

Data-driven calibration of physics-informed models of joint building/equipment dynamics using Bayesian optimization

Chakrabarty, Ankush; Maddalena, Emilio; Qiao, Hongtao; Laughman, Christopher R.

TR2021-105 September 16, 2021

Abstract

Physics-informed simulation models of heating, ventilation, and cooling (HVAC) systems play a critical role in predicting system dynamics and enabling analysis, control, and optimization of buildings and equipment. The predictive performance of these simulation models are strongly linked to calibration mechanisms: algorithms that systematically select parameter values that optimize a given calibration-cost map (e.g., L-2 error). Poorly selected parameter values typically result in large deviations between measured building data and simulated data, limiting the utility of the simulation model in subsequent design. State-of-the-art calibration methods explore the parameter space by computing numerical gradients that are susceptible to measurement noise or employing population-based search mechanisms that require exorbitant data. To improve robustness and curtail data requirements, one can ‘learn’ or approximate the calibration-cost map and subsequently leverage the topology of the approximated function to find good search directions despite noisy measurements. Concretely, we employ machine learning to construct a calibration-cost map to direct model calibration for systems with joint dynamics of buildings and HVAC equipment. The learner explores subregions of the parameter space with high uncertainty and queries the model only where collecting simulation data yields useful information. This leads to lower simulation data-requirements compared to widely used calibration mechanisms.

Building Simulation Conference 2021

© 2021 MERL. This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Data-driven calibration of joint building and HVAC dynamic models using scalable Bayesian optimization

Ankush Chakrabarty^{1,*}, Emilio Maddalena², Hongtao Qiao¹, Christopher R. Laughman¹

¹*Mitsubishi Electric Research Laboratories, Cambridge, MA, USA.*

²*École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.*

**Corresponding Author. Email: chakrabarty@merl.com. Phone: +1 (617) 758-6175.*

Abstract

High-fidelity simulation models of coupled building and HVAC dynamics are typically expressed using grey-box simulators. These simulators comprise systems of differential algebraic equations, typically informed by physics, that need to be calibrated to real-world data. In this paper, we describe a framework for calibrating simulation model parameters using scalable Bayesian optimization (BO). BO uses probabilistic learners to approximate objective functions and leverages statistical information to efficiently explore and exploit the parameter space. This approach typically requires fewer simulations than Monte-Carlo methods or population-based algorithms, without relying on gradient information. To render the Bayesian framework tractable in high-dimensional parameter spaces, we also provide a scalable BO framework employing sparse Gaussian process regression. We demonstrate that our proposed approach can simultaneously calibrate 17 parameters (including radiative emissivities, heat transfer coefficients, and thickness of walls/floors) of a Modelica model of joint building and HVAC dynamics with >85% accuracy, and 13 of those 17 with > 90% accuracy.

Key Innovations

- A novel model calibration algorithm is proposed based on Bayesian optimization that can be applied to simulation models with internal dynamics and representations that are partially (grey-box) or completely (black-box) unknown.
- The problem of joint calibration of building and HVAC dynamical models is considered.
- The algorithm uses probabilistic machine learning methods to learn a calibration-cost function with unmodeled building/HVAC dynamics.
- The model calibration mechanism is gradient-free and works with few model simulations by exploiting uncertainty quantification.
- The proposed method is also modified to perform with a large number of parameters and data points using sparse Gaussian processes.

Practical Implications

We use a timeout in the simulator to stop the simulation after 300 s of CPU time, and exclude that param-

eter combination from the training set of the learner and the admissible parameter space. We chose 300 s as our time-out because we noticed that only certain poor choices of parameters resulted in increased stiffness of the underlying dynamical equations and required over 2000 s to simulate, eventually ending in oscillatory outputs. To reduce wastage of simulation time, we timed out at 300 s, since good choices of parameters resulted in simulations terminating successfully within 200 s. Of course, if we simulated for longer, or for other models, one would have to carefully select this time-out based on empirical evidence regarding how long good simulations take.

Introduction

Simulation models of heating, ventilation, and cooling (HVAC) systems play a critical role in predicting system dynamics and enabling analysis, control, and optimization of buildings and equipment (Kim et al., 2016). Key advantages of these physics-based modeling approaches are that their structure and parameter values are often available from construction documents, and that the information encoded in their mathematical structure tends to demonstrate accurate predictive properties in comparison to more generic model structures. These predictive capabilities often come at the cost of increased nonlinearity and numerical stiffness, which can make these models difficult to solve and calibrate.

Calibration mechanisms are strongly linked to the predictive performance of a physics-based model for a given building. Initial parameters obtained from tables of physical properties or architectural drawings may deviate from the actual materials or geometry used in the built environment, while other model parameters (e.g. heat transfer coefficients) may only be derived from correlations or empirical observations. The parameter values that most accurately represent observed data must therefore be systematically identified by algorithms that optimize a given calibration-cost map, such as the well-studied coefficient of variation of root-mean squared error.

Calibration-cost functions often present numerical challenges because they are to be quite nonlinear, their sensitivities vary locally, and are seldom convex or differentiable with respect to the calibrated param-

eters. Measured data is also often corrupted by environmental and process noise, limiting the effectiveness of gradient-based methods. Population-based, gradient-free searches are scalable and effective, but incur high computational expenditure as they require an exorbitant number of simulations (Garrett and New, 2015; Asadi et al., 2019), which may not be suitable for stiff dynamical models that require long simulation times. Solely relying on dynamical estimators such as Kalman filters can also limit calibration performance due to multi-rate dynamics and poor generalizability of linearized models typically used to design these estimators (Bortoff and Laughman, 2019; Alam et al., 2018).

