A Comparison of MCMC Algorithms for the Bayesian Calibration of Building Energy Models for Building Simulation 2017 Conference

Adrian Chong[1], Khee Poh Lam[1]
[1]Center for Building Performance and Diagnostics, Carnegie Mellon University, Pittsburgh, USA

Abstract

Random walk Metropolis and Gibbs sampling are Markov Chain Monte Carlo (MCMC) algorithms that are typically used for the Bayesian calibration of building energy models. However, these algorithms can be challenging to tune and achieve convergence when there is a large number of parameters. An alternative sampling method is Hamiltonian Monte Carlo (HMC) whose properties allow it to avoid the random walk behavior and converge to the target distribution more easily in complicated high-dimensional problems. Using a case study, we evaluate the effectiveness of three MCMC algorithms: (1) random walk Metropolis, (2) Gibbs sampling and (3) No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014), an extension of HMC. The evaluation was carried out using a Bayesian approach that follows Kennedy and O'Hagan (2001). We combine field and simulation data using the statistical formulation developed by Higdon et al. (2004). It was found that NUTS is more effective for the Bayesian calibration of building energy models as compared to random walk Metropolis and Gibbs sampling.

Introduction

Detailed building energy models have been increasingly used in the analysis of building energy consumption and the evaluation of energy conservation measures. To ensure its reliability, model calibration has been recognized as an integral component to the overall analysis. A detailed description of the building's geometry, it's associated HVAC system, and the quantification of various internal loads is typically required as inputs to the model. However, detailed information is seldom available. Inarguably, uncertainty quantification becomes an important process in the use of detailed building energy models. Consequently, issues such as the calibration of input parameters, prediction accuracy, and prediction uncertainty would be of particular interest.

Many approaches for calibrating building energy models have been proposed, requiring various degrees of automation, manual tuning and expert judgment (Coakley et al., 2014). In particular, there has been increasing efforts in a Bayesian approach for the calibration of building energy models (Heo et al., 2012, 2015; Manfren et al., 2013; Chong and Lam, 2015; Li et al., 2016). This is because of its ability to quantify uncertainties in input parameters while at the same time reducing discrepancies between simulation output and physical measurements. In the Bayesian calibration of building energy models, Markov Chain Monte Carlo (MCMC) methods are a common way for sampling from the posterior distributions of the calibration parameters. Its widespread use can be attributed to its ease of use in a wide variety of problems. Two basic MCMC algorithms are random walk Metropolis and Gibbs sampling. random walk Metropolis is routinely used in the Bayesian calibration process due to its simple implementation. The random walk Metropolis algorithm (Metropolis et al., 1953) can be summarized as follows:

1. Arbitrarily select a valid initial starting point t^0.
2. Suppose $t^0, t^1, ..., t^i$ have been generated. Generate a candidate value t^* from a symmetric proposal distribution given t^i.
3. Calculate the Metropolis acceptance probability r, the probability of transitioning to the new candidate value

$$r = min\left\{\frac{p(t^*|y)}{p(t^i|y)}, 1\right\} \quad (1)$$

4. Accept and set t^{i+1} to the new candidate value with probability r or stay at the same point with probability $1 - r$.

$$t^{i+1} = \begin{cases} t^* \text{ with probability } r \\ t^i \text{ with probability } 1 - r \end{cases} \quad (2)$$

Gibbs sampling (Geman and Geman, 1984) proceeds by sampling each parameter from its conditional distribution while holding the remaining parameters fixed at their current values. To illustrate, suppose there are d parameters $t_1, t_2, ..., t_d$. At each iteration i, Gibbs sampling cycles through each parameter t_j, and samples it from its conditional distribution given

https://doi.org/10.26868/25222708.2017.336

the current value of the other parameters. This can be expressed by the following equation:

$$t_j^i \sim p(t_j | t_1^i, ..., t_{j-1}^i, t_{j+1}^{i-1}, ..., t_d^{i-1}) \quad (3)$$

where $t_1^i, ..., t_{j-1}^i, t_{j+1}^{i-1}, ..., t_d^{i-1}$ represents all other parameters at their current values except t_j.
An alternative sampling method that has been gaining interest is Hamiltonian Monte Carlo (HMC). HMC avoids the random walk behavior inherent in random walk Metropolis algorithm and Gibbs sampling by using first-order gradient information to determine how it moves through the target distribution (Hoffman and Gelman, 2014). The properties of HMC allows it to converge to the target distribution more quickly in complicated high-dimensional problems (Neal, 1993). However, HMC requires users to provide values of two hyperparameters: a step size ϵ and the number of steps L, making it difficult and time consuming to tune. To mitigate the challenges of tuning, the No-U-Turn Sampler (NUTS) was developed by Hoffman and Gelman (2014). NUTS uses a recursive algorithm to automatically tune the HMC algorithm without requiring user intervention or time consuming tuning runs was used.
Previous studies have been focused on the application of Bayesian calibration to building energy models without sufficient emphasis on the inference and assessment of convergence. Currently, the Bayesian calibration of a building energy model is considered to be complete when the model's output meets the error criteria set out by ASHRAE guideline 14 (ASHRAE, 2002). However, if the MCMC algorithm has not proceeded long enough, the generated samples may be grossly unrepresentative of the target posterior distributions (Gelman et al., 2014). In this paper, the objective is to evaluate the effectiveness of three MCMC algorithms (random walk Metropolis, Gibbs sampling and NUTS) within the Bayesian calibration framework by Kennedy and O'Hagan (2001).

