A Framework for the continuous Bayesian calibration of building energy models

Adrian Chong[1], Chao Song[1]
[1]Department of Building, School of Design and Environment, National University of Singapore, Singapore

Abstract

With the emergence of the Internet of Things (IoT), there is an opportunity to create a digital twin of a building that continuously learns and updates itself using real-time observations. Model calibration is an essential aspect of the overall process to ensure its reliability. However, the calibration of building energy models (BEM) is typically carried out only once and can quickly become outdated. Continuous Bayesian calibration reduces the effort to maintain an energy model while accounting for its uncertainty. Invariably, the model would be up-to-date for use in applications such as retrofit analysis, fault detection, and model predictive control. The present paper aims to present the concept and implementation of a framework for the continuous calibration of BEM with uncertainty. The proposed framework includes instance selection as a pre-processing step to keep the calibration process computationally tractable.

Introduction

Building performance simulation was initially developed for use during the design phase to support energy efficient design and improve occupant comfort. With the emergence of the Internet of Things (IoT), the promise of a digital twin of building energy systems is becoming a reality. Digital twins act as the bridge between the physical and digital worlds and can be defined as a virtual model that is connected to its physical counterpart through real-time data using sensors and IoT. Put differently, the digital twin of a building can be a building energy model (BEM) that continuously learns and updates itself using real-time data from sensors and the building management system (BMS). A continuous calibration approach reduces the effort needed to maintain the BEM. As a result, the model is kept current and ready to support retrofit analysis (Heo et al., 2015), fault detection (Dong et al., 2012), and model predictive controls (Zhang et al., 2018) throughout the life-cycle of the building (Chong et al., 2019).

Model calibration is an integral component of the overall analysis to ensure its reliability. According to Coakley et al. (2014), BEM calibration methods are either manual or automated. Manual calibration involves manual iterative tuning of the BEM and requires the modeler to have in-depth knowledge of the building system and its operations (Pedrini et al., 2002; Ian Shapiro, 2009). Automated approaches overcome the need for labor-intensive manual tuning by integrating mathematical methods into the calibration process (Reddy, 2006; Coakley et al., 2014). Examples of automated BEM calibration includes the use of optimization algorithms with objective functions aimed at minimizing the discrepancies between measured data and BEM predictions (Sun et al., 2016; Chaudhary et al., 2016).

Automated approaches also includes Bayesian calibration, which has been gaining increasing interest because of its ability to include uncertainty as well as expert knowledge in the form of priors (Tian et al., 2018). Also, the ability to include uncertainty is attractive because BEM calibration is an inverse problem that is ill-posed in most practical situations. The available data is usually not enough for identifying a unique solution given a large number of model input parameters. As a result, calibrating these models with limited data might lead to identifiability issues and a wide variety of model uncertainties despite the model having been calibrated (Chong and Menberg, 2018). Gaussian process (GP) models are often used as surrogates or emulator of the computationally intensive BEM to reduce the time needed to complete the calibration, which can be demanding since Bayesian calibration usually requires a large number of simulations to be run. GP models are often used because of their flexibility. Despite their higher computation costs, GP models have been shown to provide the best accuracy (Lim and Zhai, 2017). To alleviate the high computation cost of using GP emulators, several methods have been proposed, including the use of linear regression models in place of GP models (Li et al., 2016), using Hamiltonian Monte Carlo (HMC) algorithm for more efficient Markov Chain Monte Carlo sampling (Chong and Lam, 2017; Menberg et al., 2017), as well as using a smaller representative subset of the data for the Bayesian calibration (Chong et al., 2017).

Despite these advances in automated BEM calibration, the continuous updating and calibration of BEM remain mostly unexplored. Additionally, Bayesian approaches tend to be too time-consuming for use within a continuous calibration framework. Therefore, in expectation of the emergence of IoT and the digital twin technology, this paper aims to present the concept and implementation of a framework for the continuous calibration of building energy models that take into account uncertainty while being computationally tractable.