Coupled interactions between the HVAC system and the building envelope must also be accounted for during calibration. Although calibration of each of these models is often performed independently, many variables contain information about both the envelope and the HVAC system. The calibration of the joint system thus has the potential to yield more accurate parameters. The calibration of this joint model is more challenging because of the large number of states and parameters, as well as the fact that the joint system often has dynamics over a wider range of time constants than the subsystem models, implying that the algorithms should be able to perform in both data-frugal and data-abundant (equivalently, data-poor and data-rich) settings.

Machine learning techniques can be used to approximate a calibration-cost map to direct model calibration for systems with joint dynamics of buildings and HVAC equipment. The learner explores subregions of the parameter space with high uncertainty and queries the model only where collecting simulation data yields useful information. Intelligent sampling leads to lower simulation data-requirements compared to widely used calibration mechanisms (Bamdad et al., 2020). Bayesian calibration frameworks (Heo et al., 2015b; Tian et al., 2016; Berger et al., 2016) are particularly useful in this context because they provide uncertainty quantification capabilities, as they not only incorporate prior knowledge one might have of the system at hand, but also provide confidence envelopes around the nominal predictions specifying its degree of certainty without requiring a large amount of data (Li et al., 2016). The most widely used Bayesian algorithms for calibrating dynamical models and energy models fall under the umbrella of Markov chain Monte Carlo (MCMC) methods (Chong and Lam, 2017; Menberg et al., 2017; Chong et al., 2019). While these methods provide demonstrably excellent solutions for building energy calibration with few unknown parameters, they suffer from three major limitations: they typically require a large number of iterations for ‘burn-in’ and exhibit slow convergence in high-dimensional parameter spaces (Menberg et al., 2017), they require pre-

conditioning steps such as sensitivity analysis which is itself computationally expensive, and they require learning of multi-input multi-output maps via probabilistic learning methods to replace the emulator. In our work, we demonstrate that our proposed GP-based Bayesian optimization method requires few iterations to converge, does not require prior sensitivity analysis since it automatically determines relevance of parameters within the training phase, and a multi-input one-output map is learned (which reduces learning complexity) since we only seek to approximate the calibration-cost function rather than replacing the more complex underlying physics-informed model, thereby preserving fidelity.

Gaussian processes have been previously been shown to be particularly effective for calibration of building energy maps (Zhang et al., 2013; Gengembre et al., 2012; Tresidder et al., 2011). In all these case studies, however, the dynamics of the underlying system are not considered, and the number of parameters and/or parameter uncertainty is small enough such that convergence happens within 150 iterations. As pointed out by Østergård et al. (2018), the computational stability and efficiency of GP-based methods scale poorly when the number of iterations of the calibration algorithm is large. Sparse Gaussian processes (SGPs) are crucial for overcoming this important limitation (Quiñonero Candela and Rasmussen, 2005; Bauer et al., 2016).

Methods

An overview of the proposed calibration framework is presented in Figure 1.

Black-box model calibration

We denote by

$$y_{0:T} = \mathcal{M}_T(\theta)$$

a general model of the joint building and HVAC dynamics, parameterized by $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$. The admissible set of parameters Θ is assumed to be known at calibration time; for instance, Θ could denote a set of upper and lower bounds on parameters, obtained from physics or practical experience. The output vector $y_{0:T} \in \mathbb{R}^{n_y \times T}$ contains all measured outputs from the building system obtained over a time period $[0, T]$.

Note that a particular advantage of our proposed work is that we *do not require an underlying mathematical description of the dynamics at design time*. That is, $\mathcal{M}_T(\theta)$ could be a completely black-box model, where simulating $\mathcal{M}_T(\theta)$ forward with a fixed (and admissible) set of parameters θ yields a vector of outputs

$$y_{0:T} := [y_0 \quad y_1 \quad \cdots \quad y_t \quad \cdots \quad y_T],$$

with each output measurement $y_t \in \mathbb{R}^{n_y}$.

This assumption implies that our method is applicable to a wide range of models proposed in the

