



THE HONG KONG
POLYTECHNIC UNIVERSITY

香港理工大學

Pao Yue-kong Library

包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

<http://www.lib.polyu.edu.hk>

**DEVELOPMENT OF DATA-DRIVEN METHODS WITH
ENHANCED INTERPRETABILITY AND RELIABILITY FOR
OPTIMAL CONTROL AND DIAGNOSIS OF AIR-
CONDITIONING SYSTEMS**

CHEN ZHE

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Building Environment and Energy Engineering

**Development of Data-Driven Methods with Enhanced
Interpretability and Reliability for Optimal Control and
Diagnosis of Air-Conditioning Systems**

Chen Zhe

**A thesis submitted in partial fulfilment of the requirements for
the degree of Doctor of Philosophy**

August 2024

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no materials previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ **(Signed)**

Chen Zhe _____ **(Name of student)**

ABSTRACT

Enhancing the energy efficiency of building energy systems has become a key issue regarding carbon neutrality because buildings consume about 36% of energy and contribute 40% of carbon emissions. In the past decade, machine learning and data-driven modeling have been widely studied for building energy management applications, such as fault detection and diagnosis, optimal control, and demand response. Despite extensive research, the data-driven approach faces two major challenges regarding practical applications: a lack of interpretability and reliability. Interpretability refers to providing more insightful information for respective data-driven applications. Reliability aims to improve the applicability of data-driven model-related applications, such as model generalization ability, robustness and reliability in HVAC controls.

Therefore, this PhD study aims to develop data-driven methods with enhanced interpretability and reliability for optimal control and diagnosis in HVAC systems. In this study, the following research questions are addressed using novel methods with enhanced reliability and interpretability: **Q1**: How can we develop data-driven models with high interpretability and generalization ability under **data scarcity**? **Q2**: How can we address the **measurement uncertainty** issue when developing data-driven models? **Q3**: How can we resolve the **physical inconsistency** issue to enhance the interpretability and reliability of data-driven models? **Q4**: How can we achieve **reliable online control** (balancing energy efficiency and control smoothness) when deploying data-driven models in complex optimization problems? Questions Q1 to Q3 address general issues in the development of data-driven models, while Q4 addresses the application of data-driven models in complex optimization problems.

In **CHAPTER 3**, the data scarcity issue (**Q1**) is addressed by proposing a similarity learning-based method for enhanced interpretability and generalization ability, especially in typical HVAC FDD applications where only limited labeled data are available. The output of the proposed model is the similarity of the input to each fault, which provides better interpretability than conventional models that only output the probability of each fault. A temporal data-splitting method is utilized to handle the high correlation between training data and test data when a random split is adopted. The proposed method is validated in two scenarios: insufficient labeled data (scenario 1) and imbalanced labeled data (scenario 2), respectively. In scenario 1, the proposed method improves diagnostic accuracy by 45.7% compared to the baseline model

when labeled data are limited. In scenario 2, the method demonstrates superior generalization ability when dealing with imbalanced labeled data, improving fault diagnosis accuracy from 88.4% to 90.8%.

In **CHAPTER 4**, the measurement uncertainty issue (**Q2**) is addressed by probabilistic machine learning with a risk evaluation scheme for enhanced interpretability and reliability. For the typical chiller sequencing control problem, this study proposes a robust chiller sequencing control strategy to overcome the limitations faced by conventional cooling load predictions in controlling chiller plants. The strategy utilizes probabilistic cooling load predictions, with one-hour-ahead probabilistic cooling load predictions formulated as a normal distribution using the natural gradient boosting algorithm. A risk-based action evaluation scheme is designed to determine the optimal number of operating chillers by assessing the risks of possible control actions. The proposed strategy is validated through case studies based on historical operational data and an in-situ test. The case studies show that the strategy significantly enhances the reliability of the chiller plant by reducing the total switching number of chillers by up to 43.6% (from 55 to 31 in a week). The in-situ test results reveal a 56.5% decrease in total switching number compared to the rule-based strategy. The average daily energy savings is approximately 3945.1 kWh. Additionally, the coefficient of performance of the chiller plant increased by 4.2%.

In **CHAPTER 5**, the measurement uncertainty issue (**Q3**) is addressed by a new physically consistent data-driven method for enhanced interpretability and reliability. For a typical variable speed pump modeling problem, this study proposes a physically consistent, data-driven optimal sequencing control strategy to address the often-overlooked issue of optimal pump sequencing in HVAC water systems that adopt variable speed parallel pumping. By integrating physical knowledge with data-driven modeling, this approach enhances interpretability and reliability in real applications. The interpretable ElasticNet model is trained using features including pump frequency and total flow rate for different numbers of operating variable speed pumps. The optimal sequencing is subsequently determined using a physically consistent speed/frequency prediction method. The proposed strategy is validated using real building operational data, with results showing an average of 10% energy saving compared to the conventional rule-based strategy.

In **CHAPTER 6**, reliable online control (**Q4**) is achieved by a novel multi-objective optimization-based framework that enhances smoothness—an aspect often neglected in conventional deterministic and metaheuristic optimization algorithms. The distance between

two successive control actions based on Manhattan and Euclidean distances is one optimization objective, along with energy conservation as another optimization objective. The proposed strategy is validated using a typical optimization problem: optimal chiller loading. Results show that the proposed approach significantly improves control stability while maintaining energy efficiency compared to conventional optimization algorithms.

To sum up, the proposed data-driven methods improve interpretability and reliability compared to existing data-driven methods, addressing general issues in the development of data-driven models and the application of data-driven models in complex optimization problems related to fault diagnosis and optimal control in central air-conditioning systems. The new methods developed in **CHAPTER 4 and CHAPTER 5** (i.e., chiller sequencing strategy and pump sequencing strategy) have been deployed to the real chilled water system in the PolyU campus, demonstrating the energy efficiency and applicability of the proposed methods. In the future, the proposed data-driven methods are promising for large-scale deployment.

RESEARCH OUTPUT DURING PHD STUDY

Journal Papers

- [1] **Zhe Chen**, Fu Xiao*, Fangzhou Guo, and Jinyue Yan. Interpretable machine learning for building energy management: A state-of-the-art review. **Advances in Applied Energy** 2023; 9:100123.
- [2] **Zhe Chen**, Fu Xiao*, and Fangzhou Guo. Similarity learning-based fault detection and diagnosis in building HVAC systems with limited labeled data. **Renewable and Sustainable Energy Reviews** 2023;185:113612.
- [3] **Zhe Chen**, Jing Zhang, Fu Xiao*, Henrik Madsen, and Kan Xu. Probabilistic machine learning for enhanced chiller sequencing: A risk-based control strategy. **Energy and Built Environment** 2024.
- [4] **Zhe Chen**, Jing Zhang, Fu Xiao*, Kan Xu, and Yongbao Chen. Development of a probabilistic cooling load prediction-based robust chiller sequencing strategy and its real-world implementation. **Applied Energy**, accepted.
- [5] **Zhe Chen**, Jing Zhang, Fu Xiao*, Kan Xu, and Yongbao Chen. Physically consistent data-driven optimal sequencing strategy for variable speed pumps in large building chiller plants. **Submitted to Building and Environment**.
- [6] **Zhe Chen**, Fu Xiao*, and Yongbao Chen. Enhancing smoothness in model-based online optimal control through a multi-objective optimization approach. **Submitted to Energy**.
- [7] Kan Xu, **Zhe Chen**, Fu Xiao*, Jing Zhang, Hanbei Zhang, and Tianyou Ma. Semantic model-based large-scale deployment of AI-driven building management applications. **Automation in Construction** 2024;165:105579.
- [8] Ao Li, Jing Zhang, Fu Xiao*, Cheng Fan, Yantao Yu, and **Zhe Chen**. Design information-assisted graph neural network for modeling central air conditioning systems. **Advanced Engineering Informatics** 2024;60:102379.