Method

Fig. 1 shows an overview of the proposed framework for the continuous calibration, which repeatedly calibrates the model over a moving time horizon T to keep the model up to date continuously. The overall process can be summarized as follows:

1. Create training data. Training data used for the calibration is the historical data that has been collected. The data is then separated into two datasets containing only weekday and weekend data respectively.
2. Forecast weather. Forecast relevant weather parameters (X_{pred}) for the period of prediction T. For our case study, this is the outdoor dry-bulb temperature and the outdoor relative humidity, which is selected by variable importance measures of random forest regression with a threshold of 0.2.
3. Instance selection. Select k training examples from both the weekday and weekend datasets that are closest to each instance of X_{pred} using the k-nearest neighbor (k-NN) algorithm.
4. Sensitivity analysis. Sensitivity analysis is carried out using the Morris method (Morris, 1991) to select the most influential parameters for the calibration.
5. Bayesian calibration.
6. Repeat steps 1 to 5.

Case study

The case study building used in this study is an actual ten-story office building located in Pennsylvania, U.S.A. The installed HVAC system is a dual duct system. In the dual duct system, the cold and warm air are separately duct and mixed at each terminal box to achieve the desired temperature. Cooling is provided through two water-cooled chillers connected in parallel and heating through two gas boilers also connected in parallel. The model was created using EnergyPlus version 8.6 and was modeled according to the construction drawings and specifications. Seven months of daily measured electrical energy consumption was used to demonstrate the proposed continuous Bayesian calibration framework.

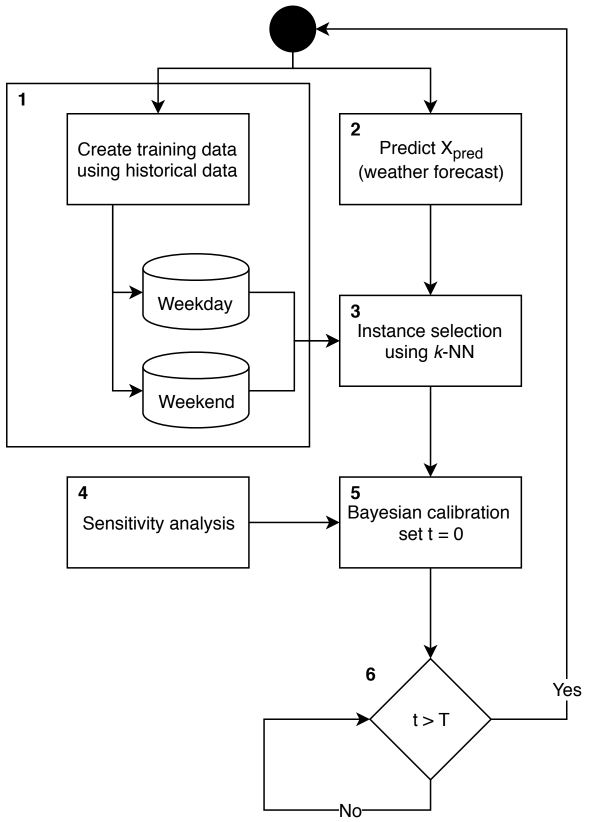

Figure 1: Overview of the proposed framework for the continuous Bayesian calibration of building energy models.

Before calibrating the model, a sensitivity analysis was carried out using Morris method (Morris, 1991) to identify influential parameters. To prevent identifiable issues (Chong and Menberg, 2018), only the top four most influential parameters were selected for the calibration and they are (1) equipment power density $[W/m^2]$, (2) cooling setpoint temperature [°C], (3) lighting power density $[W/m^2]$, and (4) fan efficiency $[-]$.

Continuous Bayesian calibration

The continuous Bayesian calibration approach presented in this section is based on Chong et al. (2019), and is proposed because it provides a flexible framework for dynamically updating model parameters while at the same time account for their uncertainties. As "new data" arrive, the "existing data" is not discarded but instead assimilated to the new data through the use of priors.

