represents the actual building, and calibration is a crucial step in this.

For the benchmarking scenario, we suggest that a model is "fitter" if the application of a benchmarking test (i.e., resulting in a pass/fail verdict) gives a certain guarantee that the actual building will have the same outcome. To apply this to our range of calibrated models means that the energy outcome of every calibration case will be tested against the benchmark value specified in the regulation. If the value is lower, the building passes. It should be remembered that the outcome of a model is a probability distribution of the value of the chosen QoI (in this case total yearly energy consumption). Therefore, the test should have a tolerance factor, which implies that we calculate the probability that the building passes. One could for instance require that this probability is at least 70%, or in the probabilistic form:

$$P(E_M < E_{bench}) \geq 0.70 \quad (9.1)$$

Where

E_M: the total predicted energy consumption (distribution) of the calibrated model [kWh/m^2/yr];

E_{bench}: the benchmarking value [kWh/m^2/yr]; we assume that E_{bench} can take any value within a given range depending on policy making.

As depicted in Figure 9.1, every model that is calibrated (following the procedures explained in this thesis) has an outcome distribution that ranges based on the fidelity of the model and the resolution of the calibrated parameters. If the model outcome has a 70% probability of being below the specified benchmarking value (this value can depend on the decision-maker), we consider that building to pass, otherwise it fails. Now since the E_{bench} can be built upon a collection of peer buildings in the given city; this value can be a stochastic variable that ranges based on the size and variation of those buildings. Moreover, the value is ultimately up to the energy saving objectives of the city's policy makers. To reflect on this, a uniformly distributed benchmarking value is chosen and a calibrated model is tested against every E_{bench}. For every model and E_{bench}, this produces a list of "pass" and "fail" results, which we denote 1 for pass and 0 for fail.

To measure the fitness of a model, we would have to construct a representative set of actual buildings for which we have established (e.g., through controlled experiments and/or submetering) the true pass or fail score. Every building in this actual set is then calibrated following the same calibration process presented in this thesis to produce a set of distributions for the model's outcome. The distribution of every building in the set is

used as a reference to measure the fitness within a specified probability range, which is defined by the number of matches and mismatches over the collection of buildings with our test model. The percentage of matches of a calibrated model can now serve as a quantified fitness level. The total number of matches of every model is then ranked to determine which model is fitter, and the model that has the highest number of matches is the fittest.

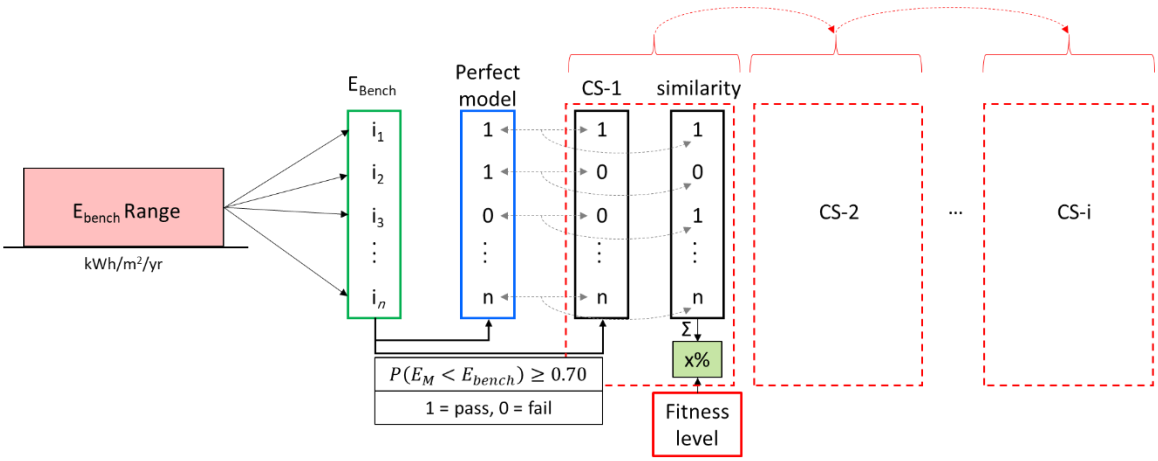


Figure 9.1 the process to establish the fitness level for a benchmarking scenario

Now, the practical limitation of this whole process lies in the difficulty of collecting all information of the set of real buildings and then build an energy model for each of one; this could be an extra effort to be considered in the future. Besides, a more in-depth assessment needs to be considered when determining which building is eligible to be deemed pass among the building stock; this adds additional complexity. To avoid this, the above approach is substituted by a simple exercise that generates a surrogate fitness level for benchmarking. In this task we use only one building (i.e., case study *two*) while assuming the true benchmarking outcome for that building is known, which is based on the outcome that is obtained by the best calibrated model. Therefore, the model with the best

calibration results (EnergyPlus-(monthly data)/CS-4/CC-1) is deemed the best model for this exercise.

9.2 Case study

This section revisits the second case study from CHAPTER 7 to apply the proposed fitness level approach. The monthly calibrated models are used to perform the fitness level determination.

As the first step, we create a range of E_{bench} values that follow a uniform distribution and ranging between 41 kWh/m^2/yr and 50 kWh/m^2/yr. They are chosen to fall within the range of the case study building used in case study *two*. This range is constrained (with 15% tolerance) by the end tails of the best model (i.e., EnergyPlus-(monthly data)/CS-4/CC-1); for simplicity we call this model "M_{fit}". Next, we determine the probability of the energy consumption with respect to the sampled E_{bench} for every model. For this process we select only ten samples for E_{bench}, and since the spread of the specified limits is relatively small, we merely list the values of E_{bench} in descending order. Following Eq. (9.1) we establish the benchmarking decisions between pass and fail results (i.e., 1 and 0) for all models, including M_{fit}. Finally, we obtain the fitness identifier by comparing the outcome of every model at a specific E_{bench} with M_{fit}. Table 9.1 and Table 9.2 present the results of this process for EPC and EnergyPlus, respectively.

Table 9.1 the fitness level outcomes for EPC models

E_{bench}	M_{fit}		EPC											
			CS-1			CS-2			CS-3			CS-4		
	$P(E_M < E_{bench})$	Pass/Fail	$P(E_M < E_{bench})$	Pass/Fail	Matches w/ M_{fit}	$P(E_M < E_{bench})$	Pass/Fail	Matches w/ M_{fit}	$P(E_M < E_{bench})$	Pass/Fail	Matches w/ M_{fit}	$P(E_M < E_{bench})$	Pass/Fail	Matches w/ M_{fit}
41	0.44	0	0.17	0	1	0.05	0	1	0.23	0	1	0.21	0	1
42	0.96	1	0.19	0	0	0.08	0	0	0.31	0	0	0.34	0	0
43	1.00	1	0.22	0	0	0.12	0	0	0.40	0	0	0.50	0	0
44	1.00	1	0.24	0	0	0.17	0	0	0.49	0	0	0.65	0	0
45	1.00	1	0.27	0	0	0.23	0	0	0.59	0	0	0.78	1	1
46	1.00	1	0.30	0	0	0.31	0	0	0.68	0	0	0.88	1	1
47	1.00	1	0.33	0	0	0.39	0	0	0.76	1	1	0.94	1	1
48	1.00	1	0.36	0	0	0.48	0	0	0.83	1	1	0.98	1	1
49	1.00	1	0.39	0	0	0.56	0	0	0.88	1	1	0.99	1	1
50	1.00	1	0.43	0	0	0.65	0	0	0.92	1	1	1.00	1	1
Total points of fitness					**1**			**1**			**5**			**7**

Table 9.2 the fitness level outcomes for EnergyPlus models

E_{bench}	M_{fit}		EnergyPlus											
			CS-1			CS-2			CS-3			CS-4		
	$P(E_M < E_{bench})$	Pass/Fail	$P(E_M < E_{bench})$	Pass/Fail	Matches w/ M_{fit}	$P(E_M < E_{bench})$	Pass/Fail	Matches w/ M_{fit}	$P(E_M < E_{bench})$	Pass/Fail	Matches w/ M_{fit}	$P(E_M < E_{bench})$	Pass/Fail	Matches w/ M_{fit}
41	0.44	0	0.00	0	1	0.00	0	1	0.00	0	1	0.44	0	1
42	0.96	1	0.00	0	0	0.00	0	0	0.00	0	0	0.96	1	1
43	1.00	1	0.00	0	0	0.01	0	0	0.15	0	0	1.00	1	1
44	1.00	1	0.00	0	0	0.43	0	0	0.80	1	1	1.00	1	1
45	1.00	1	0.01	0	0	0.97	1	1	1.00	1	1	1.00	1	1
46	1.00	1	0.02	0	0	1.00	1	1	1.00	1	1	1.00	1	1
47	1.00	1	0.05	0	0	1.00	1	1	1.00	1	1	1.00	1	1
48	1.00	1	0.09	0	0	1.00	1	1	1.00	1	1	1.00	1	1
49	1.00	1	0.15	0	0	1.00	1	1	1.00	1	1	1.00	1	1
50	1.00	1	0.23	0	0	1.00	1	1	1.00	1	1	1.00	1	1
Total points of fitness					**1**			**7**			**8**			**10**

With a 70% confidence level, the results from the two tables show that CS-2 and CS-3 in EnergyPlus model and CS-4 in EPC are the ones that have the highest value of similarities with M_{fit} with scores of 7 and 8 out of 10. This indicates that these cases are the fittest model over all competitors. Note that CS-4 in Table 9.2 is the case that is deemed the best model, which by definition is the same as M_{fit}.

Accordingly, if we take into consideration the complexity of building the energy models using the two tools, we realize that EPC provides a comparable result with less modeling effort. Based on the above analysis the policymaker could therefore decide that indeed the EPC model is fit enough to support the new city benchmarking policy thus avoiding wasting efforts on developing EnergyPlus models for all buildings in the city. In

this regard, the fitness level approach is beneficial to inform a given decision scenario with a calibrated model that lowers computational cost and has sufficient fitness.

9.3 Conclusion and remarks

Although some level of uncertainty analysis is essential in building improvement decisions and regulatory settings where energy predictions are used, current practice does not offer an adequate way to quantify the level of confidence that can be placed in a prediction with a given model. Therefore, it is hard to judge which model is adequate enough for making decisions in a given scenario. We argue in this chapter that a model fitness measure can help in making the right choice. The proposed determination of such a quantified fitness measure is introduced and applied to a simple benchmarking pass/fail decision exercise. One of the purposes of building energy performance benchmarking is to quickly identify a far less energy-efficient building among a large group of buildings. Obviously the fitness of a model will increase if it is used only to identify heavily under or over-performing buildings. This is however not yet included in our analysis but reserved for future work. At this point, the analysis points to the fact that the fitness level approach can be beneficially used for energy performance benchmarking and support the selection of an adequate model.

In summary, there are several remarks worthy of extended discussion:

- Benchmarking is chosen as an application scenario because it is a straightforward and simple exercise to test fitness. For any other scenario an approach could be developed along similar lines

- Even though we assume that benchmarking is a simple task, we cannot generalize our findings without addressing the major obstacle, which is the need to test the fitness measure against a set of real buildings. This will require a large energy modeling effort. For now it is fair to assume that there will be little appetite of funders to support this effort.

CHAPTER 10. CLOSURE

10.1 Summary and concluding remarks

There are many instances where an energy model of an existing building is needed. For instance, simulation can be implemented based on compliance testing against building energy codes; it is also advantageous for the assessment of the saving potential in retrofitting existing buildings. In both commercial and residential sectors, the importance of minimizing the discrepancy between predicted and measured energy uses is well recognized like in PM&V applications. If underdeveloped, or more specifically "uncalibrated" models are used, the large discrepancy between prediction and actual will undermine the market's confidence in energy-efficient buildings.

Calibration techniques have encountered great improvement in recent years and are well supported by ASHRAE (ASHRAE., 2009) and (ASHRAE, 2002) guidelines, new standard initiatives ASHRAE 14, (ASHRAE, 2014), and numerous theoretical and field studies. The criteria laid down in the ASHRAE guideline are useful, but more work is needed to understand the special nature of calibration of building energy models. Such criteria stipulate ranges of admissible error in the total estimated energy consumption of a building but do not address underlying aspects such as uncertainty embedded in the models or initial guesses about the value of building a model. In addition, the current methods cannot adequately support risk-conscious decision-making because they deterministically calibrate an energy model and derive a single prediction of the desired outcome. This thesis argues that adequately predicting building energy consumption is vital to the successful deployment of building energy assessments and retrofits.