Conference Presentations

[1] **Zhe Chen**, Fu Xiao*. Interpretable machine learning for AHU fault detection and diagnosis. The 14th International Conference on Applied Energy, 2022 , Bochum, Germany.

[2] **Zhe Chen**, Fu Xiao*. Robust comparison of optimization algorithms optimal chiller loading. International Conference on Sustainable Development in Building and Environment, 2023, Espoo, Finland.

Invention Patent (In Chinese)

[1] Fu Xiao, **Zhe Chen**, Jing Zhang, Wei Liao. A control method, device, equipment, and storage medium for a chiller. Granted. ZL202410250091.8

[2] Fu Xiao, **Zhe Chen**, Jing Zhang, Wei Liao. A method, device, equipment, and program product for determining the number of operating water pumps. Under review. 202411620436.0

ACKNOWLEDGEMENTS

First and foremost, I would like to extend my sincere gratitude to my supervisor, Professor Xiao Fu, for her invaluable guidance, patience, and support during my PhD study. Deep knowledge and insightful feedback were the fundamentals that helped me progress in research and become a mature researcher. I am also deeply grateful to Professor Wang Shengwei, who has given many constructive suggestions raised in group meetings and helped me in perfecting my work.

My deepest gratitude also goes to all BEAR members for their companionship during my PhD study, which has been a wonderful journey in my life. Special thanks to my best friend, Dr. Tang Hong, for her consistent and unconditional support and recognition. I am also deeply grateful to Professor Chen Yongbao for the helpful advice on academic writing. Additionally, the wisdom and profound academic experience of Dr. Zhao Wenxuan and Dr. Li Bingxu have continually inspired me. I would also like to thank Zhang Jing, Xu Kan, Wu William, and Yu Tony for their considerable support during the campus testing of the methods I have developed.

Lastly, I want to express my gratitude to my parents, Chen Wenping and Zhang Qiping, as well as our lovely puppy in heaven, Wangwang, and the new family member Wangzai. They have given me the warmth of family and provided a harbor for my life to anchor.

TABLE OF CONTENTS

ABSTRACT.....	i
RESEARCH OUTPUT DURING PHD STUDY	iv
ACKNOWLEDGEMENTS.....	vi
TABLE OF CONTENTS.....	vii
LIST OF FIGURES	xi
LIST OF TABLES.....	xiv
CHAPTER 1 INTRODUCTION	1
1.1. Background and motivation.....	1
1.1.1. Data-driven methods for optimal control and diagnosis.....	1
1.1.2. Challenge I: Lack of interpretability.....	2
1.1.3. Challenge II: Lack of reliability.....	4
1.2. Aim and objectives	4
1.3. Organization of the thesis	5
CHAPTER 2 LITERATURE REVIEW	8
2.1. Interpretability of data-driven applications.....	8
2.1.1. Classification of interpretable machine learning techniques	8
2.1.2. Ante-hoc approach	11
2.1.3. Post-hoc approach.....	15
2.2. Reliability of data-driven applications.....	20
2.2.1. Model generalization ability for fault detection and diagnosis.....	20
2.2.2. Robustness and stability for optimal control	23
2.3. Summary of research gaps	32
CHAPTER 3 A SIMILARITY-LEARNING METHOD FOR ENHANCING MODEL GENERALIZATION ABILITY AND INTERPRETABILITY	34
3.1. Proposed similarity learning-based fault diagnosis method	34
3.1.1. Overview of the proposed method.....	34
3.1.2. Model training.....	36
3.1.3. Fault diagnosis	40
3.2. Design of case studies using open dataset	41
3.2.1. Dataset description.....	41

3.2.2. Data-splitting.....	43
3.2.3. Baseline model for comparison	44
3.2.4. Scenario 1: Insufficient labeled data.....	44
3.2.5. Scenario 2: Imbalanced labeled data.....	45
3.2.6. Evaluation metric	45
3.3. Results and discussion	45
3.3.1. Scenario 1: Insufficient labeled data.....	46
3.3.2. Scenario 2: Imbalanced labeled data.....	49
3.4. Summary	51
CHAPTER 4 PROBABILISTIC MACHINE LEARNING FOR ENHANCING	
RELIABILITY AND INTERPRETABILITY ADDRESSING MEASUREMENT	
UNCERTAINTY	52
4.1. Proposed risk-based chiller sequencing control strategy	52
4.1.1. Basic principle	52
4.1.2. Overview of the proposed strategy	55
4.1.3. One-hour-ahead probabilistic cooling load prediction using natural gradient boosting.....	56
4.1.4. Enhanced cooling load-based sequencing	58
4.1.5. Risk-based action evaluation scheme	59
4.2. Test facilities and data collection.....	62
4.2.1. Chiller plant description.....	62
4.2.2. Data communication and processing	64
4.3. Robustness validation using operational data	65
4.3.1. Real building operation data	65
4.3.2. Performance of one-hour-ahead probabilistic cooling load predictions	65
4.3.3. Chiller sequencing results	68
4.3.4. Model interpretation: Risk analysis	70
4.3.5. Robustness comparison with conventional strategy under different levels of cooling load uncertainty.....	72
4.4. In-situ test of the proposed strategy	73
4.4.1. Testing day 1 and weather comparison with reference day 1	74
4.4.2. Probabilistic cooling load prediction	75
4.4.3. Robustness of sequencing actions and thermal comfort	76
4.4.4. Energy efficiency	78

4.4.5. Testing day 2 and testing day 3.....	79
4.4.6. Summary of the in-situ test.....	79
4.5. Summary.....	80
CHAPTER 5 PHYSICALLY CONSISTENT DATA-DRIVEN MODELS FOR ENHANCING INTERPRETABILITY AND RELIABILITY.....	82
5.1. Proposed physically consistent data-driven optimal sequencing control strategy.....	82
5.1.1. Derivation of operating point theorem.....	83
5.1.2. Derivation of pump power model.....	84
5.1.3. Interpretable data-driven pump modeling using ElasticNet.....	85
5.1.4. Physically consistent pump speed/frequency prediction after switching.....	86
5.2. Validation using real building operational data.....	87
5.2.1. Parallel pump system description.....	87
5.2.2. Pump power models.....	89
5.2.3. Frequency prediction after sequencing.....	90
5.2.4. Energy saving analysis.....	92
5.3. Summary.....	97
CHAPTER 6 GENERAL OPTIMIZATION FRAMEWORK FOR DATA-DRIVEN MODEL-BASED ONLINE OPTIMIZATION WITH ENHANCED CONTROL SMOOTHNESS.....	98
6.1. Proposed multi-objective optimization-based approach for online optimal control.....	98
6.1.1. Overview of the proposed approach.....	98
6.1.2. Optimization Objective 1: Energy efficiency.....	99
6.1.3. Optimization Objective 2: Control stability.....	100
6.1.4. Tolerance level-based solution selection from the Pareto front.....	101
6.2. Design of case studies.....	102
6.2.1. Typical optimization problem: optimal chiller loading.....	102
6.2.2. Data collection.....	104
6.2.3. Baseline algorithms for comparison.....	106
6.2.4. Stability metrics.....	106
6.3. Results and discussion.....	107
6.3.1. Control stability.....	107
6.3.2. Total energy consumption.....	113
6.3.3. Chiller switching details.....	114
6.3.4. The tradeoff between control stability and energy efficiency.....	118