Method

We evaluate the effectiveness of three MCMC algorithms (random walk Metropolis, Gibbs sampling and NUTS) by applying each algorithm to a Bayesian calibration approach that follows Kennedy and O'Hagan (2001). The process is as follows:

1. Build EnergyPlus model using construction drawings, design specifications and measured data
2. Conduct sensitivity analysis to reduce number of calibration parameters and avoid overfitting the model. Train a Gaussian process (GP) emulator to map the energy model's input parameters to the model output of interest.
3. Apply Bayesian calibration to the GP emulator (Kennedy and O'Hagan, 2001). The Bayesian calibration process would be repeated using different MCMC algorithms.

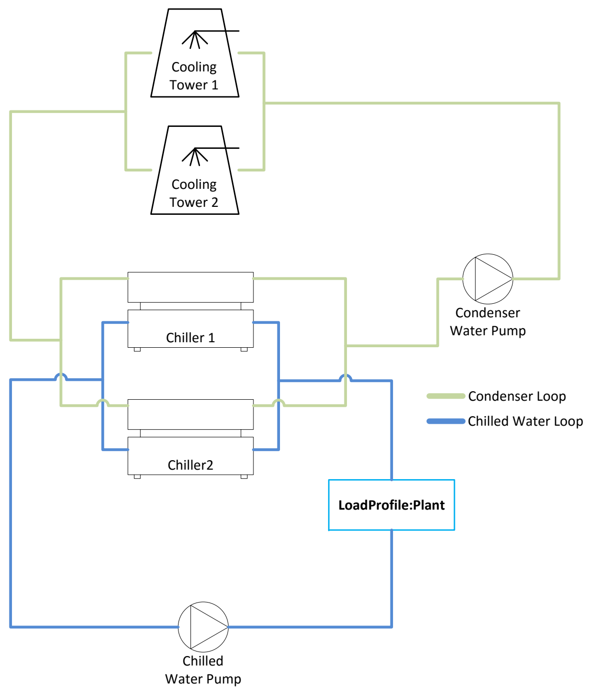

Figure 1: Diagram of cooling system modelled in EnergyPlus.

4. Compare the effectiveness of different MCMC algorithm using trace plots and Gelman-Rubin statistics to diagnose convergence to the posterior distribution.

We illustrate each step in the subsequent subsections with a case study.

EnergyPlus model

As a first step, the cooling system of a large ten-story office building located in Pennsylvania U.S.A was modelled using EnergyPlus version 8.5. The EnergyPlus model was built based on construction drawings, design specifications and site visits, and consists of the following functional parts (Figure 1): (a) Loads from cooling coil that transfers heat from air to water; (b) Two chillers connected in parallel that cools the water; (c) Chilled-water distribution pumps that send chilled water to the loads; (d) Condenser water pumps for circulation in the condenser loop; and (e) Two cooling towers in parallel that rejects heat from the chillers to the atmosphere. The EnergyPlus objects used to model these components include the LoadProfile:Plant object, the Chiller:Electric:EIR object, the Pump:VariableSpeed object and the CoolingTower:SingleSpeed object. Initial values were assigned to the model parameters based on measured data and design specifications (Table 1).
The LoadProfile:Plant object is used to simulate a scheduled demand profile when the coil loads are already known (LBNL, 2016b). This makes it possible to isolate and calibrate the HVAC system without any propagation of uncertainties due to calculation

of building loads. Hourly demanded loads were calculated based on the following equation (LBNL, 2016a):

$$Q_{load} = \dot{m}c_p(T_{in} - T_{out}) \quad (4)$$

where T_{out} and T_{in} denotes the outlet and inlet water temperature respectively; Q_{load} is the scheduled coil load; $\dot{m}$ is the mass flow rate; and c_p is the specific heat of water .
The Chiller:Electric:EIR object uses performance information at reference conditions along with three performance curves to determine the chiller's performance at off-reference conditions (LBNL, 2016a). The three performance curves are: (1) Cooling Capacity Function of Temperature Curve (CapFT) (Equation 5), (2) Energy Input to Cooling Output Ratio Function of Temperature Curve (EIRFT) (Equation 6), and (3) Energy Input to Cooling Output Ratio Function of Part Load Ratio Curve (EIRF-PLR) (Equation 7).

$$\begin{aligned} CapFT &= a_1 + b_1(T_{cw,l}) + c_1(T_{cw,l})^2 + \\ & d_1(T_{cond,e}) + e_1(T_{cond,e})^2 + f_1(T_{cw,l})(T_{cond,e}) \end{aligned} \quad (5)$$

$$\begin{aligned} EIRFT &= a_2 + b_2(T_{cw,l}) + c_2(T_{cw,l})^2 + \\ & d_2(T_{cond,e}) + e_2(T_{cond,e})^2 + f_2(T_{cw,l})(T_{cond,e}) \end{aligned} \quad (6)$$

$$EIRFPLR = a_3 + b_3(PLR) + c_3(PLR)^2 \quad (7)$$

$T_{cw,l}$ and $T_{cond,e}$ denotes the leaving chilled water temperature and entering condenser fluid temperature respectively; Q_{ref} and COP_{ref} are the chiller's capacity and coefficient of performance (COP) at reference conditions; and PLR is the chiller part-load ratio and equals $\frac{\text{cooling load}}{(Q_{ref})(CAPFT)}$. Using Equations 5 to 7, chiller power under a specific operating condition can be determined by the following equation.

$$P_{chiller} = \frac{Q_{ref}}{COP_{ref}}(CapFT)(EIRFT) \quad (8)$$

Inputs to this chiller model (Q_{ref}, COP_{ref}, regression coefficients of Equations 5, 6 and 7) were determined based on measured data using the reference-curve method that was proposed by Hydeman and Gillespie Jr (2002).
The Pump:VariableSpeed object calculates the power consumption of a variable speed pump using a cubic curve (Equation 9) (LBNL, 2016b).

