

**UNIVERSITY OF NOTTINGHAM
DEPARTMENT OF ARCHITECTURE
AND BUILT ENVIRONMENT**

Research Project ABEE3023

Impact of data length on Modelica-based energy modeling calibration

Supervisor: Zhiang ZHANG

Student: Jianru JING

ID: 20231712

Hand in: 2023/05/16

Abstract

Energy modeling studies of buildings are becoming more common as efforts to improve energy efficiency and reduce carbon emissions. Nowadays, many studies have started to focus on the accuracy of building energy models. As a result, there are many methods to improve the accuracy of the models and an increasing number of calibration methods for these building energy models. However, the accuracy and length of calibration are affected by a number of factors. Heo [1] studied the effect of calibration data intervals on accuracy. However, to minimize cost-effectiveness, this study focused on selecting representative data. On this basis, this study explores the effect of calibration data length on the simple optimization-based calibration performance of a Modelica-based MAU model using real operational data from an industrial plant. This will be achieved by 1) using K-medoids to select the representative subset with different data lengths from the whole observed data for both calibration and validation sets. 2) using Principal Component Analysis (PCA) and Kullback-Leibler (KL) divergence to analyze the representativity of different subsets. 3) using 10-fold cross-validation to iterate to calibrate and validate 4) comparing the performance of different subsets of models before and after calibration based on simple optimization, in terms of both stability of calibration parameter results and accuracy of prediction results. This study finds that more data does not necessarily lead to better calibration results. A dataset which has less representative of these data will also affect the calibration results, resulting in deviating from parameter values appropriate for representative data.

Acknowledgment

I would like to thank my final-year project supervisor, Dr. Zhiang Zhang of the Department of Architecture and Built Environment, for his guidance in my dissertation. I also would like to thank Shanghai D-TNRIUG Technology Co., Ltd for providing the operated data and models of the Tongwei industry.

Content

Abstract	1
Acknowledgment	1
1 Introduction	4
2 Literature Review	5
2.1 Modeling Method	5
2.2 Calibration Method	5
2.3 Data Processing Method Before Calibration	5
2.4 Objective function and error functions	6
3 Aim	11
4 Objectives	11
5 Methodology	11
5.1 Overview	11
5.2 Data Selection	12
5.2.1 Representative subset selection	12
5.2.2 Subset verification	13
5.2.3 Divergence comparison	13
5.3 Simple optimization-based calibration and validation	13
5.4 K-fold Cross Validation	14
5.5 Model Performance Evaluation	14
5.5.1 Calibration parameter results stability	14
5.5.2 Validation and performance comparison	15
6 Results and Discussion	15
6.1 Building and Equipment Model Description	15
6.2 Data normalization	16
6.3 Data Cleaning	17
6.4 Representative Subset Selection & Verification	18
6.5 Different Subsets Distribution	21
6.6 10-Fold Cross-Validation Results and Model Performance	21
6.6.1 Parameter calibration stability	21
6.6.2 Prediction accuracy	23
7 Conclusion	26
8 Recommendations For Further Study	27
9 Reference List	28

Content of Tables

Table 1 Commonly used thresholds	6
Table 2 Literature for different modeling methods	7
Table 3 Literature for different calibration methods	9
Table 4 Modelica-based MAU model	15
Table 5 Parameters of calibration model.....	15
Table 6 Calibration Parameter initial range.....	16
Table 7 Principal components	20
Table 8 Distributions of parameter results	22
Table 9 Suitable parameter values.....	23
Table 10 Results of MAU1 with 1000 data.....	24
Table 11 Results of MAU1 with 2000 data.....	24
Table 12 Results of MAU1 with 3000 data.....	24
Table 13 Results of MAU1 with 4000 data.....	25
Table 14 Results of MAU1 with 5000 data.....	25
Table 15 4 Principal components	31
Table 16 4 Principal components	32
Table 17 4 Principal components	33
Table 18 4 Principal components	34
Table 19 4 Principal components	35

Content of Figures

Figure 1 Overview of Data Selection.....	11
Figure 2 Overview of Calibration and Validation	12
Figure 3 Distributions of different calibration inputs.....	16
Figure 4 Normalized Values of Calibration Inputs.....	17
Figure 5 Distributions of model inputs	18
Figure 6 Calibration input normalized values.....	18
Figure 7 K-Medoids clustering cost.....	19
Figure 8 Average KL divergence.....	21
Figure 9 RMSE Compare with different data lengths.....	26

1 Introduction

Building Energy Modeling (BEM) is increasingly used to analyze and predict building energy use, Measurement and Verification (M&V) processes, and evaluate Energy Conservation Measures (ECMs) [2]. These models require a lot of computer power and a good understanding of the structure being simulated. They can simulate systems based on a range of previously unobserved initial conditions. However, when designing a building, it may be difficult to obtain all the necessary data, forcing the user to rely on various assumptions. Consequently, this can lead to deviations of simulation results from actual measurements, resulting in a widely recognized problem known as the ‘performance gap’, denoting the inconsistency between simulated and observed values. This study focuses on a small component of the HVAC system, the make-up air unit (MAU), and provides insight into the effect of calibration data length on the calibration performance of the Modelica-based MAU model. The calibration of an MAU model can be viewed as the process of adjusting the model parameters to close the ‘performance gap’. This is essential for predicting and improving their energy efficiency and operational effectiveness. While more data can theoretically improve the accuracy of models, collecting and processing large amounts of data is both time-consuming and resource-intensive, and overfitting issues can lead to less accurate predictions. Therefore, these problems should be balanced against the demand to accurately calibrate models.

Automated calibration is a process that does not require user input and is based on mathematical formulas to match simulation results with real data. The input variables need to correspond to real-world properties, even if there are differences. By using math and statistics in a computer program, calibration can be done systematically across many simulations with different parameters. The process usually ends when the ‘performance gap’ is low enough or when a time limit is reached. The calibrated outcome is the set of input values that give the lowest error rate in a specific simulation. The process is always faster and more efficient than manual calibration, especially for a huge number of parameters. However, balancing the ‘performance gap’ and the time limit is also challenging because automated calibration is iterative and requires a lot of computing power.

2 Literature Review

2.1 Modeling Method

Modeling and simulation of MAU, and other HVAC components is common practice in the design [3], [4], optimization [5], [6], and control of such systems[7], [8]. The development and application of models, especially the integration of component-level representations into plant-level systems, requires the use of a variety of models. Models can be broadly categorized into two types through modeling methods. One is the physical method, through some design-centric tools such as EnergyPlus, TRNSYS, and Modelica. There is a great need for facility information that is hardly found for these models, it will lead to a ‘performance gap’. Therefore, studies are paying more attention to the accuracy of physical models. The other is the data-driven method. Idowu et al. [9] use support vector regression (SVR), artificial neural networks (ANN), multiple linear regression, and regression trees to forecast the heat load in district heating systems. Ahmad et al. [10] compare random forest (RF) and ANN for predicting a building’s energy consumption. Bikmukhametov [11] has shown that data-driven models have better accuracy when combined with physical models. Langevin et al. [12] link data-driven and physical methods and tools to create a unified framework that can overcome the limitations of each technique by combining them. This combination is useful for addressing challenges that cannot be solved using just one of the methods mentioned before.

2.2 Calibration Method

With the development of a large number of HVAC system and component models, there is a growing focus on model calibration. Since the number of parameters is large in building HVAC model calibration, this process often uses automatic calibration, which relies on mathematical and statistical techniques. Techniques like Bayesian methods, simple optimization-based, deep neural networks (DNN), and particle swarm optimization (PSO) are utilized to enhance the accuracy and reliability of these models. The goal of the developed simple optimization-based calibration algorithm by Asadi [13] is to minimize the difference between the hourly MBE and CV (RMSE) of the measured data and the simulated results. On this basis, Gutiérrez [14] reduced the number of parameters to simplify the calibration process.

2.3 Data Processing Method Before Calibration

Commonly used input features for building HVAC models mainly include meteorological information, indoor environmental parameters, occupancy-related data, time indices, building characterization data, and historical data [15]. The predictive accuracy of the model depends heavily on having correctly selected representative training data [16]. Kristensen [17] investigated the effect of different temporal resolutions on the building energy model calibrated by the Bayesian method. Andrade-Cabrera [18] focused on the effect of different data sets on the building energy model calibrated by the PSO method. Sun [19] researched different types of data selection methods: uniform sampling, sequential sampling, preferential sampling, reversed preferential sampling, and sequential preferential sampling, without optimization for the shortcomings of these methods. Of the articles that mention data selection, techniques such as Principal Component Analysis (PCA), and Error Correction Models (ECM) play a pivotal

role in refining the data for better model performance. However, some studies on model calibration do not mention the selecting and pre-processing of representative calibration data.

2.4 Objective function and error functions

When performing calibration, various criteria need to be used to ensure that the model is properly calibrated. Therefore, a number of guidelines have published standards. For instance, ASHRAE Guideline 14 [20] states that a whole building can be considered calibrated when the simulated model has a Mean Bias Error (MBE) of 5% and Coefficient of Variation of the Root Mean Square Error (CV-RMSE) of 15% relative to monthly calibration data. These criteria must be met if the data samples period is hourly: 10% and 30%, respectfully. [13] used MBE and CV-RMSE as objective functions and error functions to meet the criteria of the ASHRAE standards. Some of the commonly used thresholds were also defined by other organizations[21].