6.4. Summary.....	121
CHAPTER 7 CONCLUSIONS	122
7.1. Summary of main contributions.....	122
7.2. Conclusions.....	123
7.3. Recommendations for future work	126
APPENDICES	128
Appendix A: Results of testing day 2 and testing day 3 in Chapter 4	128
REFERENCES	138

LIST OF FIGURES

Figure 1.1. The trade-off between model interpretability and accuracy.....	3
Figure 1.2. Enhancing interpretability and reliability: proposed vs. conventional methods	7
Figure 2.1. Classification of interpretable machine learning techniques.....	8
Figure 2.2. Ante-hoc and post-hoc interpretability	9
Figure 2.3. Global and local interpretability in a classification task.....	10
Figure 2.4. Model-agnostic and model-specific interpretable machine learning techniques..	11
Figure 2.5. Distribution of publications by year published and ante-hoc category	12
Figure 2.6. Sankey diagram depicting the connections of reviewed studies on different model dependencies, post-hoc techniques, and interpretability scopes.....	16
Figure 2.7. Random processes in metaheuristic algorithms.....	31
Figure 2.8. Local optima in optimization problems.....	32
Figure 3.1. The proposed similarity learning-based FDD method using Siamese networks..	36
Figure 3.2. Structure of the proposed Siamese network	37
Figure 3.3. Moving window sampling for multivariate time-series	38
Figure 3.4. Schematic of AHU in the ASHRAE RP-1312 dataset	42
Figure 3.5. Data-splitting for case studies.....	44
Figure 3.6. Baseline model for case studies.....	44
Figure 3.7. Average accuracy of fault diagnosis on the test set using the proposed method..	48
Figure 3.8. Performance improvement rate of the proposed Siamese networks ($N_s = 10$) ..	49
Figure 3.9. Fault diagnostic accuracy of each imbalanced fault.....	50
Figure 3.10. Fault diagnostic accuracy of remaining classes.....	50
Figure 4.1. Schematic diagram of a typical multiple-chiller system and sequencing control	53
Figure 4.2. The schematic diagram illustrating the basic idea of the proposed probabilistic cooling load prediction-based sequencing control strategy	55
Figure 4.3. Framework of the proposed chiller sequencing control strategy.....	56
Figure 4.4. Probabilistic prediction of cooling load in the form of normal distribution.....	58
Figure 4.5. Probabilistic cooling load prediction and risks of actions when switching off....	61
Figure 4.6. Probabilistic cooling load prediction and risks of actions when switching on.....	62
Figure 4.7. Diagram of the chiller plant in PolyU campus	63
Figure 4.8. Data communication between AI workstation and BMS	64
Figure 4.9. Test cooling load profile for the data experiment.....	65

Figure 4.10. One-hour-ahead probabilistic cooling load predictions.....	66
Figure 4.11. Histogram of probability integral transform.....	67
Figure 4.12. Prediction interval averaged width at different times.....	68
Figure 4.13. Control comparison of conventional and the proposed strategies.....	69
Figure 4.14. Risks of Q-based action and compromised actions on the test data.....	71
Figure 4.15. Sequencing results and risks evaluation on August 8.....	72
Figure 4.16. Impact of uncertainty level to the total switching numbers.....	73
Figure 4.17. Weather conditions on reference day 1 and testing day 1	75
Figure 4.18. One-hour-ahead probabilistic cooling load prediction on testing day 1.....	76
Figure 4.19. Chiller sequencing results on reference day 1 and testing day 1.....	77
Figure 4.20. Chilled water return temperature of three loops.....	78
Figure 5.2. Total pump power operating 1 and 2 SCHWPs	88
Figure 5.3. Power predictions of 1 SCHWP	89
Figure 5.4. Power predictions of 2 SCHWPs.....	90
Figure 5.5. Frequency prediction accuracy	91
Figure 5.6. Frequency prediction after switching from 1 to 2 SCHWPs.....	91
Figure 5.7. Frequency prediction after switching from 2 to 1 SCHWP.....	92
Figure 5.8. Energy saving from 1 to 2 SCHWPs (Min. frequency: 30 Hz).....	93
Figure 5.9. Energy saving percentage from 1 to 2 SCHWPs (Min. frequency: 30 Hz).....	94
Figure 5.10. Energy saving from 2 to 1 SCHWP (Max. frequency: 50 Hz).....	95
Figure 5.11. Energy saving percentage from 2 to 1 SCHWP (Max. frequency: 50 Hz)	95
Figure 5.12. Monthly energy saving	96
Figure 5.13. Detailed energy saving on a typical working day (21-July 2023).....	97
Figure 6.1. Overview of the proposed multi-objective optimization-based approach.....	99
Figure 6.2. Power curve against partial load ratio of chillers	104
Figure 6.3. Actual 4-week cooling load from the chiller plant at an educational building...	105
Figure 6.4. Frequency distribution of chiller plant PLR for the tested cooling load curve ..	106
Figure 6.5. Total chiller switching number of the proposed MOO-based approach	108
Figure 6.6. The difference compared to baseline algorithms in total switching numbers	109
Figure 6.7. Daily chiller switching number (random seed 0-9).....	110
Figure 6.8. Difference between maximum and minimum switching number in the day (random seed 0-9)	111
Figure 6.9. Daily average absolute change in PLR (%) (random seed 0-9)	112
Figure 6.10. Energy consumption of the proposed MOO-based approach in four weeks....	114

Figure 6.11. Daily switching number of each chiller.....	115
Figure 6.12. Partial load ratio of each chiller on August 1, 2022 (Monday)	117
Figure 6.13. Average absolute change in PLR (%) of each chiller on August 1, 2022	118
Figure 6.14. Comparison in energy efficiency and control stability.....	119
Figure 6.15. Disturbances in online control problems	121
Figure 7.1. Contributions of this PhD study	122
Figure A.1. Weather conditions on reference day 2 and testing day 2	128
Figure A.2. 1 hour ahead probabilistic cooling load prediction on testing day 2	129
Figure A.3. Chiller sequencing results of reference day 2 and testing day 2.....	130
Figure A.4. Chilled water return temperature of three loops	131
Figure A.6. Weather conditions on reference day 3 and testing day 3	133
Figure A.7. 1 hour ahead probabilistic cooling load prediction on testing day 3	134
Figure A.8. Chiller sequencing results on reference day 3 and testing day 3	135
Figure A.9. Chilled water return temperature of three loops	136