Bayesian calibration was carried out following the statistical formulation proposed by (Kennedy and O'Hagan, 2001), which explicitly models parameter uncertainty, model discrepancy, and observation error (eq. 1).

$$y(x) = \eta(x,t) + \delta(x) + \epsilon(x) \quad (1)$$

where $y(x)$ is the field observations; $\eta(x,t)$ is the BEM prediction for a given observable inputs x and calibration parameters t; $\delta(x)$ is the model discrepancy; and $\epsilon(x)$ is the observation error. The inputs and output used for the Bayesian calibration of our case study is as follows:

- Observed output $y(x)$: Measured electrical energy consumption [kWh].
- Simulation output $\eta(x,t)$: Predicted electrical energy consumption [kWh]
- Observed inputs x:
  (a) x_1: Outdoor dry-bulb air temperature [°C].
  (b) x_2: Outdoor relative humidity [%].
  (c) x_3: Solar radiation rate per area [W/m^2].
- Calibration parameters t:
  (a) t_1: Equipment power density [W/m^2].
  (b) t_2: Cooling setpoint temperature [°C].
  (c) t_3: Lighting power density [W/m^2].
  (d) t_4: AHU fan efficiency [-].

Since the BEM can be computationally expensive to evaluate, a Gaussian process (GP) model is used as an emulator or surrogate model during the iterative calibration process. Field and computer simulation data are combined in the GP model according to the framework described in (Higdon et al., 2004). Hamiltonian Monte Carlo (HMC), which is a Markov Chain Monte Carlo (MCMC) method is then used to sample from the posterior probability distributions because it is more efficient and provides better convergence (Chong and Lam, 2017; Menberg et al., 2017). Details of the Bayesian calibration and Gaussian process can be found in (Chong and Menberg, 2018).

A time-series database supports the proposed continuous Bayesian calibration method through a Representational State Transfer (REST) application program interface (API) (Fig. 2). The time-series database is used to provide optimal handling of time-series data collected real-time from the building management system (BMS). Through a RESTful API, the collected data could easily be retrieved with simple HyperText Transfer Protocol (HTTP) methods, and subsequently used for Bayesian calibration. Continuous Bayesian calibration is carried out based on the principals of a receding horizon, which uses a sliding time window. The sliding window moves forward at each sampling time and re-calibrates the BEM using the measured data and the last estimated parameter values out of the previous window as the prior knowledge. Suppose at time t with a time interval extending T time-steps into the future $t, t+1, ..., t+T$. The continuous calibration is then carried out as follows:

1. Define priors. Posteriors of the previous model are used to derive prior probability distributions through the use of maximum likelihood estimation (MLE) and Akaike information criterion (AIC).
2. Form a predictive model. Replace all unknown quantities over the time interval with their current estimates, using data available at time t.
3. Execute. Generate samples from the posterior distributions using the model from step 2.
4. Repeat. Continuously perform steps 1 to 3 after every T time-steps.

The Akaike information criterion (AIC) (Equation 2) is used to select the probability distribution that gives the "best" fit to the data (posterior samples of the previous model) (Burnham and Anderson, 2004). For this study, the data is fitted to 5 different continuous probability distributions and they include the Beta, Gamma, Lognormal, Normal, and the Weibull distribution.

$$AIC = -2 \log \mathcal{L}(\theta|x) + 2K \tag{2}$$

where θ denotes the parameters of the probability distribution and is determined using maximum likelihood estimation; X denotes the posterior samples of the previous model; n is the number of observations and k is the number of parameters to be estimated. Given the posterior samples X, maximum likelihood estimation maximizes $\mathcal{L}(p|x)$ over all possible θ.

Instance selection

The training data used for each calibration is created using all of the historical data that have been collected to date. Since this dataset would continue to increase over time, it is not suitable for use within a continuous calibration framework. In particular, the proposed Bayesian calibration method employs a GP model as an emulator, which has a runtime complexity of $\mathcal{O}(N^3)$ where N is the number of samples the GP model is trained on. Therefore, instance selection is applied to reduce the original dataset to a manageable volume, leading to a significant reduction of the computational resources that are necessary for performing the Bayesian calibration. (Chong et al., 2017) previously proposed selecting a representative subset of the full dataset using a metric known as sample quality that is based on the KullbackLeibler divergence. However, since the sampled subset is selected randomly, the resulting subset remains relatively large in order to maintain high sample quality.

In this study, to keep the Bayesian calibration computationally tractable within a continuous calibration framework, a representative subset of the full dataset is selected using the k-nearest neighbor (k-NN) algorithm (Cover and Hart, 1967). The k-NN algorithm is a non-parametric method that selects k points that are nearest to an observation. In this study, the nearest neighbors were determined using the Euclidean distance. Since the Euclidean distance is sensitive to the scale of the data (variables on a larger scale may dominate the distance calculation), min-max normalization is applied to convert the data to a $[0, 1]$ scale.