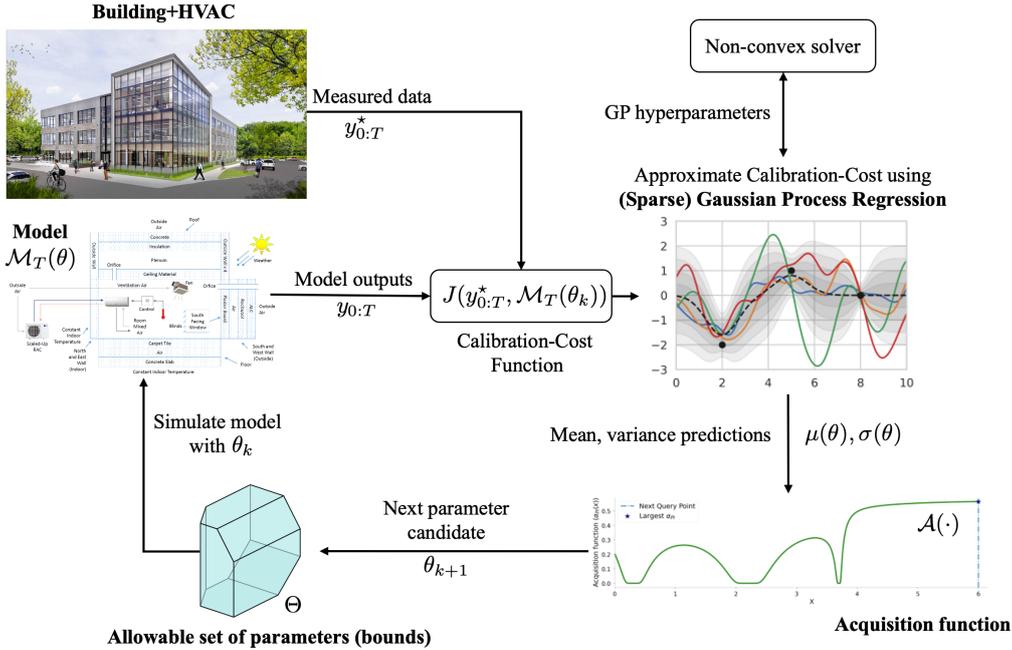


Figure 1: Schematic diagram of the proposed Bayesian optimization based calibration method.

literature. For instance, consider the commonly used (Chong et al., 2019; Heo et al., 2015a) building energy model

$$y_t = \eta(x_t, \theta) + \delta(x_t) + \epsilon(x_t),$$

where η denotes the energy prediction, δ is the model discrepancy, and ϵ is the observation error. Clearly, by recursively simulating this model from $t = 0$ to $t = T$, one can obtain a representation that conforms to our abstracted model $\mathcal{M}_T(\theta)$. Another example could be a state-space description (Guo et al., 2021) of the joint dynamics such as

$$\begin{aligned} \dot{x}_b &= f_b(x_b, x_e, u_b, w_b, \theta_b), \\ \dot{x}_e &= f_e(x_e, x_b, u_e, w_e, \theta_e), \\ y &= h(x_b, x_e, u_b, u_e, v_b, v_e) \end{aligned}$$

can be used by defining $\theta := [\theta_b, \theta_e]$ and integrating from $[0, T]$ to yield a model of the form $\mathcal{M}_T(\theta)$. Here x , u , w , and v denote states, inputs, process noise, and measurement noise, respectively, and the subscripts b and e correspond to the building and equipment, respectively.

For performing data-driven model calibration, we require measured output data $y_{0:T}^*$ from the true system. Concretely, our objective is to obtain the optimal set of parameters θ^* such that the modeling error $y_{0:T}^* - \mathcal{M}_T(\theta^*)$ is minimized (in some sense). To this end, we propose a calibration-cost function $J \equiv J(y_{0:T}^*, \mathcal{M}_T(\theta))$ and the following optimization problem to find the optimal parameters

$$\theta^* = \arg \min_{\theta \in \Theta} J(y_{0:T}^*, \mathcal{M}_T(\theta)). \quad (1)$$

While the designer is free to select any calibration-cost function J in (1), we employ

$$J := \log \left(\sum_{t=0}^T (y_t^* - y_t)^\top W (y_t^* - y_t) \right), \quad (2)$$

where W is a $n_y \times n_y$ positive-definite matrix that is used to assign importance or scale the output errors; recall $y_{0:T} = \mathcal{M}_T(\theta)$. The natural logarithm $\log(\cdot)$ promotes good numerical conditioning of the cost J by avoiding very large or very small costs.

We solve the problem (1) by intelligently extracting samples from the parameter space Θ , forward simulating the model $\mathcal{M}_T(\theta)$ from $[0, T]$ to obtain y_t , and computing the cost $J(y_{0:T}^*, y_{0:T})$; the values and history of these cost function evaluations drive future sampling and eventually yield good solutions (asymptotically, the optimal solution). This avoids dependence on the underlying description of $\mathcal{M}_T(\theta)$, making our method amenable to grey- or black-box models. The number of samples required to obtain good solutions to (1) in high-dimensional parameter spaces grows exorbitantly due to the curse of dimensionality. Bayesian optimization reduces sampling complexity by directing posterior samples based on updating prior knowledge, enabling calibration of models with a large number of parameters.

Scalable Bayesian optimization via sparse GPs

The Bayesian optimization (BO) algorithm typically consists of two steps that balance exploration and exploitation (Snoek et al., 2012). Probabilistic machine learning methods such as Gaussian processes (GPs) are used to approximate the map from the parameter space to the calibration-cost function J (Williams and

Rasmussen, 2006). By learning a probabilistic representation, one can use the approximation to generate a predictive distribution for J at each parameter θ . Furthermore, the predictive distribution is used to direct subsequent search locations, with a primary focus on sub-regions of Θ where the function most likely contains the global solution θ^* that minimizes the cost (1). A benefit of using GPs is that the predictive distributions are Gaussian, and therefore, uncertainty quantification is computationally tractable and interpretable via simple statistics such as means and standard deviations. The GPs are used to define a prior distribution over functions. The underlying assumption made is that the calibration-cost function J to be optimized has been generated from such a prior distribution, characterized by a zero mean and a kernelized covariance function $\mathcal{K}(\theta, \theta')$. The covariance function \mathcal{K} is singularly responsible for defining the characteristics of the function such as smoothness, robustness to additive noise, and so on. While many kernel functions are available, we have found (empirically) that the Matérn 3/2 function provides a good approximation of calibration-cost functions.

