

Emerging Computational Methods to Support the Design and Analysis of High
Performance Buildings

by

Kevin Cant

B.ASc., University of British Columbia, 2016

A Dissertation Submitted in Partial Fulfilment of the
Requirements for the Degree of

Master of Applied Science

in the Department of Civil Engineering

© Kevin Cant, 2022
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by
photocopying or other means, without the permission of the author.

Emerging Computational Methods to Support the Design and Analysis of High
Performance Buildings

by

Kevin Cant

B.ASc., University of British Columbia, 2016

Supervisory Committee

Dr. R. Evins, Supervisor
(Department of Civil Engineering)

Dr. P. Mukhopadhyaya, Departmental Member
(Department of Civil Engineering)

Supervisory Committee

Dr. R. Evins, Supervisor
(Department of Civil Engineering)

Dr. P. Mukhopadhyaya, Departmental Member
(Department of Civil Engineering)

ABSTRACT

This thesis presents three emerging computational methods: machine learning, gradient-free optimization, and Bayesian modelling. Each method is showcased in its ability to enable energy savings in new and existing buildings when paired with dynamic energy models. Machine learning algorithms provide rapid computational speed increases when used as surrogate models, supporting early-stage designs of buildings. Genetic algorithms support the design of complex interacting systems in a reduced amount of effort. Finally, Bayesian modelling can be leveraged to incorporate uncertainty in building energy model calibration. These methods are all readily available and user-friendly, and can be incorporated into current engineering workflows.

Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	iv
List of Tables	vii
List of Figures	viii
Acknowledgements	x
1 Introduction	1
1.1 Summary of Contributions by Author	3
2 Analysis of Feature Importance in Modeling Ground Source Heat Pump Systems Using Broad Parametric Analysis, Load Characterization and Artificial Neural Networks	5
2.1 Introduction	6
2.2 Existing Work and Contribution	6
2.2.1 Model Structure	8
2.2.2 Building Definition	8
2.2.3 Geo-exchange Field	9
2.2.4 Ground Source Heat Pump Plant	11
2.2.5 Parameter Sampling	12
2.2.6 Model Characteristics and Feature Selection	12
2.2.7 Feature Importance Selection	13
2.2.8 Simplified Feature Grouping	14
2.3 Results	15
2.3.1 Load Characterization	15

2.3.2	Feature Importance	16
2.3.3	Reduction of Features Accuracy	17
2.4	Outlook and Conclusions	18
3	Optimizing VAV Terminal Box Minimum Positions using Dynamic Simulations to Improve Energy and Ventilation Performance	21
3.1	Introduction	22
3.1.1	VAV Systems	22
3.1.2	Ventilation Rate Procedure	23
3.1.3	New research on VAV box minimums	23
3.1.4	Competing Objectives of Energy and Indoor Air Quality	24
3.1.5	Novelty of Research	25
3.2	Methods	26
3.2.1	Dynamic Simulation	26
3.2.2	Ventilation Calculation	27
3.2.3	Genetic Algorithm	28
3.3	Results	28
3.3.1	Quantifying Ventilation Performance Using Dynamic Simulations	29
3.3.2	Optimizing Energy and Ventilation by Selecting Minimum Box Position on a Zone-By-Zone Basis	30
3.4	Discussion	30
3.4.1	Limitations and Future Research	31
3.5	Conclusion	33
4	Improved Calibration of Building Models using Approximate Bayesian Calibration and Neural Networks	36
4.1	Introduction	37
4.1.1	Bayesian Calibration in Building Energy Models	40
4.1.2	Key Contributions of This Work	43
4.2	Implementation	44
4.2.1	Case Study Building	44
4.3	Results	47
4.3.1	Sampling and Neural Network	48
4.3.2	Baseline Energy Model and Standard Practice	48
4.3.3	Calibration Results	51

4.4	Discussion	58
4.4.1	Limitations and Further Work	61
4.4.2	Conclusions	62
5	Conclusions	63
	Bibliography	64

List of Tables

Table 2.1	List of building design parameters for <i>Medium Office</i>	10
Table 2.2	List of design parameters for the <i>GHE field</i>	11
Table 2.3	List of GSHP design parameters for <i>Medium Office</i>	12
Table 3.1	VRP Results	29
Table 4.1	ASHRAE Guideline 14 Error Thresholds	43
Table 4.2	Parameters incorporated in the calibration procedure.	47
Table 4.3	ASHRAE Guideline 14 error metric results for the three example cases, the baseline case, and the MAP case.	50
Table 4.4	The MAP values and associated likelihood comparison between a scenario including parameter screening, and a scenario including the full-suite of uncertain parameters.	57

List of Figures

Figure 2.1 Model Overview.	9
Figure 2.2 Annual heating and cooling load profiles with temperature of fluid leaving ground, (sample for 90 th , 50 th , and 10 th percentile total heating demand)	16
Figure 2.3 Distribution of characteristic results across full solution set (sample size: 9440)	17
Figure 2.4 Feature importance based on R2 score loss for objectives across test set	18
Figure 2.5 Predicted vs. actual objective performance across test sets . . .	19
Figure 3.1 Depiction of the DoE medium office archetype, a 3-storey core-and-shell energy model.	27
Figure 3.2 Weekday occupancy schedule for the office archetype.	28
Figure 3.3 Energy and ventilation performance comparison of potential design options.	34
Figure 3.4 Details of the NSGA-II optimized designs.	35
Figure 4.1 A summary of the main issues affecting calibration of building energy models.	38
Figure 4.2 A 3D virtual representation of the case study building.	45
Figure 4.3 Mean absolute percentage error and mean absolute error, plus one standard deviation, of the surrogate model prediction compared to the energy model outputs at identical parameter input combinations.	49
Figure 4.4 Monthly energy consumption various meters across three years of operation.	51
Figure 4.5 A comparison of prior distribution, posterior distribution, baseline case value, example cases values, and MAP value for each parameter.	52

Figure 4.6 Monthly energy consumption across three years of operation and associated posterior error distributions.	55
Figure 4.7 A point-wise comparison for each meter on its overall sensitivity to normalized annual outputs with the amount of information gain for that parameter between the prior and posterior distributions.	56
Figure 4.8 A comparison of the five most important parameters' prior and posterior distributions.	57
Figure 4.9 The error metrics for each meter from a scenario with only the five most sensitive parameters.	57
Figure 4.10 Monthly energy consumption various meters across three years of operation, plus error distributions.	59

ACKNOWLEDGEMENTS

First off, thank you to Dr. Ralph Evins for creating an environment where my half-baked ideas were built up instead of shut down, and where I was given ample opportunity to do the research that I believed was important.

Thank you to Mark Lucuik and Kyle Fridgen for the mentorship when I first joined the industry; and to Andrew Pape-Salmon, Cora Hallsworth, and Damian Stathonikos for the opportunity to keep my ideas grounded.

A huge thank you to Dave Rulff and Maddy Seattle, my time at UVic would not have been the same without your friendship. And to my colleagues in the Energy in Cities lab, time spent collaborating on projects was enjoyable and effortless.

To Robyn, I could not have done this without your support. You are a bright spot in each and every day.

And most important, to my parents Kim and Steve. Thank you for teaching me the value of hard work and instilling in me a curiosity and drive in the pursuit of knowledge.

Hard things are hard.

Barack Obama

Chapter 1

Introduction

Significant improvements need to be made in the building construction and engineering industry. Buildings account for approximately 30% of global energy consumption, and reducing emissions in new and existing buildings by 59-65% is key to meeting climate targets in British Columbia [1]. How to cost-effectively engineer high-performance buildings remains a non-trivial challenge facing the industry today.

Unfortunately, the building engineering industry is fraught with arduous and time-consuming workflows and methods. Building systems design relies on rules of thumb, standards, and conservative estimates [2]; the end result is risk-adverse designs that perform sub-optimally in order to save on up-front engineering effort and offload risks.

Energy consumption in buildings is regularly 30-50% higher than predicted during design stages [3]. This is partly due to decisions made early on in the design process. These decisions typically become locked-in as they are exceedingly expensive to remedy as projects move forward. Devising evidence during early stages of design is particularly challenging because buildings tend to be unique be-spoke developments with a vast number of design parameters and possibilities [4].

When compared to other major industries, construction is fragmented and inefficient, however, there remains a significant opportunity to leverage automation and digitization. Engineers, likewise, need to prepare and modernize their workflows. There remain opportunities to transform the manual time-consuming task effort to value-added analysis and detailed design.

Issues arise not only in the design of new buildings, but also in the operation of existing buildings. It has been estimated that up to 30% of energy consumed in the existing building stock is due to inefficient operation [5]. While the capital cost to remedy these problems is much less than major system upgrades, the up-front labour

cost to identify the inefficiencies and suggest improvements is still a barrier to uptake [6].

A further barrier to reducing energy consumption and carbon emissions is the lack of confidence surrounding building retrofit decisions. Standard modelling practice does not adequately represent the uncertainty of calibrated models [7]. It has been shown that property owners require a higher internal return on investment for energy reduction measures compared to other financial investments [8].

Advances in computational power have unlocked the capabilities of demanding analytical algorithms, including machine learning, Bayesian modelling, and gradient-free optimization. In recent years these tools have been implemented in user-friendly programming packages such as open-source Python libraries. Examples include TensorFlow for machine learning [9], Platypus for optimization [10], Eppy for EnergyPlus simulations [11], and PyMC3 for Bayesian modelling [12]. Simple integration with cloud computing now allows single users the capability of large-scale development, simulation, and analysis that would have previously required in-house supercomputing.

Machine learning is a field of artificial intelligence containing a number of algorithms that mimic how humans learn from data [13]. Their use has expanded greatly in all domains, but remains limited in the building engineering practice. There are many opportunities for machine learning in the building domain, but one promising application is the creation of surrogate models. Surrogate models can emulate the outputs from detailed dynamic energy models at a fraction of the computational cost. For example, a standard EnergyPlus model requires on the order of minutes to simulate, whereas a surrogate model only requires on the order of milliseconds.

Gradient-free optimization algorithms include genetic and evolutionary algorithms that mimic biological evolution [14]. They allow the optimization of systems where an analytical solution is difficult or impossible to determine. The methods unlock automated identification of optimal solutions of multi-objective and multi-parameter problems. This can be used to identify scenarios with complex interactive dependencies between systems, such as the optimal control of building systems.

Bayesian modelling is based on Bayes' theorem, a statistical proof that mimics human decision-making in uncertain conditions [15]. The models update based on new observations, and allow previous knowledge to be incorporated. They can be used to calibrate building models while inherently quantifying the amount of uncertainty therein.

This thesis presents an overview of recent advances in computational design and analysis tools in the form of three separate manuscripts. The first manuscript is a study of how machine learning can be used to support the early-stage design of a complex mechanical building system by utilizing an artificial neural network. The second manuscript is a case study on the simultaneous optimization of indoor air quality and energy performance of a very common mechanical system type. Finally, the third manuscript is a real-world implementation using a nascent Bayesian inference technique to calibrate a large retail building.

1.1 Summary of Contributions by Author

The main contributions by author for each manuscript are broken down by chapter as follows:

Chapter 2: Rulff, D., Cant, K., and; Evins, R. (2021). Analysis of Feature Importance in Modeling Ground Source Heat Pump Systems Using Broad Parametric Analysis, Load Characterization and Artificial Neural Networks. In eSim 2021 Conference Proceedings. Vancouver, BC; IBPSA.

K.C. and D.R. conceived of the project and wrote the manuscript. K.C. developed the case study model and neural network, and performed the feature importance. D.R. performed the ground-source heat pump post-processing and developed the fourier transform methodology. R.E. supervised the project and revised the manuscript.

Chapter 3: Cant, K., and; Evins, R. (2021). Optimizing VAV Terminal Box Minimum Positions using Dynamic Simulations to Improve Energy and Ventilation Performance. In Building Simulation 2021 Conference Proceedings. Bruges; IBPSA.

K.C. conceived of the project, developed the methodology and models, performed the analysis and wrote the manuscript. R.E. supervised the project, contributed to the discussion, and revised the manuscript.

Chapter 4: Cant, K., and; Evins, R. (2022). Improved Calibration of Building Models using Approximate Bayesian Calibration and Neural Networks. Submitted to the Journal of Building Performance Simulation.

K.C. conceived of the project, developed the methodology and models, performed the analysis and wrote the manuscript. R.E. supervised the project and revised the manuscript.

Chapter 2

Analysis of Feature Importance in Modeling Ground Source Heat Pump Systems Using Broad Parametric Analysis, Load Characterization and Artificial Neural Networks

Abstract

This paper considers the case of modeling a ground source heat pump with a range of temporal load dynamics to identify the important features used for estimating performance. Heating and cooling load profiles are generated using extensive parametric sampling of a base office building simulation, including variation of a set of parameters for heat pump system design and properties of the ground. Load characteristics are extracted from the models using aggregate output and application of Fourier Transform decomposition to describe periodic behaviour. Artificial neural networks are used to estimate the heating and cooling performance metrics of the ground source heat pump system, with significant accuracy using the full feature set ($R^2 > 0.98$). The resulting loss in accuracy due to reduced dimensionality through feature grouping is also shown, with implications for early stage design and performance modeling.

2.1 Introduction

Ground Source Heat Pump systems have the potential to significantly reduce building-sourced GHG emissions by efficiently exchanging thermal energy with a Ground Heat Exchanger (GHE) and using electricity as a low-carbon energy input where generation comes from renewable sources.

The significant variability in weather and ground conditions, along with the large number of design parameters for both the building systems and GHE field, creates a vast problem space that makes simulation of explicit configurations highly complex and generalized modeling efforts computationally intractable [16]. Additionally, the detailed information required to accurately model the GSHP system is not typically available at the early stages of design.

The objectives of this study are:

1. separate the overall problem into component sub-models (building, GSHP, GHE) and reduce dimensionality of the problem space through aggregate characterization of building heating and cooling load profiles to support generalization of results
2. identify the important features of the systems that influence GSHP performance and loading ratios through application of large scale parametric simulations
3. provide and compare the accuracy of preliminary models based on simplified feature groupings for prediction of GSHP performance using Artificial Neural Networks (ANN)

2.2 Existing Work and Contribution

The level of detail Incorporated into the model of thermal energy transfer and storage dynamics for GHE fields varies depending on the purpose of the work. Reviews of the research domain distinguish between thermal response factor methods, numerical thermal methods, artificial neural network models, and state-space models [17].

The behaviour of a GHE field as a thermal mass can be measured through the response over time to unit step heat pulses [18]. A major development by Eskilson [19] used a numerical finite-difference method to express the temperature response at the borehole wall in terms of dimensionless "g-functions". These g-functions must be

calculated directly for each GHE configuration, which can be computationally time-consuming for parametric studies [20]. This work has since been expanded upon in a variety of ways [21], though g-functions continue to be used in integrated building simulation software [22].

Work on characterizing ground temperature profiles (both surface and at various depths) based on local weather information and material properties is an ongoing area of research [23], including identifying relationships between air temperature, ground temperature and altitude [24] or investigating the impact of surface air temp fluctuations on long term vertical GHE performance [25].

Significant headway has been made with the application of machine learning to GHE modeling, due to the complexities inherent in the problem [26]. Artificial Neural Networks (ANN) have shown promise for capturing the important dynamics of the GHE field [27][28].

ANNs are used heavily for both classification and regression, and with sufficient training data can accurately predict building energy outputs, such as annual heating and cooling loads and heat pump seasonal COPs, given proper hyperparameter tuning [29]. This study adopts ANN methods to support sensitivity analysis, feature selection and to develop computationally efficient representations of not just the GHE field, but also the GSHP plant performance in response to variable thermal load profiles.

There has been some research into applying the Fast Fourier Transform to decompose load profiles into characteristic frequencies, with uses including forecasting of electrical consumption [30] and for characterizing heating demands to inform the design and sizing of thermal storage [31]. This paper extends this analysis with the use of Fourier Transform to identify important periodic features of heating and cooling profiles influencing GSHP performance.

The design of GSHP systems has a strong tradition in engineering practice, with ASHRAE publishing a detailed design guide [32][33]. Vital considerations for design include avoiding seasonal drift in ground temperatures due to unbalanced heating and cooling loads, which largely drives the need for accurate, detailed simulation of GHE fields [34]. Research into the performance of heat pump technology itself continues [35]; however, the fundamental physics of the vapour compression cycle are well understood, and performance can be reliably estimated with a reduced set of system variables including fluid temperatures, flow rates and equipment design specifications [36].

EnergyPlus is a whole building simulation program that integrates Eskilson’s g-functions to model vertical GHE fields connected to GSHP plants in a variable time-step load aggregation scheme, showing average error in predictions for heat transfer of 4-6%, and for electricity consumption of less than 3%, compared to experimental data [37]. There exists significant potential to leverage the detailed, building-integrated simulation potential of EnergyPlus for broad parametric studies of GSHP performance potential [38][39].

Methodology

2.2.1 Model Structure

The overall problem space was divided into three component sub-models for analysis (shown in 2.1): the building, the GHE field, and the GSHP plant. The building sub-model encompasses an expansive set of architectural, mechanical and electrical design parameters that are used in an hourly simulation to generate heating and cooling loads. The GHE field is represented by parameters about ground conditions and interactions with the geo-exchange system configurations, capturing dynamic heat transfer and storage behaviour in the simulation. The GSHP plant connects the building loads to the GHE field, accounting for input equipment specifications and controls to determine the electricity consumption required to handle the heating and cooling loads of the system.

All components are simulated in EnergyPlus using BESOS for access to parametric modeling and machine learning functionality, over the course of one operating year for Victoria, BC (using standard Canadian Weather for Energy Calculations (CWEC) provided by Environment Canada).

2.2.2 Building Definition

Baseline building assumptions were derived from the National Energy Code for Buildings (2015), with general inputs for building program, operating schedules, and geometry generated through work by National Resources Canada (NRCan) versions of the

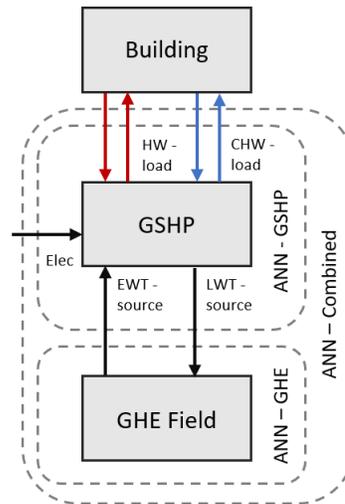


Figure 2.1: Model Overview.

Commercial Prototype Building Models originally created by the US DOE [40]. A Medium Office archetype was selected as a common platform for generating thermal load profiles and exploring the impact of a variety of system changes.

Each rectangular floorplate is $1,660 \text{ m}^2$ and represented by 5 thermal zones in the model (perimeter and core). A central VAV system with air-side economizer control and heat recovery serves each zone, with hydronic reheat coils and baseboards for perimeter heating loads. An additional server room is included with each building, acting as the space containing the additional IT loads and served by a fan coil unit connected to the central hydronic loops.