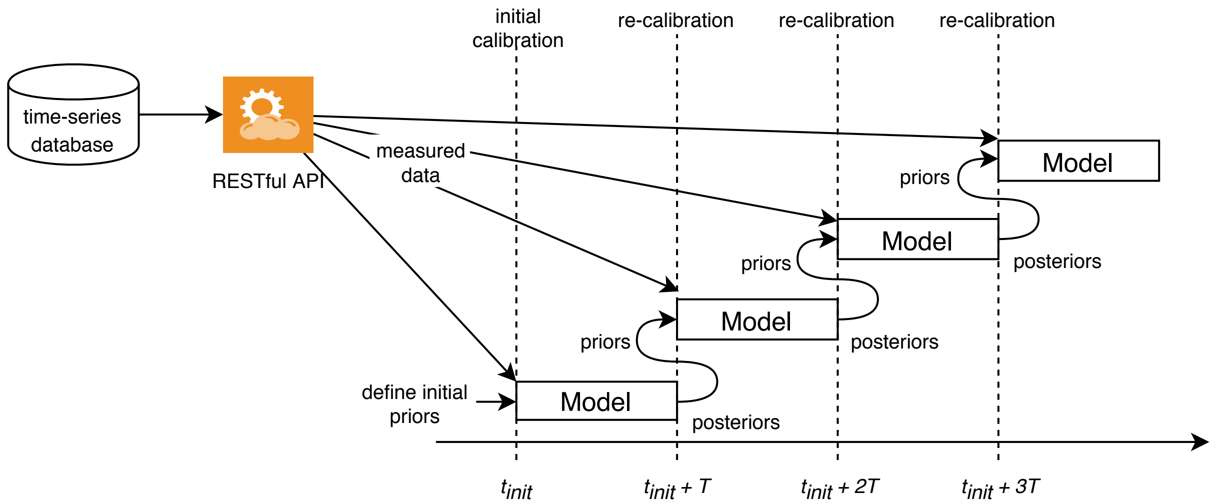

Figure 2: Continuous Bayesian calibration using a sliding time window. The posteriors of the previous calibrated model is used to derive the priors for the Bayesian calibration at the current time-step (Chong et al., 2019)

Instance selection for the continuous Bayesian calibration is carried out as follows. First, X_{pred} (outdoor dry-bulb temperature and relative humidity with respect to our case study) for the prediction period T is predicted. Next, the training data collected is separated into two datasets containing only weekday and weekend data respectively. The k-NN algorithm is then used to select a representative subset (from the weekday and the weekend datasets) that is closest to each instance of X_{pred}. Independent sampling from a weekday and weekend dataset was carried out to ensure that both weekday and weekend data are included in the resulting data used for the Bayesian calibration.

Results

We evaluate the calibration performance by comparing the calibrated GP model against the measured data. For a genuine assessment of the prediction accuracy using the proposed continuous Bayesian calibration framework, the first six months (January to June) of data was treated as historical data and the seventh month (July) as a hold-out test dataset that was not used for the calibration. A 7-day time window was used for this case study. In other words, for the month of July, we evaluate the current model's prediction accuracy using hold-out observations 7-days into the future (Fig. 2). As illustrated in Fig. 1, k-NN was used to select a representative subset as training data from a dataset comprising all historical data that has been collected at the time of the calibration. For instance, on July 7^{th}, the calibration would be carried out using training data that was selected using k-NN from a dataset comprising of data from January 1^{st} to July 6^{th}. Prediction accuracy is then evaluate using hold-out observations from July 7^{th} to July 14^{th}. The metric used for the evaluation of prediction accuracy of the hold-out observations is the coefficient of variance of the root mean squared error (CVRMSE) (eq. 3), a measure that is commonly used as an indication of the variation between the values predicted by the calibrated BEM and the observed values (ASHRAE, 2014).

$$CVRMSE[\%] = 100 \times \frac{\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \,/\, (n-1)}}{\bar{y}} \tag{3}$$

Fig. 3 and Fig. 4 shows the variation in CVRMSE and posterior predictions respectively over the test dataset using different amounts of training data (7 days, 14 days, 28 days, and 56 days). Put differently, for each of the 7 prediction points in the 7-day time window, we select $k = 1$ ($7 \times 1 = 7$ days training data), $k = 2$ ($7 \times 2 = 14$ days training data), $k = 4$ ($7 \times 4 = 28$ days training data), and $k = 8$ ($7 \times 8 = 56$ days training data) nearest neighbors respectively.