Table 1 Commonly used thresholds

Calibration Criteria	ASHRAE Guideline 14	FEMP	IPMVP
Monthly CV-RMSE	15%	15%	-
Hourly CV-RMSE	30%	30%	20%
Monthly NMBe	+/- 5%	+/- 5%	-
Hourly NMBe	+/- 10%	+/- 10%	+/- 5%

Table 2 Literature for different modeling methods

Year	Model	Input	Output	Methodology	Accuracy	Ref
2013	Building model + Heating system model	Indoor climate data, weather conditions, energy delivery, and occupancy	Indoor air temperature	EnergyPlus v7.0	R ² : 0.38/0.68 CV-RMSE: 4.13%/6.97%	[22]
2013	Building envelope models + HVAC system models + district heating system models	Building parameters, energy delivery	Energy demand and consumption	Modelica	-	[23]
2014	HR VRF	Indoor climate data, weather conditions, energy delivery, and occupancy	Energy consumption	EnergyPlus	NMBE: -6.8% CV-RMSE: 17.2%	
2015	Building heating system	Temperatures, flow rates, and water temperatures to and from the heat pump	Performance of the heating system	Simulink	-	[24]
2015	Residential HVAC system model	Weather data, thermal properties of the building, and HVAC operational parameters	Energy consumption and indoor temperature	MATLAB/Simulink	-	[25]
2015	Building energy model	Climate data, weather conditions, energy delivery, occupancy	Air temperatures, energy consumption	Energyplus	CV-RMSE: 13.1%	[26]
2016	Building model + HVAC system model	Outdoor temperature and relative humidity, solar irradiation, indoor temperature and relative humidities	Predicted temperatures	MATLAB Modelica	-	[27]

2016	Building energy model	Outdoor air temperature, occupancy data	Energy consumption	Energyplus	CV-RMSE: 25%	[28]
2017	NANDRAD FMU Modelica HVAC	Weather data, building physics parameters, user behavior data, equipment schedules	Energy consumption and indoor temperatures	Modelica	-	[29]
2020	Building energy model	Weather data, building physics parameters, user behavior data, equipment schedules	Energy consumption and indoor temperatures	TRNSYS	Before: MBE: 15.12% CV-RMSE:17.19% After: MBE: -2.76% CV-RMSE:8.40%	[21]
2020	Building model + HVAC system model	Weather data, building physics parameters, user behavior data, equipment schedules	Heating demand, cooling demand and indoor temperatures	Energyplus	Heating NMBE: 10.4% CV-RMSE: 22.7% Cooling: NMBE: 22.2% CV-RMSE: 50.1%	[30]
2021	HVAC&R system model	Real operational data, building physics parameters, equipment schedules	Equipments supply temperature	ANN	CV-RMSE: $\pm 30\%$	[31]
2023	HVAC system model	Real operational data, building physics parameters, equipment schedules	Supply air temperature, indoor air temperature	DNN	Supply air temperature: NMBE: -1.2 % CV-RMSE: 3.8 % Indoor air temperature: NMBE: 1.1 % CV-RMSE: 1.3%	[32]

Table 3 Literature for different calibration methods

Year	Calibration Input	Calibration Output	Calibration Methodology	Data Selection Method	Set size	Data resolution	Calibration Set size	Validation Set size	Accuracy	Ref
2013	Parameters	Improved predictive performance	Optimization-based calibration	-	Four day periods	44-	-	Four day periods	44-	R ² : 0.81 CV- RMSE:2.15%
2014	Chiller operational parameters data	Power consumption	NN+PSO	-	10580 set	Every five minutes	70%	30%	Chiller1: R ² : 0.99988 Error: 88% Chiller2: R ² : 0.99989 Error: 83% Chiller3: R ² : 0.99990 Error: 80% Chiller4: R ² : 0.99989 Error: 82% Chiller5: R ² : 0.99989 Error: 85%	[33]
2015	Parameters, boiler energy consumption, boiler load	Energy consumption	Bayesian calibration	-	May 1st, 2013, to December 31st, 2014	hourly	10,000	May 1st, 2014, to December 31st, 2014	CV-RMSE: 12.9%/23.5%	[2]
2016	Thermal properties, infiltration characteristics, internal load	Energy consumption metrics and peak demands	Bayesian calibration	LHD	January 2011 to November 2013	15-minute	10,000	2100	NMBE CV-RMSE	[34]

2016	Thermal performance of a detached house	Improved parameter estimates	PSO	-	5 winter months	Hourly	3 winter months	2 winter months	RMSE: 0.296 MAE: 0.442 MBE: -0.0465 %	[35]
2017	Water-cooled chiller operational data	Energy consumption of the water-cooled chiller	Bayesian calibration	Random + KL divergence	1130	Hourly	791 (70%)	339 (30%)	CV-RMSE: 9.4% NMBE: -0.7%	[36]
2017	Cooling system operational data	Energy consumption of the cooling system	Bayesian calibration	Random + KL divergence	722	Hourly	503 (70%)	219 (30%)	CV-RMSE: 6.0% NMBE: -0.3%	[36]
2017	Resistances, capacitances of a lumped parameter	Performance metrics	PSO	ECM	8 different sizes	-	8 different sizes	-	MAE MBE CV-RMSE	[18]
2020	Construction parameters	Heating demand, cooling demand, indoor temperature	Optimization-based calibration	-	Weeks/ Months/ Years	Hourly	Week/ Month/ Year	Week/ Month/ Year	Discrepancy Indicator (DI) <10%	[37]
2023	Data on metered energy, weather, schedules	Performance metrics	NN	LHS	$N \times 8760$	Hourly	90%	10%	CV-RMSE NMBE	[38]
2023	Indoor air temperature, energy demand	Energy demand	MVBEP	MVBEP	-	15 mins/hourly /daily	9 months/6 months	3 months	NMBE: 1.44% CV-RMSE: 9.01% GOF: 25%	[39]

3 Aim

This project aims to explore the effect of calibration data length on the calibration performance of a Modelica-based MAU model using real operational data from an industrial plant.

4 Objectives

- Clean the data to use for calibration and validation to eliminate incorrect data and outlier entries that can lead to inaccurate analyses and conclusions.
- Select the representative subset with different data length from the cleaned data for both calibration and validation set and verify that the subsets cover the main region of the cleaned data in the reduced dimension space.
- Analyze the difference in distribution of different subsets compared to the full set.
- Calibrate and validate the MAU model by using 10-kold cross-validation and compare the performance of different subsets of models before and after calibration based on simple optimization, in terms of both stability of calibration parameter results and accuracy of prediction results.

5 Methodology

5.1 Overview

Fig.1 shows one of the most important parts of this research, data selection and representative verification. The Scikit-learn library offers many (BBOMs) and data-driven models. The selection process utilizes the K-Medoids algorithm to initially select representative subsets. The K-Medoids clustering method is robust and accurate, particularly suitable for energy consumption analysis. It can be achieved by using the Scikit-learn library which offers many data-driven models. After selection, the next step is to verify that these subsets accurately represent the main features of the entire dataset. This verification is achieved through the application of Principal Component Analysis (PCA) for dimensionality reduction, a technique that transforms the data into a reduced dimension space [40]. In addition, for greater rigor in selecting subsets, Kullback-Leibler (KL) divergence is used to analyze the distributions of different subsets against the full dataset.

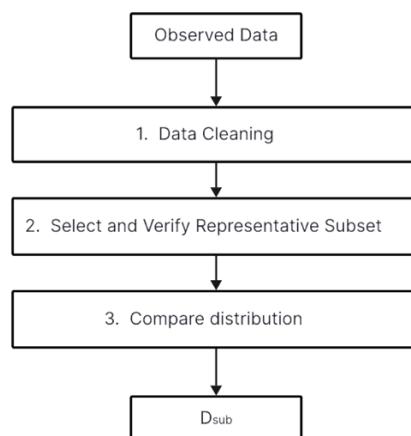


Figure 1 Overview of Data Selection

Fig. 2 shows the method for calibration and validation using subset data. After selecting different subsets of data, the MAU model is calibrated and validated using each subset individually. In this process, the K-fold cross-validation and sample-based optimization calibration methods are used. For each iteration of each subset, a portion of the data (determined by the K-fold cross-validation method) is used to calibrate the model, while the other portion of data is used to validate. Then, by using key performance metrics such as the Coefficient of Variation of the Root Mean Square Error (RMSE) and normalized Mean Bias Error (NMBE), the prediction results and parameters stability are compared.

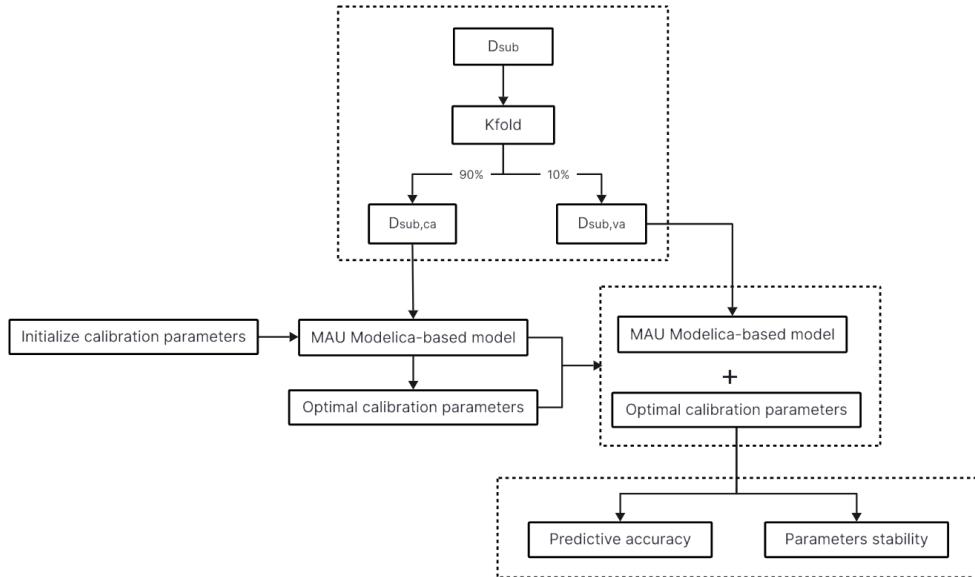


Figure 2 Overview of Calibration and Validation

5.2 Data Selection

5.2.1 Representative subset selection

The k-medoids algorithm is a clustering method for clustering the entire dataset D with n objects into a predetermined number k of clusters, where the number k is defined by the user. The cost of a subset data is the pairwise distance between points $\sum_{j \in C_i} d(i, j)$, where C_i is the cluster containing object i and $d(i, j)$ is the distance between objects i and j . Medoids for each cluster are calculated by finding object i within the cluster that minimizes $\sum_{j \in C_i} d(i, j)$. Compared with other clustering methods, the center of the clusters in k-medoids algorithm is the real data from the whole set. The medoids are those points in each cluster that have the shortest average distance (or dissimilarity) from other points. This means that they to some extent best characterize the cluster they are in. Based on python Scikit-learn package, the k-medoids algorithm works as follows:

- Initialization and Allocation: Randomly select ‘ k ’ initial medoids from the dataset and assign each object to the nearest medoid.
- Optimization: Iteratively swap medoid objects with non-medoid objects to minimize the total cost, which is defined as the sum of the differences in each cluster.
- Convergence: Repeat the optimization process until the medoids are stable and then finalize the clusters based on these medoids.

Cost decreases as the dataset expands, but the rate of decrease becomes slower. Therefore, the cost generated by the dataset is used to determine the size of each dataset. In order to get different sizes of data, the size of the data set is selected at the point where the cost is still in the decreasing stage, just becoming smooth, and already in a smooth state.

5.2.2 Subset verification

The subset validation process is a critical step in ensuring that the selected subset of calibration data represents the complete dataset and maintains the integrity of the model prediction. Apply Principal Component Analysis (PCA) to the full normalized dataset to transform it into a reduced dimension space. Prior to analysis, the dataset underwent a preprocessing stage where data normalization was carried out to ensure that each variable contributed equally to the analysis. This involved subtracting the mean and dividing by the standard deviation for each variable, achieving by Standard Scaler algorithm in python Scikit-learn library:

$$X_{norm} = \frac{X - \mu_x}{\sigma_x} \quad (1)$$

Where X is the original data, X_{norm} is the normalized data, μ_x is the mean of the data, and σ_x is the standard deviation.

PCA was conducted on the entire normalized dataset to reduce its dimensionality and identify the principal components (PCs) that explain the most significant amount of variance. PCs are essentially new axes in the data space that are linear combinations of the original variables. The PCA transformation is given by:

$$Y = X_{norm}P \quad (2)$$

Observe whether the distribution of the data covers the original dataset when the number of principle components is 2.

