Refinement of Dynamic Non-Residential Building Archetypes Using Measurement Data and Bayesian Calibration

Peter Remmen¹, Julian Schäfer¹, Dirk Müller¹
¹RWTH Aachen University, E.ON Energy Research Center, Institute for Energy Efficient Buildings and Indoor Climate, Germany

Abstract
Dynamic simulation and optimization of energy systems at urban level becomes increasingly important as an additional decision criterion for planning and operation. Essential in this context is the dynamic heating and cooling demand of buildings. Often building archetypes are used to depict the individual characteristics of buildings. While in average the demand is accurately reproduced by archetype buildings, individual buildings deviate from the statistical average.

This paper presents a methodology to refine building archetypes using real measurement data and Bayesian Calibration. The calibration uses statistical indices instead of whole time series or yearly cumulated energy demands. We apply the methodology to four real buildings, the results demonstrate the potential of the methodology by reducing the RMSE between measured and simulated heating demand up to 57%.

Introduction
Distributed urban energy systems are important to meet the increasing demands of an economical and climate-friendly energy supply in cities. In addition, an increasing interconnection of energy sectors and components in urban energy systems can be observed. For this reason, the design of distributed urban energy systems requires new planning and design methods. In this context, the dynamic behavior of buildings gains in importance. However, often only little topological and physical information about existing buildings is available. Building archetypes are a widely used approach to fill missing information with statistically or normative determined values in order to enable dynamic simulation of buildings at an urban level. Already published studies show that by using archetypes (a) the cumulated yearly energy (Masstrucci et al., 2014), (Cerezo et al., 2017) and (b) the cumulated dynamic energy demand of multiple buildings (Remmen et al., 2018) can be adequately reproduced. However, the behavior of individual buildings is often not adequately described, especially if the actual characteristics of the building differ from the statistical average. To close the gap between real energy consumption and dynamic Building Performance Simulation (BPS), calibration is key. Hong et al. (2018) names calibration as one of the ten challenges for BPS. They stress the importance of (1) accounting uncertainty of input parameters and the accuracy of BPS models, (2) need for automation in the calibration process to avoid non-unique solutions, (3) high computational costs for advanced, automated calibration methods and (4) review of current calibration criteria (e.g. from ASHRAE guideline 14) (ASHRAE, 2002).

In literature, the key challenges have already been taken up in a number of publications. Bayesian Calibration (BC) is a promising method to consider uncertainties and automate the calibration at the same time. BC for numerical models according to Kennedy and O’Hagan (2001) is used in various contexts in the field of BPS. Applications range from calibration based on annual or monthly values (Booth et al., 2013), (Heo et al., 2012) to higher resolution up to hourly data (Chong et al., 2017). Kristensen et al. (2018) applies BC and hierarchic modeling to buildings using Danish archetypes. Sokol et al. (2017) uses BC to define archetypes for an existing building stock in Kuwait. Chong et al. (2017) apply BC on buildings with high resolution measurement data, identifying the prohibitive computational times as one limitation when using daily or hourly data. Kristensen et al. (2017) show that using high resolution measurement data influences and improves the calibration results. Automated calibration techniques using Genetic Algorithms can be found in (Garrett and New, 2015). The use of statistical indices in calibration is discussed in Vogt et al. (2018).

This paper is a contribution to further improve the demand forecast of buildings for urban energy modeling, by presenting a methodology for the automated calibration of non-residential archetype building models. The calibration technique used is Bayesian Calibration, taking into account the uncertainties in physical input parameters, measurement data and the building model itself. The aim of the methodology is to calibrate the hourly dynamic behavior rather than cumulated energy consumption. In this context the
calibration runtime plays a decisive role. To this end we present a novel approach to use statistical indices as objective functions for BC.

The paper is divided into four sections. The Methodology section focuses on the use of statistical indices as objective function for BC. Further, the four buildings under investigation and their corresponding archetypes are presented. The Results section compares the calibration results using all time steps of one year against the results using statistical indices. Three different indices are tested in this context (1) coefficient of determination, (2) Mean-Absolute-Percentage-Error and (3) Median-Absolute-Percentage-Error. In the Limitations and further Research section, we question our methodology and reveal potential for further research. The paper concludes with a summary of the methodology and results presented.