Assume that we have already evaluated the objective at N_θ input samples. Let this training data be denoted by $\{(\theta_k^D, J(\theta_k^D) + \nu_k)\}_{k=1}^{N_\theta}$, where $\nu_k \sim \mathcal{N}(0, \sigma_n^2)$ is additive white noise in the measurement channel with zero-mean and unknown covariance σ_n^2 . Given a set of hyperparameters (such as length scales l in each dimension of the parameter space, kernel variance σ_0 , and noise variance σ_n) for a pre-selected kernel, one can compute the matrices

$$K_D(\theta) = [\mathcal{K}(\theta, \theta_1^D) \quad \cdots \quad \mathcal{K}(\theta, \theta_{N_\theta}^D)]$$

and (note the calligraphic \mathcal{K})

$$\mathcal{K}_D = \begin{bmatrix} \mathcal{K}(\theta_1^D, \theta_1^D) & \cdots & \mathcal{K}(\theta_1^D, \theta_{N_\theta}^D) \\ \vdots & \ddots & \vdots \\ \mathcal{K}(\theta_{N_\theta}^D, \theta_1^D) & \cdots & \mathcal{K}(\theta_{N_\theta}^D, \theta_{N_\theta}^D) \end{bmatrix}.$$

With these matrices, we can compute a posterior distribution characterized by a mean function $\mu(\theta)$ and variance function $\sigma^2(\theta)$ given by

$$\mu(\theta) = K_D(\theta)^\top \mathcal{K}_n^{-1} J(\theta), \quad (3a)$$

$$\sigma^2(\theta) = \mathcal{K}(\theta, \theta) - K_D(\theta)^\top \mathcal{K}_n^{-1} K_D(\theta), \quad (3b)$$

with $\mathcal{K}_n = \mathcal{K}_D + \sigma_n^2 I$. Now, the accuracy of predicted mean and variance is strongly linked to the selection of the kernel and the best (in some sense) set of hyperparameters such as l , σ_0 and σ_n . There are a variety of methods to obtain these hyperparameters, but the most common method involves maximizing the log-marginal likelihood function

$$\mathcal{L} = -\frac{1}{2} \log |\mathcal{K}_n| - \frac{1}{2} J(\theta)^\top \mathcal{K}_n^{-1} J(\theta) - \frac{p}{2} \log 2\pi. \quad (3c)$$

This problem is non-convex but can be solved using quasi-Newton methods or adaptive gradient methods (Williams and Rasmussen, 2006).

Unfortunately, inversion and determinant operations typically used in exact GPs result in cubic complexity with the number of data points, which implies that using exact GPs in moderate or high-dimensional parameter spaces is not practical, since finding good solutions in these spaces typically requires sampling and evaluating the calibration-cost function a large number of times. Sparse GP (SGP) techniques curtail expensive operations on $N_\theta \times N_\theta$ kernel matrices by constructing a low-rank approximation $\tilde{\mathcal{K}}_D$ of the exact matrix \mathcal{K}_D . From its definition and with the aid of the Woodbury identity and additional lemmas, the expressions describing the posterior mean and variances of SGPs can be computed efficiently by exploiting rank deficiency of the matrices, therein greatly reducing the computational complexity during training (Quiñero Candela and Rasmussen, 2005).

In particular, we consider the variational free energy (VFE) method (Titsias, 2009), which is an SGP framework that approximates the exact GP's posterior density by minimizing the mismatch between distribution using the Kullback-Leibler (KL) divergence metric. Optimizing such an objective leads to an intrinsic robustness against overfitting, since the data are not considered directly. The modified likelihood function used for VFE training is given by

$$\begin{aligned} \tilde{\mathcal{L}} = & -\frac{1}{2} \log |\tilde{\mathcal{K}}_D + \sigma_n^2 I| - \frac{1}{2} J(\theta)^\top (\tilde{\mathcal{K}}_D + \sigma_n^2 I)^{-1} J(\theta) \\ & - \frac{1}{2\sigma_n^2} \text{Tr}(\mathcal{K}_D - \tilde{\mathcal{K}}_D) - \frac{p}{2} \log 2\pi, \end{aligned}$$

for which expensive operations are performed on the low rank $\tilde{\mathcal{K}}_D$ rather than the dense \mathcal{K}_D . Here, $\text{Tr}(\cdot)$ denotes the trace operator. Maximizing $\tilde{\mathcal{L}}$ can be used not only to select kernel hyperparameters, but also to optimize the so called inducing points. Inducing points are a small set of representative points used by SGP techniques to summarise the information contained in the large set of original data, thus alleviating the associated computational burden and enabling BO at scale; for more details, we refer the reader to (Quiñero Candela and Rasmussen, 2005).