5.2.3 Divergence comparison

D_{sub} is the subset of the full set data D . In order to quantify the difference between different D_{sub} as well as the difference between D_{sub} and D , the ‘distance’ between them is measured, using the KL divergence. For each variable of the data, use entropy function of scipy to compute the KL divergence of D_{sub} with respect to D .

$$D_{KL}(P||Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right) \quad (3)$$

where P is the probability distribution of D_{sub} , Q is the probability distribution of D .

Then, average KL divergence gives a quantitative measure of the quality of D_{sub} as a representative subset of D . The value of average KL divergence gives an indication of the quality of D_{sub} . If this value is close to 0, it indicates that D_{sub} is statistically very close to D ; if this value is large, it indicates a significant difference between them.

$$\text{average KL divergence} = \frac{\sum_{i=1}^n D_{KL}(P_i||Q_i)}{n} \quad (4)$$

Where, n is the number of the variables of D_{sub} .

5.3 Simple optimization-based calibration and validation

The process of calibration is framed as one of error minimization in an optimization-based

methodology. The optimization algorithm comprises variables such as multiple model input parameters. The goal is to minimize the object function by varying the properties of these variables. The object function is to minimize the difference between the RMSE of the observed data and simulation results.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (m_i - s_i)^2}{n}} \times 100(\%) \quad (5)$$

Where m_i is the observed value, s_i is the simulated value, n is the number of observed values, and p is suggested to be equal to 1 [38].

5.4 K-fold Cross Validation

When making predictions, it is important to accurately assess the accuracy of the predictions. Training error is a simple way to estimate the prediction error, but it usually underestimates the prediction error. On the other hand, K-fold cross-validation tends to overestimate the prediction error. While the overestimation rate for leave-one-out cross-validation is usually small, the overestimation rate becomes larger for 5-fold or 10-fold cross-validation, which are commonly used for their efficiency. By combining the underestimation of training error with the overestimation of K-fold cross-validation, we can find the best estimate that balances these different estimates.

In this study, each dataset is divided into 10 subsets, and the models are looped by using 10-fold cross-validation. Each iteration consisted of calibrating the model on 9 subsets (90% of the dataset) and validating the model on the remaining subsets (10% of the dataset) to get the predicted air temperature. This process was repeated for each other of the 9 subsets, resulting in 10 different accuracy measurements. These measurements were averaged to produce an overall estimate of model accuracy. This approach ensures that each sample is used for training and testing, resulting in optimal use of the representative data and a reliable assessment of model performance.

5.5 Model Performance Evaluation

5.5.1 Calibration parameter results stability

The results of the calibration parameters are different for each fold of each calibration data. Parameter stability here refers to the consistency of the parameters when trained on different folds of data. The stability of the parameter results across folds demonstrates the sensitivity of the model to small variations in the training data, which is critical for generalizing the model to unfolded data. High variability in the parameters indicates that the model may be capturing noise rather than the underlying pattern, which can lead to overfitting. The high variability of the different folding parameters suggests that the model may be overly sensitive to specific noise patterns in the training data rather than capturing general underlying relationships. This sensitivity suggests that the details learned by the model cannot be generalized to new data, essentially memorizing the training data rather than learning from it.

5.5.2 Validation and performance comparison

Validation of prediction accuracy is based on the prediction of the calibration subset and validation subset with the parameters. The evaluation of the accuracy difference between the prediction of calibration and the prediction of validation is an effective identification of model overfitting. During the validation, Normalized Mean Bias Error (NMBE) is used to assess the accuracy of the predictions. According to the error functions, the corresponding values will be determined for each D_{sub} .

$$NMBE = \frac{1}{\bar{m}} \frac{\sum_{i=0}^n (m_i - s_i)}{n - p} \times 100(\%) \quad (6)$$

6 Results and Discussion

6.1 Building and Equipment Model Description

In the Tongwei Industry located in Chengdu, there are four production plants. MAU is one type of component in the terminal subsystem of the industrial HVAC. 12, 12, 14, and 14 MAUs were equipped in each production plant separately. On a per-production plant basis, the Modelica models were used to model the 4 MAUs. The dataset for this plant encompasses a comprehensive six-month period, spanning from June to November. Due to the extensive volume of data, which was originally recorded every minute, sampling the data at 30-minute intervals is necessary. This strategic sampling method ensures a more efficient analysis while retaining the integrity and representativeness of the entire dataset.

The Modelica-based MAU model has 10 inputs, 5 outputs, and 9 parameters.

Table 4 Modelica-based MAU model

Input	
Inlet air dry bulb temperature	TairIn
Inlet air relative humidity	HairIn
Air mass flow rate	Fair
Chilled water inlet temperature	TcwIn
Chilled water mass flow rate	Few
Hot water inlet temperature	ThwIn
Hot water mass flow rate	Fhw
Humidifier ratio	RhumwVal
Preheating water inlet temperature	TprehwIn
Preheating water mass flow rate	Fprehw
Output	
Supply air relative humidity	HairSup
Supply air temperature	TairSup
Chilled water outlet temperature	TcwRet
Hot water outlet temperature	ThwRet
Preheating water outlet temperature	TprehwRet

To calibrate the MAU model, 9 parameters have been defined. The parameters of the calibration model are:

Table 5 Parameters of calibration model

Calibration Parameter

Nominal mass flow rate of air	FairNominal
Nominal mass flow rate of preheating hot water	FprehwNominal
Nominal mass flow rate of chilled water	FcwNominal
Maximum mass flow rate of chilled water	FcwMax
Nominal mass flow rate of heating hot water	FhwNominal
Nominal UA of preheating coil	UAprehtgNominal
Nominal UA of cooling coil	UAcIgNominal
Nominal UA of heating coil	UAhtgNominal
Nominal added water mass flow rate of humidifier	FhumwNominal

Before calibration, the parameters are specified as approximate ranges, based on normal situation of the industry. Take MAU 1 as an example:

Table 6 Calibration Parameter initial range

	Range	Initialization
FairNominal	[8, 48]	0.5
FprehwNominal	2	
FcwNominal	[2, 15]	0.5
FcwMax	33.32	
FhwNominal	[11.11, 66.64]	0.5
UAprehtgNominal	2000	
UAcIgNominal	[33333, 200000]	0.5
UAhtgNominal	[3333, 30000]	0.5
FhumwNominal	7.5	

6.2 Data normalization

Before the k-medoids clustering algorithm, normalizing the data before selecting representative points is essential to ensure that all features of calibration inputs contribute equally to the calculation of distances between data points.

Fig.3 below shows the distributions of different calibration inputs:

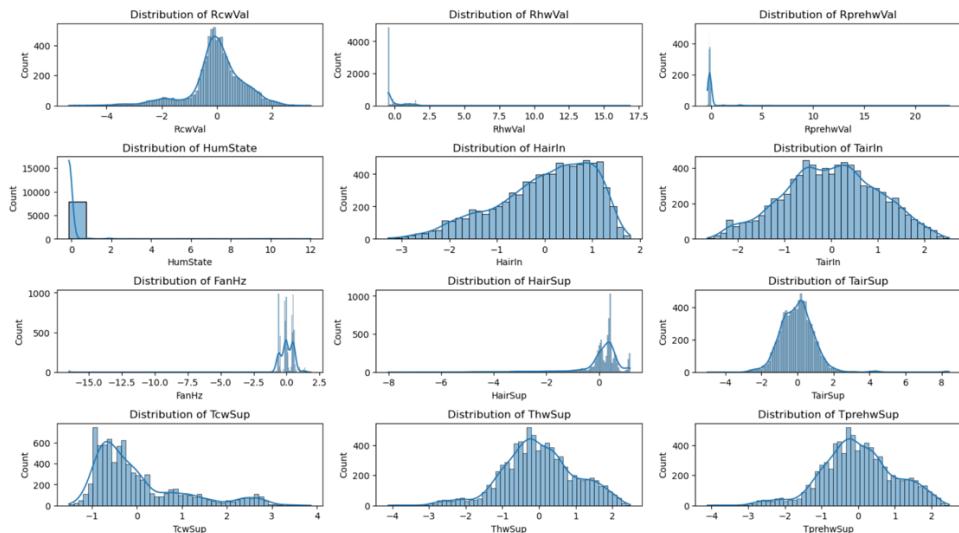


Figure 3 Distributions of different calibration inputs

Fig. 3 shows that some distributions (RhwVal, RprehwVal, HumState, FanHZ, HairSup) are skewed distribution. RhwVal and RprehwVal show positive skewness (right skew), whereas HumState, FanHZ and HairSup show negative skewness (left skew). It means there are outliers in the data, especially those that are far from the majority of the dataset. It proves that it is

necessary to do data cleaning.

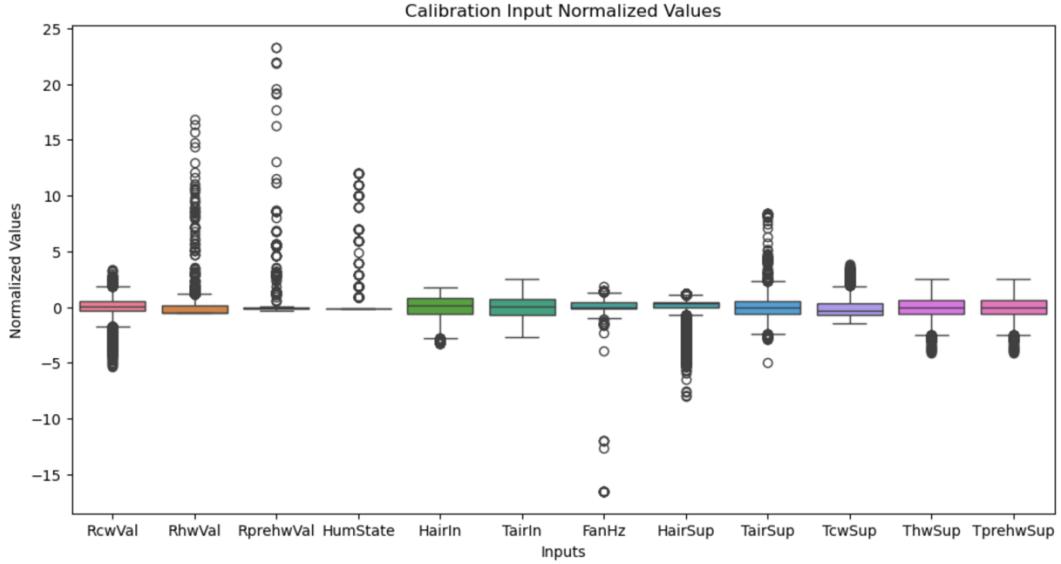


Figure 4 Normalized Values of Calibration Inputs

Fig. 4 shows the distribution of calibration input values before cleaning, giving an idea of the range, average values, and any unusual data points. Each box on the graph includes the middle value (median), the range of the middle 50% of data (IQR), and any outliers (data points far from the rest). The lines extending from the boxes show how far most data points are from the average, with points beyond a certain distance considered outliers. Some inputs have a wide range of data points with many outliers, suggesting big differences or unusual values. Other inputs have data points closer together, indicating less variation. Finding outliers in most features highlights the importance of cleaning and adjusting the data to avoid affecting analysis results. Some features have fewer outliers, showing more consistent measurements or less variation.