$$FFLP = a_5 + b_5(PLR) + c_5(PLR)^2 + d_5(PLR)^3 \quad (9)$$

where $PLR = \frac{\text{Flow Rate}}{\text{Design Flow Rate}}$. Using Equation 9, pump power is calculated by the following equation.

$$P_{pump} = (P_{design})(FFLP)(Eff_{motor}) \quad (10)$$

where P_{design} is the design power consumption and Eff_{motor} is the motor efficiency. Inputs to the pump model were assigned using measurement of flow rate and pump power consumption. We assign a value of 1 to motor efficiency because pump motor inefficiencies are already accounted for in the measurements of flow and power. We use least squares regression to compute the coefficients of Equation 9 with PLR and $FFLP$ calculated as follows:

$$FFLP_i = \frac{power_i}{max(power_1, power_2, ..., power_n)} \quad (11)$$

$$PLR_i = \frac{flow_i}{max(flow_1, flow_2, ..., flow_n)} \quad (12)$$

Sensitivity analysis

Before calibrating the model, sensitivity analysis was performed to identify the parameters that have the most influence over the model's output. The objective is to reduce the number of calibration parameters and use only important factors in the calibration process. This not only reduces computation cost but also helps mitigate overfitting. Morris method (Morris, 1991) was used to carry out the sensitivity analysis. This was executed with R sensitivity package (Pujol et al., 2016).
Ten parameters in the cooling system were selected as uncertain (Table 1). Although the set of uncertain parameters are specific to this case study, they correspond to the set of parameters typically selected as random variables for the cooling system. All parameters were assigned a uniform distribution. Pump motor efficiency was varied between 0.6 and 1.0. The remaining 8 parameters were varied ±20% of their initial values. Design fan power and nominal capacity of cooling towers 1 and 2 were modeled as a single random variable because they have the same make and specification and were installed at the same time (Table 1). On the contrary, chillers 1 and 2 have very different capacity and COP at reference conditions and hence were modeled as separate random variables. We use the modified mean μ^* proposed by Campolongo et al. (2007) and standard deviation σ

Table 1: List of model parameters and their range.

Model parameter	Symbol	Initial Value	Min	Max
Chiller 1:				
Reference Capacity	θ_1	653378	522702	784053
Reference COP	θ_2	6.86	5.49	8.23
Chiller 2:				
Reference Capacity	θ_3	243988	195190	292785
Reference COP	θ_4	2.32	1.85	2.78
Chilled water pump:				
Design Power Consumption	θ_5	18190	14552	21828
Motor Efficiency	θ_6	1.0	0.6	1.0
Condenser water pump:				
Design Power Consumption	θ_7	11592	9274	13911
Motor Efficiency	θ_8	1.0	0.6	1.0
Cooling Tower 1 and 2:				
Design Fan Power	θ_9	11592	9274	13911
Nominal Capacity	θ_{10}	549657	439726	659589

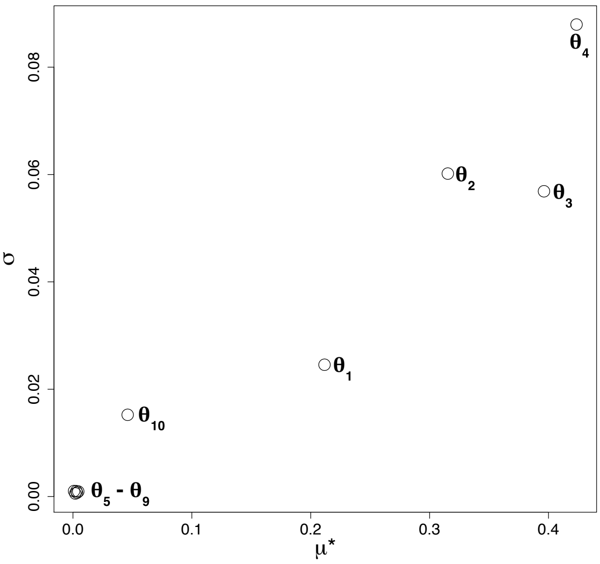

Figure 2: Sensitivity analysis (Morris method) of parameters in Table 1.

to determine which parameters are sensitive. Parameters $\theta_5 - \theta_9$ have μ^* and σ of approximately zero indicating that they are negligible parameters and should be excluded (Figure 2). Hence, only the top 5 parameters (θ_1, θ_2, θ_3, θ_4, and θ_{10}) would be used for the Bayesian calibration of the EnergyPlus model.

Bayesian calibration

A Bayesian calibration approach that follows that of Kennedy and O'Hagan (2001) was employed for this study. The formulation explicitly models uncertainty in calibration parameters, uncertainty due to discrepancy between the simulator and actual physical system, and observation errors as follows:

$$y(x) = \eta(x,t) + \delta(x) + \epsilon(x) \quad (13)$$

$\eta(x,t)$ denotes the building energy simulator output given input vector (x,t), where t represents the calibration parameters required as inputs to the energy model computed at known conditions x. Note that we make a distinction between the uncertain parameters θ and the calibration parameters t, where the calibration parameters t refer to the parameters that were selected from the set of uncertain parameters θ based on the results of the sensitivity analysis as described in the previous section. The term $\delta(x)$ is used to account for discrepancies between the simulator $\eta(x,t)$ and the actual physical system. $\epsilon(x)$ denotes observation error.

Field data and simulation data were combined using the statistical formulation developed by Higdon et al. (2004). Table 2 summarizes the data used to construct the field and simulation data for the case study. $\eta(x,t)$ denotes the output of the EnergyPlus simulation which depends on the observable inputs to the model x and the unknown calibration parameters

Table 2: Description of different parts used for Bayesian calibration of the case study.