To generate a sufficiently diverse set of heating and cooling loads, variations of a set of building design parameters was included as part of the expansive sampling of the problem space. These parameters, along with their sampled ranges, are shown in Table 2.1.

2.2.3 Geo-exchange Field

This study focuses on a vertical, closed loop, ground-coupled heat exchanger configuration, which make up 80% of installed systems in Canada [16]. The field is comprised of vertical boreholes drilled into the ground with closed piping loops and grout infill, conveying the heat transfer fluid and facilitating exchange to the surrounding soil, rock and other subsurface materials.

	Parameter	units	range
1.	Orientation	deg	(-45) - 45
2.	Wall/Roof Insulation	W/m ² K	0.7 - 3.0
3.	Slab Thickness	m	0.01 - 0.10
4.	Window U-value	W/m ² K	1.0 - 2.5
5.	Solar Heat Gain Coef.		0.2 - 0.8
6.	Window to Wall Ratio	%	20 - 90
7.	Horiz. Shading Depth	m	0 - 1
8.	Daylight Control	frac	0 - 100
9.	Lighting Power Density	W/m	3 - 15
10.	Plug Load Density	W/m	0 - 20
11.	Data Centre Load	kW	0 - 10
12.	DHW Load	L/s	0.04 - 0.2
13.	Infiltration	L/sm ²	0.1 - 0.5
14.	Ventilation Effectiveness	frac	0.5 - 1.5
15.	Cooling Setpoint	°C	22 - 26
16.	Humidification	RH	0 - 40
17.	Dehumidification	RH	50 - 100
18.	Peak Occupancy	m ² /occ	10 - 100
19.	Storeys		3 - 7

Table 2.1: List of building design parameters for Medium Office

	Parameter	units	range
20.	Ground Conductivity	W/mK	0.5 - 8.0
21.	Soil Specific Heat	J/kgK	calc'd
22.	Average Soil Surface Temp.	°C	8 - 22
23.	Average Amplitude of Temp.	°C	2 - 12
24.	Borehole Spacing	m	5 - 8
25.	Borehole Length	m	75 - 200
26.	Number of Boreholes		calc'd
27.	Ref. Field Loop Flow	m ³ /s	calc'd

Table 2.2: List of design parameters for the GHE field

The parameters representing the properties of the ground are included in Table 2.2, along with borehole design context assumptions [41]. Boreholes drilled between 75 to 250m may pass through organic sand and soils, silt and clay deposits, limestone, granite and other dense rock; therefore, ranges in properties were selected accordingly [16].

2.2.4 Ground Source Heat Pump Plant

All heating and cooling loops are connected to a central plant comprised of a multi-compressor GSHP, along with supplemental chillers and boilers. The GSHP takes priority for satisfying building heating and cooling loads, with peaking equipment scheduled to operate in sequence. Both building-side loop and ground loop connections control energy transfer using 3-way bypass flow-control valves.

The heat pump performance is calculated using EnergyPlus multi-linear regression curves that determine part load thermal output and power consumption. The coefficients assumed were typical values used in the program's sample files and described in the engineer's manual [32].

The GSHP parameters are listed in Table 2.3, with the primary independent variable being reference cooling capacity. Reference heating capacity is set to be equal to nominal cooling capacity, and flow rates through the equipment and GHE is set based on a design loop temperature difference across the heat pump. Rated COP values are adjusted to ASHRAE 90.1 testing conditions [32].

	Parameter	units	range
28.	Cooling Capacity	kW	10-200
29.	Cooling Design COP		6.2
30.	Heating Capacity	kW	calc'd
31.	Heating Design COP		3.6
32.	Flow Rates	m ³ /s	calc'd
33.	Pump Power	W	calc'd

Table 2.3: List of GSHP design parameters for Medium Office

2.2.5 Parameter Sampling

The overall parametric run involved 10,000 simulation samples using latin-hypercube sampling, which divides the design parameter space into equally large hypercubes and randomly collects samples from within each hypercube. Hourly data for each of the variables was extracted and stored for each EnergyPlus sample. These were used to calculate the hourly building thermal loads, heat pump output, loop temperatures, and electricity consumption by the equipment to satisfy heating loads and cooling loads.

2.2.6 Model Characteristics and Feature Selection

Characterizing the heating and cooling load profiles across samples is important both to understand the bounds of the scope of this study, and to narrow in on the significant information that influences GSHP performance. A set of profile characteristics have been selected to encompass the high level information that might be available to practitioners estimating GSHP performance, along with metrics capturing profile ranges and temporal dynamics. This set is not intended to represent an exhaustive collection, and future study could incorporate more granular characteristics (such as total monthly loads).

The Detailed Building Characteristics include seven derived from FFT decomposition of the hourly net heating and cooling load profiles. Each of these characteristics represents the cumulative amplitude of distinct, meaningful "bins" in the frequency domain. "Imbalance" represents correlation with a zero frequency component, therefore showing annual imbalance between heating and cooling. Stronger amplitude in

each of the other bins reflects greater contribution to total loads from periodic behaviour of the noted frequency. These features help characterize the relationships between base thermal demands, periodic loading and "noise" for each run.

Overall, these characteristics form the input parameters that will be considered to define the building loads independent of the detailed building information, operation, and weather that lead to the generation of these characteristics. Converting to these characteristics allows for agnostic comparison of GSHP performance regardless of building details, and allows for generalization of the results. The results apply for buildings with load characteristics within the bounds of this study. The original building parameters were important only for generating the heating and cooling load profiles, and are therefore considered extraneous to the following stages of the study.

2.2.7 Feature Importance Selection

The four primary output objectives for the models in this study are the proportion of annual load satisfied by the GSHP (proportion of load met) and the relative electrical input to satisfy those loads (seasonal COP, or SCOP) for each of heating and cooling, which are identified as the four 'labels'. Artificial Neural Networks (ANNs) are used to determine feature importance, and the accuracy of the ANNs depending on the selection of input features represent the relative accuracy of simplified inputs to the energy model.

Hyperparameter tuning of ANNs is the selection of inputs and architecture that defines the ANN model, such as regularization coefficient, number of nodes per hidden layer, and total number of hidden layers. In order to compare the performance of the ANN models, the overall parametric simulation dataset is split first into a training set (with 80% of total simulations), and a testing set (with 20% of total simulations). The training set is further divided using the k-folds cross validation method with three folds. K-folds cross validation works by further dividing the training set by the number of folds. For each set of hyperparameters, a unique model is trained on two-thirds of the training set, and the remaining one-third is used as the test set. This is repeated for each fold, and the average performance metrics are used to evaluate the overall hyperparameter score [42]. The hyperparameters with the highest scores are selected as the optimally tuned model and is then trained on the full training set and tested on the original testing set.

The neural network is a feed-forward neural network, trained via stochastic gradient descent to minimize R^2 score of the inputs and outputs scalarized between 0 and 1. The hyperparameters selected for tuning included number of layers, between 1 and 3; number of nodes per layer, between 20 and 400; and L2 regularization parameter alpha, between 1 and 0.001.

Hyperparameter tuning was performed for a baseline ANN that included the entire feature set and output all four labels. The optimal hyperparameters were retained and held consistent across all subsequent ANN models in the study. All model creation and modification was done using Keras for Tensorflow.

The main metric used for scoring of the models was the coefficient of determination (R^2 score). An ANN generated by all of the features is used as the baseline model. Feature importance was first estimated by removing one feature at a time and re-training the model. The average score of the new model across the four labels was compared to that of the baseline model with all features. The difference in scores was used as the metric for comparable feature importance. A larger difference in scores implied that there was a larger dependence on that feature, or that the feature was more important. However, the presence of correlated features would artificially reduce the importance metric; where if both features are removed the overall loss in score would be greater than the sum of the two individual reductions.

2.2.8 Simplified Feature Grouping

The overall feature space is grouped into different sets of simplified features. These feature sets are grouped to represent different levels of information that may be available to the designs and modellers at early stages. The key characteristics and feature groupings are summarized below:

1. Simple Building Information, including
 - Site Location
 - Total Annual Heating Load
 - Total Annual Cooling Load
 - Relative Heating HP Sizing

- Relative Cooling HP Sizing
2. Detailed Building Characteristics, including
 - Simple Building Information
 - Imbalance
 - Annual
 - Semiannual
 - Weekly
 - Daily
 - Semidaily
 - Hourly
 3. Detailed Soil Conditions, including
 - Detailed Building Characteristics
 - Ground Thermal Conductivity
 - Ground Thermal Heat Capacity
 - Average Amplitude of Surface Temperature
 4. Site Location, including
 - Average Soil Surface Temperature

2.3 Results

2.3.1 Load Characterization

Three illustrative buildings are selected based on building designs with the 90th, 50th and 10th percentiles for annual heating demand. Figures 2.2 shows the heating and cooling, along with the corresponding temperature of the flow leaving the GHE field. Some general observations can be made, such as that the outlet temperature becomes more volatile under greater cooling loads, and the field is generally warmer. However, these examples also show that even in the same location, for the same building type, total heating and cooling loads are not always correlated.

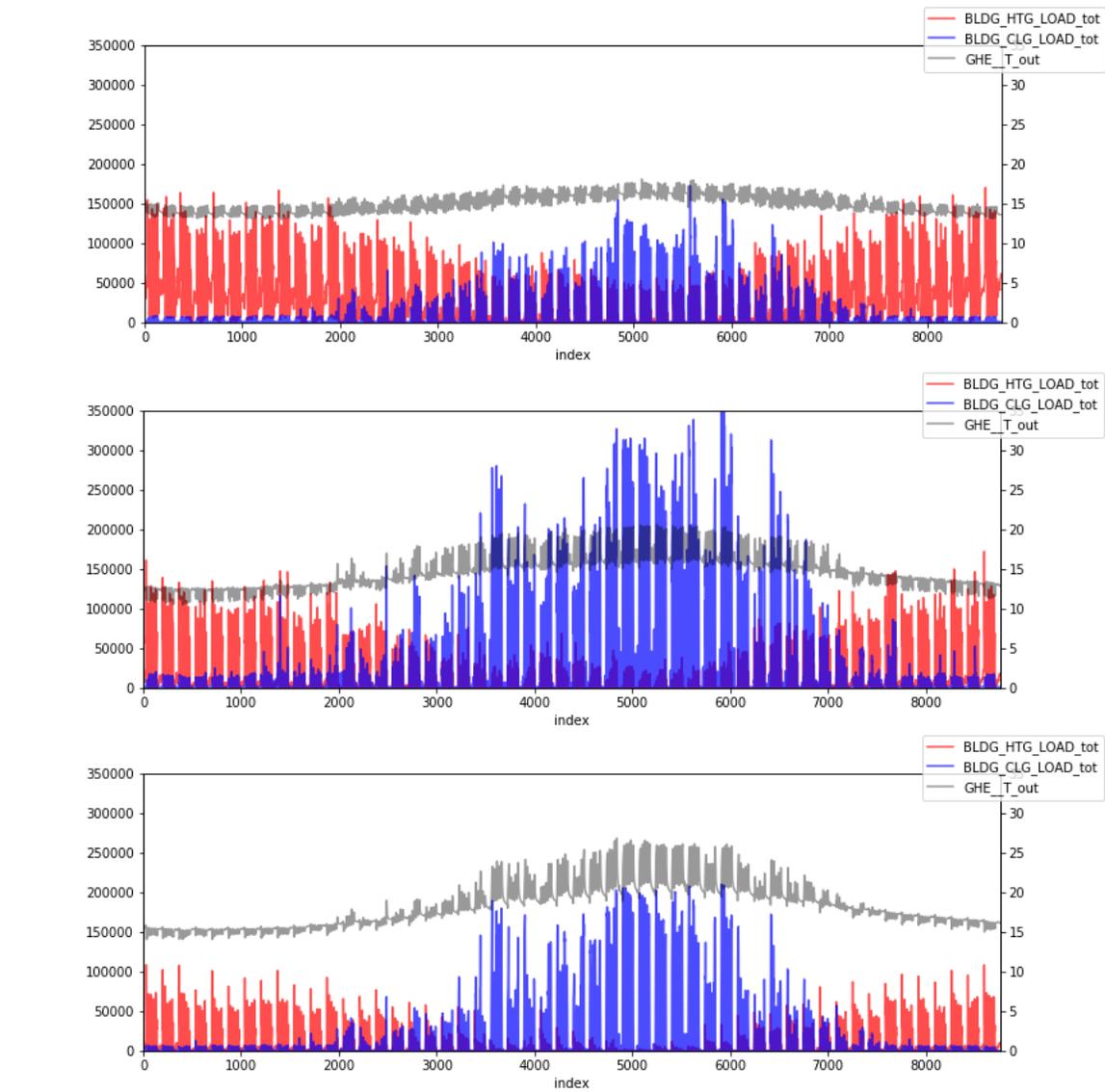


Figure 2.2: Annual heating and cooling load profiles with temperature of fluid leaving ground, (sample for 90th, 50th, and 10th percentile total heating demand)

Figure 2.3 summarizes the performance under each building characteristic for the full sample set (of 9440 building configurations), with the three buildings highlighted.

2.3.2 Feature Importance

The baseline model hypertuning resulted in an optimal model with an overall R^2 score of 0.99 on the testing set, with individual label scores shown in Figure 2.5. The

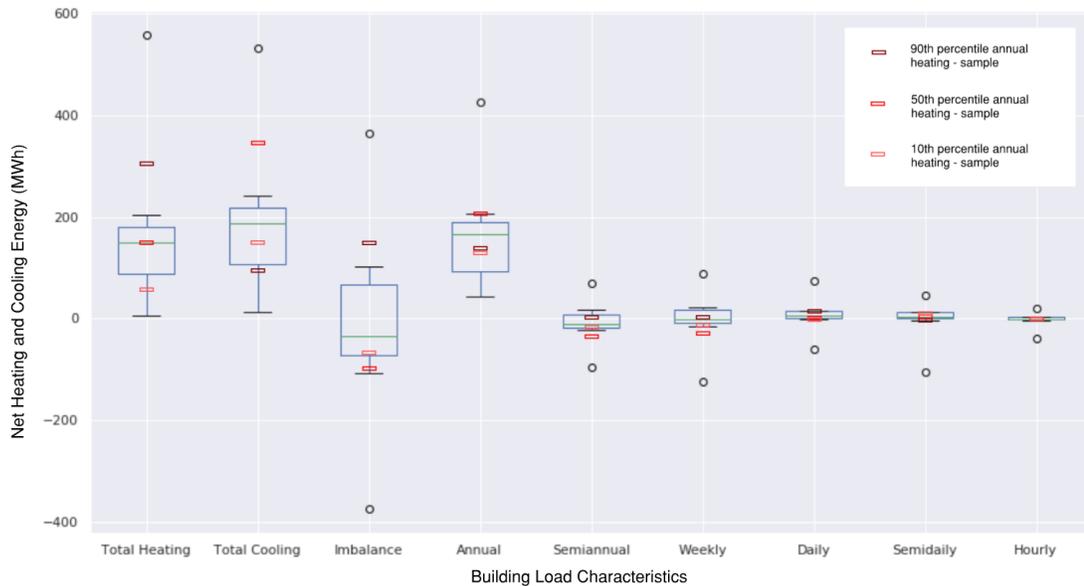


Figure 2.3: Distribution of characteristic results across full solution set (sample size: 9440)

optimal hyperparameters identified by the k-folds cross-validation procedure included two hidden layers each with 105 nodes, and an L2-regularization coefficient of 1.0. These hyperparameters were included for all future ANN development. Additional models are developed to identify the feature importance. The results are shown in Figure 2.4 for each label. The average ground surface temperature dominated the feature importance for both heating and cooling COP. Almost all other features had a negligible impact on R^2 score. Feature importance was more widespread for proportion of heating and cooling load satisfied by the GSHP.

2.3.3 Reduction of Features Accuracy

From the initial full feature baseline ANN model, three additional models were trained with subsets of features based on the original characteristic sets. Their performance estimating the objectives are compared in Figure 2.5.

A fourth model was trained only using the feature identified as "most important" (Average Soil Surface Temperature). Without sizing information, it was incapable of estimating the proportion of heating and cooling loads served, but it achieved R^2 scores of 0.74 and 0.88 for estimating cooling SCOP and heating SCOP respectively.

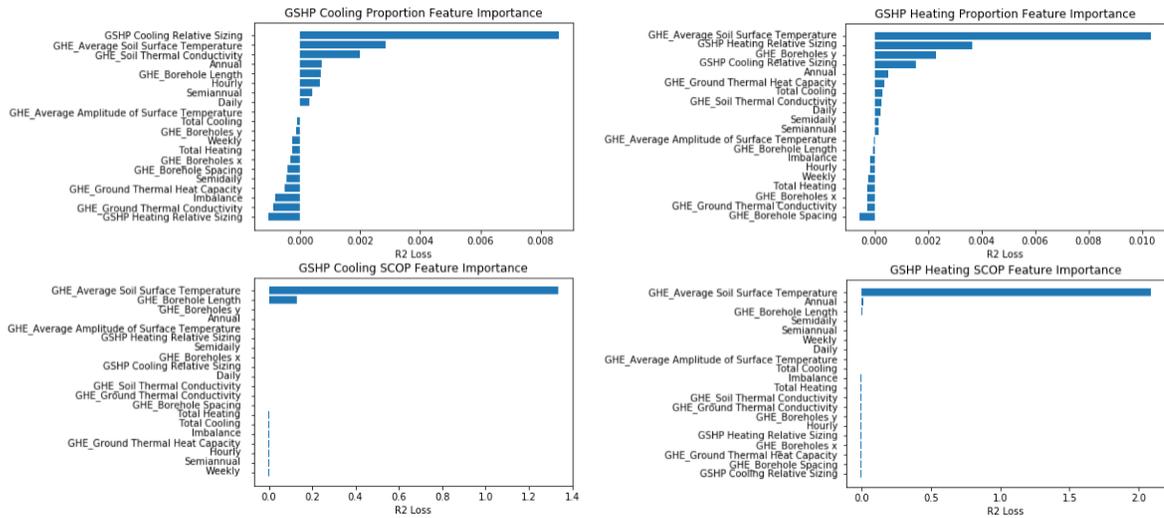


Figure 2.4: Feature importance based on R^2 score loss for objectives across test set

2.4 Outlook and Conclusions

This study has demonstrated that ANN methods can predict GSHP performance and system loading with sufficient accuracy ($R^2 > 0.98$) within a narrow scope of building types and ground conditions, and that the process identifying important features in GSHP performance modeling can be efficiently handled with those same meta-models.

Furthermore, the study provided preliminary results of a comparison of simplified models that better reflect the limited information available in early stages of design, and showed promising results for using computationally efficient ANN meta-models to replace high fidelity simulation, where only preliminary performance and loading proportion estimates are needed. As part of this approach, additional load characterization using Discrete Fourier Series decomposition was also shown to provide significant benefit for reducing prediction error.