6.3 Data Cleaning

Using data from the beginning of June through the end of November (with data from August 21st through September 1st missing), there are 8,160 data sets for each MAU. However, there are some abnormal data for each MAU in these 8160 sets of data. In the preprocessing phase of the analysis, an essential step involved the removal of outliers, which can significantly skew the results and lead to erroneous conclusions. To systematically identify and eliminate these outliers, we employed the interquartile spacing (IQR) method to capitalize on its robustness in dealing with extreme variations in the distribution of the data. IQR is defined as the difference between 75% (Q3) and 25% (Q1) of the data. Outliers are then determined based on the deviation of the outliers from this range. The data range is:

$$Q1 - 1.5 \times IQR < \text{data range} < Q3 + 1.5 \times IQR \quad (7)$$

Specifically, any data point out of the data range was considered as an outlier and is limited from further analysis. Instead of getting rid of these unusual data points, which could result in losing important information, especially when the dataset is small, they are restricted for additional examination by clipping. Any data that fall below the lower limits are set to the value of lower limits, and any values that exceed the upper limits are set to the value of upper limits. By using this clipping method, we can make sure that all data used in further analyses is kept

within a specific range, which helps to make the data more resistant to the impact of extreme outliers.

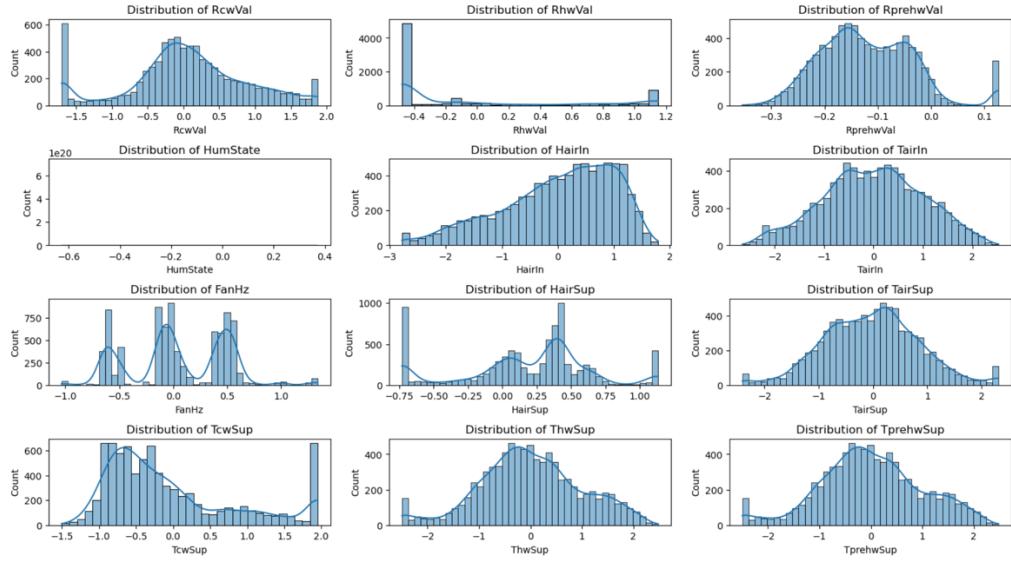


Figure 5 Distributions of model inputs

Fig. 5 shows the distributions of the cleaned data. The distribution of ‘HumState’ is empty because the value of this input is always the same after data cleaning. This is due to the fact that there is a great deal of similarity in the data even before the data is cleaned, and the data that is not the same is outside the upper and lower limits of the data.

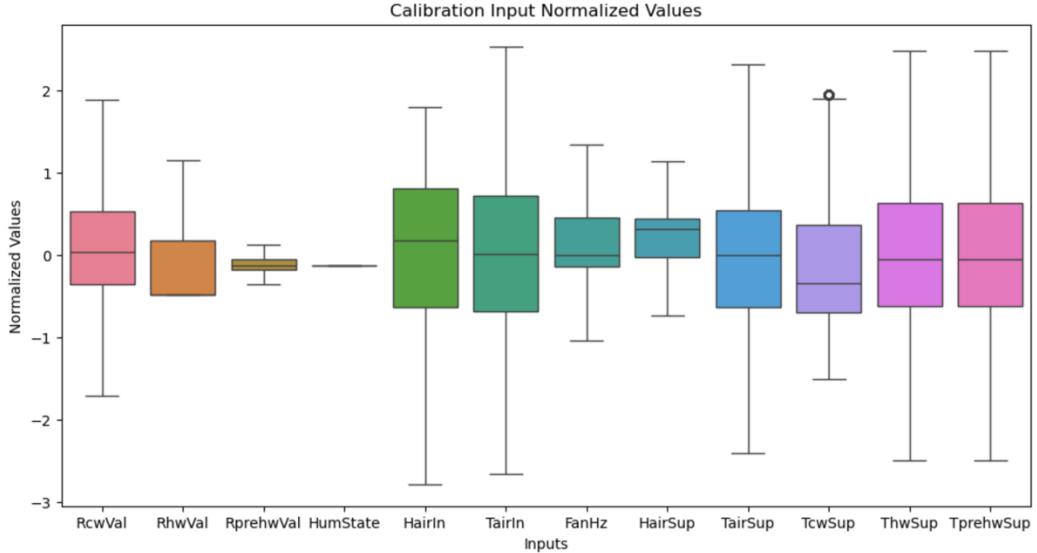


Figure 6 Calibration input normalized values

Fig. 6 shows that after data cleaning, the outliers in the calibration model inputs are significantly reduced. The figure also shows the region of the data set where most of the ranges in each data set are located. ‘RcwVal’, ‘RprehwVal’, and ‘HumState’ show signs of positive skewness (the median is closer to the bottom of the box), indicating that most of the data points are on the low end of the scale, but there are some higher values.

6.4 Representative Subset Selection & Verification

In the initial processing, the original six months of minute-by-minute data were organized into

half-hourly intervals. Take MAU1 as an example. According to clustering analysis, pairwise distance between points of the whole dataset has been shown. If the number of clusters is 5000, the pairwise distance can be reduced to 20000, while the cost-effectiveness (calibration time) will be also increased. According to Fig.3, the downward trend has become significantly slower after the number of clusters are higher than 1000. Although decrease of the total cost after data length is 2000 is slow, it becomes stable after data length is 5000. Therefore, 1000, 2000 clusters, 3000, 4000 and 5000 were chosen as the number of each representative data subsets.

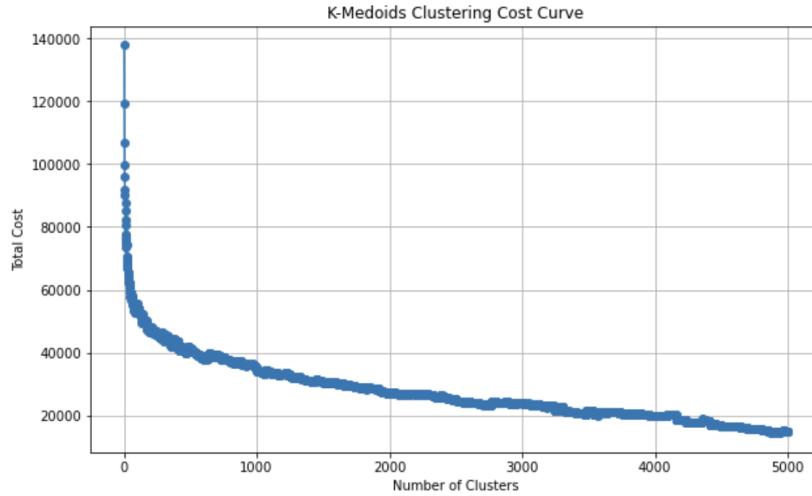
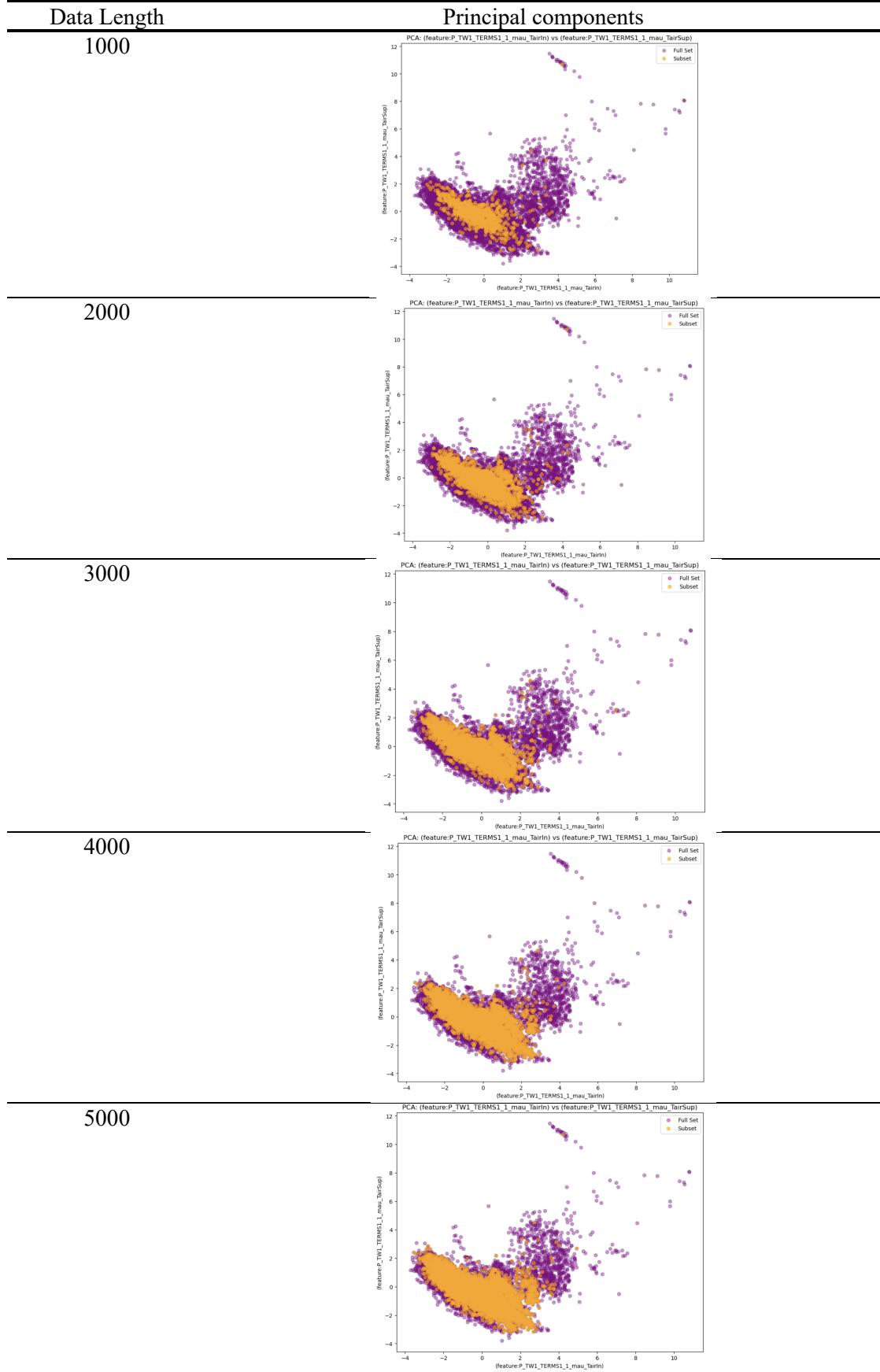