The exploration-exploitation trade-off in BO methods is performed via an acquisition function $\mathcal{A}(\cdot)$. The acquisition function uses the predictive distribution given by the SGP to compute the expected utility of performing an evaluation of the objective at each set-point θ . The next parameter at which the calibration-cost function has to be evaluated is given by $\theta_{N_\theta+1} := \arg \max \mathcal{A}(\theta)$. As this function only depends on the SGP approximated function and not on the true calibration-cost J , evaluating $\mathcal{A}(\cdot)$ is a cheap operation. In this work, we use a lower confidence bound (LCB) acquisition function given by

$$\mathcal{A}_{\text{LCB}} = \mu(\theta) - \kappa \sigma(\theta), \quad (4)$$

where κ is a scalar usually chosen ≥ 1.5 . This acquisition function estimates the expected improvement

of the steady-state power generated by the next set-point versus the current best solution. An efficient way of computing the maximum of this acquisition function is by generating random samples on Θ , computing \mathcal{A}_{LCB} for each sample, and choosing the sample maximum as the next parameter to be evaluated.

After a suitable number of iterations N_θ , the SGP is expected to learn the underlying function J and the best solution obtained thus far is deemed the best set of calibration parameters. The selection of N_θ is a design decision: it is usually informed by practical considerations such as the total amount of evaluations of J , that is, the total amount of simulations one can run with a practical time budget.

Results and Discussion

Model Description

We construct a model of the vapor compression cycle, comprising a compressor, condensing heat exchanger, electronic expansion valve, and evaporating heat exchanger. The cycle behavior is dominated by the heat exchangers over the time scales of interest, so the overall cycle model used dynamic models of the heat exchangers and algebraic (static) models of the compressor and expansion valve (Qiao et al., 2017). For the sake of simplicity, a lumped parameter method was used to characterize the dynamics of refrigerant flow in the heat exchangers. A Helmholtz equation-of-state based model was used to describe the refrigerant properties, while an ideal gas mixture was used for the moist air model.

The building models were constructed from the open-source Modelica Buildings library (Wetter et al., 2014). The room model is based on the physics-based behavior of the fundamental materials and commonly used components, while the zone air model is a mixed air single-node model with one bulk air temperature that interacts with all of the radiative surfaces and thermal loads in the room.

The building model consists of a one-story residence with nominal 2009 IECC-based construction, based on the model used by (Laughman et al., 2019). This residence has a floor area of 112.24 m² and is 2.6 m tall, and is oriented along the cardinal directions with a peak occupancy of 3 people. Each exterior wall also has a window of 1.52 m \times 2.72 m that admits solar heat gains into the spaces. A 10 cm thick concrete slab and 2 m of soil below the house was also included to characterize interactions with the thermal boundary condition under the house, which was set to a constant 21 °C. A peaked attic was also included with a maximum height of 1.5 m, so that the building model includes two thermal zones.

This building envelope model was connected to the cycle model, and a proportional integral (PI) controller was implemented on the heat pump which used the compressor frequency to regulate the room tem-

perature and the expansion valve position to regulate the evaporator superheat temperature. The controller also implemented anti-windup to maintain stability while enforcing minimum and maximum actuator limits. The connection between the HVAC system and the building envelope through mass and energy relations (e.g., coil and room temperatures, mass and specific energy of the room air) impose an acausal coupling between these models, so that data from the building model includes information about the HVAC system, and vice versa. This joint model was simulated using the Atlanta-Hartsfield TMY3 file, and included convective and radiative heat loads of 2 W/m² and a latent load of 0.6 W/m² between the hours of 8 AM and 6 PM, with weather-driven disturbances outside of these hours. This model was exported from Modelica using the Functional Mockup Interface ¹, and the resulting functional mockup unit (FMU) was imported into Python using the FMPy package² to enable seamless integration of advanced machine learning modules.

The inputs and outputs of this model were chosen to be similar to those which may be observed in a realistic experimental setting. The inputs of the heat pump include the room temperature set-point, the evaporator superheat set-point, and the indoor and outdoor fan speeds. The inputs for the building envelope model include the convective, radiative, and latent heat loads as well as the weather variables provided in the TMY3 standard. These heat loads may be estimated to reasonable accuracy via occupancy detection, load surveys, or other similar methods.

Table 1 lists the parameters of the building and the heat pump selected to evaluate the efficacy of this new calibration method. Similar numbers of parameters were selected from the envelope model and the cycle model to study the accuracy for each subsystem. These particular parameters were chosen because they are often difficult to measure in practice or to estimate from other physical quantities. For example, the refrigerant-side heat transfer coefficients (HTCs) depend on the amount of oil circulating in the pipe, the detailed configuration of tubes in the heat exchanger, and many other system and site-specific quantities. On the envelope side, the radiative emissivities in the IR and solar spectra were similarly selected because of their potential experimental variability. We then bounded the ranges of the parametric variation based on our field experience, though other alternate ranges could be easily used.

Calibration Performance

We collect ground-truth data for calibration by simulating the Modelica model from July 1-14 (thus $T = 14$ days) with the parameters of the model

¹Modelica, Functional Mockup Interface for Model Exchange and Co-Simulation, Version 2.0.1.