The most important feature influencing GSHP performance (by a significant margin) was Average Soil Surface Temperature. This aligns with most simplified models for GSHP performance, which identify correlation coefficients for inlet temperatures (from the field); however, the results from the models trained on reduced feature sets

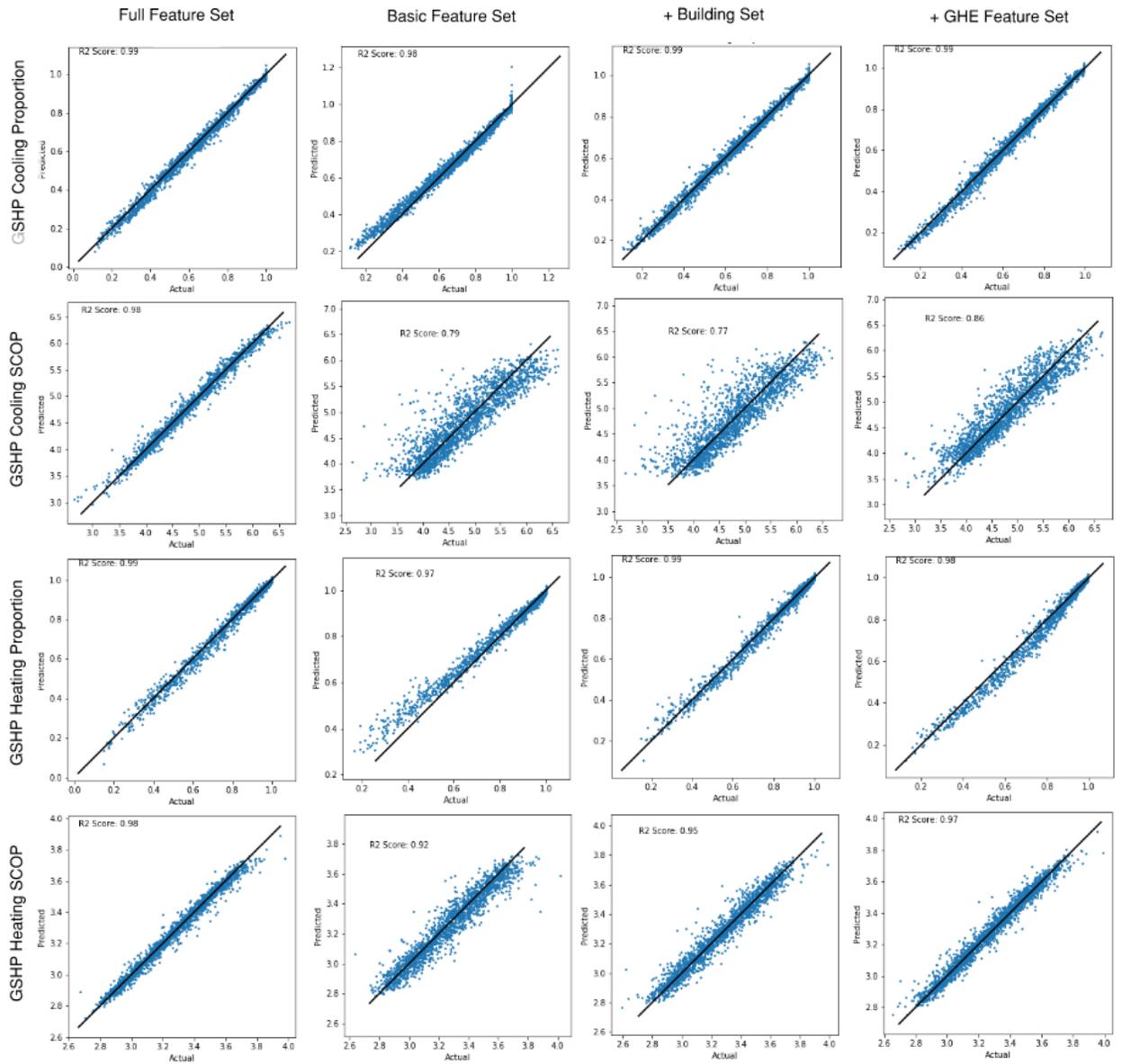


Figure 2.5: Predicted vs. actual objective performance across test sets

demonstrate that using only the most important features is not sufficient to maintain that prediction accuracy. The potential loading proportion of the GSHP plants was influenced by a more varied set of features.

There were notable limitations to the study that prevent generalize-ability. Instead, this study indicates promising direction of investigation that could be continued in a variety of ways, including:

- Including additional characteristics/features (expanding on Fourier Series load profile decomposition)

- Refining feature selection process (applying a more statistically rigorous DOE)
- Expanding scope of study (parameters, building types, weather files, etc.)
- Bringing in more detailed system modeling for design support (both GSHP plant and GHE field, over longer time horizons)

Chapter 3

Optimizing VAV Terminal Box Minimum Positions using Dynamic Simulations to Improve Energy and Ventilation Performance

Abstract

Variable air volume (VAV) systems have historically been designed following methods that result in excess energy consumption and inadequate ventilation. Designing VAV systems using heating and cooling design days and following the ASHRAE 62.1-2010 Ventilation Rate Procedure is shown to be inadequate for its intended purposes. A method for quantifying energy and ventilation performance using dynamic energy simulations is coupled with a genetic algorithm to optimize VAV terminal box minimum positions and central outside air damper minimum position. The genetic algorithm results in improved energy performance while simultaneously addressing concerns for underventilation, as shown in a case study medium office archetype energy simulation. The method provides an opportunity to improve energy consumption and indoor air quality in the existing building stock without requiring extensive capital retrofits or complex controls algorithms.

3.1 Introduction

The background and motivation for this research is presented below. First, multi-zone central variable air volume (VAV) air handling units are introduced, then the industry standard method for specifying ventilation rates is described. Next, research trends on the competing objectives of indoor air quality and energy consumption is discussed. Finally, the academic contributions of this research are defined.

3.1.1 VAV Systems

Multi-zone central variable air volume (VAV) air handling units are the most commonly installed HVAC system in North American commercial and institutional facilities [43, 44]. Simple VAV systems are comprised of a central air handling unit that is usually equipped with a variable-speed fan, heating and cooling coils, and a mixing box to maintain the air flow rate, the temperature, and ventilation, respectively. Downstream of the central unit, VAV terminal boxes maintain space temperatures by modulate dampers to vary the supply of air.

The mixing box is a set of dampers that modulate to mix return air from the occupied spaces with outside air for ventilation. Supply air temperature setpoints are set cold relative to the occupied space setpoints regardless of operation, usually at 13°C. When outside conditions are favourable the mixing box will increase the proportion of outside air to reduce cooling energy; referred to as air-side economizing. The outside air damper will have a minimum damper position setpoint such that adequate ventilation is maintained.

Terminal box dampers are likewise programmed with a minimum box position so that during periods of low loads—and therefore low airflow—there remains adequate ventilation air. Minimum box position setpoints have historically also served the role of addressing concerns that occupants feel ‘stuffy’ at a lack of adequate air movement, or feel ‘drafty’ from cold air ‘dumping’ due to a lower diffuser throw at lower airflow rates [45]. Terminal boxes can be equipped with heating coils to prevent overcooling or to provide supplemental heat as required, and are referred to as reheat boxes.

VAV terminal boxes can be classified as pressure dependent or pressure independent. Pressure independent boxes are installed with airflow measuring devices that directly calculate the amount of air passing through the box. Pressure dependent boxes do not measure the airflow and simply modulate the damper to meet thermal loads. The amount of airflow as a function of damper position is non-linear and

dependent on the relative pressure in the upstream ductwork. Newer VAV terminal boxes are more likely to be pressure independent than older boxes.

Older building automation systems (BAS) are simply collections of independently operating control hardware. Zone-level and system-level equipment on older systems do not have the capability to communicate with each other or be controlled by a central algorithm. These legacy systems lack the ability to be upgraded with some of the newer, more complex, algorithms needed for building-wide optimized operation.

3.1.2 Ventilation Rate Procedure

The most commonly followed design process for ensuring adequate ventilation from a VAV system is the Ventilation Rate Procedure (VRP) from ASHRAE Standard 62.1, Ventilation for Acceptable Indoor Air Quality, section 6.2 [46]. The standard requires that an adequate amount of outside air is supplied to the breathing zone by the VAV ‘delivering no less than the minimum ventilation rates required whenever the zones are occupied’. The minimum ventilation rate specified by ASHRAE 62.1 VRP is a summation of a floor area-based rate and an occupancy-based rate [47]. The VRP is a calculation procedure where a critical zone is determined based on worst-case conditions which define the system ventilation efficiency and required outside air rate. The standard does not dictate to the designers how to select worst-case conditions, but standard practice has been to analyze the system at cooling and heating design days during occupied periods [48, 49].

ASHRAE 62.1 also does not dictate a terminal box minimum position, however, ASHRAE standard 90.1, Energy Standard for Buildings Except Low-Rise Residential Buildings, stated that ‘the minimum airflow for the VAV reheat boxes should be set to 30% of the zone peak supply air volume or the outside air ventilation rate, or the airflow rate required to comply with applicable codes or accreditation standards, whichever is larger’. While the Standard revised the minimum from 30% to 20% in 2013 [50], setting the minimum boxes to 30% is the default and standard practice for equipment manufacturers, and it can be assumed that most of the existing building stock follows this default [48, 49].

3.1.3 New research on VAV box minimums

HVAC designers use data sheets from equipment manufacturers and guidelines to design their systems. Manufacturers give an acceptable VAV box damper control

range, typically being a 30% minimum. This default is then carried onwards for design, installation, and commissioning of systems. ASHRAE RP-1353 [51] tested a number of VAV boxes from popular manufacturers at a range of operations and found that most VAV boxes are suitably accurate down to approximately 5% damper position.

Within the HVAC design community, it is commonly assumed that a minimum amount of airflow within spaces is required to properly mix air. This eliminates cool air ‘dumping’ at low airflows and reduces ‘stuffiness’ within zones. ASHRAE RP-1515 [45] identified that those assumptions are baseless. Feeling ‘drafty’ is more likely from overcooling caused by minimum box positions set too high, and feeling ‘stuffy’ is correlated with overheating and not ventilation rates. The researchers found that reducing the minimum box positions to around 10% to 20% actually increased the thermal comfort of the occupants and reduced energy consumption by approximately 14% in a case study building.

[5] identified that retuning the minimum VAV box damper setpoints is the most effective control measure for medium and large commercial buildings. They estimate that on a per-building basis a site energy reduction of approximately 15% is possible, and nation-wide retuning of minimum damper setpoints could reduce national site energy consumption by over 6%. In addition, [52] estimate that the economic benefit of improving ventilation in the existing American building stock is potentially 40 billion USD per year.

3.1.4 Competing Objectives of Energy and Indoor Air Quality

Providing adequate fresh air to occupants of the built environment provides numerous health benefits and can significantly reduce the likelihood of the ‘sick building syndrome’ [53]. Indoor air quality can be modelled as a balance between source contaminant generation rate and ventilation dilution. Numerous challenges and uncertainties associated with contaminant source control has lead to ventilation being the main method in maintaining indoor air quality [54]. Conditioning outside air, however, can be one of the largest energy expenditures in operating buildings. This leads to an obvious tradeoff between energy consumption and indoor air quality.

There have been a number of studies that have investigated the tradeoff between energy and indoor air quality for central VAV systems. [55] studied eight ventilation

methods to quantify energy consumption (using a set of polynomial models in conjunction with BLAST models), and ventilation performance (using maximum CO₂ as the indoor air quality identifier). [44] quantified the potential energy savings of using advanced occupancy sensors to control both the zone lighting and operation of terminal boxes. [56] used the maximum CO₂ in a space as a proxy for indoor air quality while optimizing an equipment's on/off schedule. [57] and [58] used productivity models to convert indoor air quality and thermal comfort into an annual dollar value for optimizing ventilation rate and space temperature.

[59] and [60] both integrated a multi-objective algorithm directly into the HVAC control system to trade off ventilation and energy consumption. [61] developed a surrogate model to replace high-fidelity CFD models to optimize ventilation, thermal comfort, and energy consumption, using maximum CO₂ concentration as the identifier for indoor air quality.

3.1.5 Novelty of Research

To quantify indoor air quality performance, researchers tend to use either maximum CO₂ concentration in occupied zones or a productivity model. CO₂ concentrations has been proven to be a poor choice as indication of indoor air quality [54]. In fact, the use of CO₂ is likely an artefact from historical versions of the ASHRAE 62 standard [46]. There is ongoing debate and further research required to fully understand the impact of ventilation on human comfort and health; however, there is strong evidence to suggest that underventilating below the minimum requirements in ASHRAE 62.1-2010 has a significant negative impact on occupant health and contributes to the sick building syndrome [62, 53, 46]. Recent research has found a potentially quantifiable economic impact of inadequate ventilation; however, further research is recommended before it is generally applicable [53].

None of the studies on different control strategies or building optimization have considered the selection of VAV box minimums. Recent research has trended towards increasingly complex control algorithms and requires additional sensors located in each zone. These studies do not address the large existing building stock that can not implement these algorithms without extensive upgrades to their control systems [63]. In addition, the requirement for additional sensors located in all zones may significantly increase the amount of intrusive maintenance in order to ensure proper operation of the HVAC systems [64]. Finally, there should be a preference for sim-

pler control systems: a large study on high-performance UK buildings has identified complex controls systems as a major contributor to the performance gap [65]. The solution presented in this research is mainly applicable to improving the performance of the existing building stock, which likely do not have the capability to extend their control systems to include the novel higher performance control algorithms.

This paper addresses these issues in the field of optimizing the tradeoff between indoor air quality and energy consumption in two separate ways.

1. Quantification of indoor air quality performance relative to an ASHRAE 62.1-2010 baseline, along with quantifying the variation in underventilation following industry standard practice.
2. Quantifying the relative improvement in both energy consumption and underventilation a case study energy model based on an archetype building by using a genetic algorithm to retune the setpoints for terminal minimum box positions, along with a comparison to a base case following the standard ventilation rate procedure method.

The model, methods, and results are made publicly available¹.

3.2 Methods

The methods used to calculate and optimize the energy consumption and ventilation performance are documented below. First the dynamic simulation case study is presented, followed by the procedure for calculating ventilation performance, and finally the genetic algorithm for multi-objective optimization is described.

3.2.1 Dynamic Simulation

An EnergyPlus model is used to simulate the energy consumption of a building over a year. The building is a medium office, based off of the DOE medium office archetype [66], and modified for use with the Net-Zero Navigator project [4], shown in Fig. 3.1. The building is simulated with the CWEC 2016 Victoria, BC, Canada weather file. The building is equipped with a single central VAV providing supply air at a constant temperature setpoint of 13°C to VAV terminals located in each zone. The terminal

¹<https://gitlab.com/KCant/ventilation-study.git>

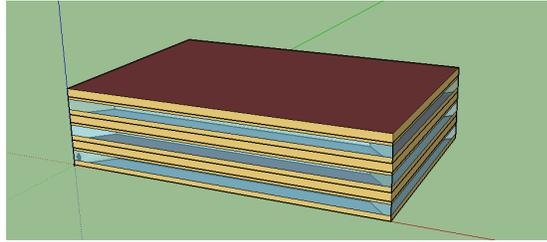


Figure 3.1: Depiction of the DoE medium office archetype, a 3-storey core-and-shell energy model.

boxes are equipped with hot water reheat coils that allow a maximum of 30% airflow or the minimum box position, whichever is larger, during heating.

Airflow rate through dampers is simplified as being directly proportional to damper position, and is independent of upstream or downstream pressure effects or non-linear damper behaviour. The minimum airflow rate through a damper is therefore equal to the minimum damper position multiplied by the nominal airflow.

3.2.2 Ventilation Calculation

Ventilation calculations follow ASHRAE 62.1-2010 Ventilation Rate Procedure (VRP). For each hour of the year a minimum specified amount of outside air is required to be supplied to each zone depending on its floor area and occupancy. The occupancy and required ventilation rates used in the simulations follow ASHRAE 62.1-2010. Design occupancy is 25 m² per person, and occupancy schedule matches the Canadian National Energy Code for Buildings (NECB) schedule type A, shown in Fig. 3.2.

The per floor area and per person ventilation rate for office zones is 0.3 L/s per m² and 2.5 L/s per occupant during all hours of occupancy. In addition, the ventilation rate effectiveness factor, the percentage of supply air that reaches the occupant breathing zone, depends on the location of diffusers, grilles, and supply air temperature. Assuming ceiling located supply and return grilles leads to a 1.0 effectiveness factor.

For each 15-minute timestep across the year (8760 hours), the amount of outside air supplied to each zone is compared to the required amount of ventilation. The quantities of underventilation is summed and compared to the nominal required ventilation amount. The relative proportion is then presented as a percentage of underventilation.

For the base VRP scenario, the design process is as follows: the minimum damper

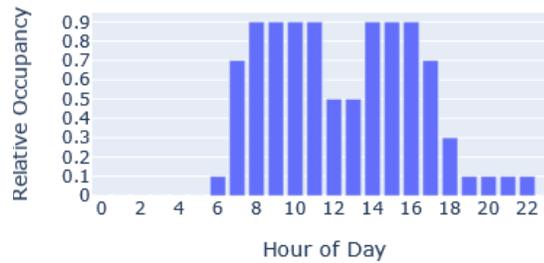


Figure 3.2: Weekday occupancy schedule for the office archetype.

position is predetermined for the terminal boxes and the simulation is run for the heating and cooling design days. For each time step in the simulation the VRP from ASHRAE 62.1-2010 section 6.2.5 is completed to calculate the central outdoor air damper position required. The highest required damper position across the design days is recorded and set as the central minimum outdoor air damper position for the simulation.

3.2.3 Genetic Algorithm

The proposed procedure emulates the multi-objective optimization described in [67], linking EnergyPlus with Python via the besos [68] suite, which connects the Eppy library [11] to the Platypus optimization library [10] using a number of Python helper functions.

A multi-objective optimization (MOO) genetic algorithm (GA) is used to optimize the performance of the building. The objectives in the MOO are to minimize the energy consumption and the amount of underventilation. Each of the 15 terminal boxes in the building are allowed a unique minimum box position, ranging from 0% (fully closed) to 100% (fully open). The MOO algorithm used is NSGA-II [69]².

3.3 Results

The results of the simulations and optimization are presented. First, the benefits of quantifying ventilation performance using dynamic simulations is shown. Next,

²For this case study, the population size is set to 100 and run for 10 generations, following the standard rule-of-thumb of a 10:1 ratio of population size to generations. Crossover and mutation operators and settings remained as the default platypus settings. The optimization algorithm is only run once and therefore no guarantee of true optimality is reached; however, the results provide a clear basis of discussion and show an improvement over standard practice. It is assumed that industry uptake of the method would not require convergence of absolute optimality.

the potential improvements of using a genetic algorithm to optimize minimum box positions is compared to the traditional approach.

3.3.1 Quantifying Ventilation Performance Using Dynamic Simulations

Table 3.1 displays the minimum outdoor air damper position, annual energy consumption, and underventilation for each set of minimum VAV box terminal positions (in 10Energy consumption ranges from 125 ekWh/m² to 198.6 ekWh/m², and underventilation ranges from 0.5% to 2.8%. In particular, the industry default value of 30% across all zones is of interest, which results in a minimum VAV central outdoor air damper position of 43% and 1.2% underventilation.