Figure 7 K-Medoids clustering cost

Five different length data subsets were sampled through K-Medoids. Then, in order to verify that the subsets cover the main region of the whole observed data, PCA method was used to take each dataset from 12 variables and descend them to have 2 principal components which are two dimensions that explain the most variance. In this case, the principal components are air supply temperature and indoor air temperature. From *Tab.5*, each length subset of data is overlaid on the full set and is mostly concentrated in the area where the full set data is denser. As the data length increases, the area of the subset covered in the dense region increases, while other regions are sampled more. According to *Tab.5*, the data selection was considered a success. If descend 12 variables to have 4 principal components, indoor air temperature, supply air temperature, indoor air humidity, and humidifier ratio are the principal components. The figures which each two combinations of four principal components were also shown in the Appendix.

Table 7 Principal components



6.5 Different Subsets Distribution

According to *Tab.5*, each representative data subset has been appropriately selected. Then, the average KL divergences for each D_{sub} were calculated. The KL divergence decreases as the subset length increases, as shown in *Fig.8*. A significant decrease in the average KL dispersion as the data subset grows from 1000 to 5000 shows that a bigger data subset better represents the overall data. This suggests that as the sample size gets bigger, the subset distribution gets closer to the overall distribution, making the two less statistically different.

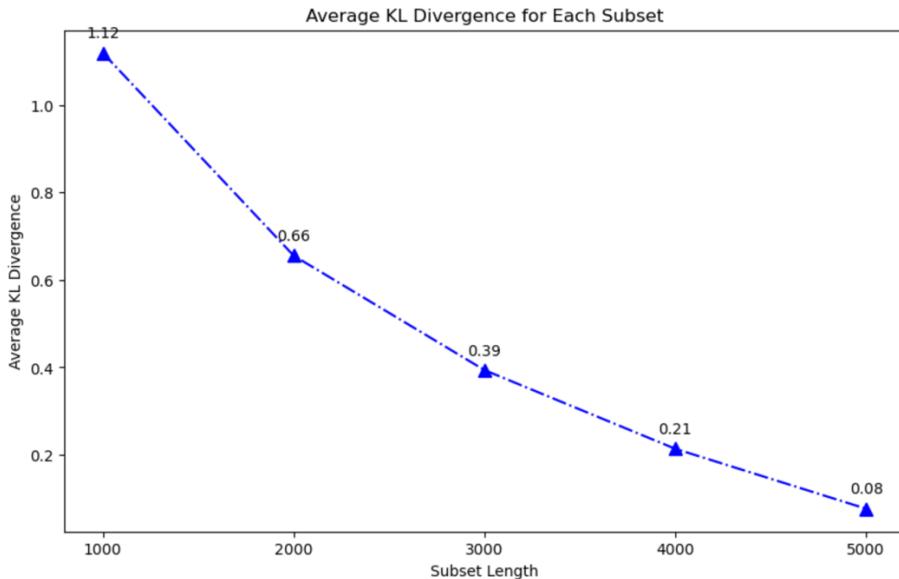


Figure 8 Average KL divergence

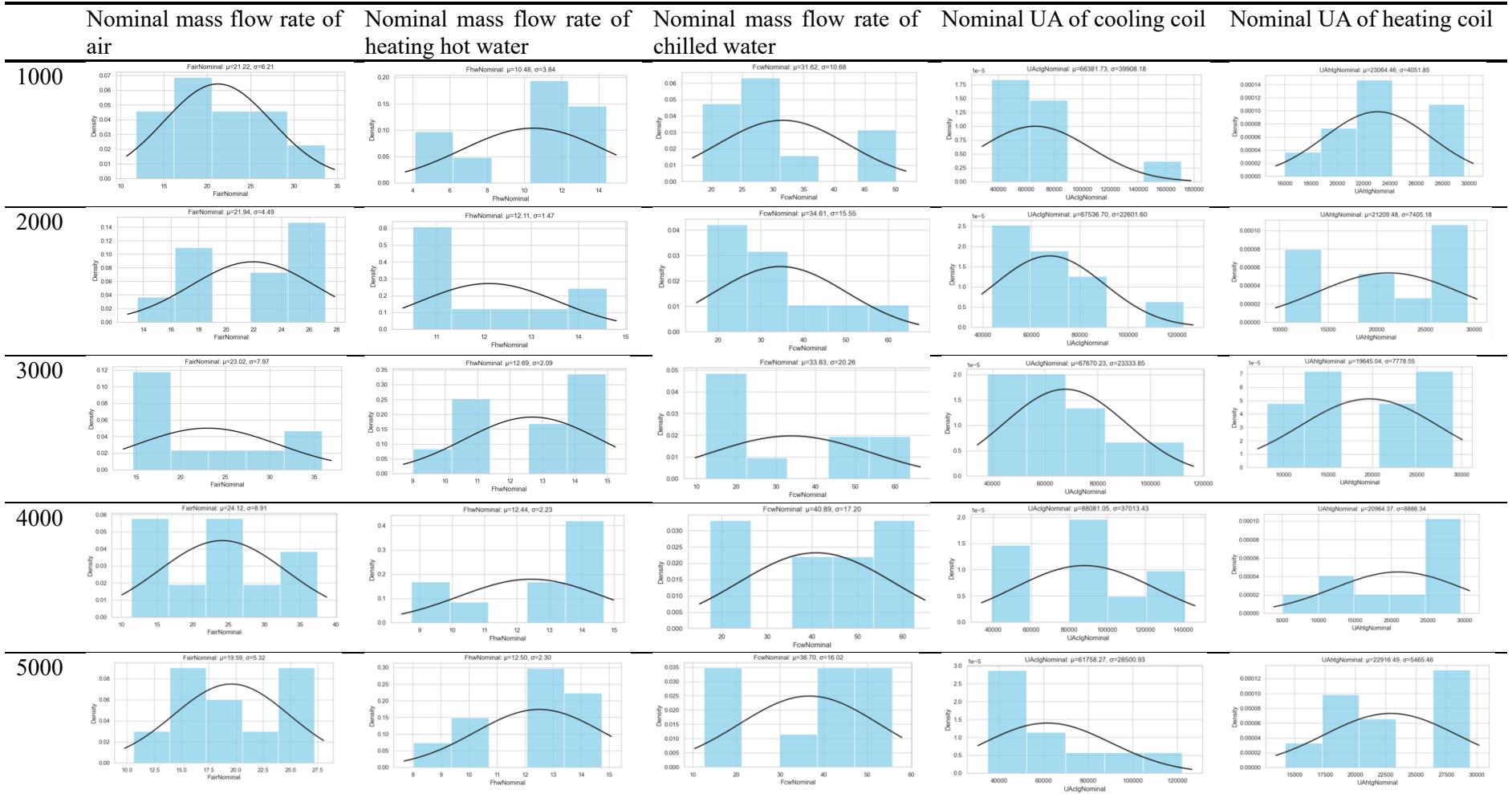
6.6 10-Fold Cross-Validation Results and Model Performance

Model performance can be evaluated by both parameter calibration stability and prediction accuracy. 10-fold cross-validation has been used for each length of data. Each calibration and validation yield 10 sets of parameter results and 10 sets of predictions based on the calibration and validation sets. The following is a detailed analysis of MAU1.

6.6.1 Parameter calibration stability

The distributions of parameters that be calculated for each data subset are shown:

Table 8 Distributions of parameter results



This MAU model has five parameters (Nominal mass flow rate of air, Nominal mass flow rate of heating hot water, Nominal mass flow rate of chilled water, Nominal UA of cooling coil, Nominal UA of heating coil) that need to be calibrated. For most parameters, calibration results are most stable at a data length of 2000. However, there are still 2 parameters that do not give the most stable calibration results at a data length of 2000. It might be caused by the different sensitivities of different parameters to the data. Some parameters might be more robust to variations in data, requiring less data to stabilize, while others might need a larger dataset to accurately capture their behavior under various operating conditions. Therefore, to maximize the stability of parameter calibration results, select the data length that allows the most stable data for the most parameters. When the data length is 1000, there may not be enough different situations included, causing big differences in calibration results when testing the model in various cross-validation sets. This could be because the model has not been exposed to a wide range of scenarios to effectively learn and apply the parameter values. When the data length is 2000, there are enough conditions to calibrate the model, so the difference between the results obtained with different calibration sets is minimized. This is the best data length for most parameters. It is probably because it balances a balance between the diversity of conditions and the manageability of data size. This length provides a comprehensive dataset that is still manageable enough to minimize noise and overfitting. When the data length is larger than 2000 (3000, 4000, 5000), the parameter calibration stability becomes less satisfied. It might be caused by the noise from a large dataset. The less representative of these data will also affect the calibration results, resulting in deviating from parameter values appropriate for representative data. Although it may seem like having more data would improve learning, including data that is not as representative or is noisy can actually lower the quality of calibration. This is especially true for parameters that are sensitive to changes in the data. On the other hand, the averages of the cross-validation results for different data lengths do not vary as large as the results for different calibration sets for the same data length. This shows that the model can handle different amounts of data well. It is a good sign that the main model setup, like the selected features and basic parameters, works well in various situations and data sizes.

In this study, the average of the parameter results obtained from the per-fold calibration of the data length with the highest stability is regarded as the more appropriate parameter.

Table 9 Suitable parameter values

Suitable parameter values	
Nominal mass flow rate of air	21.94
Nominal mass flow rate of preheating hot water	2
Nominal mass flow rate of chilled water	34.61
Maximum mass flow rate of chilled water	33.32
Nominal mass flow rate of heating hot water	12.11
Nominal UA of preheating coil	2000
Nominal UA of cooling coil	67536.70
Nominal UA of heating coil	21209.48
Nominal added water mass flow rate of humidifier	7.5

6.6.2 Prediction accuracy

RMSE and NMBE are used to assess the prediction accuracy of calibration. RMSE is an error

metric to evaluate the discrepancy between predicted and calibrated results. NMBE is an error metric determining whether a model is overvalued or undervalued.