It can be observed that the hold-out observed values fall within the prediction intervals regardless of the amount of training data (Fig. 4). However, Fig. 4 also shows that using 7, 28, and 56 days of training data produce posterior predictions with larger ranges, suggesting larger uncertainties than using 14 days of training data. This is also illustrated by the distribution of CVRMSE values calculated using samples from the posterior predictions. From the box and whisker plots (Fig. 3) for the computed CVRMSE values, it can be observed that the median CVRMSE using 14 days of training data lies below the interquartile range (the box) with 7, 28 and 56 days of training data, indicating that using 14 days of training data is likely to give lower CVRMSE and thus better prediction accuracy. The box plot of 14 training data is also comparatively shorter, denoting lower overall CVRMSE with different samples from the posterior predictions. These observations are further supported by the cumulative frequency graphs of the CVRMSE (Fig. 3), which clearly shows that using

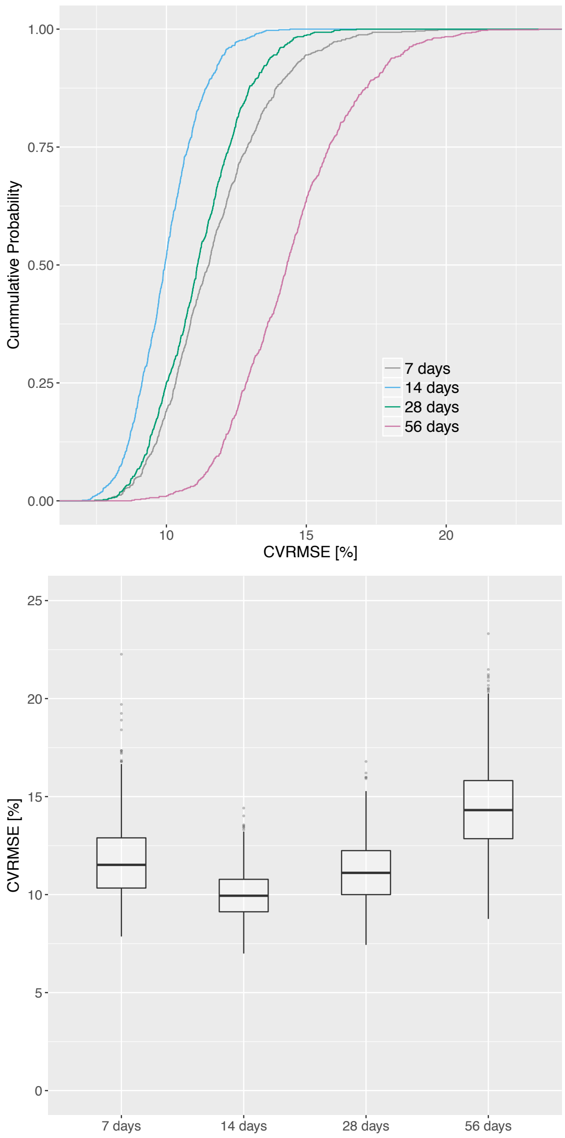

Figure 3: Cumulative distribution (top) and Box and whisker plots (bottom) showing the distribution of CVRMSE computed using samples from the posterior predictions generated by MCMC.

14 days of training data provides comparatively lower CVRMSE and thus better prediction accuracy.