LIST OF TABLES

Table 2.1. Comparison of deterministic algorithms and metaheuristic algorithms	30
Table 3.1. The architecture of LSTM subnetworks in the proposed Siamese network	38
Table 3.2. Details of typical AHU faults in the ASHRAE RP-1312 dataset	42
Table 3.3. Features used for AHU fault diagnosis in the ASHRAE RP-1312.....	42
Table 3.4. Accuracy of fault diagnosis on the test set using Siamese networks (%)	47
Table 4.1. Mapping between the Q-based action a_1 and compromised action a_2	59
Table 4.2. Specification of air-cooled/water-cooled chillers	63
Table 4.3. Specification of primary chilled water pumps and condensing water pumps.....	63
Table 4.4. Specification of cooling towers.....	63
Table 4.5. Performance of probabilistic cooling load predictions	66
Table 4.6. Control comparison of two strategies from August 7 to 13, 2022	69
Table 4.7. Sequencing details from August 7 to 13, 2022	70
Table 4.8. Weather summary on reference day 1 and testing day 1	75
Table 4.9. Chiller switching number on reference day 1 and testing day 1	77
Table 4.10. Energy consumption on reference day 1 and testing day 1 (kWh)	78
Table 4.11. Energy efficiency of chiller plant on reference day 1 and testing day 1	79
Table 4.12. Chiller switching number on reference and testing days 1-3.....	79
Table 4.13. Energy consumption on reference days 1-3 and testing days 1-3 (kWh)	80
Table 4.14. Energy efficiency of chiller plant on reference days 1-3 and testing days 1-3	80
Table 5.1. Design working conditions of the variable speed pumps.....	88
Table 6.1. Chiller performance coefficients.....	103
Table 6.2. Total chiller switching number in four weeks (random seed 0-9)	110
Table 6.3. Average absolute change of PLR (%) in four weeks (random seed 0-9).....	112
Table A.1. Weather summary on reference day 2 and testing day 2.....	129
Table A.2. Chiller switching number on reference day 2 and testing day 2	130
Table A.3. Energy consumption on reference day 2 and testing day 2 (kWh)	132
Table A.4. Energy efficiency of chiller plant on reference day 2 and testing day 2.....	132
Table A.5. Weather summary on reference day 3 and testing day 3.....	133
Table A.6. Chiller switching number on reference day 3 and testing day 3	135
Table A.7. Energy consumption on reference day 3 and testing day 3 (kWh)	136
Table A.8. Energy efficiency of chiller plant on reference day 3 and testing day 3.....	137

CHAPTER 1 INTRODUCTION

1.1. Background and motivation

1.1.1. Data-driven methods for optimal control and diagnosis

The building sector is a major contributor to global energy consumption and carbon emissions. In 2020, it accounted for 36% of global energy consumption and 37% of global CO₂ emissions [1]. Throughout the life cycle of buildings, the operation phase accounts for 80%-90% of total energy consumption [2]. Therefore, building energy management is crucial for global energy-saving and carbon neutrality. Many researchers have quantified the potential of building energy-saving and proposed plans to enhance building energy efficiency. For example, China aims to achieve a 50% reduction in building energy consumption, and one strategy is to adopt efficient equipment and smart building management systems (BMSs) [3]. In Hong Kong, buildings consume 90% of electricity, and therefore, the government plans to reduce the electricity consumption of commercial buildings by 30%-40% before 2050 [4]. In the U.S., energy consumption from the building sector can be reduced by efficient heating and advanced control. Researchers estimated that CO₂ emissions from buildings can be reduced by up to 78% by 2050 [5]. In the EU, the building sector would need to reduce its emissions by 60% to reach the EU objective of a 55% reduction by 2030 [6].

BMSs play an essential role in improving energy efficiency and flexibility during building operations. BMSs can implement various smart control strategies in building energy systems, such as heating, ventilating, and air conditioning (HVAC) systems, energy storage systems, and renewable energy systems [7,8]. Conventional control strategies such as rule-based control strategies relying on physics and experience face great challenges in tackling the complicated interactions among building energy systems [9]. Modern buildings are usually equipped with advanced metering infrastructure and numerous sensors; thus, the BMS can collect and store massive energy-related operational data. The prospect of utilizing such big data has opened up due to the advancement in machine learning algorithms. Machine learning algorithms can discover and learn new knowledge (i.e., data-driven models) from the data to support energy-efficient/energy-flexible control in the ever-changing energy market [10]. With such data-driven models, building energy systems can be monitored to make decisions autonomously with the support of big data [11].

Machine learning and data-driven modeling have effectively facilitated building energy management in various typical applications in the past decade, including cooling load prediction, equipment modeling, fault detection and diagnosis (FDD), etc.

Cooling load prediction refers to predicting the cooling demand of a building ranging from hours to months. Accurate load/power prediction is important for improving building energy efficiency and flexibility [12]. Cooling load prediction models are useful for optimizing the operation of chiller plants, e.g., chiller sequencing control [13]. Compared with physics-based load/power prediction, machine learning algorithms require only historical data instead of detailed physical information and thermal balance equations, making them easier to develop and deploy.

Modeling HVAC systems is crucial for improving energy efficiency. With the development of sensor technology, data-driven modeling is getting more and more popular [14]. Data-driven models are created by gathering performance data from real-world applications and establishing correlations between input and output variables of HVAC equipment using mathematical techniques such as statistical regression or artificial neural networks (ANN). Data-driven modeling techniques are widely adopted due to their simplicity and high accuracy, typically for modeling equipment such as chillers, piping systems, and heating/cooling coils. In addition, by combining data-driven modeling and optimization algorithms, optimization algorithms, model-based control can be achieved to improve energy efficiency and flexibility in building energy systems [15].

Machine learning has also been used to detect and diagnose faults in building energy systems, i.e., FDD [16]. Early detection of equipment faults is essential for building energy efficiency, especially for energy-intensive equipment such as chillers. Unlike knowledge-driven FDD, data-driven FDD requires less professional knowledge and can distinguish rare and unforeseen energy patterns in real operations [17], which are valuable for FDD.

Despite extensive research, the data-driven approach faces two major challenges when it comes to practical applications: a lack of interpretability and reliability.

1.1.2. Challenge I: Lack of interpretability

Model interpretability refers to the extent to which the predictions or decisions of a data-driven model can be understood by humans [18]. For example, in typical data-driven FDD, the output is simply the probability of each fault, without providing more insightful information to decision-makers [19]. Although the applications of machine learning algorithms have fully

demonstrated their values for building energy management, their broad applications are limited by a lack of interpretability [20]. In other words, most data-driven models are not transparent or explainable.