None of the solutions following the VRP technically meet the requirement that ventilation systems must be capable of delivering the minimum ventilation rates at all occupied hours of the year [47]. This is based on the interpretation that ‘capable’ includes actual hourly operation due to the designed setpoints, thermal loads, occupancy, and control algorithms.

Table 3.1: VRP Results

Min. Box Position	Min. OA Damper Pos.	Energy Cons. [ekWh/m ²]	Underventilation
10%	100%	134.2	1.3%
20%	46%	125.0	2.8%
30%	43%	135.5	1.2%
40%	30%	154.0	0.5%
50%	25%	174.7	0.7%
60%	22%	198.6	1.1%

When the central outdoor air damper minimum position is included as an additional parameter, the resulting solution space is expanded to include potential design selections that either reduce the amount of underventilation, reduce the annual energy consumption, or simultaneously reduce the energy consumption and underventilation.

3.3.2 Optimizing Energy and Ventilation by Selecting Minimum Box Position on a Zone-By-Zone Basis

The GA enabled the optimization of terminal box minimum position on a zone-by-zone basis simultaneously with the central outdoor air damper minimum position. The resulting pareto-front of nondominated solutions shown in Fig. 3.3 improved both the ventilation and energy performance of the case study building. Assuming 30% default minimum box positions and ensuring no underventilation would require a central outside air damper position of 80% to 90% and an energy consumption of 141.9 to 146.3 ekWh/m². The optimized design can achieve no underventilation and an energy consumption of 130.1 ekWh/m², a savings of 8% to 11%. Alternatively, the optimization results in a savings of 9% with no additional underventilation over the default minimum terminal positions (30%) and the VRP procedure.

Figure 3.4 presents the pareto-optimal results from the genetic algorithm. Minimum box positions mostly varied between 5% and 15% across the pareto front for east and west zones, 0% to 20% across the pareto front for north zones, 15% to 35% regardless of position on the pareto front for south zones, and between 5% and 55% for core zones. In addition, central outdoor air damper minimum position increased along the pareto front. For external zones, the optimal minimum terminal box positions rarely exceeded the default 30%, even when the focus was on minimizing underventilation, supporting recent guidelines to reduce terminal box positions below the default 30%. The trends along the pareto front can be summarized as the most effective ways to reduce underventilation at the least cost to energy consumption.

3.4 Discussion

Quantifying ventilation using dynamic hourly simulations identified that seemingly equivalent VRP-compliant designs resulted in a range of annual ventilation performance. All of the seemingly compliant designs in this study do not technically meet the requirements of ASHRAE Standard 62.1-2010 Ventilation Rate Procedure due to periods of underventilation.

If the requirement for adequate ventilation is to meet the minimum ventilation rate at any occupied period, designing the ventilation system based on a worst-case heating and/or cooling design day is not appropriate. The results of this research show that for a minimum box position of 30% set across all zones in this case study simu-

lation, the required minimum outdoor air damper position would be 90%—a number HVAC designers would likely balk at—and would look to increase the minimum position of terminal boxes in critical zones in order to reduce the central outdoor air damper minimum position to a more conservative value. The resulting annual energy consumption would increase by over 7% compared to the standard VRP calculation.

If the VRP design method is deemed valid by the authority having jurisdiction, with minimum outdoor air damper position calculated according to the expected operation on heating and cooling design-days, and underventilation of the proposed designs varied from 0.5% to 2.9%, then that would suggest that the acceptable amount of underventilation is flexible. Improving ventilation can improve occupant health and reduce the prevalence of the widely acknowledged sick building syndrome. The impact of short periods of underventilation, however, requires further research.

The design solution presented can be considered a ‘sub-optimal’ solution because there are likely still hours where the minimum box position can be improved with dynamic control. A number of studies have been completed using highly intelligent control algorithms, including neural networks, genetic algorithms, and complex rule-based controls. These algorithms require large amounts of sensors to be placed in the controlled zones and building automation systems that are capable of communicating between terminal devices and central workstations for optimal control. There is evidence that commonly-used sensors, such as carbon dioxide sensors, are likely to drift out of calibration or fail. Additionally, existing commercial buildings would likely require an upgrade to their building automation systems in order to accept these new control algorithms.

Improvements from demand-control ventilation and highly intelligent control algorithms for VAV systems has been generally in the range of 5%-30% of total energy consumption. Energy savings shown in this case study for a mild climate (ASHRAE Zone 4C) are comparable to those in the literature for novel control algorithms that would require large capital upgrades to the existing buildings stock and ongoing calibration for the vast number of required sensors, whereas the implementation effort for this retrofit would be comparable to a re-balancing project.

3.4.1 Limitations and Future Research

There are a number of limitations to the research presented that require further development and research.

The resulting setpoint selections in the study are not necessarily generalizable because they were only based off one case study simulation for a medium office building in a mild climate. In addition, airflow through dampers is simplified to not include pressure effects or non-linearities. This is especially important at minimum damper positions where the study assumes minimum damper positions result in a proportional percent of nominal airflow. Further, a common control strategy to reduce energy consumption is to reset the supply air temperature to warmer setpoints during colder ambient temperatures. An outside air temperature reset strategy may alter the impact of minimum box positions, and reduce the negative impact of higher box positions. Finally, the occupancy and internal loads in the case study building is deterministic which may not represent the actual occupancy of office buildings in use. Repeating the study for a number of different buildings, with different load conditions and space types in different climates, and with stochastic occupancy may result in more robust and generalized recommendations or guidelines.

The primary use-case for this method is the existing building stock with simple VAV systems controlled by legacy BAS equipment, in which revising the minimum box positions can be a more economical option to improve energy consumption and indoor air quality. This method, however, still requires the development of a detailed energy model and the optimization of setpoints. The relatively simple 15-zone model simulated in this case study required approximately 3 minutes per simulation. The genetic algorithm input 16 parameters and required 1,000 simulations for good convergence (although hypervolume calculations were not conducted to quantify convergence), with a total run time of under 4 hours with a 16-core processor. It is anticipated that larger, more complex buildings would result in increased simulation time and require a higher number of inputs. While the computation time increases substantially, the amount of effort saved relative to a manual design process is improved, and computational time is relatively cheap. Further improvements by replacing the dynamic energy simulation with a data-driven approach may reduce the design effort, and the use of more efficient optimization algorithms, such as Bayesian optimization, may reduce the computational burden.

Underventilation percentage is used as the metric for indoor air quality which quantifies how well the ventilation system performs compared to the minimum requirements of ASHRAE 62.1-2010, however, it does not give benefit to systems that provide additional ventilation above the minimum requirements. Additionally, it does not discriminate between large hours of small underventilation and small hours

of larger underventilation. Further research is recommended to determine a more representative metric for quantifying ventilation performance.

3.5 Conclusion

VAV systems are the most prevalent HVAC system in medium and large commercial buildings. They have historically been designed using default terminal box minimum positions, based on incorrect assumptions, and following a design-day ventilation calculation procedure. This study used dynamic hourly simulations to quantify ventilation performance to facilitate the design of systems that have equivalent or lower rates of underventilation compared to the traditional Ventilation Rate Procedure required by ASHRAE 62.1-2010, while also reducing energy consumption.

The optimization algorithm NSGA-II was used to support the design process by providing insights into the tradeoff between energy and ventilation within a large parameter set of terminal box positions. The method employed in this research can help designers implement appropriate terminal box minimum positions for simple VAV systems that reduce energy consumption and potentially improve thermal comfort while addressing concerns of underventilation. The results presented using a case-study simulation building suggest that the standard practice for implementing the ASHRAE 62.1-2010 Ventilation Rate Procedure using cooling and heating design days may not provide adequate ventilation throughout the year.

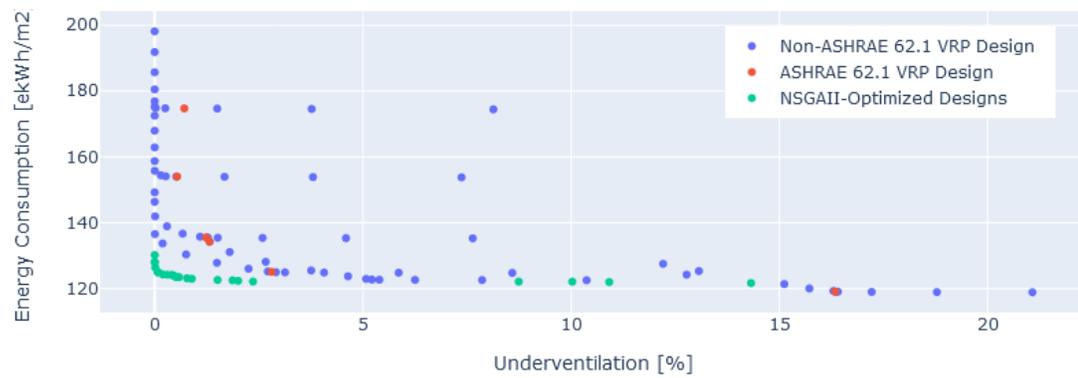


Figure 3.3: Energy and ventilation performance comparison of potential design options. A wide range of energy and ventilation performance is available for VAV terminal box minimum damper positions and central OA damper positions. Selecting each terminal box position independently can improve both energy and ventilation performance over the designs with a single minimum box selected across-the-board.

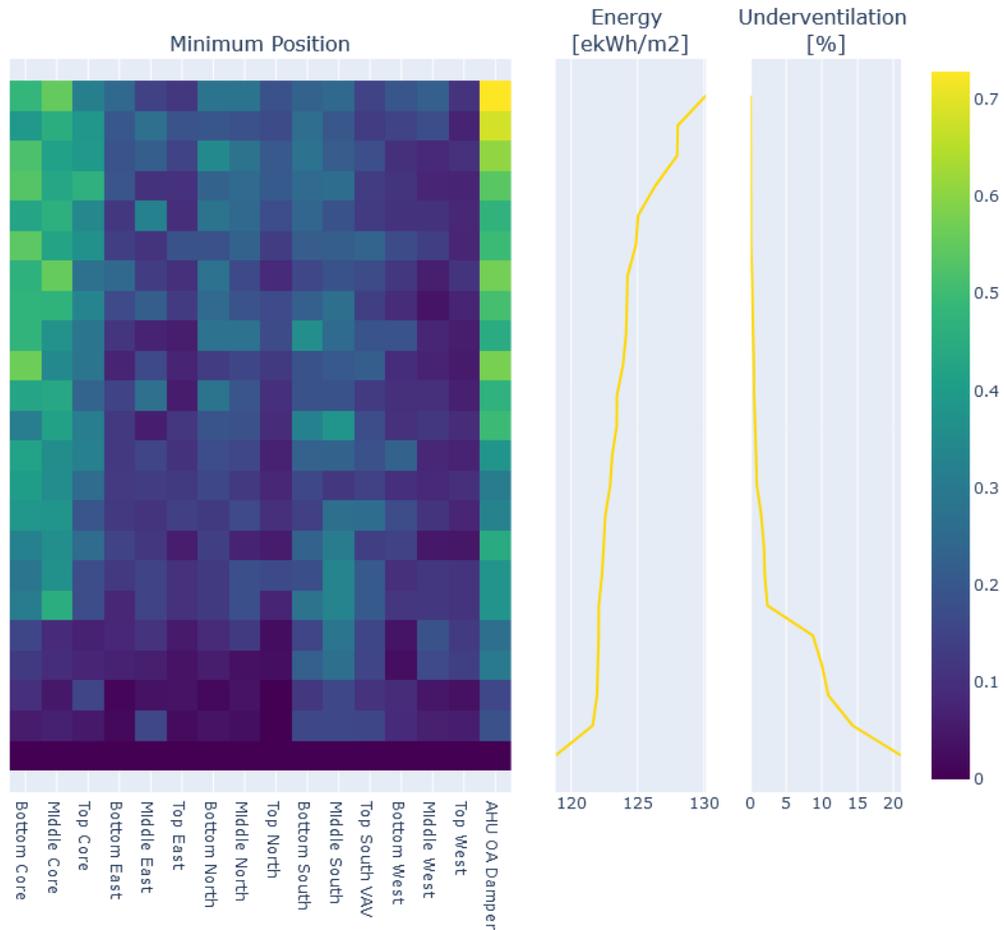


Figure 3.4: Details of the NSGA-II optimized designs. Different external load conditions result in different recommended terminal box setpoints. Designs with very low amounts of underventilation tend to be designs with core zones that have terminal box setpoints equal or above the default 30%, whereas all optimized designs have exterior zones with setpoints at or below the 30% default.

Chapter 4

Improved Calibration of Building Models using Approximate Bayesian Calibration and Neural Networks

abstract

Deep energy retrofits of buildings are crucial to meeting climate targets and depend on calibrated energy models for investor confidence. Although there remain issues in calibration, Bayesian inference can improve the rigour in standard practice and improve confidence in calibrated energy models. Approximate Bayesian computation (ABC) methods present an opportunity to calibrate energy models while inherently accounting for parameter uncertainty, and do not face the same curse of dimensionality as the current standard process for Bayesian calibration. A case study for a large, complex building is presented to demonstrate the applicability of ABC. Parameter sensitivity screening, however, is found to result in over-confidence in the resulting inference by between 14% and 85%. Finally, the presentation of posterior distributions may be misleading as independent distributions, which can misattribute the true likelihood of parameters.

4.1 Introduction

In order to meet the climate goals of Canada energy consumption and carbon emissions must be reduced from the existing building stock [70]. In British Columbia the latest climate plan requires a 59-65% reduction in building emissions [1]. The current rate of commercial building retrofit in Canada is under 1.5% per year, however, it has been argued that a deep retrofit rate of between 5% and 12% of the existing building stock per year is needed [71]. Decision makers' confidence in retrofits are undermined by the lack of quantified uncertainty in costs and savings in standard approaches [72, 73, 7]. For example, large retrofit projects that request funds from the Canadian Infrastructure Bank must be ICP Certified [74]; requiring a calibrated energy model with uncertainty levels clearly demonstrated [75]. Although there are higher up-front costs, investing in an integrated design process at has proven to provide more energy and cost savings over the life of the project [6, 76].

Calibrated energy models typically use dynamic energy modelling software, such as EnergyPlus [77], and follow ASHRAE Standard 211 [78, 79, 80]. This is a time-consuming process undertaken by qualified professionals [81, 82, 83]. The reason that existing building calibrated energy models are so time-intensive, especially compared to new building energy models, is a simultaneous lack of available documentation and validation data [7, 84, 85]. Existing buildings have inadequate information about their sub-components, such as mechanical equipment, lighting, or building envelope [86]. The deficiencies come in the form of missing as-built drawings, shop drawings or balancing reports, and many smaller changes to the building operation, equipment, or occupancy that have gone undocumented [84]. Compounding this issue is the degradation of equipment, ad-hoc changes to controls sequences, or improper installations or repairs [87, 88, 7, 85].

Further complicating this challenge is the massive number of input parameters compared to the amount of data that is available for calibration: typically on the order of thousands of input parameters compared to tens of data points [86, 88, 83, 85, 77]. This leads to an over-parameterized and under-constrained model, where the identifiability of parameters may not be possible. It produces *equifinality* in the model results: the fact that different combinations of building parameters can result in nearly identical outputs [89, 77, 7]. A number of previous studies recommend mitigating this issue by limiting the number of parameters present in the calibration process [90, 77, 91]. Recent research in high-fidelity model calibration, with smart meters

and modern building management systems [7, 92, 93], has the potential to somewhat alleviate the problem. However, there remain significant numbers of existing buildings that rely on monthly utility bills as the source of reliable calibration data [83, 77]. Indeed, the buildings that are most primed to benefit from retrofit are likely older buildings with older building management systems and meters [94].

The traditional steps for building energy model calibration are as follows [95, 86, 83]:

1. Perform an energy audit.
2. Develop a base energy model.
3. Simulate model and compare outputs to observed data.
4. Determine and tune model parameters.
5. Repeat steps 3 and 4 until calibration thresholds are achieved.

1. **Standardization.** There are established statistical criteria for energy model calibration assessment; however, the process of calibration has not been standardized, and is usually carried out on an ad-hoc basis according to the user's judgement and experience.
2. **Calibration costs.** Developing and calibrating an energy model for an existing building is a difficult and lengthy task, even more so than for a new building. The manual tuning process is highly time-consuming and requires expert judgement. The added time and expense of sub-metered data and site-collected data adds to the cost.
3. **Model complexity.** The amount of necessary input data varies according to the complexity of the energy model and building. The choice of simulation program also affects the amount of input data required.
4. **Model input data.** The amount of observed data, such as energy consumption, occupancy levels, or sub-metering can be substantial and lead to problems of data quality or data handling.
5. **Uncertainty in building models.** A deterministic approach is typically carried out with manual calibration. Not all input data affects energy consumption equally, and deterministic approaches do not provide the level of influence or confidence that parameters have on the final output.
6. **Discrepancies identification.** There are often discrepancies between simulated outputs and measured observations during calibrated simulation. While simulation experts may be able to identify the causes, the disagreements may be linked to a chain of errors in the building model definition or measurements.
7. **Automation.** No automation process has been established as standard-process in building energy model calibration. An automated process can reduce the time required to manually tune calibrated models.
8. **User's experience.** User experience is very important in building calibration. The ultimate decision is up to the expert user to determine if their building energy model is adequately calibrated, and calibration is highly dependent on the personal judgement of the analyst. A high level of domain knowledge is required to properly develop and calibrate an energy model, even with automated or systematic processes.

Figure 4.1: A summary of the main issues affecting calibration of building energy models [86].

Issues exist with the traditional method of building energy calibration. A comprehensive list adapted from Fabrizio and Monetti [86] is presented in Fig. 4.1. There have been two main advances in building energy calibration that attempt to address these issues [86, 88, 85, 77]: Automated Calibration, and Bayesian Calibration.

Automated calibration addresses issues #2 and #7, by replacing the manual tuning process with a computational method, thus reducing expensive labour effort required for model tuning. Automated calibration, however, has not been shown to improve the other issues with building energy calibration identified by [86, 96, 97, 98, 7, 99, 92, 85].

Bayesian Calibration is an alternative to automated building energy calibration. It offers all the improvements of the automated approach—namely, time savings in model tuning—and quantifies the amount of uncertainty and indeterminacy associated with the model, thus addressing issue #5 [7, 100, 85]. Furthermore, performing Bayesian calibration at an earlier stage in the project can determine if costly sub-metering is required for parameter identification, potentially addressing issue #2 [89]. Finally, and perhaps most important, Bayesian calibration can address issue #8, regarding user experience. Bayesian calibration takes into account the user’s personal judgement and provides an opportunity to quantify and analyze the impact of this judgement [91].