A possible explanation for the observed variations in prediction accuracy is that Gaussian processes are distributions over functions. Thus, they are very flexible and prone to overfitting when the dataset is small. In our case study, 7 days of training data might be too small, resulting in lower predictive performance as indicated by the higher median CVRMSE as well as the larger variability in CVRMSE (Fig. 3). On the contrary, using 28 days of training data might result in noisy inputs leading to predictions with higher CVRMSE because of the increased uncertainties. This trend of increased uncertainties in the model prediction for a given data point is further illustrated by the cumulative distribution of the CVRMSE with 56 days of training data. The median CVRMSE increases as the amount of training data used for the calibration increases from 14 to 28 to 56 days (Fig. 3). Moreover, from the figure, it can be observed that the 25^{th} percentile of the CVRMSE using 56 days of training data lies above the 75^{th} percentile of the CVRMSE using 14 days of training data, implying a difference between the groups. Therefore, Fig. 3 provides corroborating evidence of the bias-variance tradeoff when using Gaussian processes for the Bayesian calibration of BEM. Using too little data can lead to overfitting or a high-variance model. On the contrary, using too many data points can lead to noisy inputs and therefore increased bias in the model. For our case study, using 14 days of training data seems to provide a balance. It should be noted that the maximum CVRMSE obtained using 14 days of training data is also less than the more stringent threshold of 15% set by ASHRAE Guideline 14 for monthly resolution data (ASHRAE, 2014).

Fig. 5 shows the posterior predictions when only weekday (left plot) and weekend (right plot) data was used for the Bayesian calibration respectively. From the figure, it can be observed that the posterior predictions overestimate the weekend measurements when only weekday data was used for the Bayesian calibration and underestimates weekday measurements when only weekend data was used for the Bayesian calibration. This is expected since buildings tend to consume more energy during weekdays when in operation. These results indicate that for accurate predictions, it is necessary to include both weekday and weekend data in the training data (step 1 in Fig. 1) used for the BEM calibration.

Conclusion

In this paper, a framework for the continuous Bayesian calibration of building energy models was proposed. The continuous Bayesian calibration is carried out following the principals of a receding horizon, which uses a sliding time window that moves forward at each sampling time. A Bayesian approach is proposed because it provides the flexibility to assimilate "existing data" to the new data through the use of priors. This is done by passing the posterior probability distributions of the previous model as prior probability distributions into the new model.

Through an actual case study building, we show that the proposed method is able to produce posterior predictions with CVRMSE distribution (metric for prediction accuracy) that fall entirely below the error thresholds specified by ASHRAE Guideline 14 (ASHRAE, 2014). The CVRMSE distribution was computed using the samples of posterior predictions generated using MCMC. The study was carried out using a 7-day sliding time window with varying amounts of training data. Results show that Bayesian calibration gave better prediction accuracy when 14 days of training data were used as opposed to using 7,