Methodology

Bayesian Calibration with statistical indices

On the basis of the Bayesian theorem, Kennedy and O’Hagan (2001) have developed a mathematical framework for the application of Bayesian Calibration to computer models. The relation between model output \( \eta \), model error \( \delta \), measurement error \( \epsilon \) and observations \( y \) are expressed as a function of relevant boundary conditions and parameters \( x \) and the calibration parameters \( \Theta \), as shown in Equation 1. All terms in Equation 1 are modeled as random variables. A detailed description of Bayesian Calibration is not given in this paper and reference is made to (Kennedy and O’Hagan, 2001; Heo et al., 2012). For the model output \( \eta \) we use a meta-model that is modeled as a Gaussian Process with a quadratic-exponential kernel with constant fraction. The Gaussian process has been trained with 300 simulation runs of the physical model. Calibration parameters have been varied using Latin-Hypercube-Sampling. For the iterative solution of the problem, the Metropolis-Hastings algorithm is used as Markov-Chain-Monte-Carlo sampler. 15,000 iterations have been performed where the first 5,000 iterations have been discarded as burn-in.

\[
g(x) = \eta(x, \Theta) + \delta(x) + \epsilon \tag{1}
\]

One requirement of our methodology is the consideration of a dynamic calibration of building models, i.e. not monthly or annually cumulated values are used as model output \( \eta \) and observation \( y \), but hourly values. Bayesian Calibration of time series can be a computationally intensive procedure. In particular, the generation of the meta-model with use of Gaussian Process (GP) is a limiting factor. The runtime required to train the meta-model is proportional to the data points considered with the power of 3 (Gilboa et al., 2012). In the context of this work, the complexity of the GP regression increases with an increasing number of calibration parameters, the number of time steps and the number of simulated training time series.

This section presents a methodology that aims to minimize the runtime of the generation of the meta-model. For this reason, a GP regression is not performed for whole time series. Instead, an statistical index between each training time series and the real measurement data is calculated. In this way, the information of a complete time series can be recorded in a single numeric value. On the one hand this methods saves computational time for the generation of the meta-model. On the other hand a loss of information must be accepted by using the statistical index.

A discussion about using statistical indices in BPSA calibration can be found in Vogt et al. (2018). To show the potential of the methodology we limit our investigation in this paper to three different indices. In our study we use Mean-Absolute-Percentage-Error \( MAPE \) (Equation 2), Median-Absolute-Percentage-Error \( MdAPE \) (Equation 3) and Coefficient of Determination \( R^2 \) (Equation 4). In these equations \( \eta \) is the model output, \( y \) the observation, i.e. real measurement and \( N \) the number of observations (e.g. 8760 h). As Equation 2 and Equation 3 show, \( MAPE \) and \( MdAPE \) differ only by using the mean and median, respectively.

\[
MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \eta_i}{y_i} \right| \tag{2}
\]

\[
MdAPE = \text{median} \left( \left| \frac{y_i - \eta_i}{y_i} \right| \right) \tag{3}
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \eta_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{4}
\]

All three indices are scale independent, scale independent indices are suitable to evaluate results relative to the data and thus allow to compare different data sets against each other. According to Kim and Kim (2016) the \( MAPE \) is one of the most popular indicator to evaluate forecast accuracy. One disadvantage is the sensitivity to outliers. In contrast the \( MdAPE \) is more robust to these outliers and thus more robust to evaluation of real measurement data (Fildes, 1992). The advantages of \( R^2 \) are that this index is dimensionless and easy to interpret.

Although the measurement data used is already of a good quality, a data cleaning of the measurement still has to be carried out. In our case we most common error of the measurement data are extreme outliers for a short period of time (few minutes). We apply quartiles to detect these outliers. The distance between the 25 % and 75 % is called the interquartile distance and serves as a measure of when a measured value is
Table 1: Use Case buildings from Research Center in Germany.

<table>
<thead>
<tr>
<th>Building</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage</td>
<td>Office</td>
<td>Office</td>
<td>Lab</td>
<td>Lab</td>
</tr>
<tr>
<td>Year of construction</td>
<td>2006</td>
<td>2012</td>
<td>2011</td>
<td>1974</td>
</tr>
<tr>
<td>Number of storeys</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Height of storeys (m)</td>
<td>2.5</td>
<td>2.2</td>
<td>2.5</td>
<td>3.5</td>
</tr>
<tr>
<td>Net leased area (m²)</td>
<td>660</td>
<td>2170</td>
<td>3440</td>
<td>270</td>
</tr>
</tbody>
</table>

defined as an outlier (Han et al., 2012). Within our work the definition of Acuña and Rodriguez (2004) is used, where a value is considered an outlier if it exceeds three times the interquartile distance. The outliers are replaced with a moving average value with an observation period of 24 hours.