In this regard, the primary goal of this thesis is to improve the effectiveness of the calibration process such as choice of the calibration parameter set and fidelity of the model, while avoiding or at least understanding the confounding effect across the calibration parameter set. Although this obviously applies to all parameters of the building, this thesis focuses on and is de facto anchored in the crucially important façade parameters associated with infiltration. The leading parameter is the effective leakage area (ELA), which is diversified for different façades and façade elements in the BPS model. The thesis deals with the problem through: 1) building a theoretical basis for understanding the current practices in building energy model calibration and the identification of ELA (viewed in this thesis as the main calibration parameter), 2) enhancing our capability in predicting future building energy use, i.e., offer a probabilistic calibration framework that is based on repeated deterministic calibrations. The main objective of the thesis is to understand the influence of the calibration settings on the resulting calibrated model. Among these settings is the fidelity level of the model, the use of different calibration sets, the type and interval of measurement data that is used and (to some extent) the metric that is used to measure the distance between observed and model generated outcomes. With this multi-dimensional view on the calibration process, the ultimate aim is to provide in-depth information about what monitoring plan and model resolution one needs to calibrate certain parameters (e.g., ELA), such that the resulting model is fit enough to perform a given task. The resulting fitness should ideally be expressed in a quantifiable measure. Its determination is highly dependent on the calibration process, and in particular on the mentioned settings for which it is conducted.

Two case studies are provided in this thesis to demonstrate the approach. The first one is a multipurpose commercial building, and the other case study is a typical residential detached house. Both case studies were modeled with EPC, which represents the low-resolution tool, and EnergyPlus, which represents the high-fidelity tool. The results demonstrate the effectiveness of the proposed calibration process in approaching the true values of the calibration parameters with a focus on ELA.

In summary, this thesis demonstrates that:

1. The proposed calibration methods make improved use of data in constraining uncertainty and improving prediction of calibration parameters.
2. Predicting building energy consumption in a probabilistic way is the key enabler of improving predictions of uncertain parameters. This suggests that the fidelity of the underlying physical model is less important than many researchers seem to believe.
3. The lower resolution EPC calibrations can adequately optimize the energy model and derive the same consistent results as EnergyPlus models when they are supported by the same calibration set and data availability. On the other hand, EPC has not done as good in terms of the calibration parameter approximation (e.g., ELA).

Finally, one major long term mission of this thesis is the development of fitness measures that can be quantified for the use of a given model in a chosen decision scenario. A first attempt is reported calculating a fitness measure for a benchmarking scenario of existing buildings. It is used to test the concept of a fitness measure and an approach to quantify it. The results demonstrate that with the appropriate information we can measure

the reliability of the calibrated model for the given purpose, in this case expressed through the number of matches between the decision supported by a model and the true decision, albeit that for lack of broad data, the latter is for now substituted by a surrogate decision.

10.2 Recommendations for future study

Calibration is and will continue to be an important area of study for BPS applications in practice. Thus, the ultimate goal of the ongoing and future research is to establish a calibration process (and potentially a framework) that aids to understand, represent, and assess the effectiveness of calibration methods on building energy model fidelity and fitness. Some immediate future work would include:

1. More case studies are necessary to confirm the feasibility of the calibration method across larger sets of buildings of different types and various energy measures.
2. The proposed framework is adequate to incorporate all sources of uncertainty for parameter predictions. However, the sources of uncertainty in the case studies have been limited to scenario uncertainty and parameter uncertainty e.g., wind speed, physical properties, and equipment and systems performance. To correctly evaluate the variability of the calibration parameters, we need to further quantify the full spectrum of uncertainties. Hence, it is recommended to explore the use of explicit model form uncertainty quantification (MFU) to inform the calibration procedure through improved uncertainty analysis.
3. For the tested case studies we were exposed to a very limited recorded data (i.e., energy consumption and indoor temperature), the calibration modules should be

extended to calibrate models with various types and spot-monitored (potentially sparse) outcomes.

4. The calibration process is automated to run 50 samples resulting in 50 values for ELA in every step. However, this number of samples is moderately small as it might not be enough to capture all aspects of the parameter's characteristics. Increasing the number of simulations by taking advantage of powerful computation is highly recommended.
5. The proposed fitness level methodology makes a limited attempt to systematically address this issue because of time and resource constraints. It requires extended work to collect and build calibratable energy models for a large set of test buildings for the benchmarking scenario.
6. The proposed fitness measure needs to be developed for a diverse set of scenarios. Among them, good candidates are:
    a. Building diagnostics and fault detection;
    b. Determination of an optimal retrofit or upgrade package (e.g., while meeting a given IRR or RoI constraint);
    c. Guaranteeing that a minimum required energy performance will be met by the building (e.g., as part of a performance contract, where getting it wrong may lead to contractually agreed penalties).
7. The focus on CVRMSE for the discrepancy between simulated and measured values may be a limiting factor. Hence it is recommended to explore other calibration metrics to quantify the dissimilarity based on both proximity of the values and similarity of the dynamic behavior of the model outcome. We expect

that with the appropriate settings, such measures have the potential to deliver an informative calibrated model without ignoring the interdependence relationship between observed and predicted data.

APPENDIX A. WIND PRESSURE COEFFICIENT: TABLES AND GRAPHS

Table A.1 Uncertainty quantification data for CpGen (free field)

		CpGen				Database				Difference			
		Density = Free Field				Density = Free Field							
	Angle	S	E	N	W	S	E	N	W	S	E	N	W
CpGen VS Database (free field)	**0**	-0.22	-0.41	0.45	-0.41	-0.28	-0.33	0.56	-0.33	0.06	-0.08	-0.11	-0.07
	15	-0.22	-0.21	0.43	-0.38	-0.30	-0.08	0.53	-0.39	0.08	-0.13	-0.09	0.00
	30	-0.22	0.02	0.37	-0.32	-0.36	0.15	0.42	-0.43	0.14	-0.13	-0.06	0.11
	45	-0.27	0.23	0.21	-0.25	-0.40	0.36	0.25	-0.46	0.13	-0.13	-0.05	0.21
	60	-0.38	0.37	-0.04	-0.21	-0.48	0.50	0.07	-0.36	0.10	-0.12	-0.10	0.14
	75	-0.48	0.44	-0.30	-0.22	-0.53	0.56	-0.18	-0.28	0.05	-0.12	-0.12	0.06
	90	-0.51	0.45	-0.51	-0.23	-0.41	0.62	-0.43	-0.24	-0.11	-0.17	-0.09	0.02
Mean -0.080													
St Deviation 0.108													

Table A.2 Uncertainty quantification data for Swami&Chandra (free field)

		Swami				Database				Difference			
						Density = Free Field							
	Angle	S	E	N	W	S	E	N	W	S	E	N	W
Swami VS Database (free field)	**0**	-0.37	-0.44	0.60	-0.44	-0.28	-0.33	0.56	-0.33	0.09	0.11	-0.04	0.11
	15	-0.38	-0.14	0.53	-0.56	-0.30	-0.08	0.53	-0.39	0.08	0.06	0.00	0.17
	30	-0.39	0.17	0.46	-0.67	-0.36	0.15	0.42	-0.43	0.03	-0.02	-0.04	0.24
	45	-0.53	0.32	0.32	-0.53	-0.40	0.36	0.25	-0.46	0.13	0.04	-0.06	0.07
	60	-0.67	0.46	0.17	-0.39	-0.48	0.50	0.07	-0.36	0.19	0.04	-0.10	0.03
	75	-0.56	0.53	-0.14	-0.38	-0.53	0.56	-0.18	-0.28	0.03	0.03	-0.04	0.10
	90	-0.44	0.60	-0.44	-0.37	-0.41	0.62	-0.43	-0.24	0.03	0.02	0.01	0.13
Mean -0.080													
St Deviation 0.079													

Table A.3 Uncertainty quantification data for Swami&Chandra (CA = 0.10)

		Swami				Database Density = Ca10				Difference			
	Angle	S	E	N	W	S	E	N	W	S	E	N	W
Swami VS Database (non-isolated)	**0**	-0.51	-0.53	0.60	-0.53	-0.18	-0.24	0.30	-0.29	-0.33	-0.29	0.31	-0.24
	25	-0.51	0.02	0.50	-0.93	-0.23	0.12	0.22	-0.43	-0.28	-0.10	0.29	-0.50
	45	-0.73	0.32	0.32	-0.73	-0.54	0.15	0.31	-0.47	-0.19	0.17	0.01	-0.26
	70	-0.91	0.53	-0.08	-0.48	-0.48	0.17	-0.08	-0.29	-0.42	0.36	0.01	-0.19
	90	-0.53	0.60	-0.53	-0.51	-0.44	0.52	-0.56	-0.29	-0.09	0.08	0.03	-0.22
	115	0.02	0.50	-0.93	-0.51	-0.03	0.30	-0.44	-0.26	0.05	0.20	-0.49	-0.25
	135	0.32	0.32	-0.73	-0.73	0.11	0.29	-0.38	-0.54	0.21	0.03	-0.35	-0.19
	160	0.53	-0.08	-0.48	-0.91	0.35	0.04	-0.26	-0.44	0.18	-0.12	-0.22	-0.46
	180	0.60	-0.53	-0.51	-0.53	0.19	-0.15	-0.18	-0.07	0.42	-0.38	-0.33	-0.46
	205	0.50	-0.93	-0.51	0.02	0.35	-0.47	-0.28	0.11	0.15	-0.45	-0.23	-0.09
	225	0.32	-0.73	-0.73	0.32	0.27	-0.53	-0.35	0.13	0.05	-0.20	-0.38	0.19
	250	-0.08	-0.48	-0.91	0.53	-0.01	-0.31	-0.51	0.26	-0.06	-0.16	-0.40	0.28
	270	-0.53	-0.51	-0.53	0.60	-0.53	-0.29	-0.54	0.57	0.00	-0.22	0.02	0.03
	295	-0.93	-0.51	0.02	0.50	-0.50	-0.25	-0.16	0.27	-0.42	-0.26	0.18	0.23
	315	-0.73	-0.73	0.32	0.32	-0.43	-0.42	0.20	0.22	-0.31	-0.32	0.12	0.10
	340	-0.48	-0.91	0.53	-0.08	-0.27	-0.48	0.21	0.10	-0.21	-0.43	0.33	-0.18
Mean -0.123													
St Deviation 0.248													

Table A.4 Uncertainty quantification data for Swami&Chandra (CA = 0.15)

		Swami				Database Density = Ca15				Difference			
	Angle	S	E	N	W	S	E	N	W	S	E	N	W
Swami VS Database (non-isolated)	**0**	-0.51	-0.53	0.60	-0.53	-0.18	-0.39	0.45	-0.29	-0.32	-0.14	0.16	-0.24
	25	-0.51	0.02	0.50	-0.93	-0.17	0.01	0.10	-0.23	-0.34	0.01	0.40	-0.70
	45	-0.73	0.32	0.32	-0.73	-0.32	0.17	0.09	-0.44	-0.41	0.14	0.23	-0.29
	70	-0.91	0.53	-0.08	-0.48	-0.31	0.26	0.03	-0.45	-0.59	0.27	-0.11	-0.03
	90	-0.53	0.60	-0.53	-0.51	-0.13	0.23	-0.20	-0.16	-0.40	0.38	-0.33	-0.35
	115	0.02	0.50	-0.93	-0.51	-0.05	0.22	-0.36	-0.34	0.07	0.28	-0.57	-0.17
	135	0.32	0.32	-0.73	-0.73	0.01	0.02	-0.37	-0.48	0.31	0.30	-0.36	-0.25
	160	0.53	-0.08	-0.48	-0.91	0.31	-0.10	-0.37	-0.55	0.23	0.02	-0.11	-0.35
	180	0.60	-0.53	-0.51	-0.53	0.08	-0.24	-0.32	-0.13	0.53	-0.29	-0.19	-0.40
	205	0.50	-0.93	-0.51	0.02	0.37	-0.57	-0.32	-0.15	0.13	-0.36	-0.19	0.17
	225	0.32	-0.73	-0.73	0.32	0.00	-0.39	-0.28	0.09	0.32	-0.34	-0.45	0.23
	250	-0.08	-0.48	-0.91	0.53	-0.07	-0.29	-0.44	0.37	-0.01	-0.19	-0.46	0.16
	270	-0.53	-0.51	-0.53	0.60	-0.17	-0.24	-0.32	0.36	-0.36	-0.27	-0.21	0.24
	295	-0.93	-0.51	0.02	0.50	-0.29	-0.32	0.09	0.13	-0.63	-0.19	-0.07	0.38
	315	-0.73	-0.73	0.32	0.32	-0.34	-0.49	0.11	0.15	-0.39	-0.24	0.20	0.17
	340	-0.48	-0.91	0.53	-0.08	-0.23	-0.40	0.26	-0.10	-0.25	-0.51	0.28	0.02
Mean -0.126													
St Deviation 0.301													