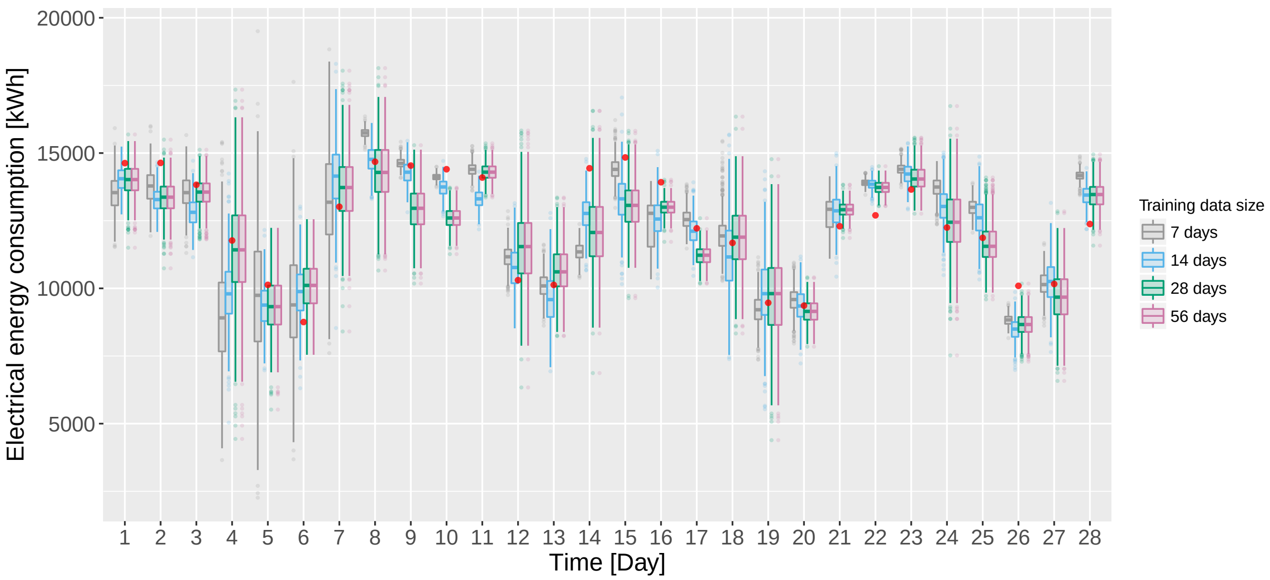

Figure 4: Box and whisker plots showing posterior predictions over test data with different amounts of training data.

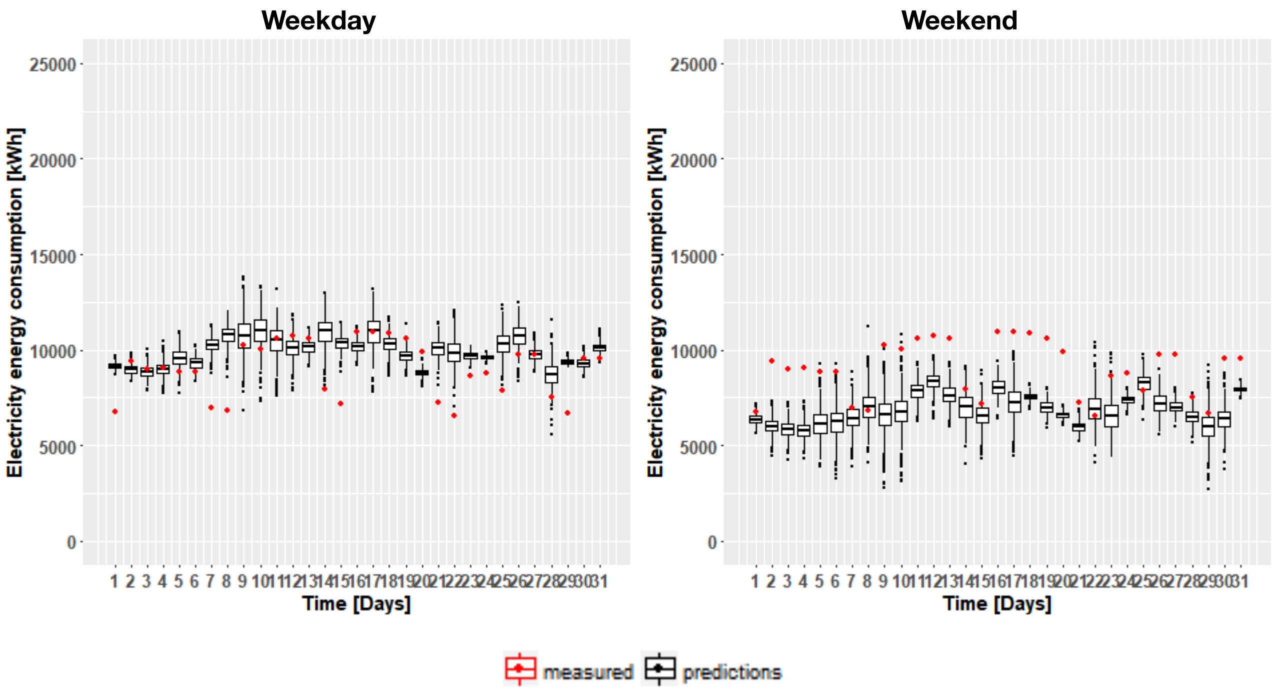

Figure 5: Posterior predictions using only weekday (left plot) and weekend (right plot) data for the Bayesian calibration.

28 and 56 days of training data respectively. This suggests a bias-variance tradeoff when varying amounts of training data is used to fit the Gaussian process model. The Gaussian process model acts as an emulator to aid the iterative Bayesian calibration process. An intuitive explanation for the observed results is that if the dataset is too small (7 days), the Gaussian process model is prone to overfitting resulting in poor generality. On the contrary, using more data (28 and 56 days) to fit the Gaussian process model might result in noisy inputs leading to greater uncertainties in the predictions and hence larger spread in CVRMSE distribution. Lastly, we show that including both weekday and weekend data in the dataset used for the Bayesian calibration is necessary for accurate posterior predictions. Therefore, it is important that we explicit sample from both weekday and weekend data when creating subsets of the data to keep the BEM calibration computationally tractable.