Symbol	Description
$y(x)$	Observed hourly cooling energy consumption at corresponding values of x
$\eta(x,t)$	Hourly cooling energy consumption prediction using EnergyPlus at corresponding values of observable inputs x and unknown calibration parameters t
x	Observed hourly cooling coil load and chilled water flow rate
t	Calibration parameters $t_1 = \theta_1$, $t_2 = \theta_2$, $t_3 = \theta_3$, $t_4 = \theta_4$ and $t_5 = \theta_{10}$ (Table 1 and Figure 2). Values were set using LHS design

t. To learn about the calibration parameters t, EnergyPlus simulations were run at the same observable inputs x in our computer design of experiments. The corresponding calibration parameters t for the simulation runs were determined using Latin Hypercube sampling (LHS) to ensure sufficient coverage of the parameter space.

Since the energy model is computationally expensive to evaluate, a key element of this approach is the use of a Gaussian Process (GP) model to carry out the inference during the MCMC sampling procedure, mapping the energy model's input parameters to the model output of interest. A mean function $\mu(x,t)$ and covariance function $Cov((x,t),(x',t'))$ is required to specify a GP model. For simplicity, we specify a mean function that is set to zero and a covariance function that follows Higdon et al. (2004) with the form:

$$Cov((x,t),(x',t')) = \frac{1}{\lambda_\eta} exp\left\{ -\sum_{j=1}^{p} \beta_j^\eta |x_{ij} - x'_{ij}|^\alpha - \sum_{k=1}^{q} \beta_{p+k}^\eta |t_{ik} - t'_{ik}|^\alpha \right\} \quad (14)$$

where λ_η is the variance hyperparameter and β^η is the correlation hyperparameter of the GP model. The discrepancy term $\delta(x)$ was also modelled as a GP model with mean function set to zero and a covariance function of the form:

$$Cov(x,x') = \frac{1}{\lambda_\delta} exp\left\{ -\sum_{k=1}^{p} \beta_k^\delta |x_{ik} - x'_{ik}|^\alpha \right\} \quad (15)$$

α was set to 2 for both covaraiance functions. Finally, observations errors $\epsilon(x)$ was modelled as Gaussian noise:

$$\epsilon(x) \sim \mathcal{N}(0, I/\lambda_\epsilon) \quad (16)$$

For estimating the calibration parameters (θ), correlation hyperparameters (β^η and β^δ), and variance hyperparameters (λ_η, λ_δ and λ_ϵ), we use MCMC to explore and generate samples from their posterior distributions.

Comparison of MCMC algorithms

We compare the effectiveness of three MCMC algorithms: NUTS (a variant of HMC) and the more commonly used random walk Metropolis and Gibbs sampling. The comparison was carried out by checking for mixing and convergence to the target distribution using the following metrics

- Trace plots: trace plots are plots of the chains versus the sample index and can be useful for assessing convergence (Gelman et al., 2014). If the distribution of points remains relatively constant, it suggests that the chain might have converged to the stationary distribution. A trace can also tell you whether the chain is mixing well.
- Gelman-Rubin statistics ($\hat{R}$): $\hat{R}$ is the ratio of between-chain variance to within-chain variance and is based on the concept that if multiple chains have converged, there should be little variability between and within the chains (Gelman et al., 2014). For convergence, $\hat{R}$ should be approximately 1 ± 0.1.

For comparison, the same number of iterations were run for all three algorithms. Four independent chains of 10,000 iterations per chain were run for each MCMC algorithm with the first 5,000 iterations (50%) discarded as warmup/burn-in to reduce the influence of the starting values. For random walk Metropolis, an additional tuning of the acceptance ratio was required. It is generally accepted that the optimal acceptance rate of the Metropolis algorithm is about 20% (Gelman et al., 1996). We used a normal proposal/jumping distribution and tuned its variance until an acceptance rate of between 20% and 25% was achieved. The random walk Metropolis took a shorter time to run than the NUTS for a single chain of 10,000 iterations. However, after considering the iterative tuning process, NUTS ran faster since approximately three to four iterations were required to achieve an acceptance rate of about 20% with random walk Metropolis. Gibbs sampling took significantly longer to run than NUTS and random walk Metropolis, since the algorithm cycles through each parameter at each iteration.

Figures 3, 4, 5 and 6 provides a visual comparison of the trace plots (10,000 iterations including warmup) with samples generated by the three different MCMC algorithms. Random walk Metropolis demonstrates bad mixing for the calibration parameters t (Figure 3), indicating that the algorithm does not sufficiently explore the parameter space. After 10,000 iterations the calibration parameters have $\hat{R}$ between 1.89 and 2.51 (Table 3), indicating that the variance between the four independent chains are still greater than the

Table 3: $\hat{R}$ of calibration parameters with different MCMC algorithms.

Parameters	Random Metropolis	Gibbs Sampling	NUTS (HMC)
t_1	1.89	1.00	1.00
t_2	2.51	1.00	1.00
t_3	1.98	1.00	1.00
t_4	2.16	1.00	1.00
t_5	1.93	1.00	1.00

Table 4: $\hat{R}$ of hyperparameters with different MCMC algorithms.

Hyper-parameters	Random Metropolis	Gibbs Sampling	NUTS (HMC)
β_1^η	3.49	1.00	1.00
β_2^η	2.73	1.00	1.00
β_3^η	7.87	1.01	1.00
β_4^η	1.57	1.01	1.00
β_5^η	1.58	1.00	1.00
β_6^η	2.94	1.01	1.00
β_7^η	1.95	1.03	1.00
β_1^δ	2.46	1.00	1.00
β_2^δ	4.05	1.00	1.00
λ_η	33.26	1.00	1.00
λ_δ	302.38	1.05	1.00
λ_ϵ	1299.79	1.46	1.00

variance within. Gibbs sampling and NUTS performs better, with the calibration parameters t achieving adequate convergence (1 ± 0.1) after 10,000 iterations. Trace plots also shows good mixing for both Gibbs sampling and NUTS.