²<https://github.com/CATIA-Systems/FMPy>

PARAMETER	VARIABLE	TRUE VALUE	UB [%]	BEST ESTIMATE	ACC [%]
Building Parameters					
Airflow infiltration rate	VFlowExt	3.368×10^{-2}	± 15	3.492×10^{-2}	96.3
Thickness of the floor	xFloorVal	1.016×10^{-1}	± 15	1.087×10^{-1}	93.1
Infrared emissivity of roof (outer)	IR-Roof-a	9.000×10^{-1}	± 15	9.253×10^{-1}	97.2
Solar emissivity of roof (outer)	Sol-Roof-a	9.000×10^{-1}	± 15	7.923×10^{-1}	88.0
Infrared emissivity of roof (inner)	IR-Roof-b	7.000×10^{-1}	± 15	7.068×10^{-1}	99.0
Solar emissivity of roof (inner)	Sol-Roof-b	7.000×10^{-1}	± 15	7.898×10^{-1}	87.2
Interior room air HTC	hInt	3.000	± 15	2.955	98.5
Exterior air HTC	hExt	1.000×10^1	± 15	1.026×10^1	97.4
HVAC Parameters					
Outdoor HEX HTC adjustment factor	pfHTC-a	1.000	± 15	9.472×10^{-1}	94.7
Indoor HEX HTC adjustment factor	pfHTC-b	1.000	± 15	9.044×10^{-1}	90.4
Indoor HEX Lewis number	Le-a	8.540×10^{-1}	± 15	9.044×10^{-1}	94.1
Outdoor HEX vapor HTC	HTC-vap-a	5.000×10^2	± 15	5.082×10^2	98.4
Outdoor HEX 2-phase HTC	HTC-2ph-a	3.000×10^3	± 15	3.107×10^3	96.4
Outdoor HEX liquid HTC	HTC-liq-a	7.000×10^2	± 15	5.968×10^2	85.3
Indoor HEX vapor HTC	HTC-vap-b	5.000×10^2	± 15	4.969×10^2	99.4
Indoor HEX 2-phase HTC	HTC-2ph-b	2.000×10^3	± 15	1.945×10^3	97.2
Indoor HEX liquid	HTC-liq-b	7.000×10^2	± 15	8.047×10^2	85.1

Table 1: Description of parameters, true values, and uncertainty expressed as a percentage from the true value. (HTC = heat transfer coefficient, HEX = heat exchanger, UB = uncertainty bound, ACC = accuracy)

set to their true values (see Table 1). The 8 measured output sequences $y_{0:T}^*$ of the model are collected at 5 minute intervals, and we do not split the data into weekdays and weekends as sometimes required (Chong et al., 2019).

Hereafter, we use GP and SGP interchangeably, since we use SGP as our learner. We initialize the SGP by choosing 100 randomly selected parameter samples from within the bounds Θ associated with each parameter (see Table 1). With each of these initial parameters samples, we simulate the Modelica model for the same time interval as the ground truth and obtain the estimated output sequence $y_{0:T}$. Subsequently, we evaluate the cost function (2) for each of the initial samples with $y_{0:T}$ and true outputs $y_{0:T}^*$. This initial collection of parameters and calibration-cost values is used to construct the initial training set of the SGP. We use Matérn 3/2 kernels with dimension-wise separate length-scales, since the admissible parameter space is not normalized. We construct the GP using the Python library `gpfLOW`, and use 500 epochs of a limited-memory BFGS (L-BFGS) solver to obtain the optimal hyperparameters for training. Unlike MCMC methods that require tens of thousands of iterations to converge, we set our BO method to run for 750 iterations, that is, the Modelica model is simulated 750 (BO iterations) + 100 (initial) = 850 times from $[0, T]$. We select the acquisition function to be a lower-confidence-bound (4) with $\kappa = 1.96$. For acquisition function maximization, we adopt a uniform random sampling approach with 10,000 samples. This sampling is cheap since it only requires evaluation of the SGP, rather than the simulation model. The specific SGP framework used is VFE with 100 inducing points.

Table 1 shows that the best estimates of the parameters after 750 iterations are quite close to the true pa-

rameter values. Indeed, 13 of the 17 parameters (6/8 building and 7/9 equipment) are captured to above 90% accuracy³, despite using only one 2-week dataset and no additional pre-processing such as sensitivity analysis or data splitting. It is noteworthy that the 4 parameters with the lowest fits, highlighted in gray in this table, were all $> 85\%$. Given the lack of heat transfer from the surfaces affected by these parameters, the relatively poor quality of these fits matches our intuition for the low sensitivity of the measured outputs to these parameters.

As per ASHRAE Guideline 14, a CVRMSE of $< 15\%$ indicates a good model fit with acceptable predictive capabilities (ASHRAE, 2014). In order to illustrate our calibration performance, we report the CVRMSE and the NMBE⁴ metrics (Ruiz and Bandera, 2017) in Table 2 for each of the model outputs. All of the parameters respect the ASHRAE guidelines in terms of the CVRMSE metric, showing the potential of our calibration mechanism and modeling approach. The highest CVRMSE is exhibited by the suction superheat, which can likely be attributed to sharp peaks produced in the signal related to large changes in the compressor speed and valve position during the rapid increase in the load caused by the morning solar load and the presence of occupants.

Although major advantages of our proposed approach are that there is no burn-in period to acquire a practical distribution and that no initial parameter guesses are required, the sequential nature of BO means that

³Accuracy of the k -th parameter θ_k is computed by $100 \times (1 - |\theta_k^{\text{true}} - \theta_k^*|/\theta_k^{\text{true}})$.