References

ASHRAE (2014). Guideline 14-2014, measurement of energy, demand, and water savings. *American Society of Heating, Refrigerating, and Air Conditioning Engineers, Atlanta, Georgia*.

Burnham, K. P. and D. R. Anderson (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research 33*(2), 261–304.

Chaudhary, G., J. New, J. Sanyal, P. Im, Z. O?Neill, and V. Garg (2016). Evaluation of "autotune" calibration against manual calibration of building energy models. *Applied Energy 182*, 115–134.

Chong, A. and K. P. Lam (2017). A comparison of MCMC algorithms for the Bayesian calibration of building energy models. In *Proceedings of the 15th IBPSA Building Simulation Conference*.

Chong, A., K. P. Lam, M. Pozzi, and J. Yang (2017). Bayesian calibration of building energy models with large datasets. *Energy and Buildings 154*, 343–355.

Chong, A. and K. Menberg (2018). Guidelines for the Bayesian calibration of building energy models. *Energy and Buildings 174*, 527–547.

Chong, A., W. Xu, S. Chao, and N.-T. Ngo (2019). Continuous-time bayesian calibration of energy models using BIM and energy data. *Energy and Buildings 194*, 177 – 190.

Coakley, D., P. Raftery, and M. Keane (2014). A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews 37*, 123–141.

Cover, T. and P. Hart (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory 13*(1), 21–27.

Dong, B., Z. O'Neill, Z. Li, D. Luo, S. Madhusudana, S. Ahuja, and T. Bailey (2012). An integrated infrastructure for real-time building energy modeling and fault detection and diagnostics. *Proceedings of SimBuild 5*(1), 448–455.

Heo, Y., G. Augenbroe, D. Graziano, R. T. Muehleisen, and L. Guzowski (2015). Scalable methodology for large scale building energy improvement: Relevance of calibration in model-based retrofit analysis. *Building and Environment 87*, 342–350.

Higdon, D., M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne (2004). Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing 26*(2), 448–466.

Ian Shapiro, P. (2009). Energy audits. *ASHRAE Journal 15*, 18–27.

Kennedy, M. C. and A. O'Hagan (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(3), 425–464.

Li, Q., G. Augenbroe, and J. Brown (2016). Assessment of linear emulators in lightweight Bayesian calibration of dynamic building energy models for parameter estimation and performance prediction. *Energy and Buildings 124*, 194–202.

Lim, H. and Z. J. Zhai (2017). Comprehensive evaluation of the influence of meta-models on Bayesian calibration. *Energy and Buildings 155*, 66–75.

Menberg, K., Y. Heo, and R. Choudhary (2017). Efficiency and reliability of Bayesian calibration of energy supply system models. In *Proceedings of the 15th IBPSA Building Simulation Conference*.

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

Pedrini, A., F. S. Westphal, and R. Lamberts (2002). A methodology for building energy modelling and calibration in warm climates. *Building and Environment 37*(8), 903–912.

Reddy, T. A. (2006). Literature review on calibration of building energy simulation programs: Uses, problems, procedures, uncertainty, and tools. *ASHRAE transactions 112*(1).

Sun, K., T. Hong, S. C. Taylor-Lange, and M. A. Piette (2016). A pattern-based automated approach to building energy model calibration. *Applied Energy 165*, 214–224.

Tian, W., Y. Heo, P. de Wilde, Z. Li, D. Yan, C. S. Park, X. Feng, and G. Augenbroe (2018). A review of uncertainty analysis in building energy assessment. *Renewable and Sustainable Energy Reviews 93*, 285–301.

Zhang, Z., A. Chong, Y. Pan, C. Zhang, S. Lu, and K. P. Lam (2018). A deep reinforcement learning approach to using whole building energy model for hvac optimal control. In *ASHRAE and IBPSA-USA SimBuild Building Performance Modeling Conference*.