Table A.5 Uncertainty quantification data for Swami&Chandra (CA = 0.30)

		Swami				Database (Density = Ca30)				Difference			
	Angle	S	E	N	W	S	E	N	W	S	E	N	W
Swami VS Database (non-isolated)	**0**	-0.51	-0.53	0.60	-0.53	-0.34	-0.19	-0.09	-0.15	-0.17	-0.34	0.69	-0.38
	25	-0.51	0.02	0.50	-0.93	-0.34	-0.01	-0.05	-0.16	-0.17	0.02	0.55	-0.77
	45	-0.73	0.32	0.32	-0.73	-0.32	-0.12	-0.06	-0.23	-0.41	0.44	0.38	-0.50
	70	-0.91	0.53	-0.08	-0.48	-0.26	-0.04	-0.13	-0.22	-0.64	0.58	0.05	-0.26
	90	-0.53	0.60	-0.53	-0.51	-0.12	-0.05	-0.19	-0.20	-0.41	0.66	-0.34	-0.31
	115	0.02	0.50	-0.93	-0.51	-0.07	-0.09	-0.23	-0.21	0.08	0.60	-0.70	-0.29
	135	0.32	0.32	-0.73	-0.73	0.03	-0.17	-0.28	-0.24	0.29	0.49	-0.45	-0.49
	160	0.53	-0.08	-0.48	-0.91	0.16	-0.22	-0.27	-0.21	0.38	0.14	-0.20	-0.69
	180	0.60	-0.53	-0.51	-0.53	0.11	-0.25	-0.13	-0.18	0.49	-0.28	-0.38	-0.35
	205	0.50	-0.93	-0.51	0.02	0.07	-0.23	-0.10	-0.14	0.43	-0.70	-0.41	0.16
	225	0.32	-0.73	-0.73	0.32	-0.02	-0.25	-0.18	-0.12	0.33	-0.48	-0.55	0.44
	250	-0.08	-0.48	-0.91	0.53	-0.01	-0.27	-0.17	-0.12	-0.07	-0.21	-0.74	0.66
	270	-0.53	-0.51	-0.53	0.60	-0.18	-0.16	-0.10	-0.07	-0.35	-0.35	-0.43	0.68
	295	-0.93	-0.51	0.02	0.50	-0.30	-0.17	-0.14	-0.03	-0.62	-0.34	0.16	0.53
	315	-0.73	-0.73	0.32	0.32	-0.30	-0.19	-0.07	-0.13	-0.43	-0.54	0.39	0.45
	340	-0.48	-0.91	0.53	-0.08	-0.32	-0.25	-0.04	-0.05	-0.16	-0.66	0.57	-0.03
Mean -0.149													
St Deviation 0.449													

Table A.6 Uncertainty quantification data for Swami&Chandra (CA = 0.50)

		Swami				Database (Density = Ca50)				Difference			
	Angle	S	E	N	W	S	E	N	W	S	E	N	W
Swami VS Database (non-isolated)	**0**	-0.51	-0.53	0.60	-0.53	-0.28	-0.26	0.01	-0.28	-0.23	-0.27	0.59	-0.25
	25	-0.51	0.02	0.50	-0.93	-0.28	-0.21	0.04	-0.31	-0.23	0.23	0.46	-0.62
	45	-0.73	0.32	0.32	-0.73	-0.26	-0.16	0.04	-0.27	-0.47	0.48	0.28	-0.46
	70	-0.91	0.53	-0.08	-0.48	-0.23	-0.13	0.09	-0.25	-0.68	0.66	-0.16	-0.23
	90	-0.53	0.60	-0.53	-0.51	-0.21	-0.14	-0.07	-0.20	-0.32	0.74	-0.46	-0.31
	115	0.02	0.50	-0.93	-0.51	-0.18	-0.06	-0.23	-0.20	0.19	0.57	-0.69	-0.31
	135	0.32	0.32	-0.73	-0.73	-0.11	-0.12	-0.24	-0.20	0.43	0.44	-0.49	-0.53
	160	0.53	-0.08	-0.48	-0.91	-0.08	-0.20	-0.25	-0.21	0.62	0.12	-0.23	-0.70
	180	0.60	-0.53	-0.51	-0.53	0.02	-0.25	-0.25	-0.16	0.59	-0.28	-0.26	-0.37
	205	0.50	-0.93	-0.51	0.02	0.14	-0.21	-0.17	-0.13	0.36	-0.72	-0.34	0.15
	225	0.32	-0.73	-0.73	0.32	0.04	-0.22	-0.19	-0.14	0.27	-0.51	-0.55	0.46
	250	-0.08	-0.48	-0.91	0.53	-0.10	-0.22	-0.18	-0.12	0.02	-0.26	-0.72	0.65
	270	-0.53	-0.51	-0.53	0.60	-0.14	-0.24	-0.16	-0.12	-0.39	-0.27	-0.37	0.72
	295	-0.93	-0.51	0.02	0.50	-0.24	-0.29	-0.17	-0.17	-0.68	-0.22	0.19	0.67
	315	-0.73	-0.73	0.32	0.32	-0.21	-0.28	-0.04	-0.23	-0.53	-0.45	0.36	0.54
	340	-0.48	-0.91	0.53	-0.08	-0.27	-0.30	0.01	-0.26	-0.21	-0.61	0.53	0.18
Mean -0.166													
St Deviation 0.459													

Table A.7 Uncertainty quantification data for CpGen (CA = 0.10)

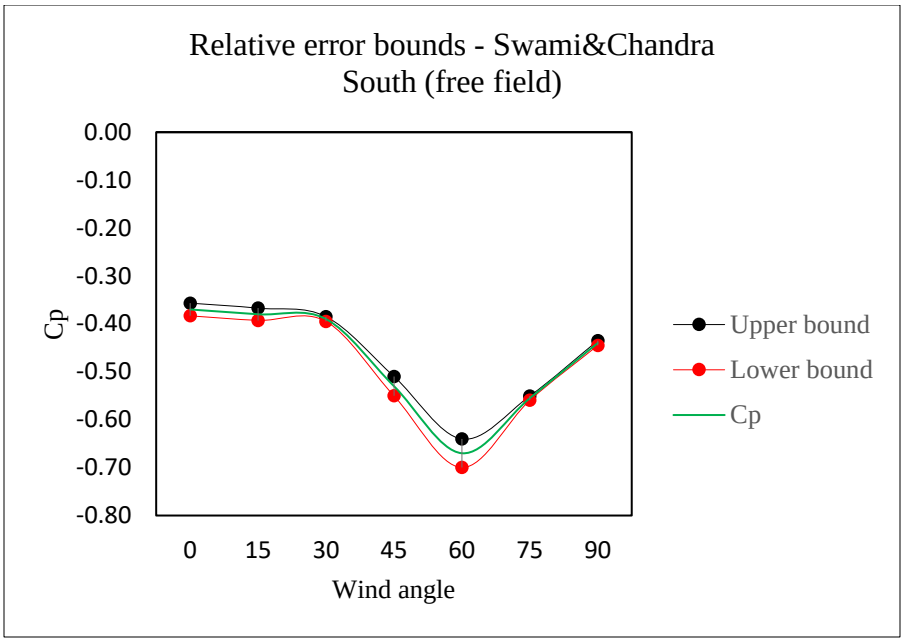
		CpGen				Database				Difference			
		Density = Ca10				Density = Ca10				S	E	N	W
	Angle	S	E	N	W	S	E	N	W				
CpGen VS Database (non-isolated)	**0**	-0.38	-0.77	0.68	-0.70	-0.18	-0.24	0.30	-0.29	-0.19	-0.52	0.38	-0.41
	25	-0.36	-0.11	0.60	-0.60	-0.23	0.12	0.22	-0.43	-0.14	-0.23	0.39	-0.18
	45	-0.50	0.49	0.39	-0.42	-0.54	0.15	0.31	-0.47	0.04	0.34	0.08	0.05
	70	-0.83	0.48	-0.38	-0.38	-0.48	0.17	-0.08	-0.29	-0.35	0.31	-0.30	-0.09
	90	-0.80	0.76	-0.66	-0.42	-0.44	0.52	-0.56	-0.29	-0.36	0.24	-0.09	-0.13
	115	-0.15	0.77	-0.84	-0.32	-0.03	0.30	-0.44	-0.26	-0.12	0.47	-0.40	-0.06
	135	0.37	0.12	-0.38	-0.36	0.11	0.29	-0.38	-0.54	0.26	-0.17	0.00	0.18
	160	0.73	-0.37	-0.35	-0.67	0.35	0.04	-0.26	-0.44	0.38	-0.42	-0.09	-0.22
	180	0.74	-0.83	-0.40	-0.76	0.19	-0.15	-0.18	-0.07	0.55	-0.68	-0.21	-0.69
	205	0.60	-0.61	-0.37	-0.11	0.35	-0.47	-0.28	0.11	0.25	-0.14	-0.09	-0.22
	225	0.39	-0.42	-0.48	0.46	0.27	-0.53	-0.35	0.13	0.12	0.11	-0.13	0.33
	250	-0.32	-0.40	-0.80	0.47	-0.01	-0.31	-0.51	0.26	-0.30	-0.08	-0.29	0.21
	270	-0.74	-0.40	-0.89	0.96	-0.53	-0.29	-0.54	0.57	-0.21	-0.11	-0.35	0.39
	295	-0.76	-0.37	-0.21	0.71	-0.50	-0.25	-0.16	0.27	-0.25	-0.12	-0.05	0.44
	315	-0.46	-0.42	0.35	0.35	-0.43	-0.42	0.20	0.22	-0.04	-0.01	0.16	0.13
	340	-0.36	-0.64	0.73	-0.25	-0.27	-0.48	0.21	0.10	-0.09	-0.17	0.53	-0.35
Mean -0.123													
St Deviation 0.287													

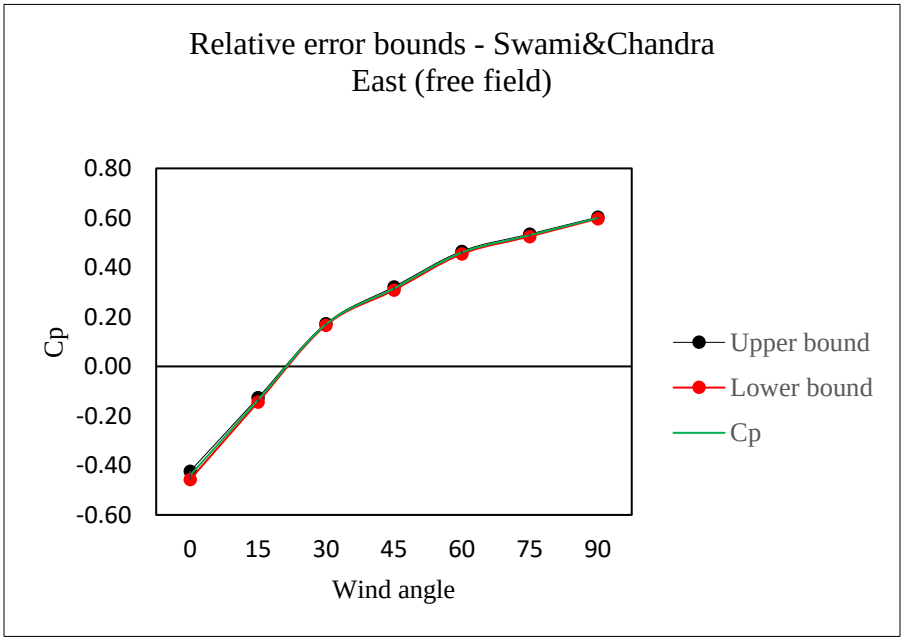
Table A.8 Uncertainty quantification data for CpGen (CA = 0.30)