Bayesian Modelling

Statistical inference is about trying to learn what we cannot easily observe through what we can observe. Statistical inference can be performed using Bayesian modelling, which integrates prior knowledge that gets conditioned on observations in a statistically consistent manner. Bayesian modelling is founded on Bayes’ theorem, which is based on **conditional probability**: the likelihood of an event happening due to the occurrence of a separate event or outcome. Bayes’ theorem is stated mathematically as:

$$P(\theta | D) = \frac{P(D | \theta) P(\theta)}{P(D)}. \quad (4.1)$$

In this case, θ is a parameter, or set of parameters, of interest, and D is observed data. Hereinafter, the term $P(\theta | D)$ is referred to as the *posterior*, the term $P(D | \theta)$ is referred to as the *likelihood*, the term $P(D)$ is referred to as the *evidence*, and the term $P(\theta)$ is referred to as the *prior*. In the context of building energy model calibration,

the parameters are characteristics of the building that are difficult to ascertain. The likelihood is a stochastic model relating input parameters to output data. The priors incorporate the existing knowledge of the parameters before inference, usually from a combination of subject matter expertise, existing drawings or reports, or previously measured or inferred data. The evidence takes the form of a scaling factor which is used to ensure the posterior is a proper probability distribution with a summation of 1 [101].

Bayesian modelling can be broken down into three major steps [102]:

1. Setting up a full probability model, including joint probabilities of all observable and unobservable quantities. This model should be consistent with scientific or engineering knowledge of the problem at hand and the observed data.
2. Calculating the posterior by conditioning the model on observed data.
3. Evaluating the posterior and associated implications of the model. The resulting posterior and conclusions should be reasonable. If the resulting posterior is not reasonable, then there is an error either in the definition of the models or prior, or the calculation and conditioning on observed data.

In most applications the likelihood is a statistical parametric model. A stochastic parametric model, in this sense, is a model where a stochastic output can be defined from a finite set of parameters [29]. Examples of parametric models include normal distributions (parameterized by the mean and standard deviation), or the binomial distribution (parameterized by the number of experiments and number of successes). When performing Bayesian inference on parametric models, the parameters that define the model are themselves inferred. For example, if a normal distribution is used for the parametric model, the posterior distribution of the mean and the standard deviation parameters would be inferred from the priors and observations.

4.1.1 Bayesian Calibration in Building Energy Models

As opposed to stochastic parametric models, building energy simulations are complex deterministic models. There is no randomness involved and the output of the model will always be the same if the inputs are unchanged. The parameters of interest are characteristics of the building and not characteristics of the model itself. Standard Bayesian inference cannot be applied directly to deterministic building energy models, because the likelihood term, $P(D)$ does not have a stochastic output and the numerator of Bayes' theorem cannot be calculated.

The Kennedy and O’Hagan (KOH) method [103] is often used to solve this problem [100, 91]; it was developed to incorporate deterministic computer models, like building energy simulations, into Bayesian inference. The error between the model outputs and observed data points following calibration includes all forms of uncertainty remaining in the model. Examples of these uncertainties, given by Kennedy and O’Hagan [103] are:

- parametric uncertainty,
- model inadequacy,
- residual variability,
- parametric variability,
- observation error, and/or
- code uncertainty.

The majority of work to date in building energy model calibration has followed the Kennedy and O’Hagan approach, and most apply Gaussian Processes (GPs) as surrogate models because of their ability to emulate both the simulation outputs and the associated uncertainty [100, 92, 91]. Surrogate models, or meta-models, are data-driven representations of the complex dynamic model that emulate the output of the dynamic model at a fraction of the computational cost, which is crucial because developing a robust posterior distribution can require thousands of iterations. Because GPs can effectively model physics-based computer models they are used as the likelihood function during Bayesian inference [100, 91]. Although Gaussian Processes are regarded as accurate surrogate models, their usefulness is limited to a relative few number of parameters because they suffer greatly from the *curse of dimensionality*: the computational demand increases exponentially with the size of the model [90, 104, 91]. Neural networks, on the other hand, have been shown to be able to provide similar levels of accuracy while accommodating a greater numbers of parameters [105, 92].

Approximate Bayesian Computation (ABC), or likelihood-free inference, is an emerging method which is used for situations where a probabilistic likelihood function is difficult or impossible to define [106]. ABC methods still follow the same ideology as the seminal KOH method, however, the use of GPs is replaced with a generic model and a predefined distance metric to approximate the likelihood function.

The most basic ABC algorithm is the ABC-Rejection algorithm [107]. The user defines a simulation model, prior distributions of parameters, a distance metric, a

tolerance value, and the minimum size of the posterior set. The ABC-Rejection algorithm is then as follows:

1. Sample a set of parameter points from the prior distribution.
2. Create a set of outputs by running the set of parameter points through the simulation model.
3. Compare each output from the set of outputs with the observed data via the distance metric.
4. If the resulting distance metric is within the tolerance value then accept the associated parameter values, otherwise reject the associated parameter values.
5. Repeat steps 1-4 until the minimum size of the posterior set has been achieved.

It has been proven that the resulting posterior from ABC-Rejection algorithms can closely match that of the true posterior [108]. The choice of tolerance value is used to balance computational efficiency (time) with posterior precision. The tolerance and distance metric combination, similar to Kennedy and O’Hagan’s method, should be used to incorporate additional sources of uncertainty unaccounted for in the parameters. For this reason it is generally not recommended to choose a tolerance value close to zero, meaning only accepting parameters which result in outputs close to the measured data, unless the user is very confident in the total error of their modelling system [109].

An alternate and more computationally efficient algorithm for ABC is Sequential Monte Carlo, or ABC-SMC [110, 108]. This algorithm is similar in form and function to the ABC-Rejection algorithm, with the addition of a tempering value that progressively morphs from sampling the prior to sampling the posterior distribution [110]. ABC-SMC has been proven to be robust in sampling multi-modal posterior distributions and converges asymptotically to the true posterior faster than ABC-Rejection [109].

ABC-SMC has recently been employed for model inference in biological systems [111]), epidemiology [108], chemical networks [112], and other situations where a deterministic model may be advantageous. ABC has been used for building energy calibration by [113] using an EnergyPlus archetype as a case study to compare the impact of different machine learning algorithms as surrogates in the ABC process, and found that machine learning models with ABC methods can provide reliable parameter estimation with fast computation. In addition, they recommend further research towards the application of the Sequential Monte Carlo method of ABC.

Calibration Threshold Metrics

ASHRAE Guideline 14 provides calibration error metrics to assess the level of calibration [114, 115]: the coefficient of variation of the root mean square error (cvRMSE), and the normalized mean bias error (NMBE).

The maximum allowable errors to be considered calibrated, according to ASHRAE Guideline 14, are shown in Table 4.1. Recent research suggests cvRMSE is the more representative measure of calibration for building energy consumption [116], however, for the purposes of this research we will follow ASHRAE Guideline 14 thresholds.

Table 4.1: ASHRAE Guideline 14 Error Thresholds

Metric	Maximum Error
cvRMSE	15%
NMBE	5%

Consistent with interpretations made by Kennedy and O’Hagan, ASHRAE Guideline 14 incorporates allowable cvRMSE and NMBE errors with the understanding that a model with perfect parameter outputs can still result in a non-perfect match of outputs to measured data [115].

4.1.2 Key Contributions of This Work

This paper documents a case study of Approximate Bayesian Calibration for a large mixed-use retail building. This is the first known demonstration of ABC in a real world building energy model calibration. In addition, it is the first demonstration of the Sequential Monte Carlo method of ABC for building energy calibration. The case studies proves that ABC-SMC with neural network surrogate models can be applied to real-world scenarios without the curse of dimensionality associated with the KOH method. Finally, an analysis of parameter sensitivity compared to information gain from observed data and the impact of re-sampling from mutually independent posterior distributions are presented. The unintended consequences from sensitivity analysis pre-screening is discussed along with the interpretation of posterior distributions and their associated errors.

4.2 Implementation

This section highlights the case study presented in this manuscript. It describes the building of interest and the detailed energy model developed, including the parameters of uncertainty and their associated prior distributions, the surrogate model, and the development, simulation, and analysis environment employed.

4.2.1 Case Study Building

The case used as example in this research is a large indoor mixed-use retail centre. The building is approximately 500,000 ft² with a large central atrium and approximately 100 unique retail units of various sizes. In addition, it houses a cafeteria with a number of fast-food restaurants, a separate sit-down restaurant, a number of office spaces, and a fitness gym. The building is located in a Mediterranean climate (Köppen Csb, ASHRAE 4C).

The building is equipped with a central water-cooled chiller providing chilled water to two-pipe fancoils located in retail units and in the mall atrium. Retail units have electric heating installed as required by the tenant. Lighting in tenant spaces is installed, maintained, and controlled by each individual tenant and the type and amount of lighting varies considerably by tenant. The common space lighting is installed, maintained, and controlled by the building owner, and is predominately LED. Natural gas is consumed by tenant cooking equipment, a number of gas-fired rooftop units, and domestic hot water in the fitness gym.

Electricity for the chilled water plant, common-space lighting, common-space plug loads, and common-space fancoil units is paid for by the property manager. Electricity and natural gas consumed in each retail unit for lighting, plug loads, fancoil units, and gas equipment is paid for by each respective tenant. Monthly common electricity consumption, monthly total tenant electricity consumption, and monthly total natural gas consumption has been provided between the years 2017 to 2019, inclusive, for 108 total data points. It has been noted that increasing the temporal resolution of data may improve the precision of posteriors [117], however, monthly utility data is still the most commonly available temporal resolution of data and thus representative of the industry as a whole. Hourly weather data for the three year period corresponding to the utility data months was generated as an .EPW file from the NASA Power Project [118].

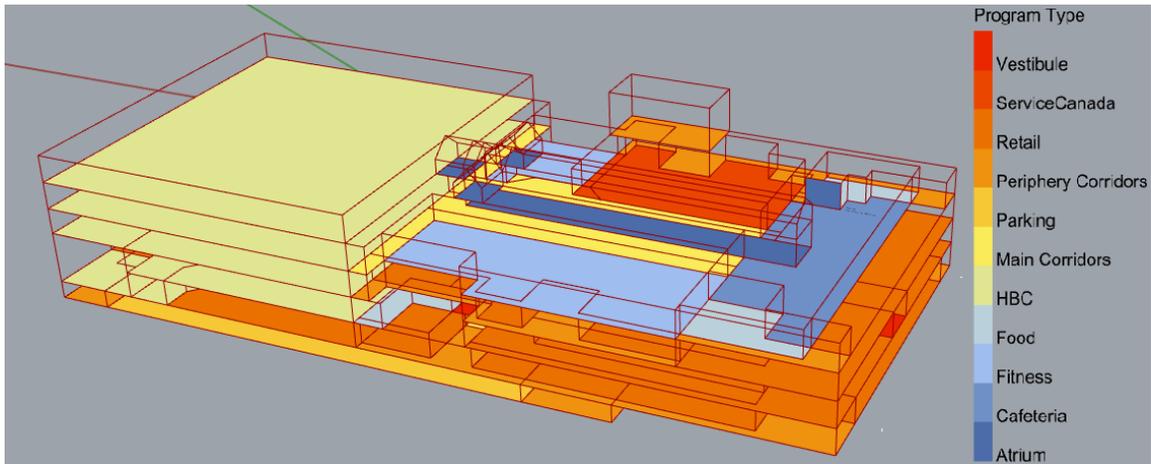


Figure 4.2: A 3D virtual representation of the case study building. The geometry of the building was incorporated into the model using Rhinoceros 7 3D modelling software and translated into a .idf file using Ladybug tools. The building loads and equipment were input on a zone-level basis, and the zones are coloured according to their zone-type.

Dynamic Energy Model

The simulation engine used in this case study is EnergyPlus version 9.5 [119], a well-validated simulation engine and the most commonly used engine in recent building energy model calibration research [77]. The geometric and basic information of the model was generated using Rhino 7 and the Ladybug modelling platform for Grasshopper [120].

The building has been modelled with 6 floors, 71 total zones, with 11 different zone-types. A visual overview of the model is shown in Fig. 4.2. The model includes multiple mechanical system types and process loads. The relative complexity of the model necessitates the calibration of a large number of uncertain parameters, which may pose problems with the pre-existing Bayesian calibration guidelines that recommend a minimal number of parameters. The authors could not find other samples in the literature of whole-building Bayesian calibration that account for the complexity of building, the quantity of meters, and the sources of uncertainty. The authors intentionally selected this case study to demonstrate the applicability of the proposed methods in difficult calibration scenarios.

Parameters

The parameters were chosen based on subject matter expertise and following the energy audit as likely parameters that would influence the monthly energy profile. The allowable range and prior distribution for each parameter was based on information gained during the preliminary energy audit, through site investigations and document review.

All priors used in the analysis follow beta distributions due to their ability to model a diverse set of probability distributions while remaining within a $[0,1]$ bound. This gives the potential to specify both the shape and confidence level of the prior without risking the sampling to enter a physically infeasible value.

Surrogate Model

The surrogate model is a multi-level perceptron neural network using the Keras Python library as an API for the TensorFlow machine learning platform [9]. The model is built with two dense layers and includes batch normalization and dropout between each layer. The inputs to the neural network are the continuous building parameters in question. The outputs are monthly energy consumption for each utility over the course of three years, for 108 total outputs. Inputs and outputs to the surrogate are normalized between $[0, 1]$ by removing the mean and scaling to unit variance. Data is split based on an 80/20 training/test split, and the neural network was trained for 1,000 epochs using the Adam training algorithm. The model was trained to minimize the mean squared error loss.

Analysis and Simulation Environment

The handling of simulations, samplings, surrogate model development and training, inference, and analysis was completed using Jupyter Notebooks version 3.2.8 running Python version 3.8.10. The handling of simulations, parameterization of the simulations, and subsequent sampling was completed using the BESOS library [122]. Surrogate model development, training, and analysis was completed using Keras version 2.6.0. Bayesian inference was completed using PyMC3 version 3.11.4 [12].

Table 4.2: Parameters incorporated in the calibration procedure. The parameter name, units, maximum and minimum allowable values, and normalized beta distribution parameters are provided.

Parameter	Units	Range	Alpha	Beta
Wall R-Value	ft ² ·°F·h/BTU	1-10	3	2
Wall Infiltration	m ³ /s-m ²	5e-5 - 5e-4	2	2
Roof R-Value	ft ² ·°F·h/BTU	18-40	3	5
Glazing U-Value	W/m ² K	2-4	2	3
Glazing SHGC	-	0.2-0.8	2	2
Common Lights	Multiplier*	0.1-2.0	2	3
Tenant Lights	Multiplier*	0.1-2.0	2	3
Lighting Schedule	factor**	0-1	2	3
Common Plug Loads	Multiplier*	0.3-3.0	2	2
Tenant Plug Loads	Multiplier*	0.3-3.0	2	2
Ventilation Rate	Multiplier*	0.5-2.0	3	2
Chiller Nominal COP	W/W	4-7	2	2
Cooking Energy	W/m ²	5-200	1	1
Elevator Energy	kW	20-40	2	2
RTU Gas Efficiency	%	60%-90%	2	2
DHW Gas Efficiency	%	60%-98%	2	2
FCU Fan Pressure	Pa	75-500	1	1
FCU Variable Speed	%***	30%-90%	10	2
DOAS Fan Pressure	Pa	250-1000	3	2
Cooling Setpoint	°C	20.5-22	2	2
Heating Setpoint	°C	18-20	2	2
Vestibule Infiltration	m ³ /s-m ² ****	2.5e-4 - 2.5e-2	2	2
Roof Temperature	°C*****	0-3	2	2

* The values are based on NECB 2015, following a zone-by-zone basis.

These values are multiplied by the parameter value for each zone.

** The base lighting schedule follows NECB 2015 schedule-type C. A value of 0 represent the baseline schedule. A value of 1 results in a schedule of always-on. The schedule is varied proportionally between the baseline and always-on schedules on an hour-by-hour basis.

*** A parameter representing the proportion of fan-coil units that have been converted to variable speed.

**** The main doors to the vestibule were noted to remain open throughout the summer.

This parameter estimates the level of infiltration due to the always-open doors.

***** The roof is a large, flat, dark membrane. These types of roofs can exhibit micro-climatic temperature increases relative to the ambient, which increases the air temperature at ventilation units [121]

4.3 Results

The results for the case study calibration procedure are presented in the following sections. First, the surrogate model is analyzed to ensure the regression predictions are reasonable. Next, a set of cases is used to compare the Bayesian inference approach and the standard industry practice, The calibration results for the case study are then

presented. Finally, the risks of selecting only the most sensitive parameters for the calibration are illustrated and the risks of misattributing the posterior distributions as mutually independent distributions are presented.

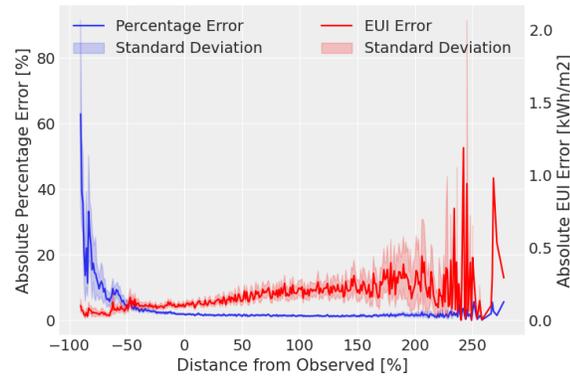
4.3.1 Sampling and Neural Network

For training the surrogate model, 300 samples are generated following a Latin hypercube sampling procedure, which provides an even distribution of samples across the parameter ranges in Table 4.2. Approximately 10 samples per parameter is a standard rule-of-thumb for reasonable surrogate model accuracy [91]. The model is trained on 240 samples and tested on the remaining unseen 60 samples. The model converged at approximately 400 epochs on [0,1] normalized training data. Total accuracy for monthly energy predictions on the test dataset is a mean absolute error (MAE) of 0.3 kWh/m² per month or a 3.5% mean absolute percentage error (MAPE). Important for this study, as shown in Fig. 4.3a and 4.3b, the accuracy at values near to the observed data is around 2%, or 0.1 kWh/m².

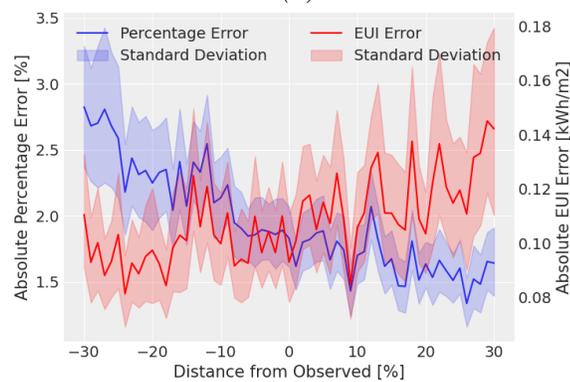
4.3.2 Baseline Energy Model and Standard Practice

The baseline energy model was developed from observations on site and what was deemed as a reasonable starting point for each parameter. The baseline model is simulated, as would happen in a traditional calibration process, and compared to the observed energy meter data. Monthly energy consumption for the three models is shown in Fig. 4.4. The baseline energy model with default values does not meet any of the minimum error metrics in order to be considered calibrated. In a standard calibration procedure, the user would begin iterating the parameters based on intuition and subject-matter expertise in order to reduce the error gap.