⁴CVRMSE (Coefficient of Variation of the Root Mean Square Error) measures the variability of the errors between measured and simulated values. NMBE (Normalized Mean Bias Error) is the normalized average of the error sequence. We refer the reader to (Ruiz and Bandera, 2017) for their mathematical definitions.

OUTPUT	CVRMSE	NMBE	MBE
Room Temp.	< 0.1%	< 0.1%	< 0.1 K
Room Humidity	1.01%	11.50%	0.08
Ceiling Temp.	< 0.1%	< 0.1%	< 0.1 K
Attic Temp.	< 0.1%	< 0.1%	< 0.1K
Suction Superheat	6.03%	0.62%	0.03 K
Ambient Temp.	< 0.1%	< 0.1%	< 0.01 K
Compressor Freq.	0.99%	1.41%	0.43 Hz
Exp. Valve Pos.	< 0.1%	< 0.1%	0.05

Table 2: Output calibration performance metrics.

the calibration performance is dictated strongly by the quality of the initial SGP model. We therefore tested our proposed approach for robustness to initial conditions in which the calibration mechanism was run 50 times, with different initial random seeds (that is, different samples were extracted for the initial GP construction). The results of these simulation studies are shown cohesively in Figure 2. The median (horizontal orange line), quartiles (horizontal box lines), and range (whisker ends) for the best parameter set obtained over 50 runs are shown using boxplots, with the true parameter value shown with a ‘*’. We deduce from these plots that the best parameter estimates are close to their true values, with (predictably) the worst calibration performance exhibited by the inner roof parameters. Interestingly, the liquid heat transfer coefficients do not exhibit significant decline over runs, but the Lewis number does. It is likely that this variation can be attributed to the time varying moisture removal rate of the evaporating heat exchanger and the dependence of the indoor relative humidity on both ambient conditions and internal latent loads.

Conclusions

In this paper, we developed a Bayesian optimization methodology for calibrating black-box models without exorbitant simulations. We utilized sparse GPs as surrogate models for the calibration-cost function, and demonstrated the accuracy and efficacy of our proposed approach on a Modelica model of a real building with 17 tunable parameters, exhibiting over 90% accuracy on 13 of 17 parameters. We showed that the outputs we estimate are mostly < 1% CVRMSE and < 1% NMBE. We also perform further testing on our proposed method and demonstrate that the method is robust to the initial set of samples used to construct the initial sparse GP model. In future, we plan to test this method on real-world data which will be noisy and with larger model mismatch.

References

Alam, M. et al. (2018). Applying extended Kalman filters to adaptive thermal modelling in homes. *Advances in Building Energy Research* 12(1), 48–65.

Asadi, S. et al. (2019). Building energy model calibration using automated optimization-based algorithm. *Energy and Buildings* 198, 106–114.

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

Bamdad, K., M. E. Cholette, and J. Bell (2020). Building energy optimization using surrogate model and active sampling. *Journal of Building Performance Simulation* 13(6), 760–776.

Bauer, M., M. van der Wilk, and C. E. Rasmussen (2016). Understanding probabilistic sparse gaussian process approximations. *NeurIPS* 29, 1533–1541.

Berger, J. et al. (2016). Bayesian inference for estimating thermal properties of a historic building wall. *Building and Environment* 106, 327–339.

Bortoff, S. A. and C. R. Laughman (2019). An Extended Luenberger Observer for HVAC Application using FMI. In *Modelica*, pp. 157–015.

Chong, A. and K. P. Lam (2017). A comparison of MCMC algorithms for the bayesian calibration of building energy models. In *Proc. of the 15th IBPSA Building Simulation Conference*, Volume 4.

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.

Garrett, A. and J. New (2015). Scalable tuning of building models to hourly data. *Energy* 84, 493–502.

Gengembre, E. et al. (2012). A Kriging constrained efficient global optimization approach applied to low-energy building design problems. *Inverse Problems in Science and Engineering* 20(7), 1101–1114.

Guo, Z. et al. (2021). Aggregation and data driven identification of building thermal dynamic model and unmeasured disturbance. *Energy and Buildings* 231, 110500.

Heo, Y. et al. (2015a). Evaluation of calibration efficacy under different levels of uncertainty. *Journal of Building Performance Simulation* 8(3), 135–144.

Heo, Y. et al. (2015b). Scalable methodology for large scale building energy improvement: Relevance of calibration in model-based retrofit analysis. *Building and Environment* 87, 342–350.

Kim, D. et al. (2016). System identification for building thermal systems under the presence of unmeasured disturbances in closed loop operation. *Building and Environment* 107, 169–180.

Laughman, C. et al. (2019). Modeling and control of radiant, convective, and ventilation systems for multizone residences. In *Proc. of Building Simulation 2019*, pp. 1956–1963.