The trade-off between model accuracy and model interpretability limits the power of machine learning [21,22], as shown in **Figure 1.1**. For example, ANN usually consists of an input, output, and hidden layers. Increasing the number of hidden layers of an ANN model can often improve its accuracy in modeling complicated systems, but the interpretability of the model decreases [23]. The model becomes “deeper” and “darker”, making it more difficult for users to understand and interpret the modeling process and results. Meanwhile, the existing ready-to-interpret models, such as linear regression, lack good prediction performance. Researchers made great efforts to improve the prediction performance of black-box models, such as ANN and support vector machine (SVM), but generally overlooked their interpretability in building energy management. **Figure 1.1** also shows the two major approaches to addressing the trade-off between model accuracy and model interpretability: ante-hoc and post-hoc approaches.

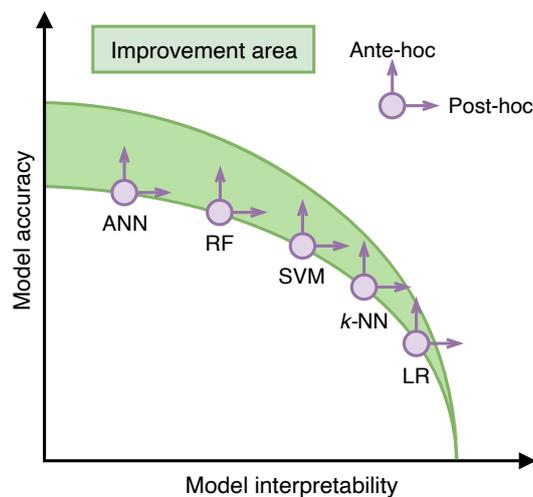


Figure 1.1. The trade-off between model interpretability and accuracy

The lack of interpretability also challenges the mass deployment of data-driven models in real-world applications [24]. First, during the training process, the training data are usually incomplete; therefore, the trained data-driven models need to tackle out-of-distribution data after deployment [25]. During the training process of data-driven models, physical knowledge and information are usually ignored compared with physics-based modeling. Therefore, decision-makers may find the data-driven models untrustworthy if the models are not trained on complete operational data and the real-world performance is worse than on the training data. Second, because of the black-box nature, data-driven models produce output without any

explanations. Decision-makers usually need insights into how and why the black-box models produce such predictions so that they can understand, check and apply the models. Generally, there is still significant skepticism in the building industry about the broad application of machine learning because there is a mismatch between training and deployment environments. Additionally, in practical applications, it is also necessary to provide reasonable interpretations that help operators to understand the decisions made by data-driven models.

1.1.3. Challenge II: Lack of reliability

Previous data-driven models and their applications in building energy management have also lacked reliability. First and foremost, the generalization ability is key to the performance of data-driven models when new data is fed in, i.e., after the models have been trained based on historical data and deployed for online applications. This is especially true in the application of FDD, where historical labeled data are usually limited or unbalanced. When data-driven models are trained on such datasets, they are often overfitted or biased, which impairs their reliability in actual applications. Additionally, the reliability of data-driven models faces the challenge of high autocorrelation in time-series data, which can lead to overestimation of model performance.

In optimal control of building energy systems, the challenges of data-driven model-based applications lie in robustness and stability. For example, cooling load prediction-based sequencing control for multiple-chiller plants tends to be reliable but suffers from uncertainties from the sensors and weather forecasts. Therefore, it is necessary to consider the uncertainty of cooling load prediction to make more robust control actions. Furthermore, the application of data-driven model-based optimization must consider the stability of online control. Deterministic algorithms provide predictability but can suffer from rigidity, while metaheuristic algorithms face instability due to their inherent randomness.

1.2. Aim and objectives

The lack of interpretability and reliability are the main challenges in applying data-driven models to HVAC systems. Among them, interpretability is an important foundation for reliability. The ultimate goal of improving interpretability and reliability is to enhance the applicability and scalability of data-driven models in practical applications, enabling machine learning techniques to be applied to practical projects on a large scale rather than stopping at the research stage. Therefore, this study aims to develop data-driven methods with **enhanced**

reliability and interpretability, and to validate their **applicability** using data experiments and in-situ tests. This aim can be accomplished by addressing the following four research questions:

Q1: How can we develop data-driven models with high interpretability and generalization ability under **data scarcity**?

Q2: How can we address the **measurement uncertainty** issue when developing data-driven models?

Q3: How can we resolve the **physical inconsistency issue** to enhance the interpretability and reliability of data-driven models?

Q4: How can we achieve **reliable online control** (balancing energy efficiency and control smoothness) when deploying data-driven models in complex optimization problems?

Questions **Q1** to **Q3** address general issues in the development of data-driven models, while **Q4** addresses the application of data-driven models in complex optimization problems.

1.3. Organization of the thesis

The thesis consists of 7 chapters. A summary of the content in each chapter is presented as follows:

CHAPTER 1 introduces the background and motivation for data-driven methods aimed at enhancing interpretability and reliability in optimal control and diagnosis of air-conditioning systems. It discusses two major challenges faced by conventional data-driven models in HVAC systems: lack of interpretability and lack of reliability. The chapter also presents the aim, main objectives, and the organization of the thesis.

CHAPTER 2 presents a comprehensive review of research on data-driven building energy management and the two major challenges: the lack of interpretability and reliability. The significance of interpretability and reliability of data-driven model-based applications in building energy management are detailed. The main research gaps are identified from the literature review.

CHAPTER 3 addresses the data scarcity issue (**Q1**) and presents a similarity learning-based fault detection and diagnosis method for building HVAC systems. The output is the similarity towards each fault, which provides better interpretability. The temporal data-splitting method is adopted to tackle the issue of the high correlation of training data and test data using random split. Two case studies were conducted to test the effectiveness of the proposed method, with

insufficient labeled data and imbalanced data, respectively. Two case studies are conducted utilizing an open dataset for AHU FDD to validate the proposed method.

CHAPTER 4 addresses the measurement uncertainty issue (**Q2**) and presents a probabilistic machine learning for enhanced interpretability and reliability for the typical measured cooling load-based chiller sequencing control problem. One-hour-ahead probabilistic cooling load prediction in the form of a normal distribution is made using natural gradient boosting (NGBoost). An online risk-based action evaluation module is designed to determine the number of operating chillers and simultaneously assess the risks in the process and the reliability of the strategy. The proposed control strategy determines the number of operating chillers based on the probability distributions of cooling loads in the following hour. The proposed strategy is validated using historical operational data experiments and an in-situ test.

CHAPTER 5 addresses the measurement uncertainty issue (**Q3**) and presents a novel physically consistent data-driven method for enhanced interpretability and reliability using the modeling of variable speed pumps as an example. For different numbers of operating variable speed pumps, features including pump frequency and total flow rate are used to train the interpretable ElasticNet data-driven model. The optimal sequencing is then determined using a physically consistent speed/frequency prediction method. The energy performance of the proposed strategy is validated using real building operational data.

CHAPTER 6 achieved reliable online control (**Q4**) by proposing a multi-objective optimization-based framework for enhanced smoothness when data-driven models are used in complex optimization problems to overcome the instability issue often found in conventional deterministic and metaheuristic algorithms. The distance between two successive control actions is one optimization objective alongside energy conservation. The control stability of the proposed strategy is validated using a typical optimization problem: optimal chiller loading.

CHAPTER 7 summarizes the main contributions of this PhD study and gives recommendations for future research and applications.