As expected, with random walk Metropolis, the correlation hyperparameters β_{1-7}^η (Figure 4) and $\beta_{1,2}^\delta$ (Figure 5) of the GP model do not appear to be stable. It is also clear that for β_1^η, β_2^η, β_3^η, β_1^δ and β_2^δ the different chains have not converged to a common distribution. On the contrary, trace plots of samples generated by Gibbs sampling and NUTS indicates rapid mixing for the correlation hyperparemeters β_{1-7}^η (Figure 4) and $\beta_{1,2}^\delta$ (Figure 5). Comparing $\hat{R}$ for the correlation hyperparemeters, Table 4 shows that after 10,000 iterations, the samples generated by random walk Metropolis have not converged yet ($1.5 < \hat{R} < 7.9$). However, samples generated by Gibbs sampling and NUTS have converged adequately with $\hat{R}$ within 1.0 ± 0.1 for all β^η and β^δ.

Figure 6 shows the trace plots for the variance hyperparameters λ_η, λ_δ, and λ_ϵ. From the figure, it can be observed that with random walk Metropolis, the chains are moving very slowly (due to low acceptance rates), and after 10,000 iterations the parallel sequences still have not converged to a common distribution. Gibbs sampling performs better, showing rapid mixing for λ_η and slower but adequate mixing for λ_δ. However, the trace plots show that λ_ϵ is moving very slowly through the parameter space, advancing to the target distribution only after 5,000 iterations. The small step size also suggests poor mixing and that more iterations are needed to achieve adequate convergence. On the contrary, NUTS is able

Table 5: $\hat{R}$ of calibration parameters and hyperparameters with 50, 500 and 2000 iterations and 4 independent chains using NUTS and Gibbs sampling. Values exceeding 1 ± 0.1 are in red font

	Number of Iterations					
	50		500		2000	
	NUTS	Gibbs	NUTS	Gibbs	NUTS	Gibbs
t_1	1.00	1.05	1.00	1.03	1.00	1.07
t_2	1.10	1.06	1.00	1.01	1.00	1.01
t_3	1.06	1.20	1.01	1.01	1.00	1.01
t_4	1.00	1.49	1.00	1.04	1.00	1.01
t_5	1.03	1.04	1.00	1.01	1.00	1.00
β_1^η	1.03	7.22	1.00	1.77	1.00	2.91
β_2^η	1.01	1.84	1.00	1.20	1.00	1.43
β_3^η	0.98	1.43	1.00	1.40	1.00	1.49
β_4^η	0.98	1.30	1.00	1.03	1.00	1.02
β_5^η	1.01	1.53	1.00	1.01	1.00	1.04
β_6^η	1.02	1.57	1.00	1.05	1.00	1.05
β_7^η	1.02	1.77	1.00	1.25	1.00	1.08
β_1^δ	0.97	1.29	1.00	1.00	1.00	1.02
β_2^δ	1.06	1.02	1.00	1.06	1.00	1.00
λ_η	1.00	1.39	1.00	1.02	1.00	1.01
λ_δ	1.11	17.97	1.00	2.15	1.00	1.42
λ_ϵ	0.98	179.00	1.00	38.66	1.00	11.08

generate samples of the variance hyperparameters effectively, as shown by the rapid mixing within each independent chain. Additionally, after 10,000 iterations, samples generated by random walk Metropolis still have very large $\hat{R}$, suggesting that the algorithm needs to be run much longer (Table 4). $\hat{R}$ for samples generated by Gibbs sampling perform much better. However, $\hat{R} = 1.46$ for λ_ϵ suggests that slightly more iterations is required before adequate convergence can be achieved. Samples generated by NUTS performs the best, having $\hat{R} = 1.00$ for all the variance hyperparameters.

To summarize, using random walk Metropolis results in very poor performance. After 10,000 iterations, none of the calibration parameters and hyperparameters achieved adequate convergence. $\hat{R}$ values were larger than 1.1 and the trace plots also show poor mixing. To achieve convergence with random walk Metropolis, the step size of the jumping distribution needs to be further tuned. Bad mixing may also be due to strong correlations in the parameter space.

Gibbs sampling show significantly better performance with all calibration parameters and hyperparameters achieving adequate convergence, with the exception of λ_ϵ. The trace plot and $\hat{R}$ of λ_ϵ suggests that it is close to converging to the posterior distribution. Hence running Gibbs sampling for slightly more iterations should result in adequate convergence for all calibration parameters and hyperparameters. NUTS shows the best performance with $\hat{R}$ values of exactly 1.00 for all calibration parameters and hyperparamters, indicating adequate convergence. Trace plots of all samples generated by NUTS also show rapid mixing. Furthermore, NUTS require a lot less iterations to converge (Table 5) due to the rapid mixing in the chains . After 50 iterations, $\hat{R}$ is already close to 1.00 for all calibration parameters and GP hyperparameters. With 500 iterations, the four independent chains have certainly achieved adequate convergence. In comparison to Gibbs sampling, after 50 iterations, almost all parameters have not converged (Table 5). After 500 and 2000 iterations, the calibration parameters t have adequately converged but several GP hyperparameters still have $\hat{R}$ larger than 1.1, indicating the variance between the four chains are still larger than the variance within. This suggests that it is harder to achieve adequate convergence for the GP hyperparameters and that users should pay greater attention to the assessing convergence of these hyperparameters.