Comparison of model performance on calibration and validation sets

Table 10 Results of MAU1 with 1000 data

Data Length = 1000	RMSE		NMBE	
	Calibration	Validation	Calibration	Validation
1	0.66	0.64	-2.94%	-2.91%
2	0.58	0.53	-0.81%	-0.98%
3	0.64	0.68	1.24%	1.25%
4	0.57	0.60	-0.44%	-0.16%
5	0.59	0.60	-1.03%	-0.44%
6	0.63	0.61	1.61%	1.07%
7	0.59	0.59	-0.46%	-1.10%
8	0.58	0.57	-0.50%	-1.19%
9	0.62	0.61	-2.31%	-2.24%
10	0.66	0.75	0.19%	1.07%
Average	0.612	0.618	-	-

The results of MAU1 with 1000 data show that the RMSE of the calibration subset and the validation subset are similar, the average RMSE of the validation subset is a little higher than the average RMSE of the calibration subset. It means that the overfitting does not happen, and the calibration results can be used for new conditions well. The average NMBE is negative. It means that the prediction results are always undervalued.

Table 11 Results of MAU1 with 2000 data

Data Length = 2000	RMSE		NMBE	
	Calibration	Validation	Calibration	Validation
1	0.66	0.58	-0.96%	-0.71%
2	0.61	0.61	0.45%	0.29%
3	0.60	0.57	-0.48%	-0.35%
4	0.61	0.61	-1.24%	-1.23%
5	0.64	0.65	0.31%	0.51%
6	0.59	0.60	-0.59%	-1.01%
7	0.60	0.66	-1.06%	-1.04%
8	0.58	0.59	0.03%	0.04%
9	0.59	0.59	0.19%	-0.45%
10	0.63	0.66	0.44%	1.27%
Average	0.611	0.612	-	-

The results of MAU1 with 2000 data are similar to the results of MAU1 with 1000 data. The average RMSE of results with 2000 data is a little lower than the average RMSE of results with 1000 data. The subset covers more conditions in the calibration. However, the prediction results are still undervalued.

Table 12 Results of MAU1 with 3000 data

Data Length = 3000	RMSE		NMBE	
	Calibration	Validation	Calibration	Validation
1	0.64	0.60	-0.34%	-0.37%
2	0.57	0.62	-0.74%	-0.13%
3	0.61	0.59	-0.83%	-0.87%

4	0.68	0.64	-0.28%	0.04%
5	0.70	0.72	0.98%	1.01%
6	0.69	0.70	1.21%	1.35%
7	0.73	0.74	1.61%	1.43%
8	0.64	0.66	-0.07%	-0.56%
9	0.62	0.61	-0.84%	-0.98%
10	0.60	0.58	-0.37%	-0.70%
Average	0.648	0.646	+	+

The average RMSE of both calibration and validation are increased. The prediction results by 3000 data are overvalued.

Table 13 Results of MAU1 with 4000 data

Data Length = 4000	RMSE		NMBE	
	Calibration	Validation	Calibration	Validation
1	0.69	0.71	-1.63%	-1.80%
2	0.71	0.66	-2.08%	-1.88%
3	0.61	0.64	-0.55%	-0.45%
4	0.63	0.63	-0.10%	-0.16%
5	0.67	0.67	-1.72%	-1.47%
6	0.68	0.68	-0.50%	-0.60%
7	0.69	0.69	1.50%	1.45%
8	0.63	0.62	-0.97%	-0.85%
9	0.7	0.71	0.01%	-0.31%
10	0.67	0.67	-0.52%	-0.39%
Average	0.668	0.668	-	-

The average RMSE still increase as the increase of calibration and validation data. When data length is 4000, the prediction results are undervalued in general.

Table 14 Results of MAU1 with 5000 data

Data Length = 5000	RMSE		NMBE	
	Calibration	Validation	Calibration	Validation
1	0.65	0.63	-0.98%	-0.68%
2	0.66	0.68	-0.90%	-0.89%
3	0.73	0.73	0.29%	0.31%
4	0.64	0.64	-0.68%	-0.53%
5	0.71	0.74	0.69%	0.65%
6	0.65	0.65	-1.68%	-1.51%
7	0.73	0.73	-2.71%	-2.69%
8	0.70	0.71	-1.80%	-2.36%
9	0.64	0.66	-0.30%	-0.21%
10	0.64	0.61	-0.43%	-0.38%
Average	0.675	0.678	-	-

When data length is 5000, the average RMSE is the highest, and the prediction results are also undervalued in general.

For this model, the calibration performance is the best when the data length is 2000. At shorter data lengths, the calibration RMSE remains relatively stable and low. However, as the data length increases, particularly from 2000 to 5000 data points, there is a noticeable and steady rise in the RMSE. This suggests that initially increasing the amount of data helps to refine the model by providing more training examples that may cover a wider range of conditions, thus enhancing the model prediction ability. At shorter data lengths, the calibration RMSE remains relatively stable and low. However, as the data length increases, particularly from 2000 to 5000

data points, there is a noticeable and steady rise in the RMSE. This suggests that initially increasing the amount of data helps to refine the model as it provides more training examples that may cover a wider range of scenarios, thus improving the model performance. When data length increases from 1000 to 2000, *Fig. 9* shows that after a certain point, just adding more data without changing the calibration model can make it work less effectively.

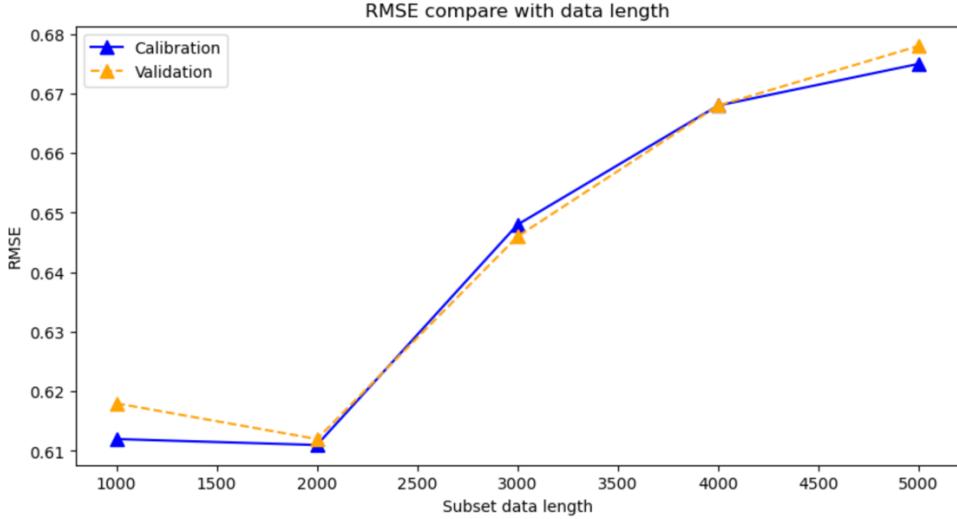


Figure 9 RMSE Compare with different data lengths

7 Conclusion

This study explores the effect of different data lengths on the calibration performance of a Modelica-based Make-Up Air Unit (MAU) model using real operational data from an industrial plant. The key objective was to determine the optimal data length that provides the most reliable calibration results without overfitting or introducing excessive noise into the model. The representative data selection and 10-fold cross-validation for calibration and validation of the MAU model are the significant pre-process of comparing and analyzing the model performance from two aspects: parameter results stability and MAU model prediction results accuracy. The study employed various data lengths and examined their effects using techniques like K-medoid clustering, Principal Component Analysis (PCA), and Kullback-Leibler (KL) divergence to assess the representativeness of data subsets. Before the application of K-medoid clustering, data standardization is used to make each input equally sensitive to prevent inconsistent effects of different data inputs on model calibration. Based on the interquartile range (IQR) method, the outliers were changed to upper and lower limits. Therefore, the effect of extreme values on the calibration process is mitigated. 5 subsets of cleaned data of different lengths were selected for each MAU model as inputs for calibration and validation. Each subset spread from data-dense to data-sparse areas.

The MAU models are calibrated and validated using 10-fold cross-validation to confirm the model performance on different data sets. The method compares the stability of calibrated parameters with the accuracy of model predictions. Statistical parameters such as RMSE and NMSE are used to assess the accuracy and reliability of the model results. The results show that simply adding calibration data does not necessarily improve model accuracy. To keep a

balance between the representativeness of the data set and the efficiency of the model, a certain length of data is necessary. If the length of data is too low, the model cannot be calibrated under different conditions well, and the parameter results may be not suitable for some other conditions. It means the models cannot predict new conditions to be very close to real observed data. On the other hand, too much data will make the model too complex, and too much noise interferes with the models. This will reduce the accuracy of predictions with new data as well. The study results show that the data length of 2000 will lead to a good balance between providing enough and more general conditions to calibrate for better-predicted parameters and avoiding overfitting problems which will influence the prediction of models. As measures such as the root mean square error (RMSE) coefficient and the normalized mean bias error (NMBE) of both calibration and validation predicted results show that the data length provides more stable calibration parameters and the most reliable predictions among the 5 different data lengths in this study.

This study provides a basis for a better understanding and application of small-scale calibrations (MAU calibrations) for energy models. The knowledge gained from studying the optimal data length for calibration is useful not only for creating energy models, but also for other areas such as system simulation and predictive modeling. By continuing to explore and utilize this knowledge, it is possible to significantly improve the use of energy models.

8 Recommendations For Further Study

1. Calibration and validation for larger model

Further research can determine how to properly calibrate and validate larger models of more complex HVAC systems. The methodology developed in this study focuses on Modelica-based MAU models. In fact, MAU components should connect with other components and combine into more complex systems. They have different inputs, outputs, and parameters. Extending this research to other component models and larger models may highlight new challenges and opportunities for model calibration, such as the ability to handle higher computational demands and greater data variability.

2. Integration of Real-Time Data and Continuous Calibration

Future research could explore methods that combine real-time data with continuous calibration. The calibration in this study was only able to calculate fixed parameter values. Periodically updating the model with new data could help create adaptive models that are better adapted to changes in environmental or operational factors. This approach could produce more reliable adaptive models that remain accurate even when system behavior or external conditions change.

3. Integration with IoT devices:

Integrating IoT devices in building management systems offers a promising avenue for automating the collection and processing of calibration data. Investigating how IoT can be used to streamline the calibration process could make energy modeling practices more dynamic and adaptable.