For illustrative purposes, three separate examples were selected to represent what results may be expected following a standard calibration procedure. Examples 1 through 3 were selected by iterating the design variables until the NMBE and cvRMSE of all three energy meters were within acceptable limits per ASHRAE Guideline 14. Also included for comparison are the baseline pre-calibration case and maximum a posteriori (MAP) estimate. The MAP estimate is the value most likely value for each parameter following Bayesian inference (inference results are described in further detail in Section 4.3.3). The error metrics are shown in Table 4.3 for each case, metric,



(a)



(b)

Figure 4.3: Mean absolute percentage error and mean absolute error, plus one standard deviation, of the surrogate model prediction compared to the energy model outputs at identical parameter input combinations. The surrogate model is used to predict monthly energy consumption for all three meters across all three years for 108 total outputs. The distance from observed x -axis value is the distance from the surrogate prediction output compared to the building's observed data consumption for that month and meter.

and energy meter. Note that the MAP does not perform significantly better than the other three ‘calibrated’ examples in terms of calibration metrics.

Table 4.3: ASHRAE Guideline 14 error metric results for the three example cases, the baseline case, and the MAP case. A cvRMSE value less than 15% and a NMBE value less than 5% are both required in order to be considered as calibrated. All three examples plus the MAP could be considered as calibrated models, whereas the baseline requires significant tuning.

Case	Meter	cvRMSE	NMBE
Baseline	Tenant Electricity	84.4%	77.0%
	Common Electricity	20.7%	-9.6%
	Natural Gas	75.2%	71.8%
Example 1	Tenant Electricity	7.1%	-4.2%
	Common Electricity	11.5%	-2.7%
	Natural Gas	10.6 %	-0.5%
Example 2	Tenant Electricity	8.1%	-2.9%
	Common Electricity	11.9%	-3.5%
	Natural Gas	8.2%	1.9%
Example 3	Tenant Electricity	5.3%	0.2%
	Common Electricity	11.2%	-1.6%
	Natural Gas	11.2%	-4.4%
MAP	Tenant Electricity	5.6%	1.1%
	Common Electricity	11.7%	-4.2%
	Natural Gas	6.5%	-0.4%

The parameter values for each example case are shown in Fig. 4.5, along with the baseline and MAP cases. Parameter values varied considerably for many parameters, indicating the overparameterization inherent in building energy model calibration and the equifinality that cannot be escaped with building energy simulations. In a standard process, the user may have settled at any of the design options. Also shown in Fig. 4.5 is the likelihood of each parameter according to the prior and the posterior distributions. A number of the parameters in each example are identified as having a low likelihood of being the actual parameter, even though their selection can still meet the calibration thresholds. Without performing the inference procedures, the user would have no indication that their single selection of calibrated model may indeed be a very unlikely candidate.

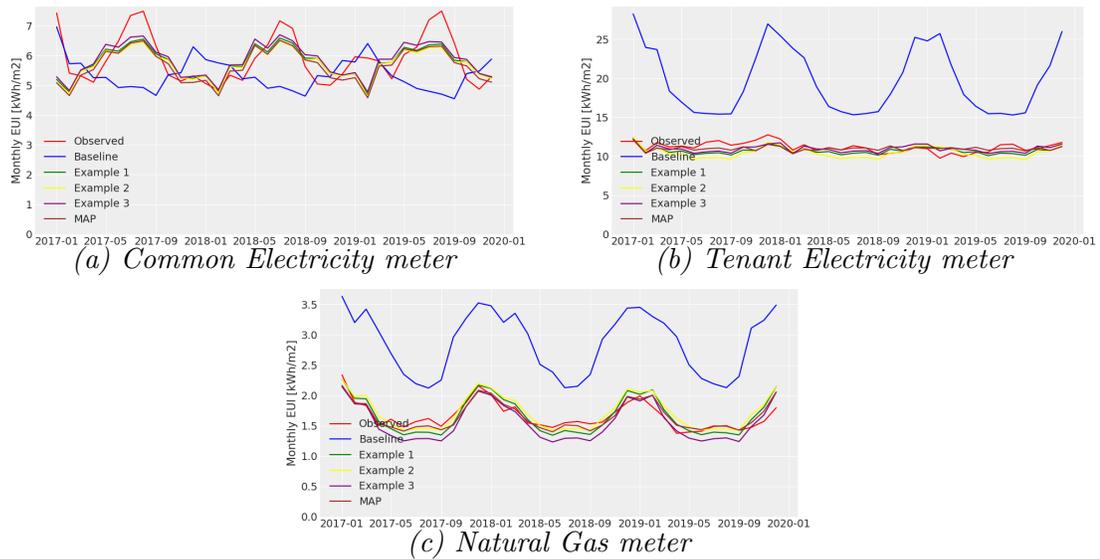


Figure 4.4: Monthly energy consumption various meters across three years of operation. The observed data is the calibration target, whereas the baseline is the best-guess of building characteristics before the calibration procedure. Three examples show alternate manually-tuned calibrated model outputs. The maximum a-posterior (MAP) is the combination of parameter values with the highest likelihood following the Bayesian inference.

4.3.3 Calibration Results

The surrogate model was put through the Bayesian calibration process. Convergence metrics are then assessed to determine posterior representation. Following that, a comparison of prior and posterior parameter distributions are presented, and then monthly energy consumption and error metrics are shown.

ABC-SMC Posterior Convergence Checks

It is important to analyze the performance of the ABC-SMC algorithm. In theory, if the algorithm is allowed to run indefinitely, it will eventually converge on the ‘true posterior’. In reality, due to computational time constraints, we run multiple simultaneous inference chains and define the length of each chain. In order to assess the convergence of the posterior distributions a common metric is the Gelman-Rubin statistic (\hat{R}). The Gelman-Rubin statistic compares the variation within each chain with the variation between each chain. This results in a unitless ratio, which trends towards a value of 1 with improved convergence [101]. An acceptable Gelman-Rubin statistic value is subjective and may change depending on the needs of each project,

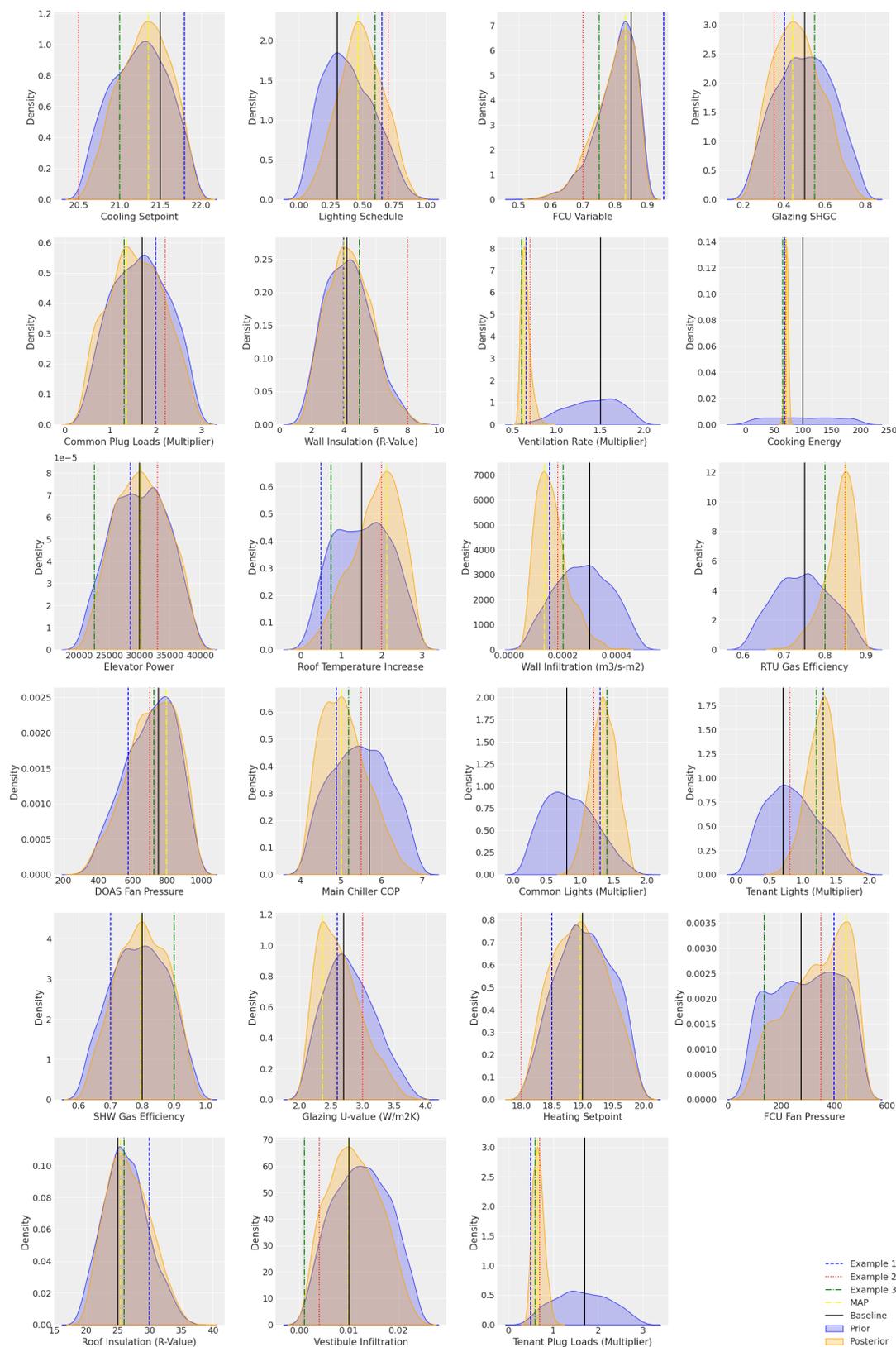


Figure 4.5: A comparison of prior distribution, posterior distribution, baseline case value, example cases values, and MAP value for each parameter. The prior distribution of each parameter represents the knowledge of the user before completing the Bayesian inference steps. The posterior distribution of each parameter represents the knowledge of the parameter after the prior has been conditioned on the observed data. The ranges between example cases show the vast range of possible combinations that can all be considered as calibrated. While most example cases trended from the baseline to a higher likelihood value, there are a number of examples with parameters that exhibit very low likelihoods, but still meet the ASHRAE Guideline 14 targets.

however, a standard rule-of-thumb number is to accept convergence if the Gelman-Rubin statistic is below a value of 1.1 [102]. The Gelman-Rubin statistic for all parameters in our posterior distribution chains are all equal or lower than a value of 1.01, indicating very good convergence has been achieved, and that our posteriors are highly representative of the ‘true’ posterior.

Inference

A comparison of the prior distributions and posterior distributions are presented in Fig. 4.5. The difference between prior distributions and posterior distributions represents the amount of information that is contained within the observed data (i.e., monthly energy consumption). Parameters that observed distinct changes to the shape of the distribution from prior to posterior are those model parameters that are highly influenced by the available observed data. This is in contrast to a sensitivity analysis. A parameter may have a large influence on the final model outputs, but the likelihood of that parameter may not be influenced heavily by the observed data for a variety of reasons. One such reason is the issue of identifiability or overparameterization in building energy models.

While identifiability of parameters due to overparameterization presents problems in other calibration methods, Bayesian inference inherently quantifies the amount of unidentifiability that is present. A parameter distribution that remains unacceptably uncertain in the posterior means that the combination of prior knowledge and observation data is insufficient. The practitioner then has the following three options:

- obtain additional observations,
- improve the prior knowledge of the parameter in question, or
- improve the prior knowledge of other influential parameters.

An example of a parameter that was highly influenced by the observations was the cooking energy parameter. The audit and prior knowledge of the subject were weak, and therefore a flat prior across a large parameter range was selected. In addition, while there are interactive effects by many characteristics of the building contained in other parameters, only three parameters directly related to gas consumption: cooking energy, rooftop unit efficiency, and domestic hot water efficiency. Following inference, only a small subset of the cooking energy parameter could result in a gas consumption

that matched the magnitude and shape of the monthly gas consumption between 2017 and 2019.

This can be compared to the DOAS Fan Pressure parameter, in which no noticeable change occurred between the prior and posterior distributions. Essentially, the parameter did not interact with the model in such a way that the observed data could only be achieved with a subset of the parameter values, even though the outputs were moderately sensitive to that parameter. One such possibility is that sensitivity analysis was conducted across the entire range of a parameter's possible values. If a parameter's impact of outputs was non-linear, the sensitivity analysis may show a large impact on outputs, but at the 'true' range of possible impacts, the parameter has a small impact. Of course, without performing inference, the user would not know which range of values to perform the sensitivity analysis before-hand. For more information on sensitivity analysis refer to Section 4.3.3.

The pre-calibration monthly energy distribution (based on prior distributions), the post-calibration monthly energy distribution (based on posterior distributions), and the observed monthly energy is shown for the common electricity meter in Fig. 4.6a the tenant electricity meter in Fig. 4.6b, and the natural gas meter in Fig. 4.6c. Overall, the posteriors matched the observed data much better, both in terms of magnitude and seasonality. Note that the priors assumptions were developed before the first round of simulations were completed.

The error metrics following calibration for this case study are presented in Fig. 4.6d. The median error values for all meters for both cvRMSE and NMBE are well within the allowable ranges set forth by ASHRAE Guideline 14. All cvRMSE error values, except for a select few outliers in natural gas consumption, fall below the 15% allowable threshold. The NMBE for all meters falls within +/- 5% for values between the 25th and 75th percentile of the posterior. The values on the extreme ends for NMBE for natural gas and the common electricity meter would not fall within the allowable threshold of ASHRAE Guideline 14. The addition of further meters and tightening of the tolerance value would reduce the NMBE and cvRMSE margins.

Sensitivity Analysis Screening

The amount of information gain for each parameter is measured by comparing the relative entropy between the prior and posterior distributions. The statistical distance between the prior and posterior is calculated using the Kullback-Leibler Divergence,

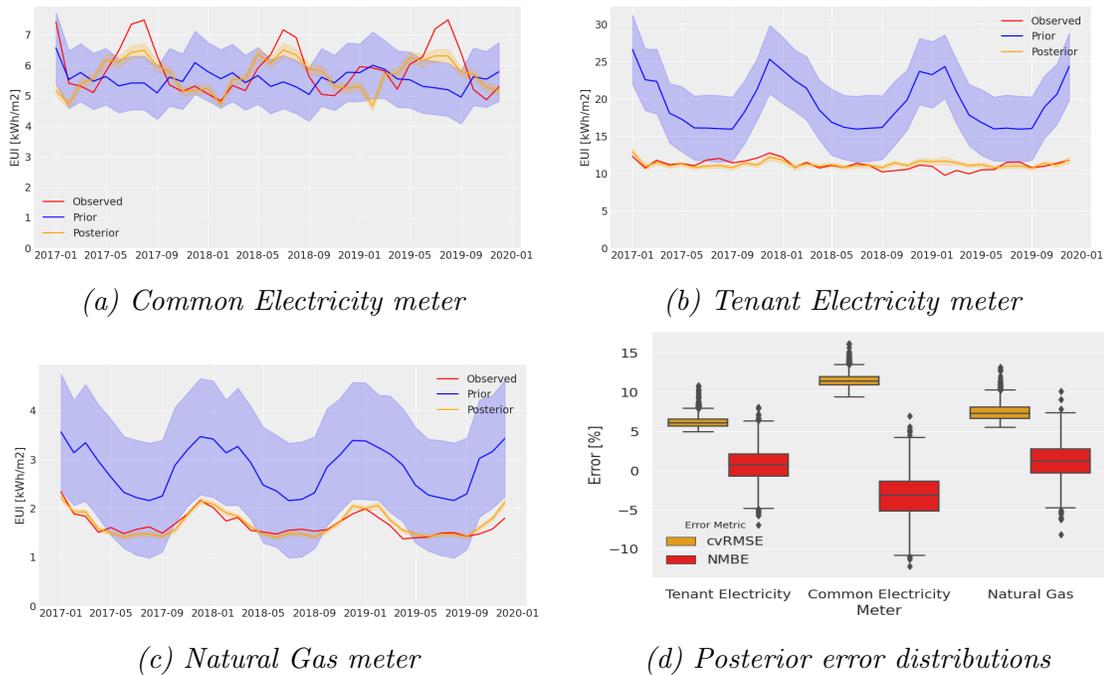


Figure 4.6: Monthly energy consumption across three years of operation and associated posterior error distributions. The observed data represents the ground-truth data that is the target for calibration. The prior distribution of monthly energy consumption is the mean predicted energy consumption from the energy model after independent sampling from the prior parameter distributions, along with a standard deviation of predictions. The posterior is the distribution of monthly energy predictions of parameter combinations following the ABC-SMC procedure. Each distribution represents 2000 samples. The posterior distributions matched the observed data in both seasonality and magnitude much better than the priors. The posteriors could not adequately match the peak summer common electricity consumption, indicating that the combination of parameters and model could never adequately quantify summer consumption and further parameterization would be required.

or KL Divergence, in units of bits. The greater the value of KL Divergence, the greater the distribution has changed. A value of 0 indicates that the two distributions, the prior and posterior, contain identical amounts of information [123].

A sensitivity analysis was done using the Sobol global sensitivity method [86] [124] to rank the most important parameters. Sensitivity was calculated as the impact of parameters on total annual energy consumption for each meter individually. Therefore, if a parameter had a very high impact on one meter but a very low impact on other meters, it would still result in a high relative sensitivity. The KL Divergence and the total sensitivity are compared in Fig. 4.7. The relative amount of information gain appears to be proportional to the relative sensitivity, at least in order of magni-

tude, and a R^2 score of 0.54 indicates that total sensitivity is somewhat correlated to the amount of information gain.

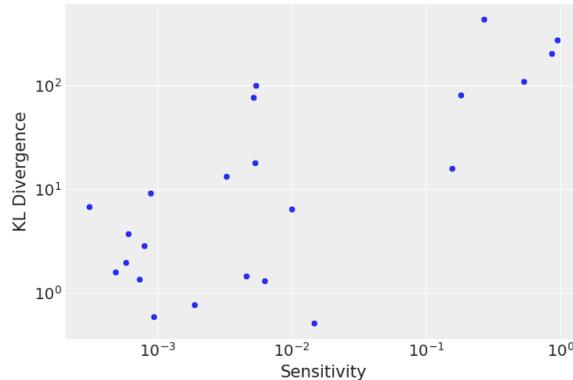


Figure 4.7: A point-wise comparison for each meter on its overall sensitivity to normalized annual outputs with the amount of information gain for that parameter between the prior and posterior distributions. Sensitivity analysis was calculated following the Sobol global variance method, whereas information gain was calculated according to the KL Divergence. Note the log-log scale axes, which are used to visualize the order-of-magnitude impacts. There is a general trend between sensitivity and information gain; however, there remains a few parameters that are orders of magnitude less important but contain similar amounts of information gain.