Conclusion

The effectiveness of three MCMC algorithms (random walk Metropolis, Gibbs sampling and NUTS) were evaluated in this paper. An EnergyPlus model was first built. This is followed by using sensitivity analysis to reduce the number of calibration parameters. Measured data and simulation data were then combined using the statistical formulation developed by Higdon et al. (2004), which closely follows Kennedy and O'Hagan (2001) Bayesian calibration approach. Since the energy model is computationally expensive to evaluate, a key element of this approach is the use of a Gaussian Process (GP) emulator. Each of the three MCMC algorithms was separately used to estimate the posterior distributions of the calibration parameters (t) and the GP hyperparameters (β^η, β^δ, λ_η, λ_δ and λ_ϵ).

From the trace plots and Gelman-Rubin statistics ($\hat{R}$) of the samples generated by each algorithm, it was found that NUTS is able to achieve adequate convergence to the posterior distribution the fastest, with significantly lesser number of iterations. This study showed that using NUTS, it is possible to achieve adequate convergence with significantly lesser number of iterations (approximately 500 iterations) and no hand tuning at all. Random walk Metropolis showed the poorest performance with none of the parameters showing convergence after 10,000 iterations. Gibbs sampling showed significant improvements in sampling effectiveness as compared to random walk Metropolis but may require large number of iterations to achieve convergence. In conclusion, this study showed that for the Bayesian calibration of building energy models, compared to the commonly used random walk Metropolis and Gibbs sampling, NUTS was able to more effectively generate samples from the posterior distributions of the calibration parameters and GP hyperparameters.

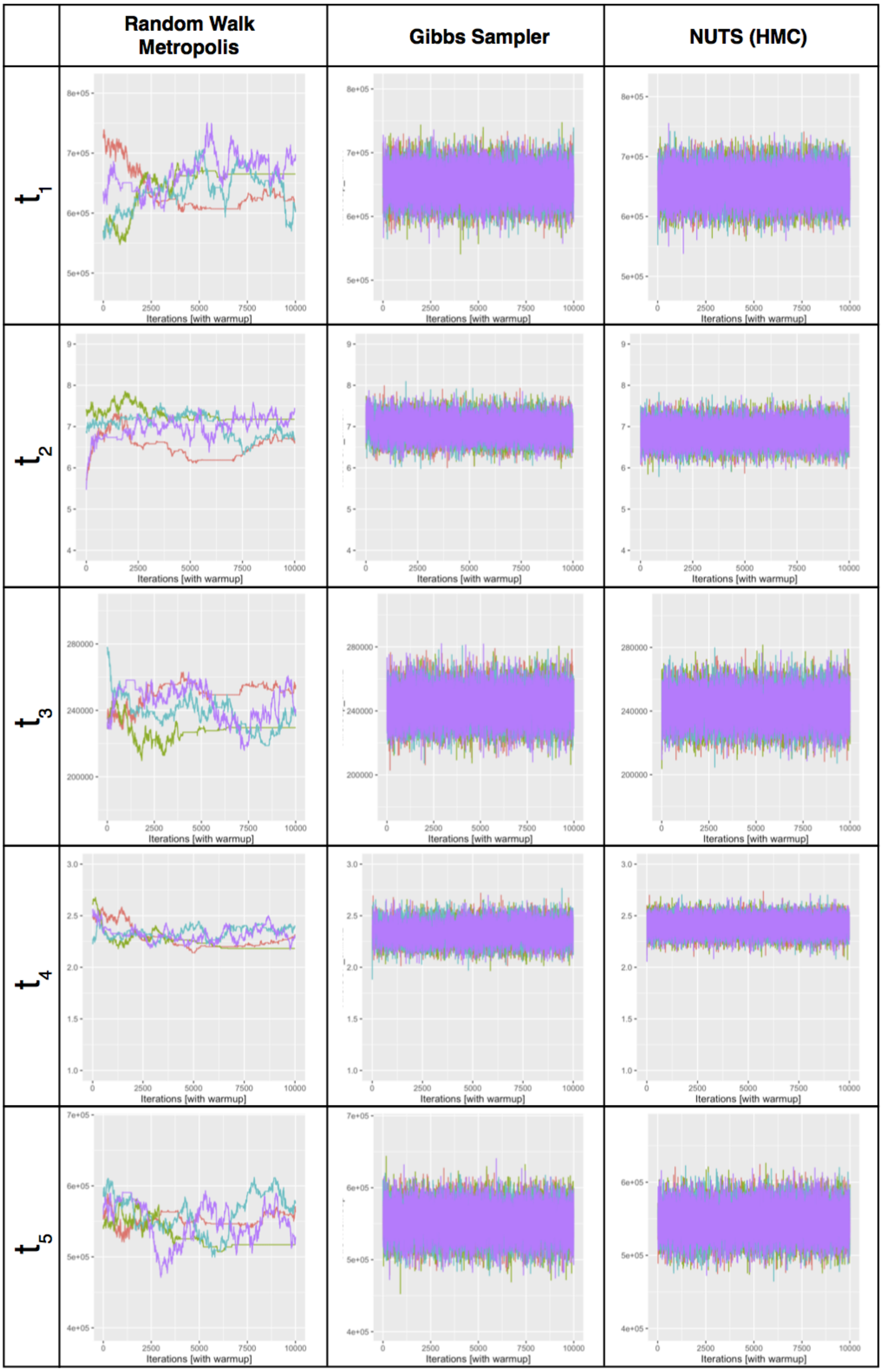

Figure 3: Trace plot of calibration parameters t (Table 1). Four independent chains of 10000 iterations per chain were run for each MCMC algorithm.