**Use Case buildings and building archetypes**

In this section, the buildings investigated, the archetypes used and the selected calibration parameters are presented in more detail. Four buildings of a Research Center in Germany serve Use Case buildings, which can be characterized on the basis of data shown in Table 1. All four buildings are connected to a local heating network and the heat demand from 2014 is known in hourly resolution from measurements at the substation of the building. Also weather data is recorded during this year.

To model the buildings as archetypes we use the open source software tool TEASER (Remmen et al., 2018), we use the Reduced-Order-Model implemented in Modelica from AixLib (Mueller et al., 2016). The assignment of the buildings to the archetypes is also shown in Table 1. TEASER distinguishes between residential buildings and non-residential buildings as archetypes. The non-residential archetype buildings are divided into 6 (office buildings, Office) and 7 (laboratory buildings, Lab) representative usage zones. Individual usage conditions are assigned to each zone, more details on the used typology can be found in (Remmen et al., 2018). A large amount of typological parameters and further model assumptions are estimated from the 4 input data. The calibration of all parameters is not possible with Bayesian Calibration. For this reason, six relevant parameters were selected in a preliminary investigation, which will be used in the following as calibration parameters \( \Theta \). The parameters as well as the archetype values are given in Table 2 for all buildings. Due to the archetype approach, the values of the buildings do not differ. In Table 2 the assumed prior distributions of the individual parameters are also given, where uni(min, max) corresponds to a uniform distribution and tri(a,c,b) corresponds to a triangular distribution.

With the help of the methodology presented, a total of four investigations will be carried out. As a reference, a calibration is used in which all time steps of the present measurement series are used to calibrate the building model. In the second step, the meta-model is trained with statistical indices, which are calculated between the simulation results and the measurement data of the same time period. A calibration is carried out for each of the considered indices. In order to compare the results of the four investigations, a new simulation is carried out for each building after calibration. The expected value \( E \) of the posterior distribution is used for the calibration parameters. In order to ensure a fair comparison of the indices, a different index is used to evaluate the simulation results. The root-mean-square-error (RMSE) is a well excepted parameter to compare simulation with measurement data. The RMSE is a scale dependent parameter and is therefore well suited to compare the four investigations for each building. As an additional indicator, we use the computation time for the creation of the meta-model.

**Results**

At first the results of the calibration with all available measurement points is presented. A calibration is carried out for all four buildings. In the following this calibration is referred to as conventional calibration. Figure 1 shows the posterior distribution, as well as the trace plot for Building D. For the sake of simplicity, in the posterior distribution only the value range that was actually sampled is shown. According to Lunn et al. (2012), the small value range of the distribution as well as the course of the trace plot indicate a good convergence of the Hastings-Metropolis sampler. For the other buildings but a similar behavior can be observed.

The posterior distributions resulting from the first calibration are evaluated on the basis of the expected value. The results of the evaluation are shown in Figure 2. The difference between the expected value of the posterior distribution and the original parameters are also given, where uni(min, max) corresponds to a uniform distribution and tri(a,c,b) corresponds to a triangular distribution.

### Table 2: Calibration parameters \( \Theta \) and prior distribution for calibration. Office share is the proportion of office space of the total usable floor area.

<table>
<thead>
<tr>
<th>Building</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Prior dist.</th>
</tr>
</thead>
<tbody>
<tr>
<td>set temp. (deg C)</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>tri(17,21,25)</td>
</tr>
<tr>
<td>office share (-)</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>uni(0,1)</td>
</tr>
<tr>
<td>infil- tration (1/h)</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>tri(0,1,2)</td>
</tr>
<tr>
<td>setback (K)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>tri(0.2,5,5)</td>
</tr>
<tr>
<td>heating start (h)</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>tri(3,5,7)</td>
</tr>
<tr>
<td>heating end (h)</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>tri(16,18,20)</td>
</tr>
</tbody>
</table>

---

Proceedings of the 16th IBPSA Conference  
Rome, Italy, Sept. 2-4, 2019  
4684
Figure 1: Posterior distributions and traceplot of office share calibration parameter of Building D - conventional calibration. Note that only the range of the sampled values is displayed.