It is common practice to perform a sensitivity analysis to remove parameters that have less impact on the final results. Removing a parameter from the inference procedure is equivalent to setting that parameter to a value with complete certainty. To illustrate the impact of parameter screening the five most sensitive parameters, representing 93% of total variance, were selected. The priors and observed values remain consistent among both cases, and the parameters removed from the inference are set to their default values from the baseline model. In Fig. 4.8 the resulting posterior from the five most sensitive parameters is compared with the resulting posterior from all possible parameters. The difference in posterior maximum a-posteriori values and their associated levels of likelihood are identified in Table 4.4. The resulting range of calibration error metrics, shown in Fig. 4.9 have increased for some meters, however, the overall scores remain reasonably calibrated.

The results speak to the potential outcome when parameters deemed less important are neglected during the inference process. In general, the confidence in parameters increased, sometimes significantly, towards values that are indeed less likely when the full prior parameter set is considered. The resulting calibration error metrics are only slightly worse than with the full-suite of parameters.

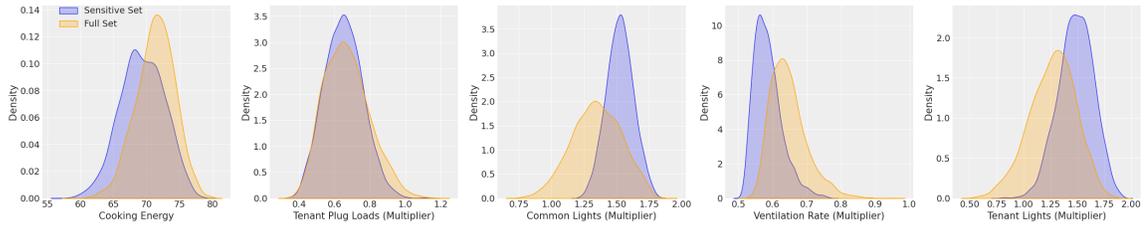


Figure 4.8: A comparison of the five most important parameters' prior and posterior distributions. One scenario completed the same inference procedure with only the five most sensitive parameters, and the other scenario contained all uncertain parameters that were included in the case study. The MAP value is compared both in the value of the parameter and the posterior likelihood of that variable. Of the five parameters, only the ventilation rate multiplier and tenant plug load had MAP changes of less than 5%. The remaining parameters shifted by between 10% and 15%. The likelihood at the MAP increased for four parameters by 17% to 89%, except for the Cooking parameter which decreased by 19%.

Table 4.4: The MAP values and associated likelihood comparison between a scenario including parameter screening, and a scenario including the full-suite of uncertain parameters.

Parameter	Maximum A Posteriori % Difference	Likelihood % Difference at MAP
Cooking	-4.7%	-19.0%
Tenant Plug Load	0.6%	17.3%
Common Lighting	14.7%	88.6%
Ventilation Rate	-10.4%	31.4%
Tenant Lighting	12.1%	24.0%

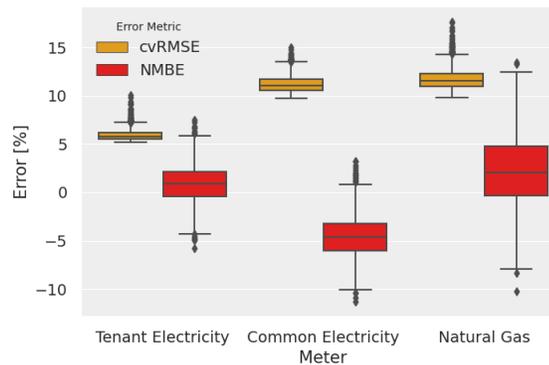


Figure 4.9: The error metrics for each meter from a scenario with only the five most sensitive parameters. The overall error metrics predominately meet the calibration thresholds.

Posterior vs. Independent Distributions

It is important to note that the posteriors generated from Bayesian calibration techniques are presented as independent distributions for each parameter, but they are in fact mutually dependent distributions. The posteriors presented are smoothed kernel density estimates of a large number of individual simulations that meet the tolerance requirements set out in the inference algorithm. A number of parameters that impact the observed data in a similar way may have regions of reasonable likelihood in the posterior that, in combination, are indeed highly unlikely. An example in our case study are the lighting power density and lighting schedule in our example. There may be cases where values that are reasonably likely in the posterior distributions from the parameters cannot both be present at once while meeting calibration thresholds, and the posterior distributions on their own do not adequately represent this.

To illustrate this, Figs. 4.10a, 4.10b, and 4.10c compare the monthly predicted energy consumption from the inference posterior chains with samples that were drawn independently from each posterior along with the observed energy consumption for each meter. The spread is significantly higher for each meter if the posteriors are considered as independent. The ASHRAE Guideline 14 error metrics also exhibit a much wider range, shown in Fig. 4.10d for mutually independent posterior samples, with significantly more combinations of parameters exceeding the thresholds.

4.4 Discussion

The Bayesian method presents an alternative to standard practice. One of the major differences is a shift from a deterministic to a stochastic viewpoint, in which unknown parameters are not selected as a single value that meets the calibration targets, but as likelihood distributions containing many options that meet the calibration targets. This viewpoint can be carried forward throughout the process if retrofits are intended and final savings can be presented in a similar manner with uncertainty accounted for. This may improve the retrofit process by communicating the level of confidence to which estimates are obtained.

The process outlined in this case study is not intended to replace the expert user with a machine, or to automate the calibration entirely. There are steps that are completed programmatically, such as sampling and modifying the simulation input, that would otherwise be far too time-consuming to be completed in a step-by-step

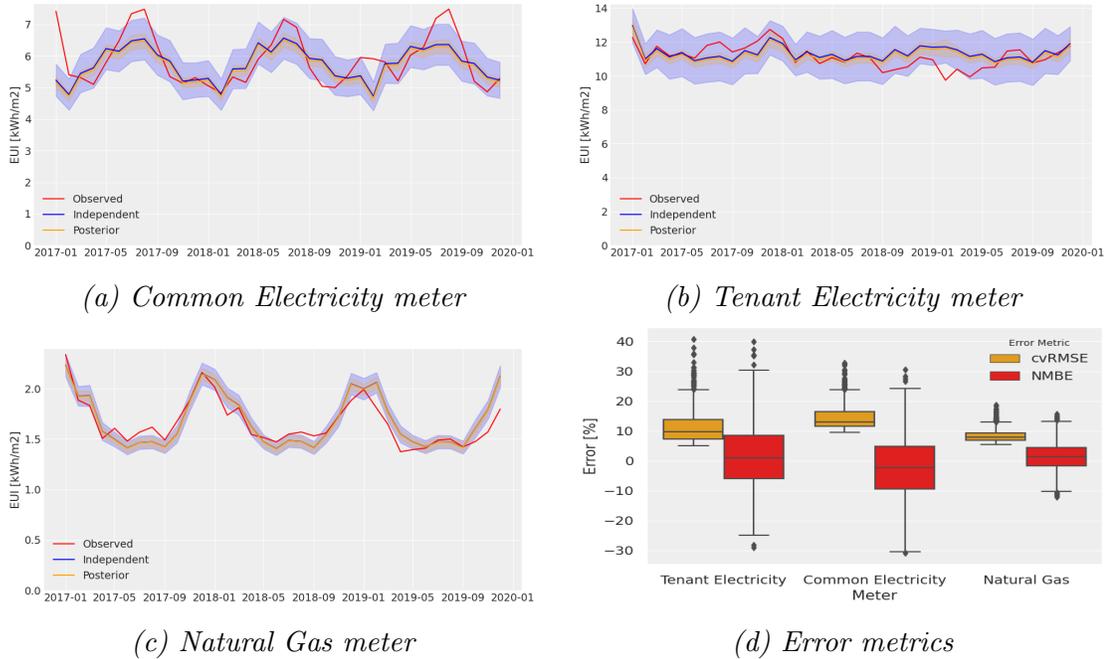


Figure 4.10: Monthly energy consumption various meters across three years of operation, plus error distributions. The mean monthly energy predictions, along with a standard deviation, for each meter, is compared for the posterior samples from the ABC-SMC procedure and from a re-sampling of the individual parameter posterior distributions. Both posterior distributions follow similar patterns, but the independent samples result in a much wider standard deviation. The spread is also represented in the larger spread of error metrics. This is the result of combinations of parameters that may not meet calibration thresholds, but may both be individually possible.

manual process. This also has the side benefit of increasing the possible number of parameters that are investigated. In particular, the inclusion of the neural network in the approximate Bayesian computation approach compared to the Gaussian Processes in the standard KOH calibration approach remove the computational barrier associated with increasing the number of uncertain parameters.

The process outlined in this study is intended to improve the rigour of the calibration process, not by removing the expert user’s bias and intuition, but by quantifying it. The inference exercise advises the user how much information about the parameters is contained by the interaction between the observed data and the model. If there is no shift between the prior and posterior during the inference, then the user cannot use that calibration process, as it stands, to improve their understanding of that parameter. Without performing inference, for example if one were to follow the standard process of manual iteration, there would be no indication of how likely that single choice is.

During the calibration process, a tolerance threshold is identified. This tolerance threshold contains the amount of allowable error between the predicted and observed values, and contains all forms of error associated within the analysis. It includes errors from model simplification, errors from the simulation itself, and in our case, errors from the surrogate model, among others. If we identify that tolerance to be, for example, 5%, then we are declaring that even if all parameter selections were exactly true, the predicted results may still be off from the observed data points by 5% because of the remaining associated errors. It is the author's recommendation to retain the posterior distribution for subsequent steps, however, selecting the maximum a posteriori of each parameter is a best selection if a single combination of parameters is needed.

In completing Bayesian inference the posterior outcomes should never be a surprise or appear incorrect to the user. If that is the case, then there is either an error with the simulation, the observed data, or the user simply did not adequately incorporate their true prior beliefs in the problem. That is not to say that the user should trust implicitly the results from the inference, but it may hint at a need to provide more information to the problem. This may involve the addition of further observed data streams, such as monthly peak demand, or to revisit the site to collect more detailed information of a particular subsystem.

Analyzing discrepancies between the predicted and observed data points can also be insightful to the users. If a posterior distribution has converged satisfactorily (e.g., with a \hat{R} score close to 1), and there remains significant errors between predicted and observed data points, one can confidently say that the current model and parameter combinations statistically cannot explain the observations to a quantifiable degree. To illustrate, in the case study above, summer electricity consumption could not be matched following the calibration procedure, therefore no combination of the parameters and their ranges included, as they associate with the model, could explain the summer electricity consumption. If a close match to summer electricity consumption was desired, the user would need to modify or add further parameters to account for this variation. The addition of further parameters would likely decrease the confidence in the remaining parameters, but would be a truer indication of the actual uncertainty associated with those parameters.

The author strongly recommends to include as many parameters as is feasible, and to select priors as close as possible to the user's true prior beliefs when performing Bayesian inference. Counter to guidance in [89] and [91], among others, the goal of

the process is not to end up with the narrowest posterior distributions of parameters, but to quantify uncertainty in parameters as accurately as possible based on the existing prior beliefs and observations. Priors should not be selected as strong, weak, or uninformed unless that is the true indication of the users prior beliefs [101]. In addition, parameters should not be removed with an intent to narrow down the posterior distributions of the remaining parameters. If the users have a true uncertainty in parameters they should remain, as is, in the procedure. Sensitivity analysis can be used to winnow down parameters to ensure important parameters are not neglected, but as demonstrated in this case study, can have the unintended result of promoting an incorrect posterior with improperly associated confidence. Sensitivity analysis should be used as a method of selecting which parameters are included if there is a capacity limit either due to labour or computation limitations.

Finally, it is worth mentioning that the parameter posteriors resulting from the ABC-SMC process are not independent distributions. They are mutually dependent and if the posteriors are presented in a way which may be interpreted as independent distributions, then the associated error metrics should be independently resampled from the posterior distributions. As was illustrated in the case study above, the resulting error metrics are likely to be significantly worse.

For further discussion on the benefits and limitations of Bayesian analysis, the authors recommend referring to [101].

4.4.1 Limitations and Further Work

This study presented a case of Bayesian calibration applied to a large retail building. Contributions relating to the feasibility of the method; the impact of sensitivity screening; and the mutual dependence of posteriors was presented, however, further research is recommended on a number of related topics.

The distance metrics, summary scores, and tolerance thresholds were not analyzed in detail in this study. They likely have a significant impact on computational effort and accuracy, but establishing guidelines and recommendations require further study.

The case study incorporated three years of energy consumption across three different meters. During the inference process, all 108 data points were used for calibration simultaneously, and with equal weighting. There may be benefits to a cascading calibration, with posteriors from each historic year acting as priors for the subsequent year's calibration.

This case study presented the Bayesian calibration stage as a final step in a project. There may be potential to incorporate these calibrations in an iterative way to reduce the overall effort required. For example, there may be parameters of interest that are difficult to measure. Inferring these values with greater confidence may be likely through further confidence in other parameters that are easier to observe. This has the potential benefit of focusing one's efforts on collecting only the necessary data points to adequately calibrate the model. The final result may be less overall effort during the energy audit and energy model development stages.

The posterior distributions are presented for each parameter independently in this case study, however, there may be further information to be gained by analyzing the set of posterior parameter combinations themselves. A clustering analysis, for example, may elucidate further information on parameter groupings.

4.4.2 Conclusions

The results presented above for a case study of a large mixed-use retail centre show the applicability of performing approximate Bayesian computation for building energy calibration. The process leverages a neural network surrogate for time-efficient calculations without the curse of dimensionality associated with Gaussian Processes.

The selection of parameters and priors still depends on an expert user's subjectivity and experience, and does not replace the expert user. In addition, the steps leading up to the calibration, including reviewing data, building a model, and performing an energy audit, are still necessary. The inference process does, however, inform the user how much information can be gained from observed data compared to prior beliefs gathered.

A sensitivity analysis can support the selection of parameters, and there is some indication that the amount of information gain is proportional to the sensitivity of that parameter, but it is not advised to winnow parameters down below what is needed based on other limitations.

Bayesian inference provides a much-needed step in the calibration process of buildings that improves the rigour and will improve the confidence that both users and the industry at wide will have with building energy calibration. It provides a way to quantify and defend the selection of model parameters based on the available data, and can be used to support and guide the calibration process.

Chapter 5

Conclusions

This thesis presented an overview of three emerging computational methods that can be readily implemented into building engineering methodologies and workflows. These have the potential to reduce the engineering burden that is required to design or retrofit buildings to be low-energy and low-carbon.

In Chapter 2, an artificial neural network was used to support early-stage design decision making. This research demonstrates that machine learning can identify the important parameters to be understood, as well as the relative impact of that understanding. Furthermore, it enables a much quicker computational speed which enables rapid iteration under a lack of detailed information.

In Chapter 3, a genetic algorithm was used to optimize the performance of a archetypical existing building model. The paper shows that significant energy and indoor air quality improvements can be made to existing buildings. Utilizing a genetic algorithm side-steps the significant time requirement that would be needed for a designer to manually iterate over the multitude of interacting parameters.

In Chapter 4, Bayesian inference is used to calibrate an energy model for a real-world complex building. This research provides evidence that Approximate Bayesian Computation is a viable method for incorporating uncertainty in building calibration.

The applications of computational methods have all been applied to reduce building energy consumption and leverage building energy models, but their usefulness in the construction industry does not end there. Future work is recommended to investigate how these methods can be applied to different facets of the industry.

Overall, there are ways to improve engineering efficiency in the industry. Designing and operating buildings to be low-energy and low-carbon is essential for global sustainability, and a reduction in the up-front effort is one less barrier to doing so.

Bibliography

- [1] CleanBC. Roadmap to 2030. Technical report, Government of British Columbia, 2021.
- [2] João Ribeiro, Jan Mischke, Gernot Strube, Erik Sjödin, Jose Luis Blanco, Rob Palter, Jonas Biörck, David Rockhill, and Timmy Andersson. The next normal in construction. Technical report, McKinsey and Co, 2020.
- [3] Chris Van Dronkelaar, Mark Dowson, E Burman, Catalina Spataru, and Dejan Mumovic. A review of the energy performance gap and its underlying causes in non-domestic buildings. *Frontiers in Mechanical Engineering*, 1:17, 2016.
- [4] Paul Westermann, David Rulff, Kevin Cant, Gaelle Faure, and Ralph Evins. Net-Zero Navigator. *Preprint on EnerArxiv*, 2020.
- [5] Nicholas E.P. Fernandez, Srinivas Katipamula, Weimin Wang, YuLong Xie, Mingjie Zhao, and Charles D. Corbin. Impacts of Commercial Building Controls on Energy Savings and Peak Load Reduction. Technical Report PNNL-25985, 1400347, May 2017.
- [6] Daniel S. Bertoldi. Deep energy retrofits using the integrative design process: Are they worth the cost. Master’s thesis, The University of San Francisco, 2015.
- [7] D. Coakley, P. Raftery, and M. Keane. A review of methods to match building energy simulation models to measured data. *Renewable and sustainable energy reviews*, 37:123–141, 2014.
- [8] Travis Gliedt and Christina E Hoicka. Energy upgrades as financial or strategic investment? energy star property owners and managers improving building energy performance. *Applied Energy*, 147:430–443, 2015.

- [9] M. Abadi, P. Barham, et al. TensorFlow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pages 265–283, 2016.
- [10] David Hadka. Project-Platypus, 2015.
- [11] Santosh Philip. Eppy, September 2016.
- [12] J. Salvatier, T. Wiecki, and C. Fonnesbeck. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:e55, 2016.
- [13] Jinming Zou, Yi Han, and Sung-Sau So. Overview of artificial neural networks. *Artificial Neural Networks*, pages 14–22, 2008.
- [14] Seyedali Mirjalili, Jin Song Dong, Ali Safa Sadiq, and Hossam Faris. *Genetic Algorithm: Theory, Literature Review, and Application in Image Reconstruction*, pages 69–85. Springer International Publishing, Cham, 2020.
- [15] Koosha Khalvati, Seongmin A. Park, Saghar Mirbagheri, Remi Philippe, Mariateresa Sestito, Jean-Claude Dreher, and Rajesh P. N. Rao. Modeling other minds: Bayesian inference explains human choices in group decision-making. *Science Advances*, 5(11):eaax8783, 2019.
- [16] Jasmin Raymond. Colloquium 2016: Assessment of subsurface thermal conductivity for geothermal applications. *Canadian Geotechnical Journal*, 55(9):1209–1229, 2018. ISBN: 0008-3674 Publisher: NRC Research Press.
- [17] Ercan Atam and Lieve Helsen. Ground-coupled heat pumps: Part 1 – literature review and research challenges in modeling and optimal control. *Renewable and Sustainable Energy Reviews*, 54:1653–1667, 2016.
- [18] Min Li and Alvin C. K. Lai. Review of analytical models for heat transfer by vertical ground heat exchangers (ghes): A perspective of time and space scales. *Applied Energy*, 151:178–191, 2015.
- [19] P Eskilson. Thermal analysis of heat extraction boreholes. 1987.
- [20] H. Yang, P. Cui, and Z. Fang. Vertical-borehole ground-coupled heat pumps: A review of models and systems. *Applied Energy*, 87(1):16–27, 2010.