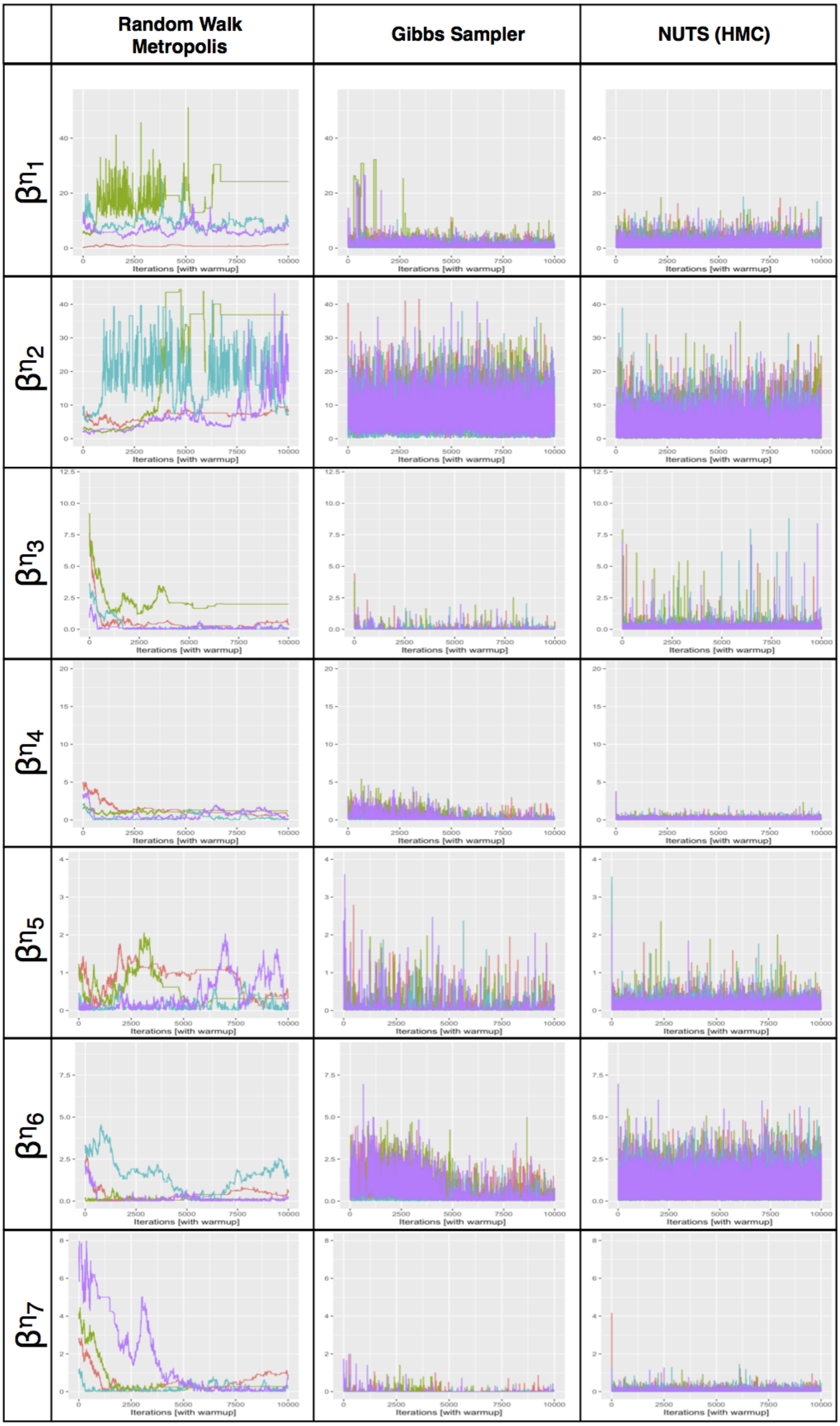

Figure 4: Trace plots of correlation hyperparameters β_1^η to β_7^η. Four independent chains of 10000 iterations per chain were run for each MCMC algorithm.

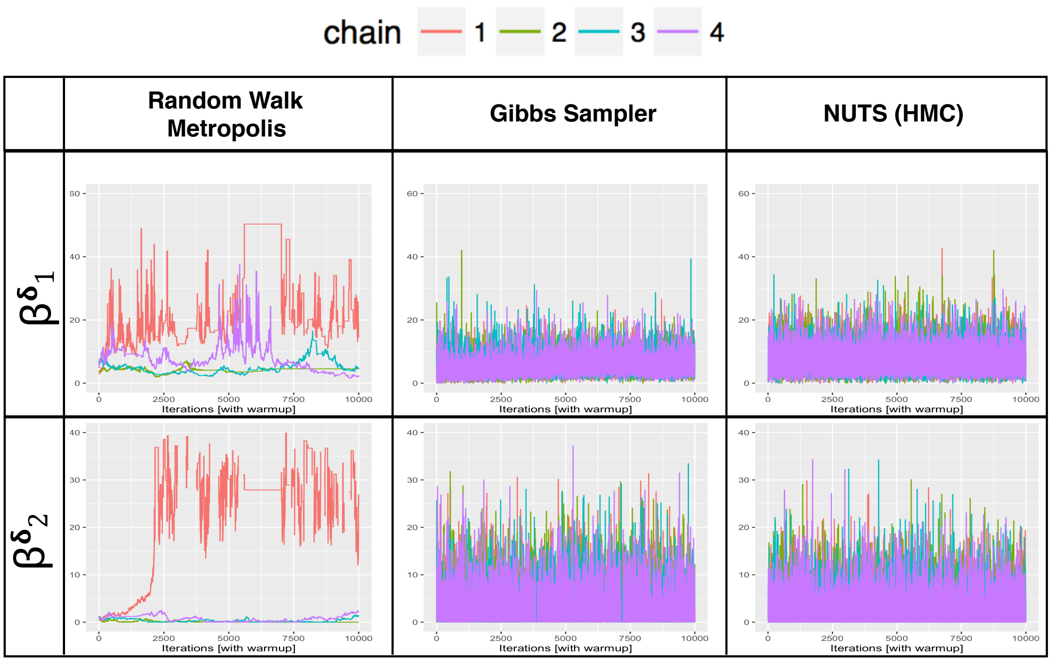

Figure 5: Trace plots of correlation hyperparameters β_1^δ and β_2^δ. Four independent chains of 10000 iterations per chain were run for each MCMC algorithm.