The organization of the main chapters of the thesis and the comparison of the proposed methods with their respective convention methods in terms of interpretability and reliability are illustrated in **Figure 1.2**. In this study, interpretability refers to providing more insightful information for respective data-driven applications. Furthermore, reliability aims to improve the applicability of data-driven model-related applications, such as model generalization ability, robustness and control in HVAC controls.

Research question	Conventional method	Proposed method
Q1: Data scarcity (CHAPTER 3)	Probability of faults	Similarity to faults like k -NN
	Lack of generalization ability with limited labeled data	Improved generalization ability using similarity learning
Q2: Measurement uncertainty (CHAPTER 4)	Point prediction without uncertainty	Probabilistic prediction with risk interpretation
	Lack of control robustness	Improved control robustness using risk evaluation
Q3: Physical inconsistency (CHAPTER 5)	Pure data-driven pump curves	ElasticNet interpretable model integrating physical rules
	Simulation based on manufacture curves and affinity laws	Data experiment and in-situ test
Q4: Reliable Online Control (CHAPTER 6)	Lack of control smoothness	Improved control smoothness

 Interpretability
  Reliability

Figure 1.2. Enhancing interpretability and reliability: proposed vs. conventional methods

CHAPTER 2 LITERATURE REVIEW

This chapter presents a comprehensive review of research on data-driven building energy management and the two major challenges: the lack of interpretability and reliability. Section 2.1 details the significance of model interpretability for building energy management, introduces the taxonomy of interpretable machine learning techniques, and then reviews the literature that adopts ante-hoc and post-hoc approaches. Section 2.2 details the importance of reliability of data-driven model-based applications in building energy management, especially in fault detection and diagnosis and model-based control. Section 2.3 summarizes the main research gaps identified from the literature review.

2.1. Interpretability of data-driven applications

2.1.1. Classification of interpretable machine learning techniques

According to different criteria, i.e., application stage, interpretability scope, and model dependency, techniques for interpretable machine learning can be categorized into different groups [26], as shown in **Figure 2.1**. For example, the physics-informed neural network (PINN) is an ante-hoc, global, and model-specific technique that integrates physical rules into neural networks [27]. Detailed classifications of interpretable machine learning techniques are provided in Sections 2.1.1.1-2.1.1.3. Then, Sections 2.1.2 and 2.1.3 introduce the popular interpretable machine learning techniques based on two application stages: ante-hoc and post-hoc.

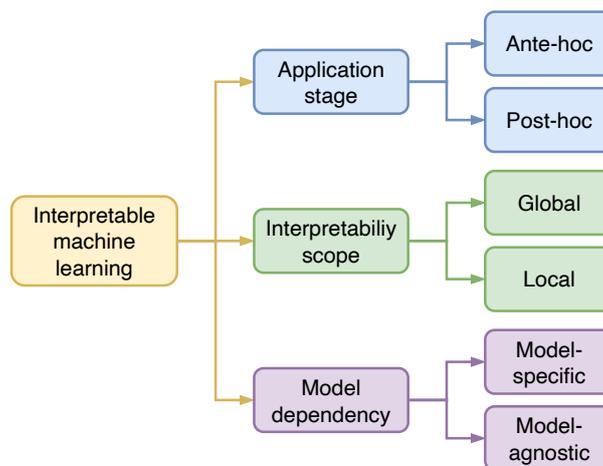


Figure 2.1. Classification of interpretable machine learning techniques

2.1.1.1. Application stage

First, interpretable machine learning techniques can be classified according to when the techniques are adopted in building data-driven model. Ante-hoc interpretable machine learning techniques are applied during the model training process, and post-hoc interpretable machine learning techniques are applied after training. **Figure 2.2** shows how ante-hoc and post-hoc interpretable techniques are applied at different stages in the model training process.

Ante-hoc interpretable data-driven models are usually self-explanatory. Therefore, data-driven models developed using ante-hoc techniques are also called intrinsic or transparent models. For example, linear regression is a simple ante-hoc model for predicting a continuous outcome variable based on one or more predictor variables. Linear regression is self-explanatory because it makes predictions using a linear combination of the input variables, which can be easily understood and explained [26]. Although linear regression has high interpretability according to **Figure 1.1**, it is too simple to address complicated problems in building energy management [28]. In this paper, a variant of linear regression named generalized additive models (GAMs) is reviewed. GAMs have strong flexibility and interpretability in regression and classification tasks [29]. As shown in **Figure 2.2**, post-processing is used to evaluate each input's impact according to the parameters of the intrinsic model. For example, the coefficients of GAMs can be used to evaluate input features' positive or negative effects.

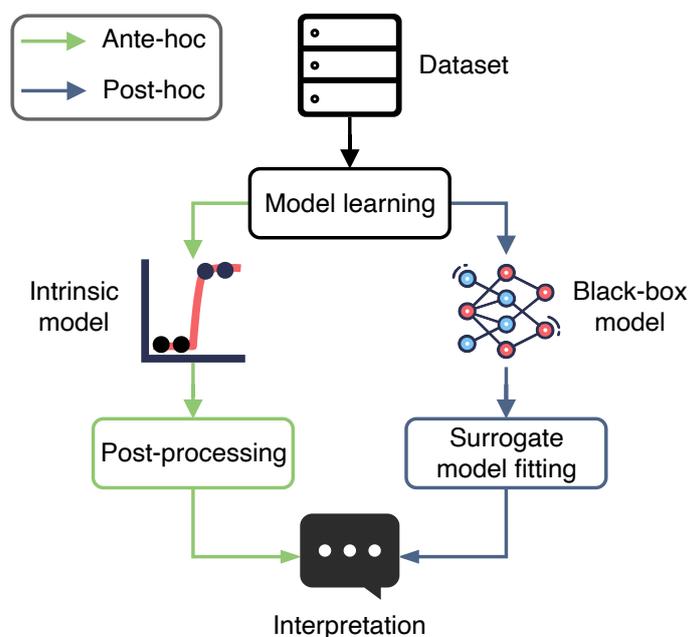


Figure 2.2. Ante-hoc and post-hoc interpretability

Post-hoc interpretable machine learning techniques are applied to black-box models after training. They are used to interpret and understand the dependency and significance of specific input features over the output by fitting surrogate models without the need to understand the internal structures. Post-hoc interpretable machine learning techniques generate interpretation by examining the interrelationship between input features and the predictions.

2.1.1.2. Interpretability scope

Interpretability scope refers to the scope of model output that needs to be interpreted. As the classification problem shown in **Figure 2.3**, global interpretation explains a data-driven model based on a full view of the model structures and parameters. In contrast, local interpretation explains each prediction individually.