Figure 2: Difference between expected values and archetype parameters - conventional calibration.

Figure 3: Posterior distributions and trace of calibration parameters of Building D - calibration with MdAPE as statistical index.

Nonetheless, Figure 4 summarizes the differences between the expected value of the posterior distribution with the original archetype parameter exemplary for the statistical index MdAPE for all buildings. The comparison of the two figures Figure 2 and Figure 4 clearly shows that the parameters office share, heating start, and heating end.
infiltration rate, setback and, with some limitations, the set temperature show the same tendencies as the conventional calibration. The calibrated set temperature for building C remains almost constant. For the Buildings A, B and D a clear reduction of the set temperature up to 1.5 Kelvin is noticeable. The conventional calibration showed the same tendencies for Buildings A and D, whereas Buildings B and C remained a constant set temperature. Both in conventional calibration and calibration with MdAPE the infiltration rate and the setback of all for buildings are increased or decreased, respectively. In both calibration the office share is lowered for Buildings A and B and increased for C and D. As already mentioned, the parameters heating start and heating end are not changed by the calibration.

In the next step, all calibrated buildings from all investigations (conventional and three statistical indices) have been simulated using the expected values and the RMSE between real measurement and simulation data has been computed. Excerpts of the simulated time series for Building D are shown in Figure 5 and Figure 6. Figure 5 shows the time series for conventional calibration and Figure 6 the time series for calibration using the statistical index MdAPE. The qualitative perception that the results improve significantly to the non-calibrated case can be confirmed by the calculation of the RMSE over the whole investigated period.

The results illustrated in Figure 7 show, with the exception of two calibrations, a minimization of the RMSE. The diagram shows for Building A and D that each of the calibrations computes a comparable RMSE.

It is noticeable that conventional calibration does not show any further minimization of the RMSE compared to the indice based calibration for Buildings A and D. For buildings Buildings B and C, conventional calibration as well as calibration using the $R^2$ and MdAPE leads to approximately the same improvement in the RMSE, whereas the MAPE shows worse results. In the following, we will examine in more detail why the MAPE in some cases leads to a deterioration of the RMSE in the calibrated case.

Figure 8 shows a time series for the building with an
erroneous calibration. Furthermore, between hours 725 and 750 an non-typical behavior is visible. This is due to a failure either of the building energy system or the measurement equipment. While NaN or extremely high values have been filtered before the investigation, zero values were not taken further into consideration. Fildes (1992) shows that the MAPE is susceptible to outliers. Looking at the measured data of Building C in more detail, it is often the case that the heat demand suddenly falls to zero over a period of several hours. This leads to high MAPE values and misleading meta-models. Figure 9 shows the time series of the calibration of the building Building C with a consideration of the index MdAPE. The time series of the calibrated building model shows a significant improvement compared to an application of the MAPE. The improved result can be justified by the fact that the MdAPE is more resistant to outliers (Trapero et al., 2013). Buildings A and D do not show any of these outliers.

In summary, it can be stated that results of a Bayesian Calibration based on statistical indices are comparable to those of conventional Bayesian calibration. The used statistical index has a relevant influence on the result of the calibration. When using the MAPE, for example, outliers can cause distortion of the data stored in the meta-model. The distortion is reflected in an incorrect estimation of posterior distribution and thus also in incorrect simulation time series. The expected values of the calibration with statistical indices show same tendencies as the conventional calibration. The investigation shows that time-dependent parameters, such as the heating time, cannot be calibrated with the statistical indices under consideration. The paper shows that the computing time can be significantly reduced by using statistical indices and thus the methodology presented enables dynamic calibration on the basis of hourly measured values on an urban level.

**Limitations and further Research**

In this study, we performed a Bayesian calibration using statistical indices. The applied methodology is a first step to calibrate buildings on an urban level and therefore offers further potential for improvement. The limitations and envisaged developments are described in more detail in this section.

- **Validation and verification** Although the calibrated time series and calculated RMSE values clearly indicate realistic simulations, the adjustment of the calibration parameters is currently difficult to evaluate. A calibration of a physical building model is only successful if the calibration parameters approach the realistic values. In order to verify the new values of the calibration parameters, these values must be recorded on the real buildings. This is particularly difficult for time-variant parameters. Future research must therefore focus on new possibilities for validating the calibration of dynamic building models.