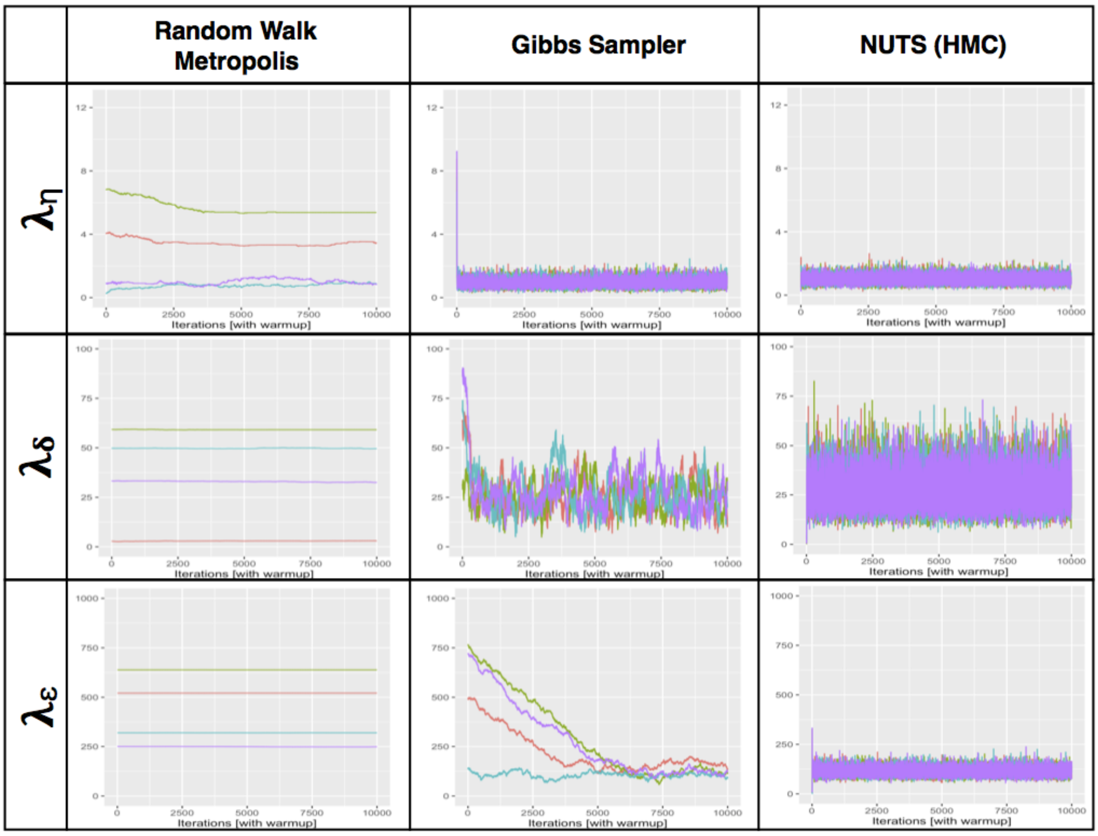

Figure 6: Trace plots of variance hyperparameters λ_η, λ_δ, and λ_ϵ. Four independent chains of 10000 iterations per chain were run for each MCMC algorithm.

References

ASHRAE (2002). Guideline 14-2002, measurement of energy and demand savings. *American Society of Heating, Ventilating, and Air Conditioning Engineers, Atlanta, Georgia*.

Campolongo, F., J. Cariboni, and A. Saltelli (2007). An effective screening design for sensitivity analysis of large models. *Environmental modelling & software 22*(10), 1509–1518.

Chong, A. and K. P. Lam (2015). Uncertainty analysis and parameter estimation of hvac systems in building energy models. In *Proceedings of the 14th IBPSA Building Simulation Conference*.

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.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2014). *Bayesian data analysis*, Volume 2. Taylor & Francis.

Gelman, A., G. O. Roberts, W. R. Gilks, et al. (1996). Efficient metropolis jumping rules. *Bayesian statistics 5*(599-608), 42.

Geman, S. and D. Geman (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence 6*, 721–741.

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.

Heo, Y., R. Choudhary, and G. Augenbroe (2012). Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings 47*, 550–560.

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.

Hoffman, M. D. and A. Gelman (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research 15*(1), 1593–1623.

Hydeman, M. and K. L. Gillespie Jr (2002). Tools and techniques to calibrate electric chiller component models/discussion. *ASHRAE Transactions 108*, 733.

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.

LBNL (2016a). Energyplus engineering reference: the reference to energyplus calculations. *US Department of Energy*.

LBNL (2016b). Energyplus input output reference: the encyclopedic reference to energyplus input and output. *US Department of Energy*.

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.

Manfren, M., N. Aste, and R. Moshksar (2013). Calibration and uncertainty analysis for computer models–a meta-model based approach for integrated building energy simulation. *Applied energy 103*, 627–641.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics 21*(6), 1087–1092.

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

Neal, R. M. (1993). Probabilistic inference using markov chain monte carlo methods.

Pujol, G., B. Iooss, A. J. with contributions from Khalid Boumhaout, S. D. Veiga, J. Fruth, L. Gilquin, J. Guillaume, L. Le Gratiet, P. Lemaitre, B. Ramos, T. Touati, and F. Weber (2016). *sensitivity: Global Sensitivity Analysis of Model Outputs*. R package version 1.12.2.