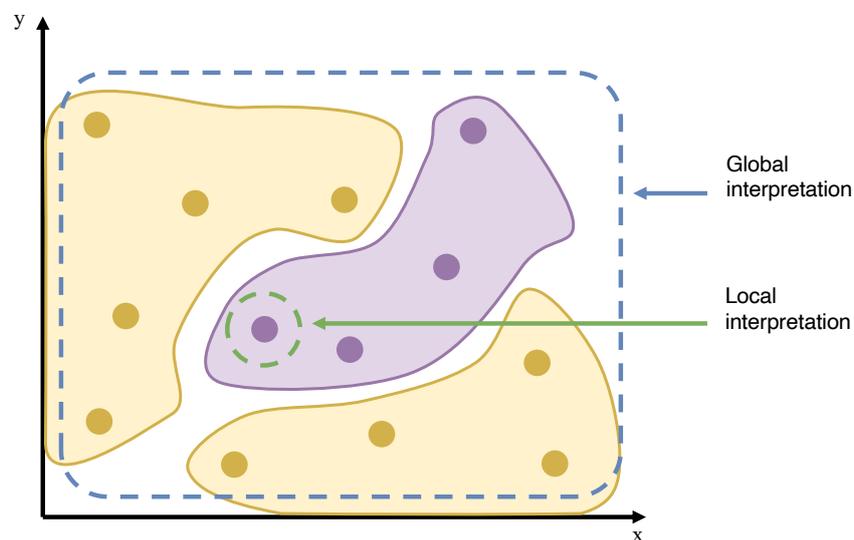


Figure 2.3. Global and local interpretability in a classification task

Global interpretable machine learning techniques aim to provide a holistic understanding of the data-driven model by measuring the global effects of the input features on the model prediction. They require only the black-box models and the entire training data. Global interpretation helps decision-makers gain a macro-level understanding of the data-driven model, including the most influential input features. In the context of FDD, global interpretability helps explain which features are most significant in predicting equipment faults.

Local interpretable machine learning techniques provide a transparent understanding of the model prediction for a specific input sample. Instead of focusing on the importance of global features, local methods focus on the contribution of each feature to a prediction sample and require both the black-box model and prediction sample. Local interpretation is important for

decision-makers to trust the output or correct the wrong output. In the context of FDD, local interpretability helps explain which features contribute the most to the prediction sample, such as high supply air temperature in the air handling unit operation.

2.1.1.3. Model dependency

Model dependency refers to whether the interpretable machine learning technique can be applied to any data-driven model or specific models. Some interpretation techniques treat the data-driven models as black-box models, and these techniques are applicable to any data-driven model or are independent of the type of data-driven model. Therefore, these techniques are model-agnostic, as **Figure 2.4** (a) illustrates. Other techniques can only be applied to interpret certain types of data-driven models and are thus called model-specific techniques, as shown in **Figure 2.4** (b).

Model-agnostic techniques can be applied to any data-driven model because they require only the input and output of the data-driven model without considering its inner structures. Therefore, most post-hoc interpretable machine learning techniques are model agnostic. For example, LIME is a post-hoc model-agnostic tool that can approximate any data-driven model locally.

Model-specific techniques can dig into the specific characteristics or architecture of the, providing in-depth interpretability that may not be possible with model-agnostic methods. For example, the attention mechanism is usually employed in neural networks to improve interpretability as a model-specific technique.

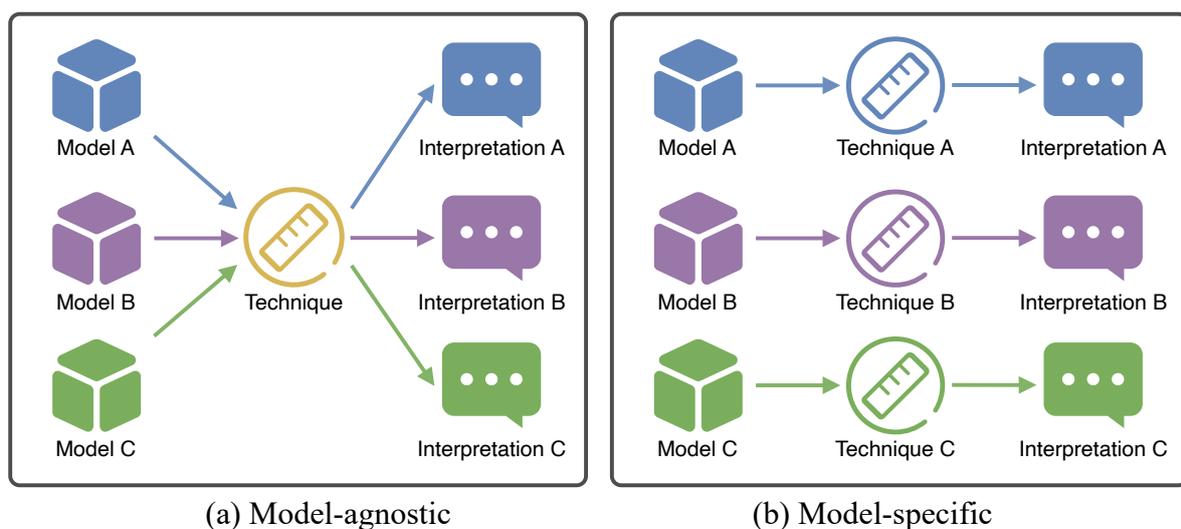


Figure 2.4. Model-agnostic and model-specific interpretable machine learning techniques

2.1.2. Ante-hoc approach

40 papers that adopted the ante-hoc approach to improve model interpretability for building energy management were reviewed in this section. These papers are further divided into four categories according to the specific ante-hoc techniques adopted: modified neural networks, attention mechanism, clustering and feature extraction, and generalized additive models (GAMs). **Figure 2.5** shows the number of studies in each category from 2018 to 2022.

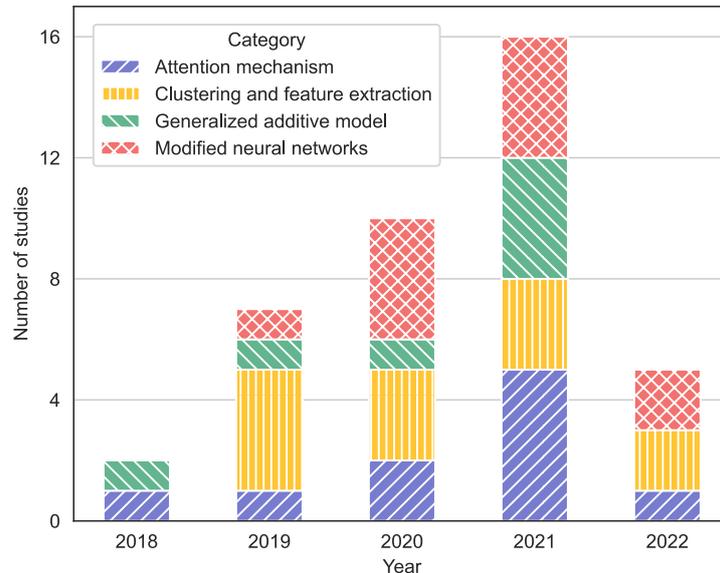


Figure 2.5. Distribution of publications by year published and ante-hoc category

2.1.2.1. Modified neural networks

ANN has become popular in building energy management in the past decades. It is well known that ANN is dark for users. A typical approach to improving ANN's interpretability is modifying neural networks' structure, which generates the so-called modified neural networks with enhanced interpretability.