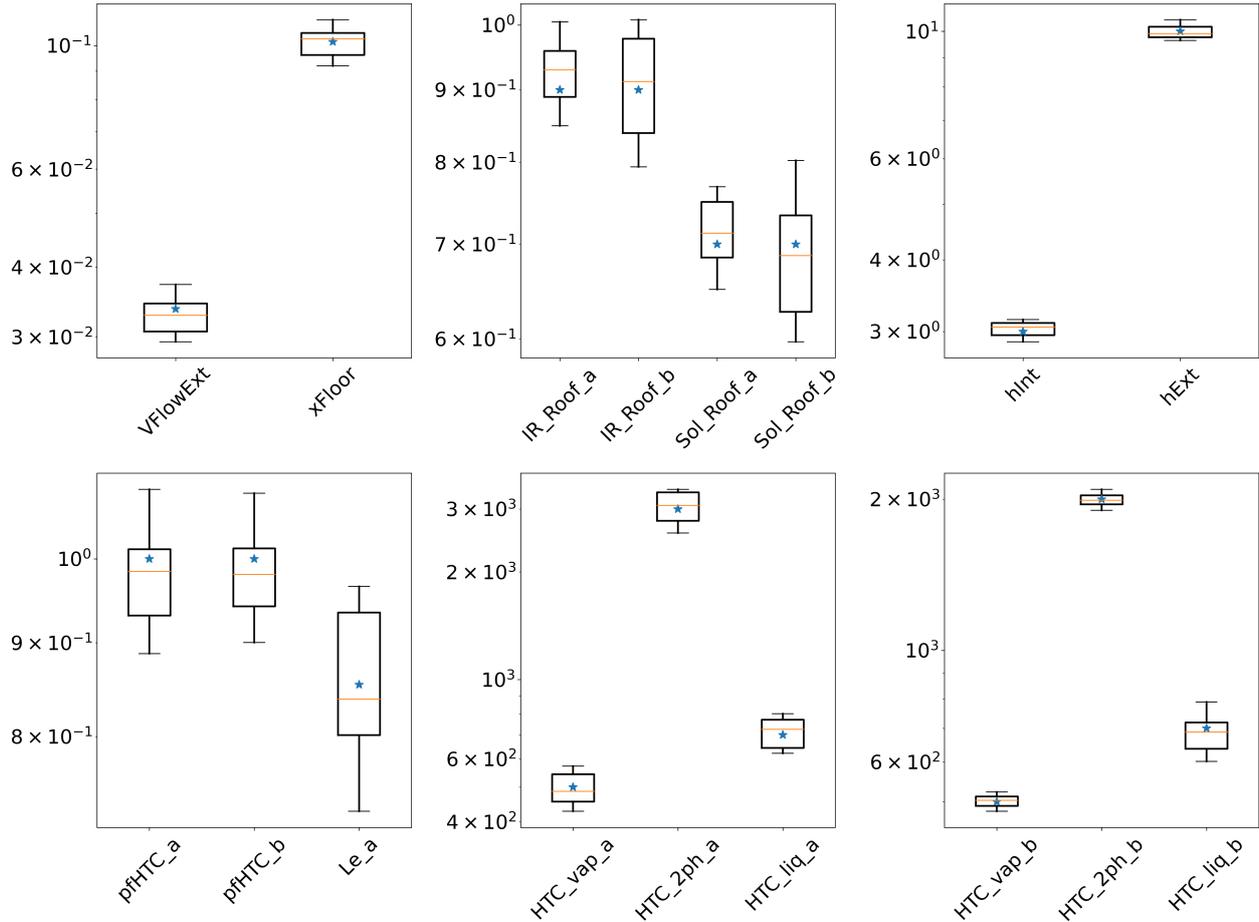


Figure 2: Boxplots of median (orange line) and interquartile range (boxes) of all parameters along with true values (star).

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.

Menberg, K., Y. Heo, and R. Choudhary (2017). Efficiency and Reliability of Bayesian Calibration of Energy Supply System Models. *Proc. of the 15th IBPSA Building Simulation Conference* (1), 1212–1221.

Østergård, T., R. L. Jensen, and S. E. Maagaard (2018). A comparison of six metamodeling techniques applied to building performance simulations. *Applied Energy* 211, 89–103.

Qiao, H. et al. (2017). Dynamic Characteristics of an R-410A Multi-split Variable Refrigerant Flow Air-conditioning System. In *12th IEA Heat Pump Conf.*

Quiñonero Candela, J. and C. E. Rasmussen (2005). A unifying view of sparse approximate Gaussian process regression. *JMLR* 6(Dec), 1939–1959.

Ruiz, G. R. and C. F. Bandera (2017). Validation of calibrated energy models: Common errors. *Energies* 10(10), 1587.

Snoek, J., H. Larochelle, and R. P. Adams (2012). Practical bayesian optimization of machine learning algorithms. *NeurIPS* 25, 2951–2959.

Tian, W., S. Yang, Z. Li, S. Wei, W. Pan, and Y. Liu (2016). Identifying informative energy data in bayesian calibration of building energy models. *Energy and Buildings* 119, 363–376.

Titsias, M. (2009). Variational learning of inducing variables in sparse Gaussian processes. In *AISTATS*, pp. 567–574.

Tresidder, E., Y. Zhang, and A. I. Forrester (2011). Optimisation of low-energy building design using surrogate models. In *Proc. of Building Simulation*, pp. 1012–1016.

Wetter, M. et al. (2014). Modelica buildings library. *Journal of Building Performance Simulation* 7(4), 253–270.

Williams, C. K. and C. E. Rasmussen (2006). *Gaussian Processes For Machine Learning*, Volume 2. MIT press Cambridge, MA.

Zhang, R. et al. (2013). Optimal selection of building components using sequential design via statistical surrogate models. In *Proc. of the 13th IBPSA Building Simulation Conference*, pp. 2584–2592.