- [21] Stephane Bertagnolio, Michel Bernier, and Michaël Kummert. Comparing vertical ground heat exchanger models. *Journal of Building Performance Simulation*, 5(6):369–383, 2012.
- [22] Georgios Florides and Soteris Kalogirou. Ground heat exchangers—a review of systems, models and applications. *Renewable Energy*, 32(15):2461–2478, 2007.
- [23] Messaoud Badache, Parham Eslami-Nejad, Mohamed Ouzzane, Zine Aidoun, and Louis Lamarche. A new modeling approach for improved ground temperature profile determination. *Renewable Energy*, 85:436–444, 2016. ISBN: 0960-1481 Publisher: Elsevier.
- [24] Sarah Signorelli and T. Kohl. Regional ground surface temperature mapping from meteorological data. *Global and Planetary Change*, 40(3-4):267–284, 2004. ISBN: 0921-8181 Publisher: Elsevier.
- [25] Asal Bidarmaghz, Guillermo A. Narsilio, Ian W. Johnston, and Stuart Colls. The importance of surface air temperature fluctuations on long-term performance of vertical ground heat exchangers. *Geomechanics for Energy and the Environment*, 6:35–44, 2016. ISBN: 2352-3808 Publisher: Elsevier.
- [26] Shiyu Zhou, Dawei Liu, Shubo Cao, Xiaoping Liu, and Yucheng Zhou. An application status review of computational intelligence algorithm in gshp field. *Energy and Buildings*, 203:109424, 2019.
- [27] Sang Ku Park, Hyeun Jun Moon, Kyung Chon Min, Changha Hwang, and Suduk Kim. Application of a multiple linear regression and an artificial neural network model for the heating performance analysis and hourly prediction of a large-scale ground source heat pump system. *Energy and Buildings*, 165:206–215, 2018.
- [28] Ling Zhou, Yanjun Zhang, Zhongjun Hu, Ziwang Yu, Yinfei Luo, Yude Lei, Honglei Lei, Zhihong Lei, and Yueqiang Ma. Analysis of influencing factors of the production performance of an enhanced geothermal system (egs) with numerical simulation and artificial neural network (ann). *Energy and Buildings*, 200:31–46, 2019.
- [29] P. Westermann and R. Evins. Surrogate modelling for sustainable building design—a review. *Energy and Buildings*, 198:170–186, 2019.

- [30] E. González-Romera, M. A. Jaramillo-Morán, and D. Carmona-Fernández. Monthly electric energy demand forecasting with neural networks and Fourier series. *Energy Conversion and Management*, 49(11):3135–3142, 2008.
- [31] Sebastian Pinnau and Cornelia Breilkopf. Determination of thermal energy storage (tes) characteristics by fourier analysis of heat load profiles. *Energy Conversion and Management*, 101:343–351, 2015.
- [32] Stephen P. Kavanaugh and Kevin D. Rafferty. *Geothermal Heating and Cooling: Design of Ground-source Heat Pump Systems*. ASHRAE, 2014. Google-Books-ID: 8u4OogEACAAJ.
- [33] Mohammadamin Ahmadfard and Michel Bernier. Modifications to ashrae’s sizing method for vertical ground heat exchangers. *Science and Technology for the Built Environment*, 24(7):803–817, 2018.
- [34] Stuart J. Self, Bale V. Reddy, and Marc A. Rosen. Geothermal heat pump systems: Status review and comparison with other heating options. *Applied Energy*, 101:341–348, 2013.
- [35] K. J. Chua, S. K. Chou, and W. M. Yang. Advances in heat pump systems: A review. *Applied Energy*, 87(12):3611–3624, 2010.
- [36] NRCan. Ground-source heat pump project analysis. Ministry of Natural Resources, Canada, May 2001.
- [37] Daniel E. Fisher, Simon J. Rees, S. K. Padhmanabhan, and A. Murugappan. Implementation and validation of ground-source heat pump system models in an integrated building and system simulation environment. *HVAC&R Research*, 12(sup1):693–710, 2006. Publisher: Taylor & Francis.
- [38] Daeho Kang and Heejin Cho. Modeling of central ground-source heat pump system in energyplus. 2016.
- [39] Sankaranarayanan Padhmanabhan. Modeling, verification and optimization of hybrid ground source heat pump systems in energyplus. 2005.
- [40] Canmet-Energy. *BTAP*, 2019. <https://github.com/canmet-energy/btap>.

- [41] Patrick Lambert. Adapting Geothermal Flow to Heat Pump's Requirement, 2013.
- [42] Juan D Rodriguez, Aritz Perez, and Jose A Lozano. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE transactions on pattern analysis and machine intelligence*, 32(3):569–575, 2009.
- [43] Stephen Taylor. Making VAV Great Again. *ASHRAE Journal*, pages 64–71, August 2018.
- [44] Jian Zhang, Robert G. Lutes, Guopeng Liu, and Michael R. Brambley. Energy Savings for Occupancy-Based Control (OBC) of Variable-Air-Volume (VAV) Systems. Technical Report PNNL-22072, 1063080, January 2013.
- [45] Edward Arens, Hui Zhang, Tyler Hoyt, Soazig Kaam, Fred Bauman, Yongchao Zhai, Gwelen Paliaga, Jeff Stein, Reinhard Seidl, Brad Tully, Julian Rimmer, and Jørn Toftum. Effects of diffuser airflow minima on occupant comfort, air mixing, and building energy use (RP-1515). *Science and Technology for the Built Environment*, 21(8):1075–1090, November 2015.
- [46] Andrew Persily. Challenges in developing ventilation and indoor air quality standards: The story of ASHRAE Standard 62. *Building and Environment*, 91:61–69, September 2015.
- [47] ASHRAE. ANSI/ASHRAE Standard 62.1-2010. Ventilation for acceptable indoor air quality. 2010, 2010.
- [48] Barry Abramson and Lung-Sing Wong. Application of ASHRAE Standards and procedures in LEED-EB certification. *ASHRAE Transactions*, 117(1), January 2011.
- [49] Tom Hudson. Building simulation for LEED[R] EQc2 credit. *ASHRAE Journal*, 49(9), September 2007.
- [50] ASHRAE. ASHRAE 90.1-2013 Energy Standard for Buildings Except Low-Rise Residential Buildings. Technical report, American Society of Heating Refrigerating and Air Conditioning Engineers, 2013.
- [51] Ran Liu, Jin Wen, Adam Regnier, Xiaohui Zhou, and Curtis Klaassen. Stability and Accuracy of VAV Box Control at Low Flows. Technical Report RP-1353,

- American Society of Heating Refrigerating and Air Conditioning Engineers, 2012.
- [52] William J. Fisk, Douglas Black, and Gregory Brunner. Changing ventilation rates in U.S. offices: Implications for health, work performance, energy, and associated economics. *Building and Environment*, 47:368–372, January 2012.
- [53] Behrang Chenari, João Dias Carrilho, and Manuel Gameiro da Silva. Towards sustainable, energy-efficient and healthy ventilation strategies in buildings: A review. *Renewable and Sustainable Energy Reviews*, 59:1426–1447, June 2016.
- [54] Steven J. Emmerich, Kevin Y. Teichman, and Andrew K. Persily. Literature review on field study of ventilation and indoor air quality performance verification in high-performance commercial buildings in North America. *Science and Technology for the Built Environment*, 23(7):1159–1166, October 2017.
- [55] Ke Yu-Pei, Stanley A Mumma, and Dennis Stanke. Simulation Results and Analysis of Eight Ventilation Control Strategies in VAV Systems - ProQuest. *ASHRAE Transactions*, 103:381, 1997.
- [56] Andrew Kusiak and Mingyang Li. Optimal decision making in ventilation control. *Energy*, 34(11):1835–1845, November 2009.
- [57] Changzhi Dai, Li Lan, and Zhiwei Lian. Method for the determination of optimal work environment in office buildings considering energy consumption and human performance. *Energy and Buildings*, 76:278–283, June 2014.
- [58] Tom Ben-David, Adams Rackes, L. James Lo, Jin Wen, and Michael S. Waring. Optimizing ventilation: Theoretical study on increasing rates in offices to maximize occupant productivity with constrained additional energy use. *Building and Environment*, 166:106314, December 2019.
- [59] Farhad Mofidi and Hashem Akbari. Integrated optimization of energy costs and occupants’ productivity in commercial buildings. *Energy and Buildings*, 129:247–260, October 2016.
- [60] Nabil Nassif, Stanislaw Kajl, and Robert Sabourin. Optimization of HVAC Control System Strategy Using Two-Objective Genetic Algorithm. *HVAC&R Research*, 11(3):459–486, July 2005.

- [61] Liang Zhou and Fariborz Haghighat. Optimization of ventilation system design and operation in office environment, Part I: Methodology. *Building and Environment*, 44(4):651–656, April 2009.
- [62] Thad Godish and John D. Spengler. Relationships Between Ventilation and Indoor Air Quality: A Review. *Indoor Air*, 6(2):135–145, 1996.
- [63] Ethan Png, Seshadhri Srinivasan, Korkut Bekiroglu, Jiang Chaoyang, Rong Su, and Kameshwar Poolla. An internet of things upgrade for smart and scalable heating, ventilation and air-conditioning control in commercial buildings. *Applied Energy*, 239:408–424, April 2019.
- [64] Mike Schell. Proven Energy Savings with DCV Retrofits. *HPAC Engineering*, pages 41–47, February 2001.
- [65] Jason Palmer, Nicola Terry, and Peter Armitage. Building Performance Evaluation Programme: Findings from non-domestic projects. Technical report, Innovate UK, January 2016.
- [66] M. Deru, K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, B. Liu, M. Halverson, D. Winiarski, M. Rosenberg, M. Yazdanian, J. Huang, and D. Crawley. U.S. Department of Energy Commercial Reference Building Models of the National Building Stock. Technical Report NREL/TP-5500-46861, 1009264, February 2011.
- [67] S. Papadopoulos and E. Azar. Optimizing HVAC operation in commercial buildings: A genetic algorithm multi-objective optimization framework. In *2016 Winter Simulation Conference (WSC)*, pages 1725–1735, December 2016.
- [68] Theodor Christaanse, Paul Westermann, Will Beckett, Gaelle Faure, and Ralph Evins. *besos: Building and Energy Simulation, Optimization and Surrogate-modelling*, 2021.
- [69] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002.
- [70] Environment and Climate Change Canada. Pan-Canadian framework on clean growth and climate change. Technical report, The Government of Canada, 2017.

- [71] B. Haley and R. Torrie. Canada’s climate retrofit mission. Technical report, Efficiency Canada, 2021.
- [72] M. Krieske, H. Huafen, and T. Egnor. The scalability of the building retrofit market: A review study. In *2014 IEEE Conference on Technologies for Sustainability (SusTech)*, pages 184–191, 2014.
- [73] E. Franconi, K. Field, and M. Deru. Building performance modeling for gaining investor confidence. In *13th Conference of International Building Performance Simulation Association*, 2013.
- [74] Trevyr Meade. Canada infrastructure bank uses investor confidence ”investor ready energy efficiency” (iree) certification for 2 billion in energy-efficiency investments, 2021.
- [75] EDF. Energy performance protocol - targeted commercial. Technical report, Environmental Defense Fund, 2013.
- [76] C. Fluhrer, E. Maurer, and A. Deshmukh. Achieving radically energy efficient retrofits: The Empire State Building example. *ASHRAE Transactions*, 2010.
- [77] A. Chong, Y. Gu, and H. Jia. Calibrating building energy simulation models: A review of the basics to guide future work. *Energy and Buildings*, 253:111533, 2021.
- [78] ASHRAE. Ashrae standard 211: Standard for commercial building energy audits. Technical report, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 2018.
- [79] Rocky Mountain Institute. Deep energy retrofits: An emerging opportunity. Technical report, RMI, AIA, 2013.
- [80] Liu G., B. Liu, et al. Advanced energy retrofit guide office buildings. Technical report, PNNL, 2011.
- [81] Luka Matutinovic. Using calibrated energy models to help understand, manage and improve existing building energy performance. In *eSIM 2014 Conference Proceedings*, 2014.

- [82] M. Shamsi, W. O’Grady, et al. A generalization approach for reduced order modelling of commercial buildings. *Energy Procedia*, 122:901–906, 2017.
- [83] Z. O’Neill and B. Eisenhower. Leveraging the analysis of parametric uncertainty for building energy model calibration. In *Building simulation*, volume 6, pages 365–377. Springer, 2013.
- [84] L. Sanhudo, N. Ramos, et al. Building information modeling for energy retrofitting—a review. *Renewable and Sustainable Energy Reviews*, 89:249–260, 2018.
- [85] D. Hou, I. Hassan, and L. Wang. Review on building energy model calibration by Bayesian inference. *Renewable and Sustainable Energy Reviews*, 143:110930, 2021.
- [86] E. Fabrizio and V. Monetti. Methodologies and advancements in the calibration of building energy models. *Energies*, 8(4):2548–2574, 2015.
- [87] M. Ardehali, T. Smith, et al. Building energy use and control problems: an assessment of case studies. *ASHRAE Transactions*, 109:111, 2003.
- [88] T. Hong, J. Langevin, and K. Sun. Building simulation: Ten challenges. *Building Simulation*, 11(5):871–898, October 2018.
- [89] D. H. Yi, D. W. Kim, and C. S. Park. Parameter identifiability in Bayesian inference for building energy models. *Energy and Buildings*, 198:318–328, 2019.
- [90] M. Manfren, N. Aste, and R. Moshksar. Calibration and uncertainty analysis for computer models—a meta-model based approach for integrated building energy simulation. *Applied energy*, 103:627–641, 2013.
- [91] A. Chong and K. Menberg. Guidelines for the Bayesian calibration of building energy models. *Energy and Buildings*, 174:527–547, 2018.
- [92] C. Calama-González, P. Symonds, et al. Bayesian calibration of building energy models for uncertainty analysis through test cells monitoring. *Applied Energy*, 282:116118, 2021.
- [93] F. Roberti, U. Oberegger, and A. Gasparella. Calibrating historic building energy models to hourly indoor air and surface temperatures: Methodology and case study. *Energy and Buildings*, 108:236–243, 2015.

- [94] Kevin Cant and Ralph Evins. Optimizing vav terminal box minimum positions using dynamic simulations to improve energy and ventilation performance. In *Proceedings of the 17th IBPSA Conference*, 2021.
- [95] T. Agami Reddy. Literature review on calibration of building energy simulation programs: uses, problems, procedures, uncertainty, and tools. *ASHRAE transactions*, 112:226, 2006.
- [96] N. Zibin, R. Zmeureanu, and J. Love. Automatic assisted calibration tool for coupling building automation system trend data with commissioning. *Automation in Construction*, 61:124–133, 2016.
- [97] S. Sansregret, K. Lavigne, et al. Excalibbem: Tool development to calibrate building energy models. In *eSIM 2014 Conference Proceedings*, 2014.
- [98] G. Chaudhary, J. New, et al. Evaluation of “autotune” calibration against manual calibration of building energy models. *Applied energy*, 182:115–134, 2016.
- [99] S. Asadi, E. Mostavi, et al. Building energy model calibration using automated optimization-based algorithm. *Energy and Buildings*, 198:106–114, 2019.
- [100] M. Riddle and R. Muehleisen. A guide to Bayesian calibration of building energy models. In *Building simulation conference*, 2014.
- [101] R. van de Schoot, S. Depaoli, et al. Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1):1–26, 2021.
- [102] A. Gelman, J. Carlin, et al. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.
- [103] M. Kennedy and A. O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.
- [104] P. Westermann and R. Evins. Using Bayesian deep learning approaches for uncertainty-aware building energy surrogate models. *Energy and AI*, 3:100039, 2021.

- [105] R. Edwards, J. New, et al. Constructing large scale surrogate models from big data and artificial intelligence. *Applied energy*, 202:685–699, 2017.
- [106] M. Sunnåker, A. Busetto, et al. Approximate Bayesian computation. *PLoS computational biology*, 9(1):e1002803, 2013.
- [107] T. Toni and M. Stumpf. Tutorial on ABC rejection and ABC SMC for parameter estimation and model selection. *arXiv preprint arXiv:0910.4472*, 2009.
- [108] T. Kypraios, P. Neal, and D. Prangle. A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation. *Mathematical Biosciences*, 287:42–53, 2017. 50th Anniversary Issue.
- [109] Richard D. Wilkinson. Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. *Statistical applications in genetics and molecular biology*, 12(2):129–141, 2013.
- [110] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.
- [111] T. Toni, D. Welch, et al. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202, 2009.
- [112] G. Molyneux and A. Abate. ABC SMC 2: Simultaneous inference and model checking of chemical reaction networks. In *International Conference on Computational Methods in Systems Biology*, pages 255–279. Springer, 2020.
- [113] C. Zhu, W. Tian, et al. Uncertainty calibration of building energy models by combining approximate Bayesian computation and machine learning algorithms. *Applied Energy*, 268:115025, 2020.
- [114] A. Garrett and J. New. Suitability of ashrae guideline 14 metrics for calibration. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2016.
- [115] J. Haberl, D. Claridge, and C. Culp. Ashrae’s guideline 14-2002 for measurement of energy and demand savings: How to determine what was really saved by the retrofit. 2005.

- [116] S. Martinez, P. Eguia, et al. A performance comparison of multi-objective optimization-based approaches for calibrating white-box building energy models. *Energy and Buildings*, 216:109942, 2020.
- [117] M. Kristensen, R. Choudhary, and S. Petersen. Bayesian calibration of building energy models: Comparison of predictive accuracy using metered utility data of different temporal resolution. *Energy Procedia*, 122:277–282, 2017.
- [118] Adam H. Sparks. nasapower: a NASA POWER global meteorology, surface solar energy and climatology data client for r. *Journal of Open Source Software*, 3(30):1035, 2018.
- [119] D. Crawley, L. Lawrie, et al. EnergyPlus: creating a new-generation building energy simulation program. *Energy and Buildings*, 33(4):319–331, 2001.
- [120] M. Roudsari, M. Pak, et al. Ladybug: a parametric environmental plugin for Grasshopper to help designers create an environmentally-conscious design. In *Proceedings of the 13th international IBPSA conference held in Lyon, France Aug*, pages 3128–3135, 2013.
- [121] A. Green, L. Gomis, et al. Above-roof air temperature effects on HVAC and cool roof performance: Experiments and development of a predictive model. *Energy and Buildings*, 222:110071, 2020.
- [122] P. Westermann, T. Christiaanse, et al. BESOS: Building and energy simulation, optimization and surrogate modelling. *Journal of Open Source Software*, 6(60):2677, 2021.
- [123] Jonathon Shlens. Notes on kullback-leibler divergence and likelihood. *arXiv preprint arXiv:1404.2000*, 2014.
- [124] Ilya M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280, 2001.