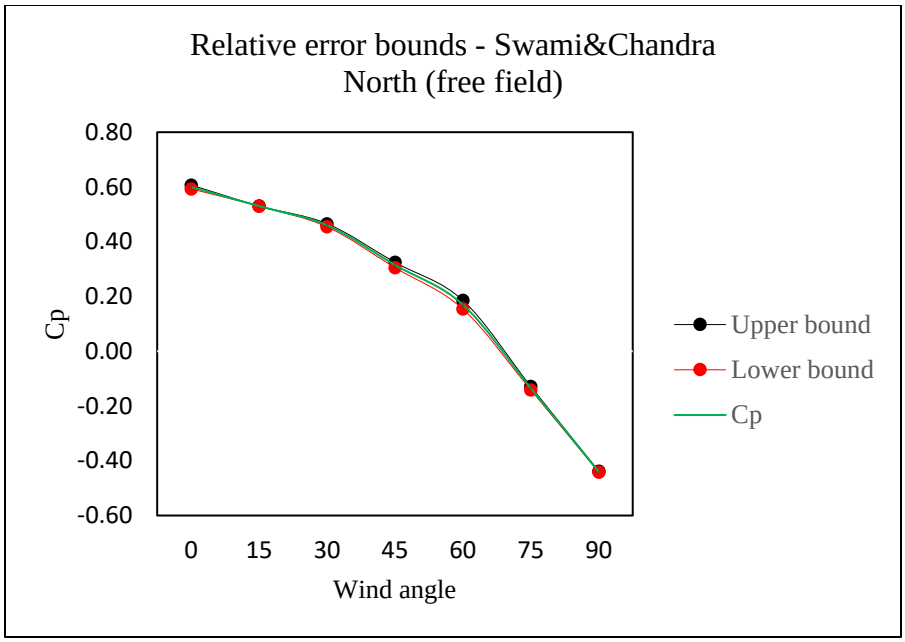
		CpGen				Database				Difference			
		Density = Ca30				Density = Ca30				S	E	N	W
	Angle	S	E	N	W	S	E	N	W				
CpGen VS Database (non-isolated)	**0**	-0.39	-0.56	0.35	-0.67	-0.34	-0.19	-0.09	-0.15	-0.04	-0.37	0.44	-0.52
	25	-0.36	-0.13	0.74	-0.59	-0.34	-0.01	-0.05	-0.16	-0.02	-0.12	0.78	-0.43
	45	-0.48	0.49	0.37	-0.43	-0.32	-0.12	-0.06	-0.23	-0.16	0.61	0.43	-0.20
	70	-0.82	0.47	-0.42	-0.38	-0.26	-0.04	-0.13	-0.22	-0.55	0.52	-0.29	-0.16
	90	-0.90	0.68	-0.66	-0.40	-0.12	-0.05	-0.19	-0.20	-0.78	0.73	-0.47	-0.19
	115	-0.21	0.78	-0.82	-0.38	-0.07	-0.09	-0.23	-0.21	-0.15	0.87	-0.60	-0.17
	135	0.37	0.28	-0.50	-0.45	0.03	-0.17	-0.28	-0.24	0.34	0.45	-0.22	-0.22
	160	0.57	-0.28	-0.38	-0.67	0.16	-0.22	-0.27	-0.21	0.41	-0.06	-0.10	-0.45
	180	0.59	-0.80	-0.38	-0.72	0.11	-0.25	-0.13	-0.18	0.47	-0.54	-0.25	-0.54
	205	0.61	-0.61	-0.36	-0.13	0.07	-0.23	-0.10	-0.14	0.54	-0.39	-0.26	0.01
	225	0.25	-0.43	-0.48	0.46	-0.02	-0.25	-0.18	-0.12	0.26	-0.19	-0.30	0.58
	250	-0.32	-0.40	-0.84	0.47	-0.01	-0.27	-0.17	-0.12	-0.31	-0.13	-0.68	0.59
	270	-0.74	-0.42	-0.83	1.17	-0.18	-0.16	-0.10	-0.07	-0.56	-0.26	-0.73	1.24
	295	-0.73	-0.29	-0.06	0.33	-0.30	-0.17	-0.14	-0.03	-0.43	-0.12	0.09	0.36
	315	-0.34	-0.44	0.41	-0.01	-0.30	-0.19	-0.07	-0.13	-0.04	-0.24	0.48	0.12
	340	-0.39	-0.64	0.87	-0.33	-0.32	-0.25	-0.04	-0.05	-0.07	-0.39	0.91	-0.28
Mean -0.149													
St Deviation 0.455													

Table A.9 Uncertainty quantification data for CpGen (CA = 0.50)

		CpGen				Database				Difference			
		Density = Ca50				Density = Ca50				S	E	N	W
	Angle	S	E	N	W	S	E	N	W				
CpGen VS Database (non-isolated)	**0**	-0.38	-0.69	0.76	-0.69	-0.28	-0.26	0.01	-0.28	-0.10	-0.43	0.74	-0.40
	25	-0.36	-0.10	0.67	-0.58	-0.28	-0.21	0.04	-0.31	-0.09	0.11	0.63	-0.28
	45	-0.46	0.40	0.35	-0.42	-0.26	-0.16	0.04	-0.27	-0.20	0.56	0.31	-0.15
	70	-0.77	0.73	-0.37	-0.37	-0.23	-0.13	0.09	-0.25	-0.55	0.86	-0.46	-0.12
	90	-0.89	0.78	-0.88	-0.39	-0.21	-0.14	-0.07	-0.20	-0.68	0.92	-0.82	-0.19
	115	-0.22	0.68	-0.71	-0.37	-0.18	-0.06	-0.23	-0.20	-0.04	0.75	-0.48	-0.17
	135	0.35	0.40	-0.46	-0.42	-0.11	-0.12	-0.24	-0.20	0.46	0.52	-0.22	-0.22
	160	0.71	-0.23	-0.36	-0.62	-0.08	-0.20	-0.25	-0.21	0.79	-0.03	-0.11	-0.41
	180	0.76	-0.69	-0.38	-0.69	0.02	-0.25	-0.25	-0.16	0.74	-0.44	-0.13	-0.52
	205	0.67	-0.58	-0.36	-0.10	0.14	-0.21	-0.17	-0.13	0.53	-0.37	-0.19	0.03
	225	0.35	-0.42	-0.46	0.40	0.04	-0.22	-0.19	-0.14	0.31	-0.20	-0.28	0.53
	250	-0.37	-0.37	-0.77	0.73	-0.10	-0.22	-0.18	-0.12	-0.27	-0.15	-0.59	0.84
	270	-0.88	-0.39	-0.88	0.78	-0.14	-0.24	-0.16	-0.12	-0.74	-0.15	-0.73	0.90
	295	-0.71	-0.37	-0.22	0.69	-0.24	-0.29	-0.17	-0.17	-0.47	-0.08	-0.05	0.85
	315	-0.46	-0.42	0.35	0.40	-0.21	-0.28	-0.04	-0.23	-0.26	-0.14	0.40	0.62
	340	-0.36	-0.62	0.71	-0.23	-0.27	-0.30	0.01	-0.26	-0.10	-0.32	0.70	0.02
Mean -0.166													
St Deviation 0.481													


Relative error bounds - Swami&Chandra
South (free field)
0.00
-0.10
-0.20
-0.30
-0.40
-0.50
-0.60
-0.70
-0.80
Cp
0
15
30
45
60
75
90
Wind angle
Upper bound
Lower bound
Cp


Relative error bounds - Swami&Chandra
East (free field)
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
Cp
0
15
30
45
60
75
90
Wind angle
Upper bound
Lower bound
Cp



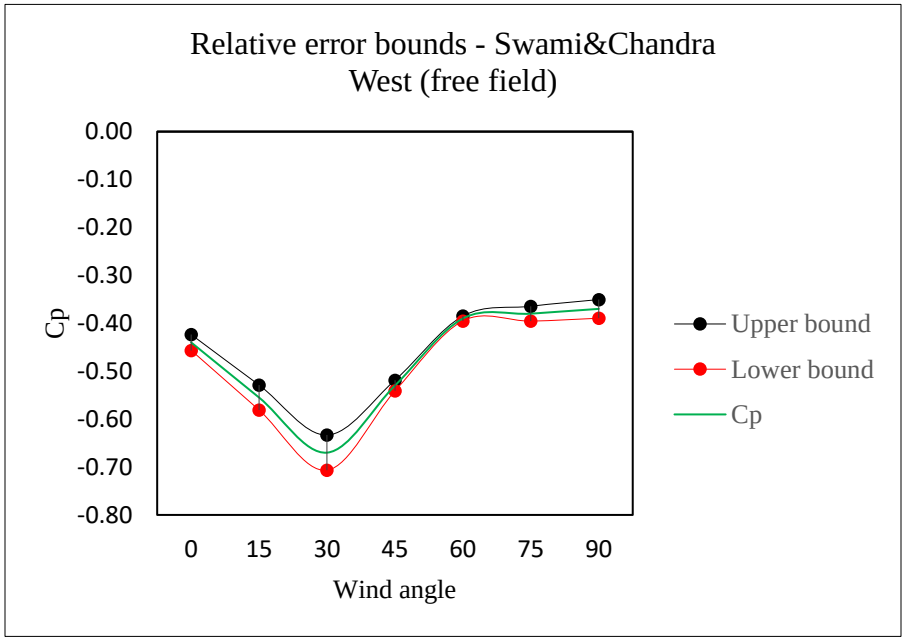
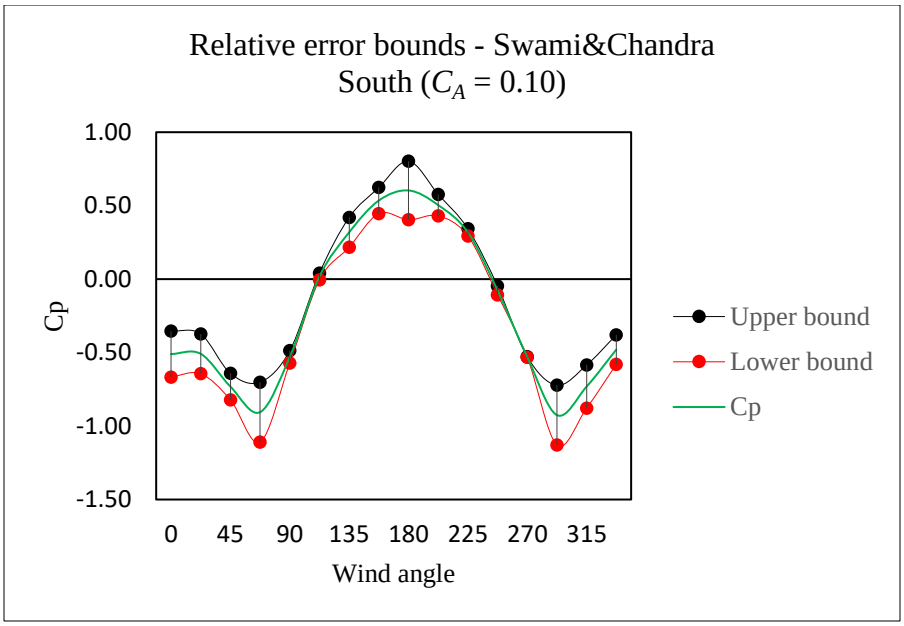
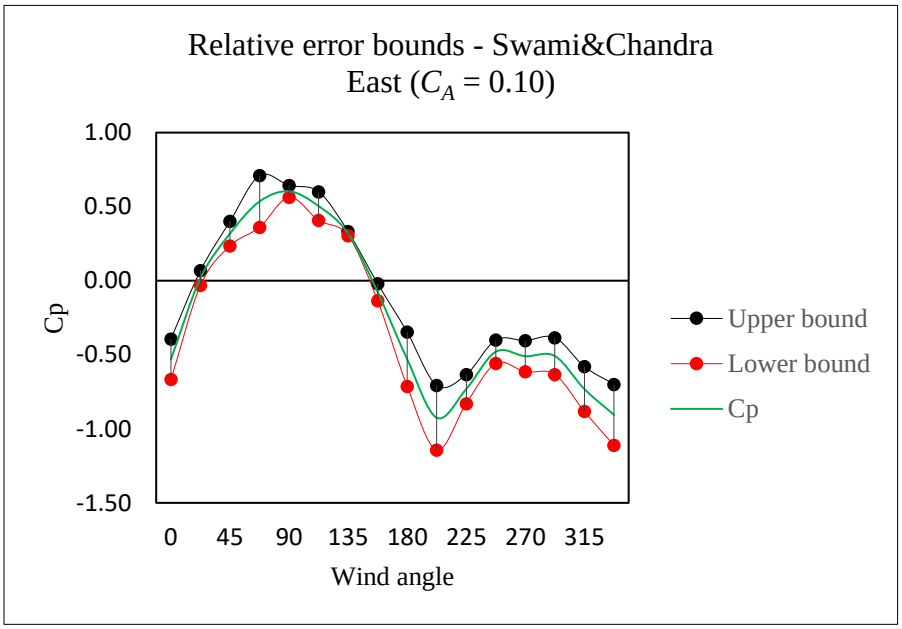