9 Reference List

- 1 Heo Y, Graziano DJ, Guzowski L, et al. Evaluation of calibration efficacy under different levels of uncertainty. *Journal of Building Performance Simulation* 2015;8:135-44.
- 2 Chong A, Lam KP. Uncertainty Analysis And Parameter Estimation of Hvac Systems in Building Energy Models. *Building Simulation Conference Proceedings* 2015.
- 3 Kircher K, Shi X, Patil S, et al. Cleanroom energy efficiency strategies: Modeling and simulation. *Energy and Buildings* 2010;42:282-9.
- 4 Zhao Y, Li N, Tao C, et al. A comparative study on energy performance assessment for HVAC systems in high-tech fabs. *Journal of Building Engineering* 2021;39:102188.
- 5 Wang K-J, Lin CJ, Dagne TB, et al. Bilayer stochastic optimization model for smart energy conservation systems. *Energy* 2022;247:123502.
- 6 Zhao W, Li H, Wang S. A generic design optimization framework for semiconductor cleanroom air-conditioning systems integrating heat recovery and free cooling for enhanced energy performance. *Energy* 2024;286:129600.
- 7 Shan K, Wang S. Energy efficient design and control of cleanroom environment control systems in subtropical regions – A comparative analysis and on-site validation. *Applied Energy* 2017;204:582-95.
- 8 Zhuang C, Wang S, Shan K. Probabilistic optimal design of cleanroom air-conditioning systems facilitating optimal ventilation control under uncertainties. *Applied Energy* 2019;253:113576.
- 9 Idowu S, Saguna S, Åhlund C, et al. Applied machine learning: Forecasting heat load in district heating system. *Energy and Buildings* 2016;133:478-88.
- 10 Ahmad MW, Mourshed M, Rezgui Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings* 2017;147:77-89.
- 11 Bikmukhametov T, Jäschke J. Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. *Computers & Chemical Engineering* 2020;138:106834.
- 12 Langevin J, Reyna JL, Ebrahimigharehbaghi S, et al. Developing a common approach for classifying building stock energy models. *Renewable and Sustainable Energy Reviews* 2020;133:110276.
- 13 Asadi S, Mostavi E, Boussaa D, et al. Building energy model calibration using automated optimization-based algorithm. *Energy and Buildings* 2019;198:106-14.
- 14 Gutiérrez González V, Ramos Ruiz G, Fernández Bandera C. Empirical and comparative validation for a building energy model calibration methodology. *Sensors* 2020;20:5003.
- 15 Sun Y, Haghhighat F, Fung BC. A review of the-state-of-the-art in data-driven approaches for building energy prediction. *Energy and Buildings* 2020;221:110022.
- 16 Guyon I, Elisseeff A. An introduction to variable and feature selection. *Journal of machine learning research* 2003;3:1157-82.
- 17 Kristensen MH, Choudhary R, Petersen S. Bayesian calibration of building energy models: Comparison of predictive accuracy using metered utility data of different temporal resolution. *Energy Procedia* 2017;122:277-82.
- 18 Andrade-Cabrera C, Burke D, Turner WJN, et al. Ensemble Calibration of lumped

- parameter retrofit building models using Particle Swarm Optimization. *Energy and Buildings* 2017;155:513-32.
- 19 Sun Y, Haghhighat F, Fung BCM. Trade-off between accuracy and fairness of data-driven building and indoor environment models: A comparative study of pre-processing methods. *Energy* 2022;239:122273.
- 20 American Society of Heating R, Engineers A-C. *ASHRAE guideline: Measurement of energy and demand savings*. Ashrae 2002.
- 21 Tüysüz F, Sözer H. Calibrating the building energy model with the short term monitored data: A case study of a large-scale residential building. *Energy and Buildings* 2020;224:110207.
- 22 Farhang T, Robert Z, Matthias S, et al. Simulation model calibration: an optimization-based approach. *Proceedings of BauSim Conference 2012: 4th Conference of IBPSA-Germany and Austria*: IBPSA-Germany and Austria 2012:386-91.
- 23 Burhenne S, Wystrcil D, Elci M, et al. Building performance simulation using Modelica: analysis of the current state and application areas. 2013.
- 24 Lim S, Park S, Chung H, et al. Dynamic modeling of building heat network system using Simulink. *Applied Thermal Engineering* 2015;84:375-89.
- 25 Kassas M. Modeling and simulation of residential HVAC systems energy consumption. *Procedia computer science* 2015;52:754-63.
- 26 Royapoor M, Roskilly T. Building model calibration using energy and environmental data. *Energy and Buildings* 2015;94:109-20.
- 27 Perera DWU, Winkler D, Skeie NO. Multi-floor building heating models in MATLAB and Modelica environments. *Applied Energy* 2016;171:46-57.
- 28 Liang X, Hong T, Shen GQ. Improving the accuracy of energy baseline models for commercial buildings with occupancy data. *Applied Energy* 2016;179:247-60.
- 29 Nicolai A, Paepcke A. Co-Simulation between detailed building energy performance simulation and Modelica HVAC component models. *Modelica* 2017;132:006.
- 30 Guyot D, Giraud F, Simon F, et al. Building energy model calibration: A detailed case study using sub-hourly measured data. *Energy and Buildings* 2020;223:110189.
- 31 Kim H, Sumiyoshi D, Choi Y. Development of flow rate and equipment simulation model for commercial building HVAC&R system by data-driven method. *Journal of Building Engineering* 2021;44:103303.
- 32 Ra SJ, Kim J-H, Park CS. Real-time model predictive cooling control for an HVAC system in a factory building. *Energy and Buildings* 2023;285:112860.
- 33 Chen C-L, Chang Y-C, Chan T-S. Applying smart models for energy saving in optimal chiller loading. *Energy and Buildings* 2014;68:364-71.
- 34 Li Q, Augenbroe G, Brown J. Assessment of linear emulators in lightweight Bayesian calibration of dynamic building energy models for parameter estimation and performance prediction. *Energy and Buildings* 2016;124:194-202.
- 35 Andrade-Cabrera C, Turner WJ, Burke DJ, et al. Lumped parameter building model calibration using particle swarm optimization. *3rd Asia conference of International Building Performance Simulation Association (ASim2016), Jeju island, Korea, 27-29 November 2016* 2016.
- 36 Chong A, Lam KP, Pozzi M, et al. Bayesian calibration of building energy models with large datasets. *Energy and Buildings* 2017;154:343-55.

- 37 Martínez S, Eguía P, Granada E, et al. A performance comparison of multi-objective optimization-based approaches for calibrating white-box building energy models. *Energy and Buildings* 2020;216:109942.
- 38 Herbinger F, Vanden Hof C, Kummert M. Building energy model calibration using a surrogate neural network. *Energy and Buildings* 2023;289:113057.
- 39 Alrobaie AS, Krarti M. Measurement and Verification Building Energy Prediction (MVBEP): An interpretable data-driven model development and analysis framework. *Energy and Buildings* 2023;295:113321.
- 40 Kumar Mohapatra S, Mishra S, Tripathy HK, et al. A sustainable data-driven energy consumption assessment model for building infrastructures in resource constraint environment. *Sustainable Energy Technologies and Assessments* 2022;53:102697.