- **Selection of calibration parameters** Within the framework of Bayesian Calibration, a number of assumptions must be made which significantly influence the result of the calibration. An elementary assumption is the selection of the most sensitive calibration parameters. In this study we used the same six calibration parameters for all four buildings. However, in reality the sensitivity of different buildings might vary, even though they are assigned to the same archetype. In our future work will develop rules for sensitive parameters of different building archetypes.

- **Selection of prior distribution** Another important assumption is the choice of prior distributions, as this directly influences the sampling of posterior distributions. In this paper, the calibration parameters are represented by uni-

---

**Figure 8:** Time series for measurement data, non-calibrated and calibrated simulation data of Building C - calibration with MAPE as statistical index.

**Figure 9:** Time series for measurement data, non-calibrated and calibrated simulation data of Building C - calibration with MdAPE as statistical index.

As already described, calibration on the basis of statistical indices has advantages over conventional calibration in terms of runtime, since only one GP regression has to be performed. At this point, the runtime improvement should be put into a quantifiable context. The execution of a GP regression took 1.65 seconds using a conventional computer with an i5 processor (4 cores, 1.80 GHz) and 8 GB RAM (2400 MHz). This value is compared to the runtime of the conventional creation of the meta-model, which lasts approx. 15,000 seconds (250 minutes) if 8760 time steps are considered.
form and triangle distributions. This assumption can be improved by a stronger consideration of prior knowledge by means of an individually adapted prior distribution. We will focus on how this prior knowledge can be obtained using (a) the analysis of measurement series using classification algorithms and (b) the posterior distributions from previous calibrations of identical building archetypes.

- **Optimization of Bayesian Calibration**
  Within the scope of this work, the main focus was not on the optimization of BC. However, many investigations show that the choice of the hyperparameters of the meta-model, the distribution of the model error ($\delta$) and measurement error ($\epsilon$) as well as the selection of the MCMC sampler can have a considerable influence on the result of the calibration (Chong et al., 2017; Heo et al., 2015; Lim and Zhai, 2017; Menberg et al., 2018). These parameters will have to be investigated in future studies. This could contribute to improve the convergence in the sampling process. In this context a detailed investigation has to be carried out to determine the best statistical indices for the calibration, as all indices have advantages and disadvantages.

- **Workflow automation**
  The presented methodology is fully automated and uses the 3DCityDB (Yao et al., 2018) to store the semantic building data, a time series database (InfluxDB (InfluxData, 2019)) to store the measurement and simulation data. Scikit-learn is used to create the meta model (Pedregosa et al., 2011). For the implementation of the Bayesian calibration itself, the Python module PyMC3 is used (Salvatier et al., 2016). In order to guarantee automation in the sense of the previously described extensions (selection of prior distribution, selection of calibration parameters), the workflow and software tools must be extended process-specifically. For this purpose, we aim for an API by means of object-relational mapping to the 3DCityDB. The method presented and the software components used are to be made open source.

**Conclusion**

The presented paper applies a Bayesian calibration to building archetypes according to Kennedy and O’Hagan (2001) using statistical indexes to generate the meta model. The focus is on calibration evaluation based on hourly measurements. Three different indexes are used $R^2$, Mean-Absolute-Percentage-Error $MAPE$ and Median-Absolute-Percentage-Error $MdAPE$ and compared to a calibration based on all available measured values. The presented methodology is applied to real measurement and building data of a Research Center in Germany. The conventional calibration shows for each of the four considered buildings a significant reduction of the RMSE compared to the archetype. In the context of the $RMSE$, the results for the $MdAPE$ and $R^2$ indices are comparable to those obtained using conventional methods. In each of the cases examined, the $RMSE$ for at least one building can be minimized by more than 50%. In contrast to this, the use of the $MAPE$ achieves faulty calibrations, which can be explained by outliers in the measurement data. Furthermore, it was found that time-varying parameters such as heating start and end cannot be calibrated with the use of statistical indices. The biggest advantage of the presented method is the massive reduction of the runtime for the generation of the meta-model, since only one value has to be considered for the GP regression. Future work in this area will concentrate on the selection of calibration parameters for different archetypes, the determination of the posterior distribution and further process automation.

**Acknowledgment**

Grateful acknowledgment is made for financial support by BMWi (German Federal Ministry of Economic Affairs and Energy), promotional references 03ET1352A.

This work was supported by the Helmholtz Association under the Joint Initiative “Energy System 2050 - A Contribution of the Research Field Energy”.

**References**