Figure A.1 Results of Cp from Swami&Chandra model with uncertainty bands under free field conditions


Relative error bounds - Swami&Chandra
South (C_A = 0.10)
1.00
0.50
0.00
-0.50
-1.00
-1.50
Cp
0
45
90
135
180
225
270
315
Wind angle
Upper bound
Lower bound
Cp


Relative error bounds - Swami&Chandra
East (C_A = 0.10)
1.00
0.50
0.00
-0.50
-1.00
-1.50
Cp
0
45
90
135
180
225
270
315
Wind angle
Upper bound
Lower bound
Cp

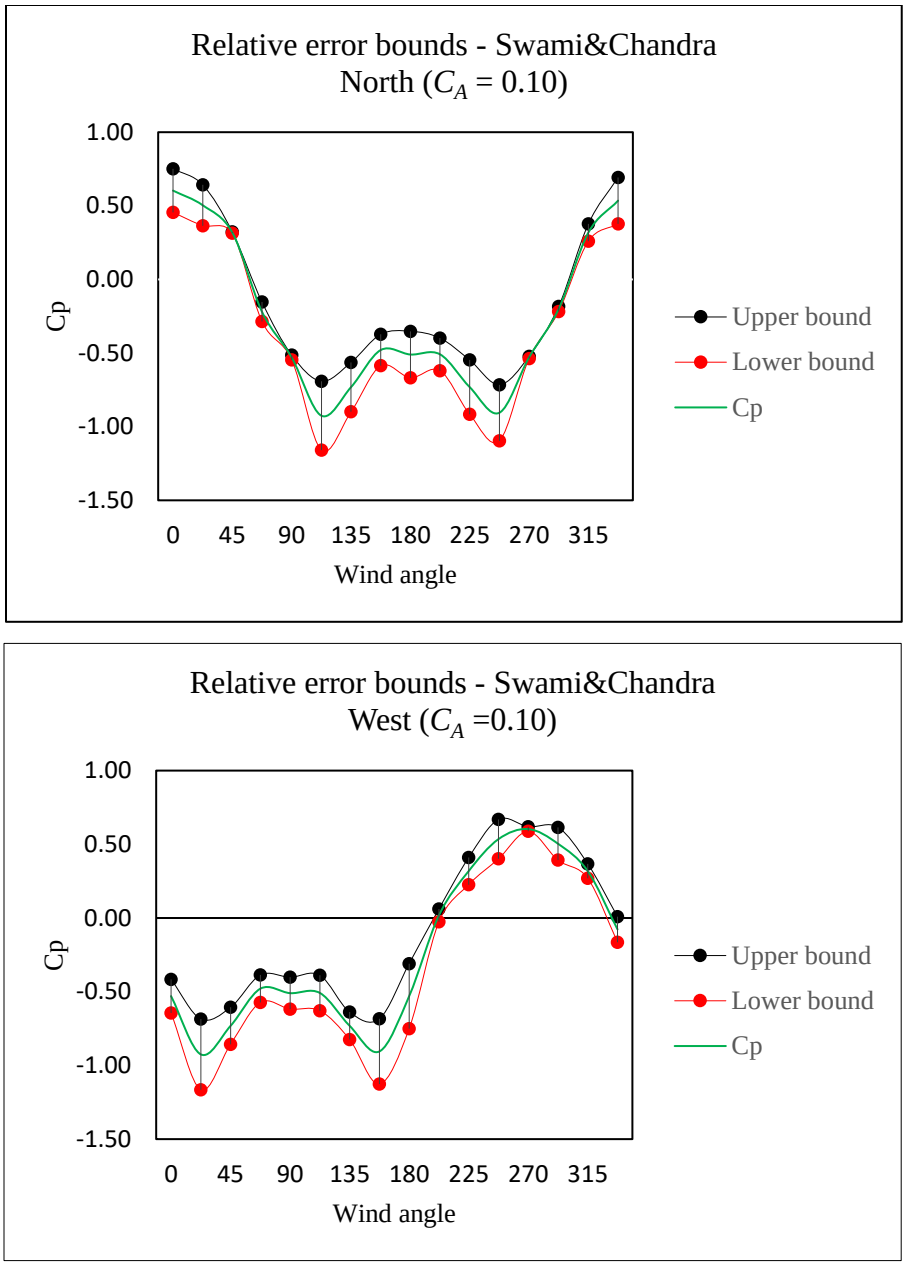
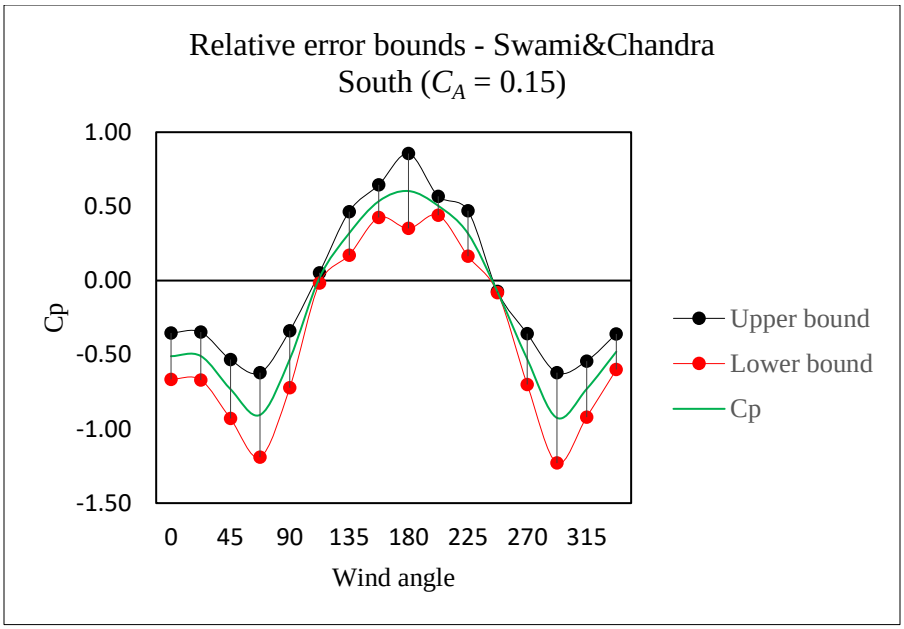
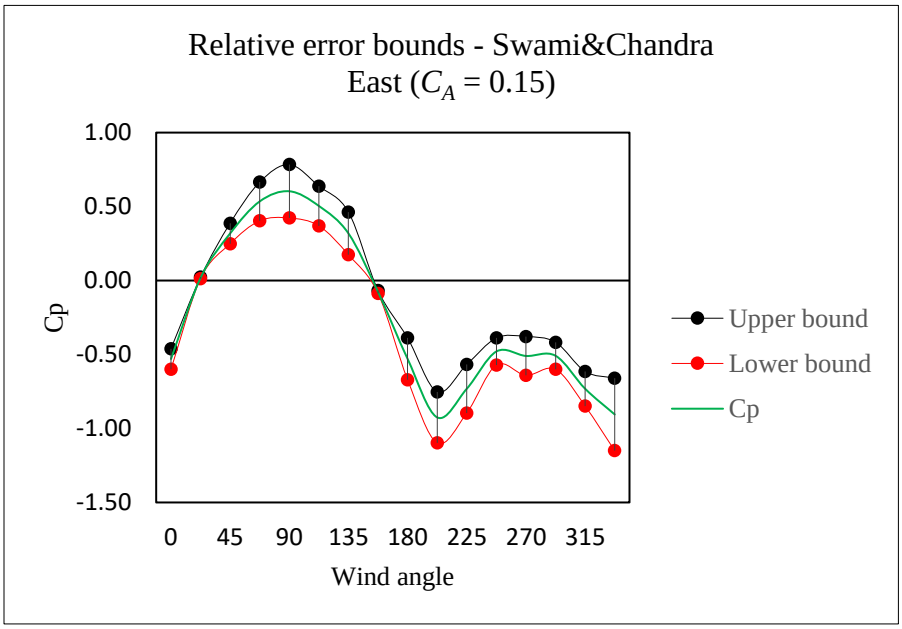


Figure A.2 Results of Cp from Swami&Chandra model with uncertainty bands under CA = 0.10


Relative error bounds - Swami&Chandra
South (C_A = 0.15)
1.00
0.50
0.00
-0.50
-1.00
-1.50
Cp
0
45
90
135
180
225
270
315
Wind angle
Upper bound
Lower bound
Cp


Relative error bounds - Swami&Chandra
East (C_A = 0.15)
1.00
0.50
0.00
-0.50
-1.00
-1.50
Cp
0
45
90
135
180
225
270
315
Wind angle
Upper bound
Lower bound
Cp