10 Appendix

Table 15 4 Principal components

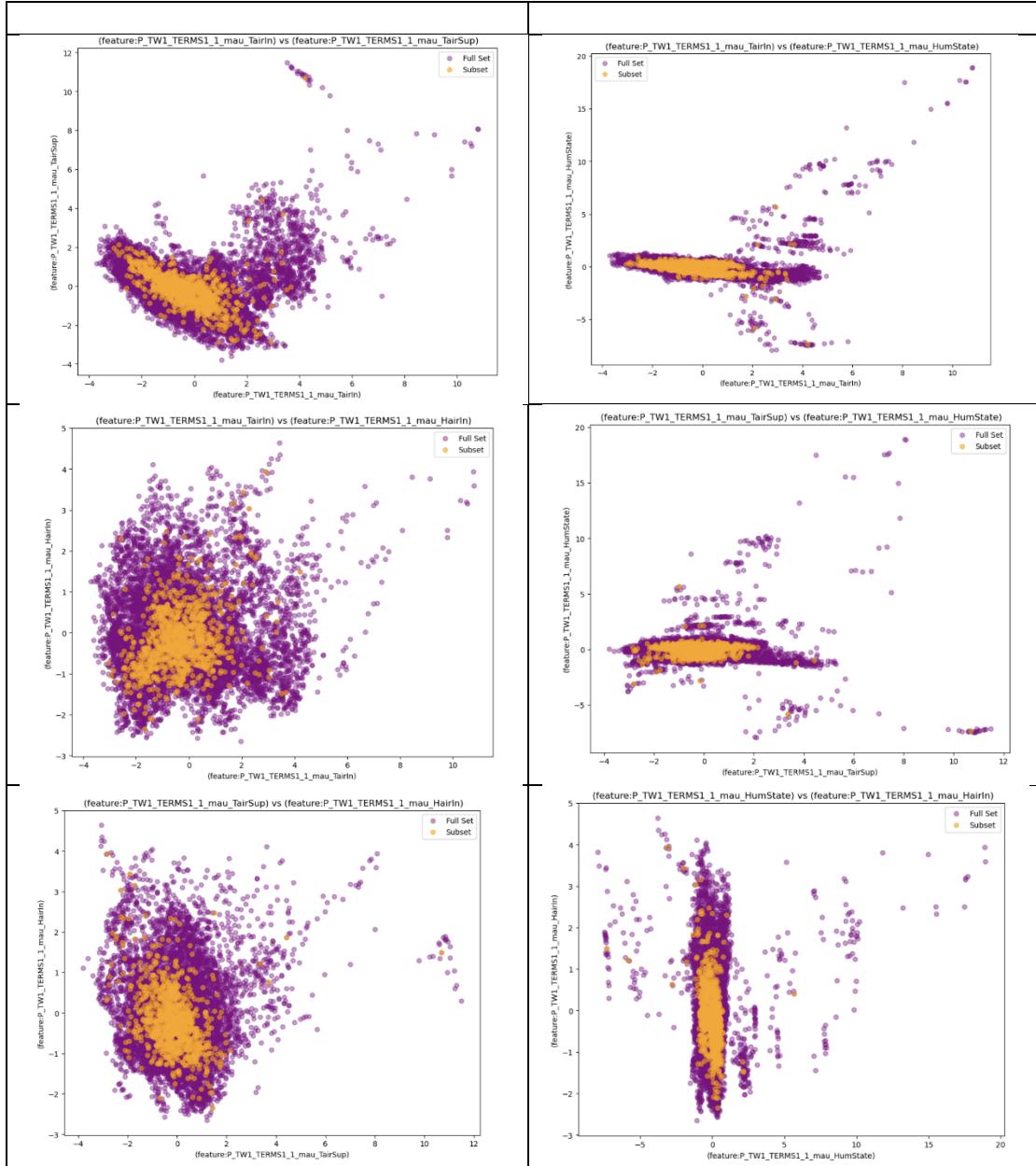


Table 16 4 Principal components

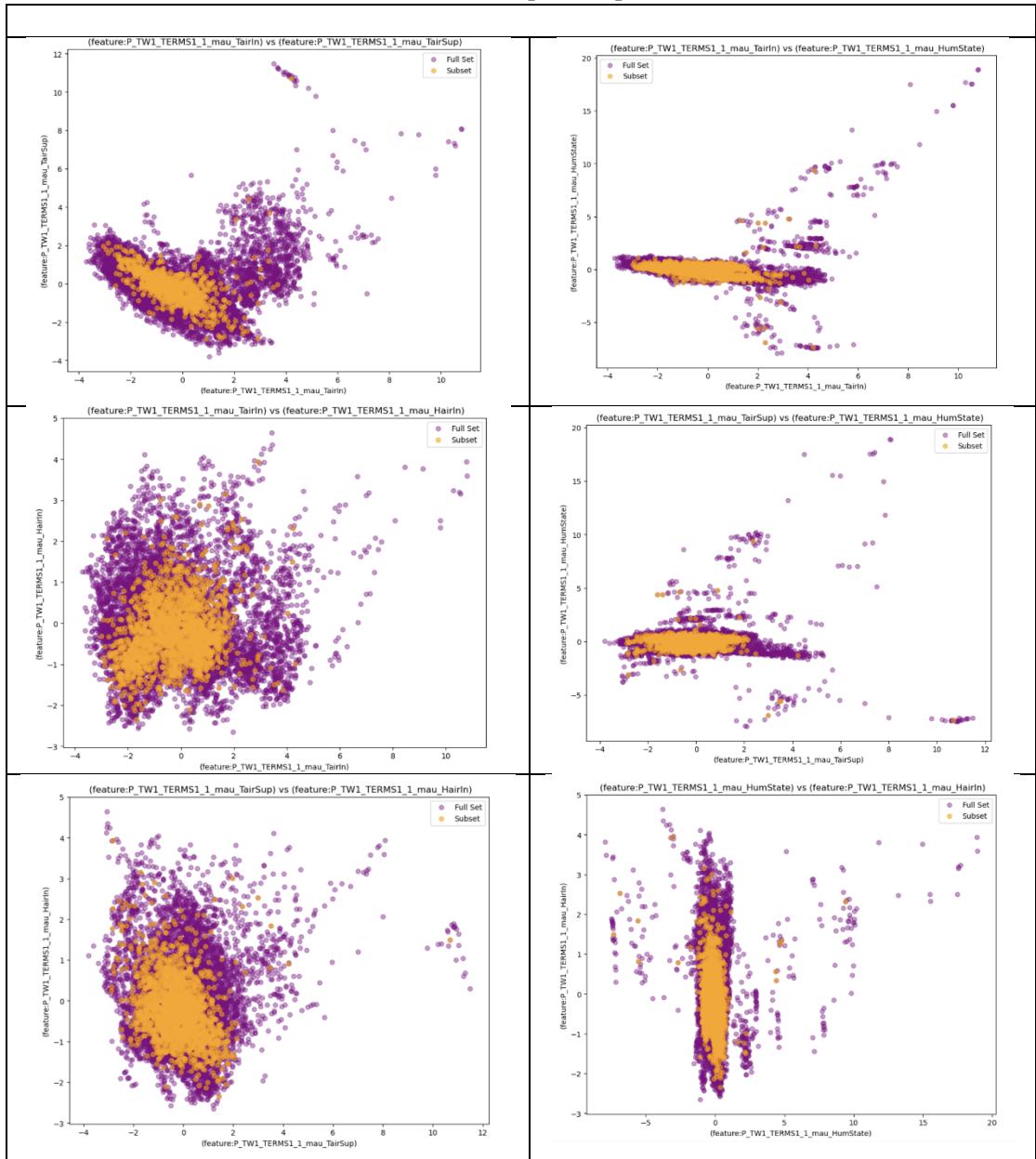


Table 17 4 Principal components

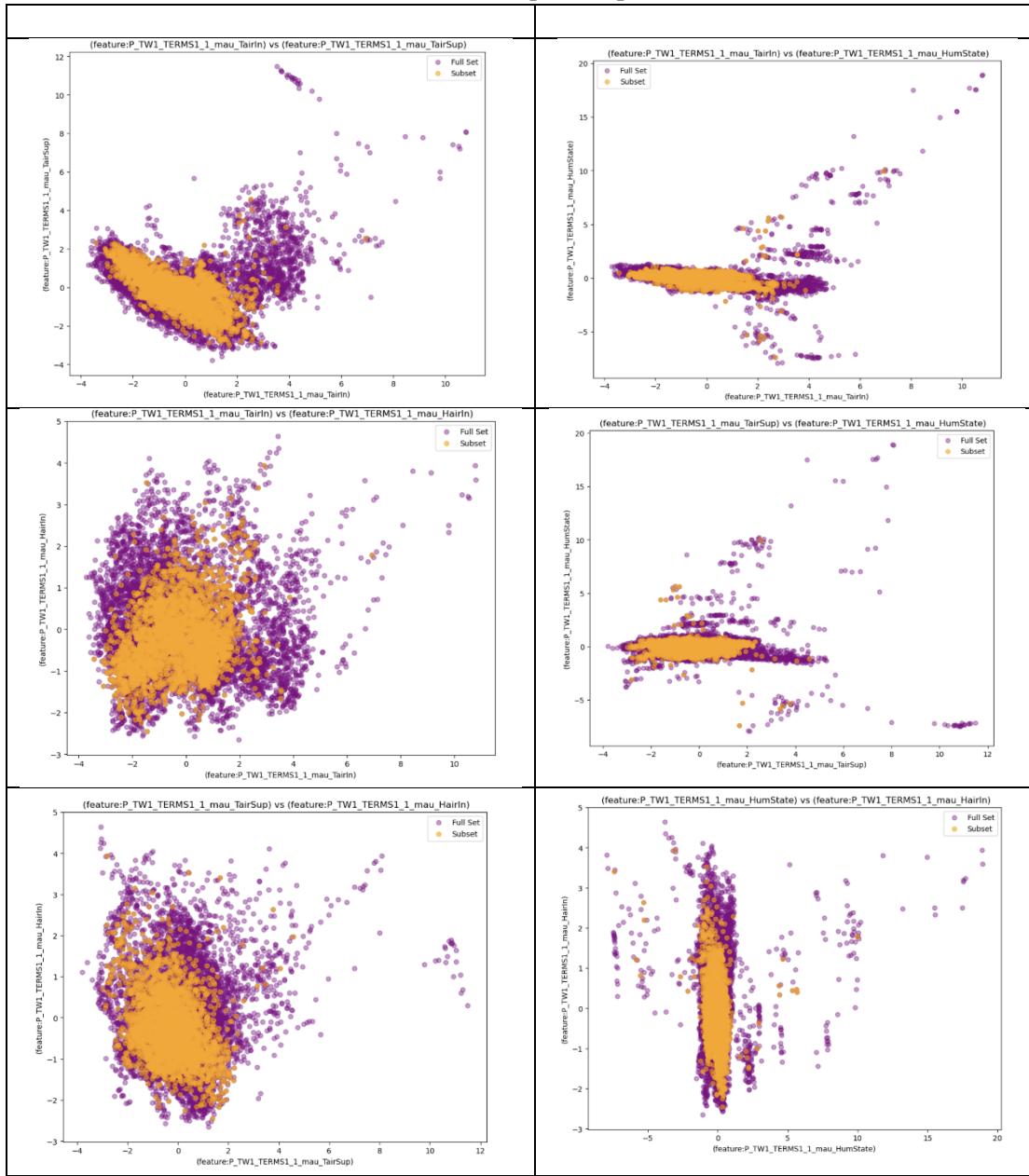


Table 18 4 Principal components

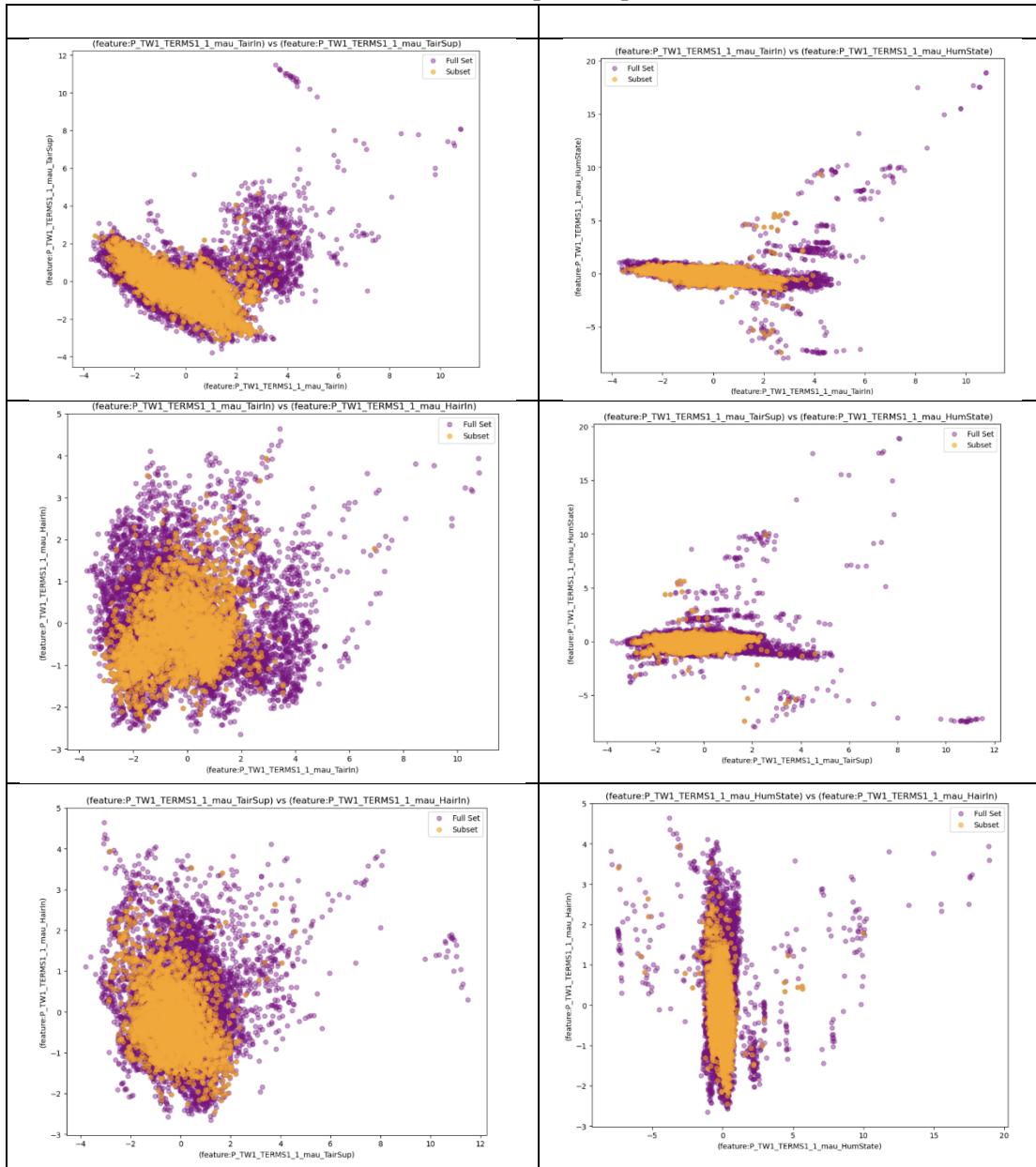


Table 19 4 Principal components