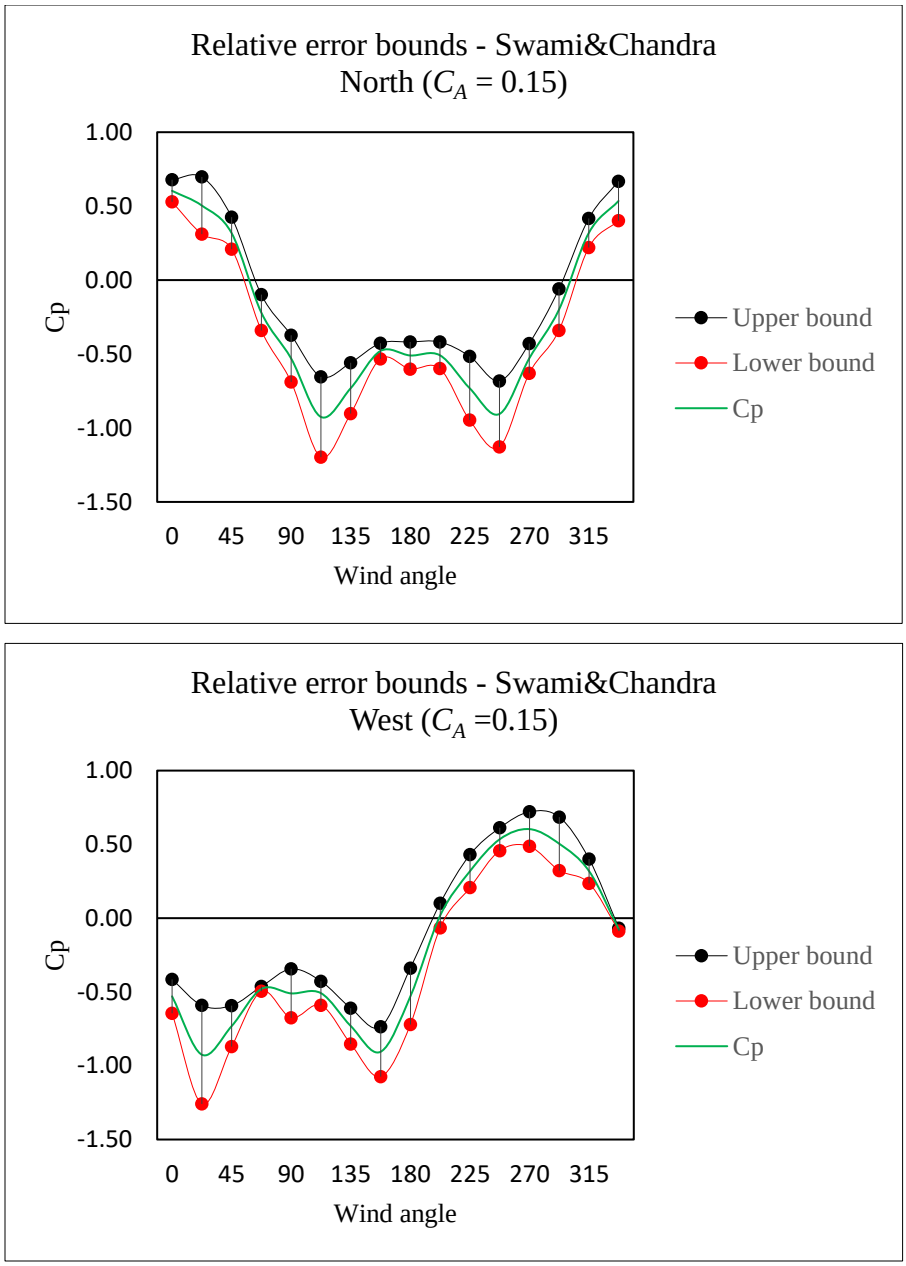
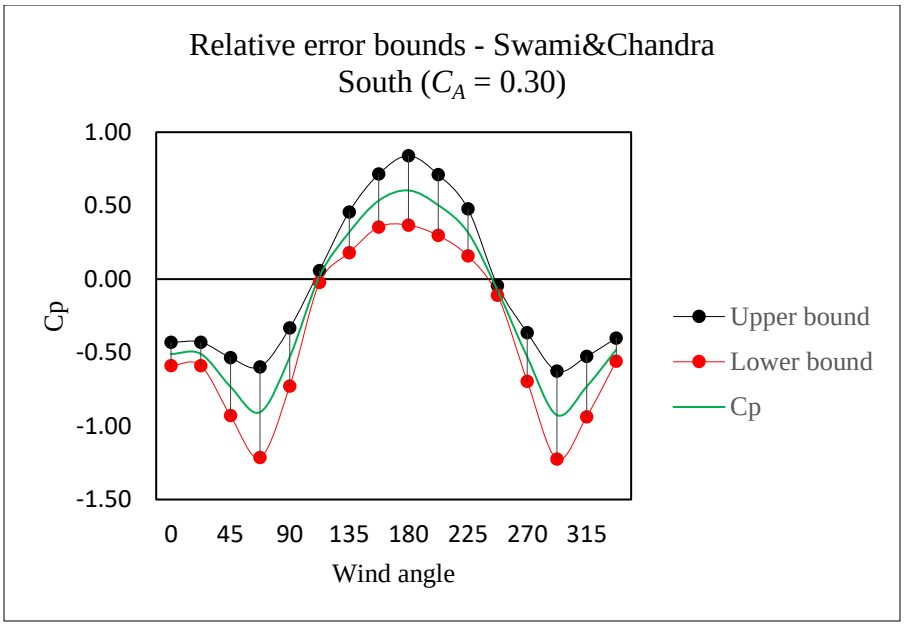
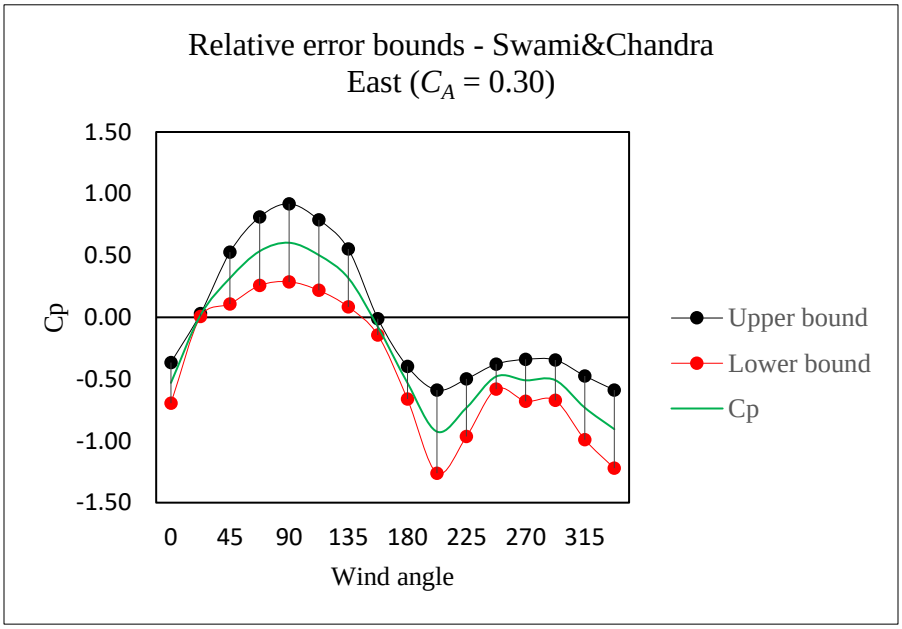
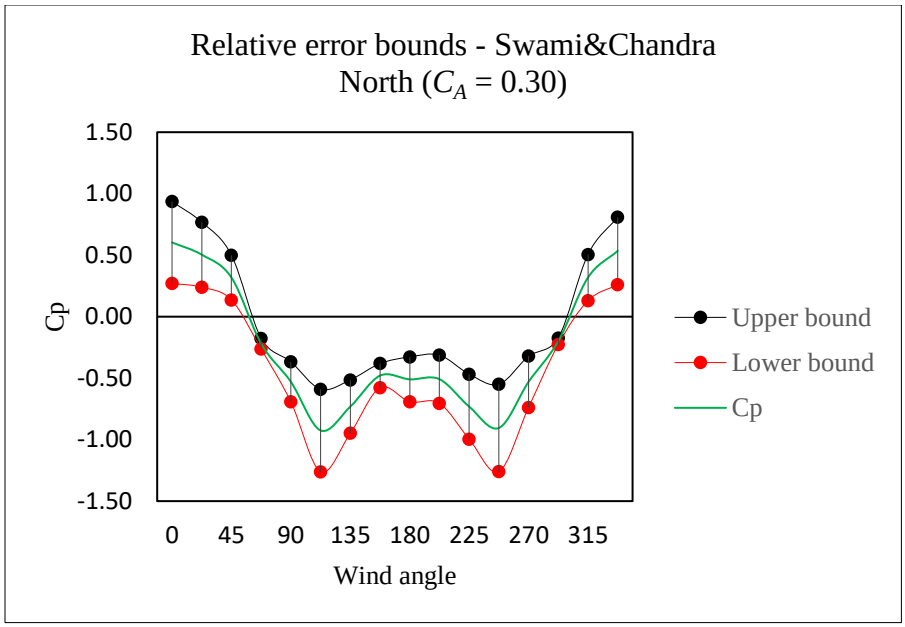


Figure A.3 Results of Cp from Swami&Chandra model with uncertainty bands under CA = 0.15


Relative error bounds - Swami&Chandra
South (C_A = 0.30)
1.00
0.50
0.00
-0.50
-1.00
-1.50
Cp
0
45
90
135
180
225
270
315
Wind angle
Upper bound
Lower bound
Cp


Relative error bounds - Swami&Chandra
East (C_A = 0.30)
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
Cp
0
45
90
135
180
225
270
315
Wind angle
Upper bound
Lower bound
Cp



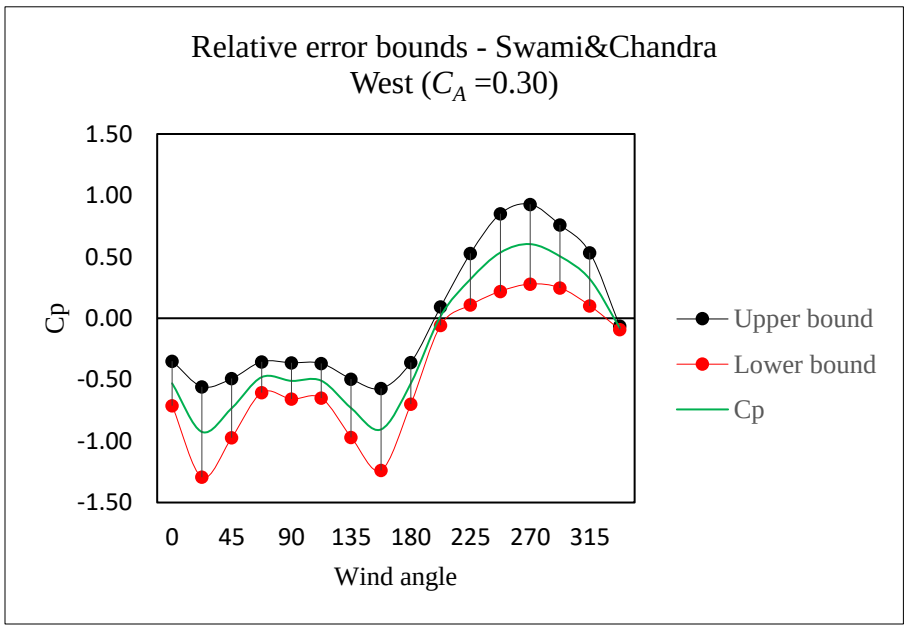
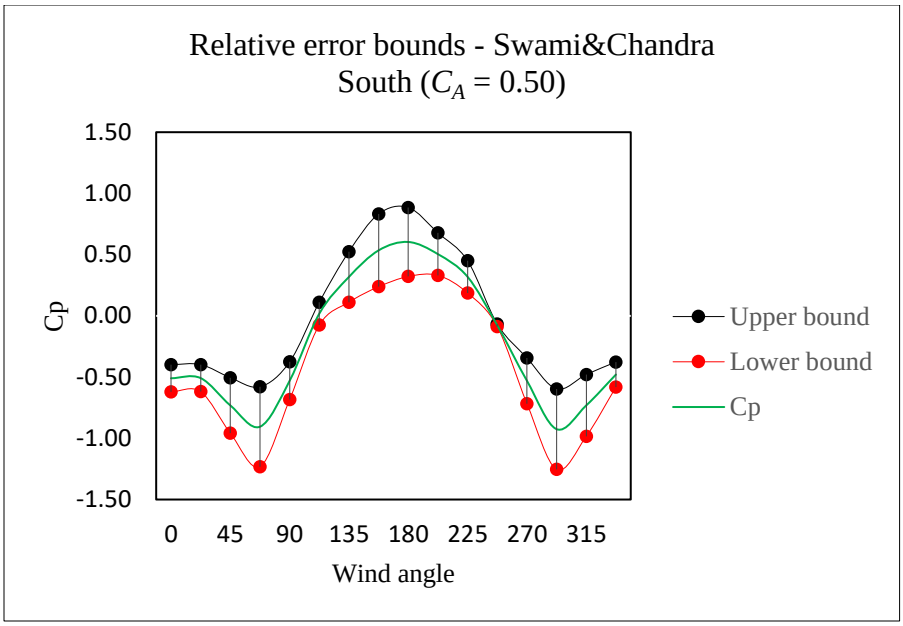
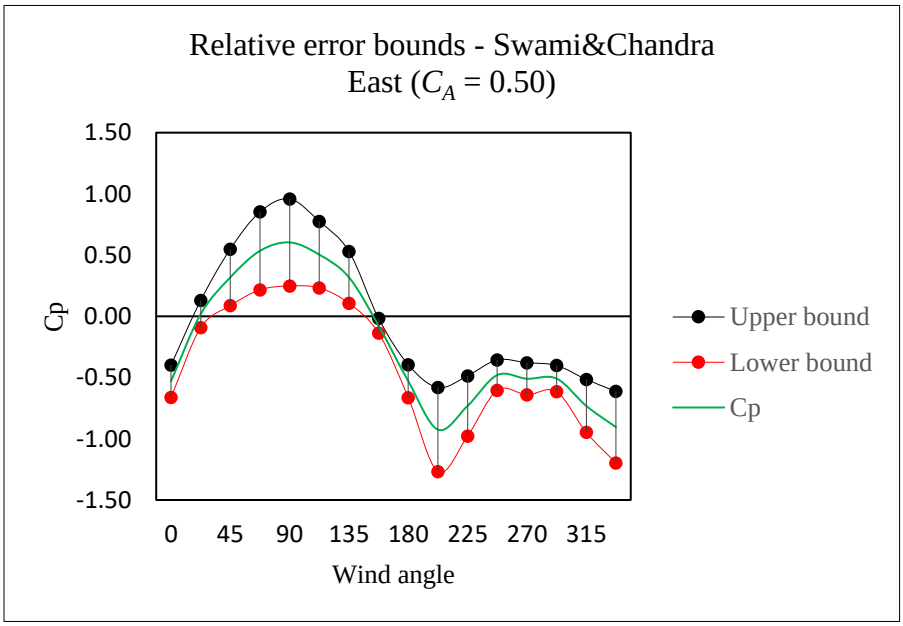
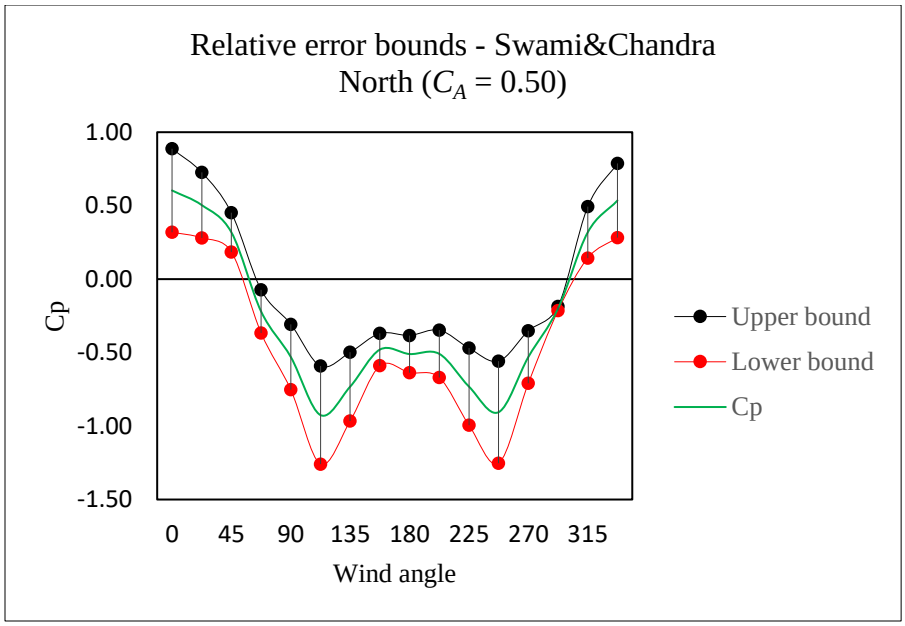


Figure A.4 Results of Cp from Swami&Chandra model with uncertainty bands under CA = 0.30


Relative error bounds - Swami&Chandra
South (C_A = 0.50)
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
Cp
0
45
90
135
180
225
270
315
Wind angle
Upper bound
Lower bound
Cp


Relative error bounds - Swami&Chandra
East (C_A = 0.50)
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
Cp
0
45
90
135
180
225
270
315
Wind angle
Upper bound
Lower bound
Cp



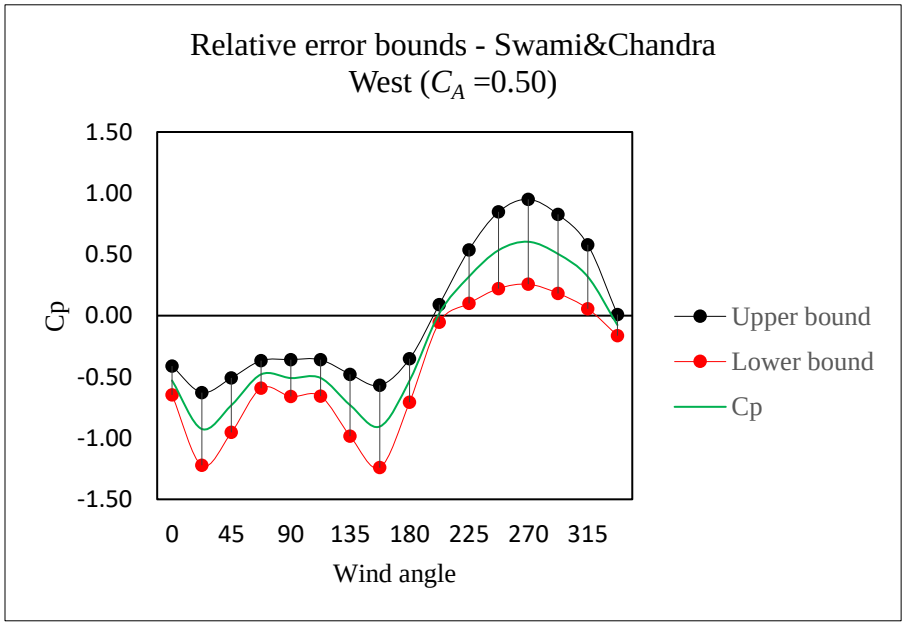


Figure A.5 Results of Cp from Swami&Chandra model with uncertainty bands under CA = 0.30

REFERENCES

Afonso, C. (2013). Tracer gas technique for measurement of air infiltration and natural ventilation: case studies and new devices for measurement of mechanical air ventilation in ducts. *nternational Journal of Low-Carbon Technologies, Volume 10*(Issue 3), 188-204.

AIVC. (1992). Air infiltration and ventilation glossary. AIC-TN-36-1992 ISBN 0 946075 58 1.

ASHRAE. (1997). 1997 ASHRAE handbook : Fundamentals. Chapter 25. ISBN 1883413443 9781883413446.

ASHRAE. (2001). *2001 ASHRAE Fundamentals Handbook.* Atlanta GA.

ASHRAE. (2002). Guideline 14-2002: Measurement of Energy and Demand Savings.

ASHRAE. (2004). *Procedures for Commercial Building Energy Audits.* ASHRAE.

ASHRAE. (2014). Guideline 14-2014: Measurement of Energy, Demand, and Water Savings.

ASHRAE. (2009). 2009 ASHRAE Handbook - Fundamentals.

ASTM, E. (2011). Standard Test Method for Determining Air Change in a Single Zone by Means of a Tracer Gas Dilution.

Augenbroe, G., Heo, Y., & Choudhary, R. (2011). RISK ANALYSIS OF ENERGY-EFFICIENCY PROJECTS BASED ON BAYESIAN CALIBRATION OF

BUILDING ENERGY MODELS. *12th Conference of International Building Performance Simulation Association*, (pp. 2579-2586). Sydney.

Bedford, T., & Cooke, R. (2001). Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines. *32*(1-4), 245–268. Retrieved from https://link.springer.com/article/10.1023/A:1016725902970

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer.

Blanchard, M., & Desrochers, G. (1984). Generation of Autocorrelated Wind Speeds for Wind Energy Conversion System Studies. *Solar Energy, 33*(6), 571-579.

Bronson, D. J., Hinchey, S. B., Habrel, J. S., & O'Neal, D. L. (1992). A procedure for calibrating the DOE-2 simulation program to non-weather-dependent measured loads. ASHRAE Transactions 98(1).

Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An Effective Screening Design For Sensitivity Analysis of Large Models. *Environmental Modelling & Software, 22*(10), 1509-1518. Retrieved from https://doi.org/10.1016/j.envsoft.2006.10.004

Carroll, W. L., & Hitchcock, R. J. (1993). DE-93-14-2 -- Tuning simulated building descriptions to match actual utility data. In A. Transactions (Ed.), *ASHRAE Transactions. 99*, pp. 429-439. Denver: ASHRAE.

Chong, A. (2018). *Datasets, Bayesian Calibration of Building Energy Models for Large Datasets.* Carnegie Mellon University. Thesis. Retrieved from https://doi.org/10.1184/R1/6714953.v1

Chong, A., & Lam, K. (2015). Uncertainty Analysis and Parameter Estimation of HVAC Systems In Building Energy Models. (pp. 2788-2795). Hyderabad, India: ibpsa. Retrieved from http://www.ibpsa.org/proceedings/BS2015/p2114.pdf

Chouakria, A., & Nagabhushan, P. (2007). Adaptive Dissimilarity Index for Measuring Time Series. *Springer*.

CONTAM. (2015). CONTAM User Manual.

Costola, D., Blocken, B., & Hensen, J. (2009). Overview of pressure coefficient data in building energy simulation and airflow network programs.

Costola, D., Blocken, B., Ohba, M., & Hensen, J. (2010). Uncertainty in airflow rate calculations due to the use of surface-averaged pressure coefficients. Energy and Buildings 42 881–888.

Crawley, D., Hand, J., Kummert, M., & Griffith, B. (2008). Contrasting the Capabilities of Building Energy Performance Simulation Programs. *Building and environment*, 231-238. Retrieved from https://strathprints.strath.ac.uk/6555/6/strathprints006555.pdf

d'Ambrosio, F., Dell'Isola, A. M., & G.Ficco, F. T. (2011). Experimental analysis of air tightness in Mediterranean buildings using the fan pressurization method. *Elsevier: Building and Environment*, 16-25.

de Wilde, P., van der Vordan, M., Augenbroe, G., & Kaan, H. (2001). The Need for Computational Support In Energy-Efficient Design Projects In The Netherlands. *Seventh International IBPSA Conference*, (pp. 513-519). Rio de Janeiro.

de Wit, M. S. (2001). *Uncertainty in Predection of Thermal Comfort in Buildings.* Delft: Delft University of Technology.

Deru, M., & Burns, P. (2003). Infiltration and Natural Ventilation Model for Whole-Building Energy Simulation of Residential Buildings. NREL is a U.S. Department of Energy LaboratoryOperated by Midwest Research Institute. DE-AC36-99-GO10337.

Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM Computing Surveys (CSUR), 27*(3). Retrieved from https://dl.acm.org/doi/10.1145/212094.212114

Djunaedy, E., Loomans, M., & Hensen, J. (2003). Toward external coupling of building energy and airflow modeling programs. *ASHRAE Transactions*, 771-787.

Djuric, N., Novakovic, V., & Frydenlund, F. (2008). Heating system performance estimation using optimization tool and BEMS data. *Energy and Buildings, 40*, 1367-1376. Retrieved from https://doi.org/10.1016/j.enbuild.2007.12.006

DOE. (2017). *EnergyPlus Input-Output Reference.*

Dunn, G., & Knight, I. (2005). Small power equipment loads in UK office environments. *Energy and Buildings, 37*(1), 87-91. Retrieved from https://doi.org/10.1016/j.enbuild.2004.05.007

EIA. (2019). *Annual energy outlook 2019 with projections to 2050.*

Emmerich, S. J., & Persily, A. K. (2013). Analysis of U. S. Commercial Building Envelope Air Leakage Database to Support Sustainable Building Design.

Emmerich, S. J., McDowell, T. P., & Anis, W. (2007). Simulation of the Impact of Commercial Building Envelope Airtightness on Building Energy Utilization. ASHRAE Trans. 113(2):379-399.

Emmerich, S., & Persily, A. (1998). Energy impacts of infiltration and ventilation in USoffice buildings using multizone airflow simulation. in: Proceedings of IAQ and Energy, Vol. 98.

Fabrizio, E., & Monetti, V. (2015). Methodologies and advancements in the calibration of building energy models. Energies, 8(4), 2548–2574. https://doi.org/10.3390/en8042548.

Gibbs, A. L., & Su, E. W. (2002). On Choosing and Bounding Probability Metrics. *International Statistical Review, 70*(3), 419-435.

Gowri, K., Winiarski, D. W., & Jarnagin, R. E. (2009). Infiltration Modeling Guidelines for Commercial Building Energy Analysis. Pacific Northwest National Laboratory.

Habrel, J., Bronson, D., & O'Neal, D. (1993c). Impact of using measured weather data versus TMY weather data in a DOE-2 simulation. ASHRAE Transactions 101(2).

Hall, I., Prairie, R., Anderson, H., & Boes, E. (1978). Generation of a typical meteorological year.

Hejab, M., & Parsloe, C. (1992). *Space Allowances for Building Services Distribution Systems: Detail Design Stage.* Bracknell, England: Building Services Research and Information Association.

Heo, Y. (2011). Bayesian calibration of building energy models for energy retrofit decision-making under uncertainty. Georgia Institute of Technology.

Heo, Y., & Zavala, V. M. (2012). Gaussian process modeling for measurement and verification of building energy savings. Energy and Buildings, 53, 7–18. https://doi.org/10.1016/j.enbuild.2012.06.024.

ISO-12569. (2017). Thermal performance of buildings and materials -- Determination of specific airflow rate in buildings -- Tracer gas dilution method". 91.120.10 Thermal insulation of buildings. 91.120.10 Thermal insulation of buildings.

ISO-13790. (2008). *Energy performance of buildings — Calculation of energy use for space.* Switzerland.

Jianhua, L. (1991). Divergence Measures Based on the Shannon Entropy. *IEEE Transactions On Information Theory, 37*(1), 145-151. Retrieved from https://www.cise.ufl.edu/~anand/sp06/jensen-shannon.pdf

Judkoff, R., & Neymark, J. (2006). Model validation and testing: the methodological foundation of ASHRAE Standard 140. In the ASHRAE 2006 Annual Meeting. Quebec City, Canada.

Judkoff, R., Polly, B., & Neymark, J. (2016). A Method to Test Model Calibration Techniques. Presented at the 2016 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, California, August 21–26, 2016.

Kaplan, M. B., Caner, P., & Vincent, G. W. (1992). Guidelines for energy simulation of commercial buildings. Proceedings from the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, Vol. 1.

Kaplan, M. B., McFerran, J., Jansen, J., & Pratt, R. (1990a). Reconciliation of a DOE2.1C model with monitored end-use data for a small office building. ASHRAE Transactions 96(1):981–993.

Kaplan, M., Jonse, B., & Jansen, J. (1990b). DOE-2.1C model calibration with monitored end-use data. Proceedings from the ACEEE 1990 Summer Study on Energy Efficiency in Buildings, Vol. 10.

Kim, S., & Augenbroe, G. (2013). Uncertainty in Developing Supervisory Demand-side Controls in Buildings: A Framework and Guidance. *Automation in Construction, 35*, 28-43.

Knoll, B., Phaff, J. C., & de Gids, W. F. (1995). Pressure simulation program, Proceedings of the Conference on Implementing the Results of Ventilation Research. AIVC.

Krarti, M. (2014). *ENERGY AUDIT OF BUILDING SYSTEMS - AN ENGINEERING APPROACH 2nd edition.* CRC Press - Taylor & Francis Group.

Kullback, S., & Leibler, R. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 79-86. Retrieved from https://projecteuclid.org/download/pdf_1/euclid.aoms/1177729694

Kurnitski, J., Saari, A., Kalamees, T., Vuolle, M., Niemelä, J., & Tark, T. (2011). Cost optimal andnearly zero (nZEB) energy performance calculations for residential buildingswith REHVA definition for nZEB national implementation. Energy and Buildings 43 3279-3288.

Lee, B. (2014). *A Pragmatic Value-Driven Approach to Design with Applications to Energy-Conscious Buildings.* Georgia Tech Theses and Dissertations database. Retrieved from https://smartech.gatech.edu/handle/1853/53094

Lee, B., Sun, Y., Augenbroe, G., & Paredis, C. (2013). Towards Better Prediction of Building Performance: A Workbench To Analyze Uncertainty In Building Simulation. *13th Conference of International Building Performance Simulation Association*, (pp. 1231-1238). France.

Lee, B., Sun, Y., Hu, H., Augenbroe, G., & Paredis, C. (2012). A Framework for Generating Stochastic Meteorological Years for Risk-conscious Design of Building. *SimBuild.* Retrieved from https://www.researchgate.net/publication/261913957

Li, Q. (2017). ADDRESSING DATA INFORMATIVENESS IN RISK-CONSCIOUS BUILDING PERFORMANCE SIMULATION APPLICATIONS. PhD thesis. Georgia Institute of Technology.

Li, Q., Augenbroe, G., & Muehleisen, R. (2017). A Framework for Empirical Validation of Building Performance Simulation under Uncertainty. *15th IBPSA Conference* (pp. 1770-1779). San Francisco: International Building Performance Simulation Association. Retrieved from https://www.semanticscholar.org/paper/A-Framework-for-Empirical-Validation-of-Building-Li-Augenbroe/98a8fb78cf1370b028e03c15a8c6ba68f6c26335

Macdonald, I. (2002). *Quantifying the effects of uncertainty in building simulation.* United Kingdom: University of Strathclyde. Retrieved from http://www.esru.strath.ac.uk/Documents/PhD/macdonald_thesis.pdf

Machairas, V., Tsangrassoulis, A., & Axarli, K. (2014). Algorithms for optimization of building design: A review. *Renewable and Sustainable Energy Reviews, 31,* 101–112. Retrieved from https://doi.org/10.1016/j.rser.2013.11.036

McKenna, C., & Munis, R. (1989). *Reference Guide for Building Diagnostics Equipment and Techniques.* Washington, D.C.: US Army Corps of Engineering.

Moon, H. J. (2005). Assessing mold risks in buildings under uncertainty. PhD. Thesis, Georgia Institute of Technology.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2):161-174.

Mustafaraj, G., Marini, D., Costa, A., & Keane, M. (2014). Model calibration for building energy efficiency simulation. *Applied Energy, 130,* 72-85. Retrieved from https://doi.org/10.1016/j.apenergy.2014.05.019

Ng, L. C., Emmerich, S. J., & Persily, A. K. (2014). An improved method of modeling infiltration in commercial building energy models. NIST Technical Note 1829.

Ng, L. C., Musser, A., Persily, A. K., & Emmerich, S. J. (2013). Multizone airflow models forcalculating infiltration rates in commercial reference buildings. Energy andBuildings. 58 11–18.

Ng, L., Musser, A., & Persily, A. (2012). Airflow and Indoor Air QualityModels of DOE Reference Commercial Buildings.

Nguyen, A.-T., Reiter, S., & Rigo, P. (2014). A review on simulation-based optimization methods applied to building performance analysis. *Applied Energy, 113,* 1043-1058. Retrieved from https://doi.org/10.1016/j.apenergy.2013.08.061

Niu, J. L., Zhang, L. Z., & Zuo, H. G. (2002). Energy savings potential of chilled-ceiling com-bined with desiccant cooling in hot and humid climates. Energy and Buildings. 34 487–495.

Palisade. (n.d.). Retrieved from Palisade Softwares: https://www.palisade.com/risk/

Persily, A. (1982). Repeatability and accuracy of pressurization testing. Proceedings of the ASHRAE/DOE Conference Thermal Performance of the Exterior Envelopes of Buildings II, Las Vegas, NY, December 6-9.

Persily, A. K. (1998). Airtightness of Commercial and Institutional Buildings: Blowing Holes in the Myth of Tight Buildings.

Price, P. N., Chang, S. C., & Sohn, M. D. (2004). Characterizing Buildings for Airflow Models: What Should We Measure. Indoor Environment Department, Lawrence Berkeley National Laboratory, Berkeley, CA.

Quan, Y., Tamura, Y., Matsui, M., Cao, S., & Yoshida, A. (2007). TPU Aerodynamic Database for Low-rise Buildings. *12th International Conference on Wind Engineering.* Cairns. Retrieved from http://wind.arch.t-kougei.ac.jp/system/eng/contents/code/tpu

Raftery, Keane, & Costa. (2011). Calibrating whole building energy models: Detailed case study using hourly measured data. Energy and Buildings, 43(12), 3666–3679.

Raftery, P., & Keane, M. (2011). Visualising patterns in building performance data. In Proceedings of the 12th International Conference of the International Building Performance Simulation Association.

Raftery, P., Keane, M., & O'Donnell, J. (2011). Calibrating whole building energy models: An evidence-based methodology. Energy and Buildings, 43(9), 2356–2364. https://doi.org/10.1016/j.enbuild.2011.05.020.

Ramos Ruiz, G., Fernández Bandera, C., Gómez-Acebo Temes, T., & Sánchez-Ostiz, A. (2016). Genetic algorithm for building envelope calibration. *Applied Energy, 168*, 691-705. Retrieved from https://doi.org/10.1016/j.apenergy.2016.01.075

Reddy, T. A. (2006). Literature review on calibration of building energy simulation programs: Uses, problems, procedures, uncertainty, and tools. ASHRAE Transactions, 112(1), 226–240. https://doi.org/Article.

Reddy, T. A., Maor, I., Jian, S., & Panjapornpon, C. (2006). Procedures for reconciling computer-calculated results with measured energy data. ASHRAE Research Project 1051- RP.

Reddy, T., Maor, I., & Panjapornpon, C. (2007). Calibrating detailed building energy simulation programs with measured data. Part I: General methodology (RP-1051).

Roberti, F., Oberegger, U., & Gasparella, A. (2015). Calibrating historic building energy models to hourly indoor air and surface temperatures: Methodology and case study. *Energy and Buildings, 108,* 236-243. Retrieved from https://doi.org/10.1016/j.enbuild.2015.09.010

Robertson, J. J., Polly, B. J., & Jon, C. M. (2015). Reduced-order modeling and simulated annealing optimization for efficient residential building utility bill calibration. *Applied Energy, 148,* 169-177. Retrieved from https://doi.org/10.1016/j.apenergy.2015.03.049

Royapoor, M., & Roskilly, T. (2015). Building model calibration using energy and environmental data. *Energy and Buildings, 94,* 109-120. Retrieved from https://doi.org/10.1016/j.enbuild.2015.02.050

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., . . . Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer.* Wiley.

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity Analysis in Practice: A Guide to Scientific Models.* New Delhi, India: John Wiley & Sons, Ltd.

Sherman, M. (1980). *Air Infiltration in Building.* University of California, Energy & Environment Division. Berkeley, CA 94720: Lawrence Berkeley Laboratory.

Sherman, M. (1998). The Use of Blower-door Data. Laurence Berkley Laboratory, Report LBL-35173, Berkley; CA.

Sherman, M. H., & Grimsrud, D. T. (1980). "MEASUREMENT OF lNFllTRATlON USING FAN PRESSURlZATlON AND WEATHER DATA". Energy and Environment Division, Lawrence Berkeley Laboratory, University of California.

Sherman, M., & Grimsrud, D. (1980). Measurement of Infiltration Using Fan Pressurization And Weather Data. *Air Infiltration Centre on Instrumentation and Measurement Techniques. LBL-10852*, pp. 277-322. England: Lawrence Berkeley Laboratory.

SimLab. (2008). *EU Science HUB*. Retrieved from European Commission: https://ec.europa.eu/jrc/en/samo/simlab

Sun, J., & Reddy, T. (2006). Calibration of Building Energy Simulation Programs Using the Analytic Optimization Approach (RP-1051). *HVAC&R Research, 12*, 177-196. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/10789669.2006.10391173

Sun, Y. (2014). Closing the building energy performance gap by improving our predections. PhD. Thesis. Georgia Institute of Technology.

Sun, Y. M., Heo, Y., Tan, M., Xie, H., Wu, C., & Augenbroe, G. (2014). Uncertainty Quantification of Microclimate Variables in Building Energy Models. Journal of Building Performance Simulation, 7(1), 17-32. doi: 10.1080/19401493.2012.757368.

Swami, M. V., & Chandra, S. (1988). Correlations for pressure distribution on buildings and calculation of natural-ventilation airflow. ASHRAE Transactions, 94, 243–266.

Thormark, C. (2002). A low energy building in a life cycle—its embodied energy, energy need for operation and recycling potential. Building and environment. 37 429–435.

Trubiano, F. (2013). *Performance Based Envelopes: A Theory of Spatialized Skins and the Emergence of the Integrated Design Professional.* MDPI AG.

Wang. (2013). Investigation of the Impact of Building Entrance Air Curtain on Whole Building Energy Use.

Wang, L., & Zhong, Z. (2014). An approach to determine infiltration characteristics of building entrance equipped with air curtains. Energy Building. 75 312–320.

Wang, L., & Zhong, Z. (2014). Whole building annual energy analysis of air curtain performance in commercial building. eSim Building Performance Simulation Conference.

Wang, Q. (2016). Accuracy, validity and relevance of probabilistic building energy models. Georgia Institute of Technology.

Wang, Q., Augenbroe, G., & Sun, Y. (2014). The Role of Construction Detailing and Workmanship in Achieving Energy Efficient Buildings. *Construction Research Congress*, (pp. 2224-2233).

Wyss, G., & Jorgensen, K. (1998). A user's guide to LHS: Sandia's Latin Hypercube Sampling Software. Sandia National Laboratories, Albuquerque, NM.

Yang, Z., & Becerik-Gerber, B. (2015). A model calibration framework for simultaneous multi-level building energy simulation. *Elsevier: Applied Energy*, 415-431.

Younes, C., & Abi Shdid, C. (2013). A methodology for 3-D multiphysics CFD simulation of air leakage in building envelopes,. Energy and Buildings 65 146–158.