There are two ways to modify the structure of neural networks in general. First, elements with physical meaning can be directly added to the models. Shan et al. integrated the gravitational model (GRA) with the gated recurrent units (GRU) model for building energy consumption prediction [30]. In the proposed GRA-GRU model, linear model GRA and non-linear model GRU were ensembled. The weights of the two models were determined using mutual information and weighted entropy. Wang et al. proposed a direct explainable neural network (DXNN) using the ridge function instead of the widely-used sigmoid activation function as the kernel function [31]. Considering the polynomial characteristic of the ridge function, the mathematical relationship between the model input and output can be directly obtained. The DXNN was used for solar irradiance forecasting, and the results showed that the output is a

quadratic function of input features. Zhang et al. combined a deep belief network with Takagi-Sugeno-Kang fuzzy classifier to generate interpretable fuzzy rules for indoor occupancy detection [32]. Kim modified the conventional convolutional neural network (CNN) and proposed interpretable CNN (I-CNN) for indoor human activity detection by adding temporal convolution and pooling layers into the CNN. The author demonstrated that the proposed I-CNN could rank the importance of sensor signals and improve the performance of I-CNN [33]. Chen and Zhang used domain knowledge to obtain the average trend of district load reflecting the periodic patterns so that the data-driven model could predict irregular local load fluctuations [34].

Second, domain knowledge can be used to guide the design of neural networks or improve the training process of neural networks, making the neural networks and training processes physically explainable. Chen et al. used domain knowledge to guide the design of model input/output and the structure of neural networks for air-conditioner modeling [35]. To make the control signal predicted by the deep Q network trustworthy and aligned with domain knowledge, Yu et al. integrated *a priori* knowledge into the searching strategy. They concluded that the knowledge-based search strategy could significantly reduce training time [36]. The modified LSTM proposed in [37] also used thermal dynamics to guide the design of a recurrent neural networks (RNN) model for building thermal modeling, which can learn interpretable dynamic models from measurement data. Di Natale et al. proposed a physically consistent neural network by incorporating domain knowledge into black-box models for building thermal modeling, and the proposed approach was proved to be physically interpretable [38].

2.1.2.2. Attention mechanism

The attention mechanism was first introduced by Bahdanau et al. to improve the performance of the encoder-decoder model for machine translation [39]. Inspired by the cognitive attention process, the attention mechanism can improve the interpretability of encoder-decoder models by stressing some parts of the input features in making predictions while weakening the rest features based on the context vectors. Because encoder-decoder models deal with time-series data, the attention mechanism can consider the temporal dependency of time-series data [40]. It is an ante-hoc approach because it is embedded into the prediction model [41].

Many studies have used the attention mechanism to analyze temporal dependency in time-series data in both regression and classification tasks. According to the individual attention matrix of input samples, Li et al. analyzed the temporal dependency of time-series data and

removed redundant features for chiller fault diagnosis [42]. It could provide local interpretation of the importance of sensors on the fault diagnosis resulting from the encoder-decoder network. Attention weight heatmap was used in [43] to explore the features emphasized in the LSTM model for day-ahead daily load prediction. Results showed that one day-ahead load is the most important feature. Similarly, average attention patterns in [44] demonstrated that the impact of historical features on model output exhibited 24-hour periodicity, indicating a strong relationship between energy consumption and the hour of the day. Li et al. adopted the attention mechanism in the ANN model for building cooling load prediction. They found that the most recent energy consumption data had the most significant influence on the next-hour cooling load prediction [45]. In [46,47], spatiotemporal attention values were almost evenly distributed across all input time steps for zone air temperature prediction because air temperature had faster thermal dynamics than the building envelope.

Technically, the attention mechanism generates local interpretations because it treats each input sample individually. Nevertheless, many studies treat the average of attention values as global interpretability. For example, Guo et al. [48] stated that the average attention values of LSTM could represent the importance of input features, which conform to the domain knowledge.

2.1.2.3. Clustering and feature extraction

Unlike modified neural networks and attention mechanisms embedded into black-box models (neural networks), clustering and feature extraction techniques do not change the structure of the original black-box models. The clustering and feature extraction techniques improve the interpretability of data-driven models by clustering raw data into several groups with human-interpretable characteristics (e.g., interpretable rules) or extracting interpretable features.

Bhatia et al. proposed a novel clustering technique named axis-aligned hyper-rectangles [49] for clustering simulated building thermal design data. Compared with other clustering techniques, it could generate hyper-rectangle boundaries that can be described with interpretable rules. It was employed to extract interpretable rules, like the range of window-to-wall ratio, to assist the design of building envelopes in different climate zones [49]. Some studies have compared the performance of clustering techniques with black-box models. Grimaldo and Novak claimed that the interpretable machine learning approach did not sacrifice the model's accuracy [50]. Their case study showed that the k-nearest neighbors (kNN) algorithm had similar accuracy for load prediction compared with sophisticated data-driven

models such as RF and gradient boosted trees. The kNN algorithm is interpretable because it is model-free and makes predictions according to the nearest neighbors of a sample.

Visualizing the results of clustering techniques can improve interpretability as it shows intuitive differences among different clusters. Grimaldo and Novak used kNN and decision trees (DT) to predict building energy consumption on similar days. They then developed a smart energy dashboard visualizing energy consumption of similar days to help users understand the prediction results [51]. They also presented a radar chart to compare the similarity of weather parameters in the same prediction task [52].

2.1.2.4. Generalized additive models

Generalized additive models (GAMs) have gained increasing attention recently owing to their model interpretability. GAMs are a variant of generalized linear models that can model the non-linear additive effects of each feature [26]. The general structure of GAMs is defined as:

$$g(\mathbb{E}(y \mid \mathbf{x})) = \mathbf{w}_0 + f_1(\mathbf{x}_1) + \dots + f_i(\mathbf{x}_i) \quad (2.1)$$

where $g(\cdot)$ is the link function that connects the estimated mean $\mathbb{E}(y \mid \mathbf{x})$ to the sum of additive effects, \mathbf{w}_0 is the model intercept, and $f_i(\cdot)$ is the additive effect function (e.g., linear, cubic spline) for the feature \mathbf{x}_i to be estimated.

Compared with linear models, GAMs are more flexible and can incorporate irregular and volatile effects to improve flexibility in handling high-resolution data [53]. Bujalski and Madejski used GAMs to predict heat production in a combined heat and power plant system [54]. The results showed that ambient air temperature, solar radiation, and hour of the day had different impacts on the heating load. For example, outdoor air temperature showed a negative linear relation with heating load prediction, while solar radiation showed a negative exponential relationship. In addition, GAMs were also applied to identify operational patterns of HVAC systems [55] and perform sensitivity analysis of input features in thermal comfort modeling [56], thermal energy storage modeling [57], distributed PV power prediction [58], and short-term energy prediction in buildings [59].

2.1.3. Post-hoc approach

51 papers that adopted the post-hoc approach were reviewed in this section. In **Figure 2.6**, the Sankey diagram shows the connections in three dimensions, namely model dependency, post-hoc technique, and interpretability, along with the number of papers in each subdivided category. Because some studies adopted more than one post-hoc technique, the total number

of papers in each dimension is greater than the total number of papers adopting the post-hoc approach (i.e., 51).

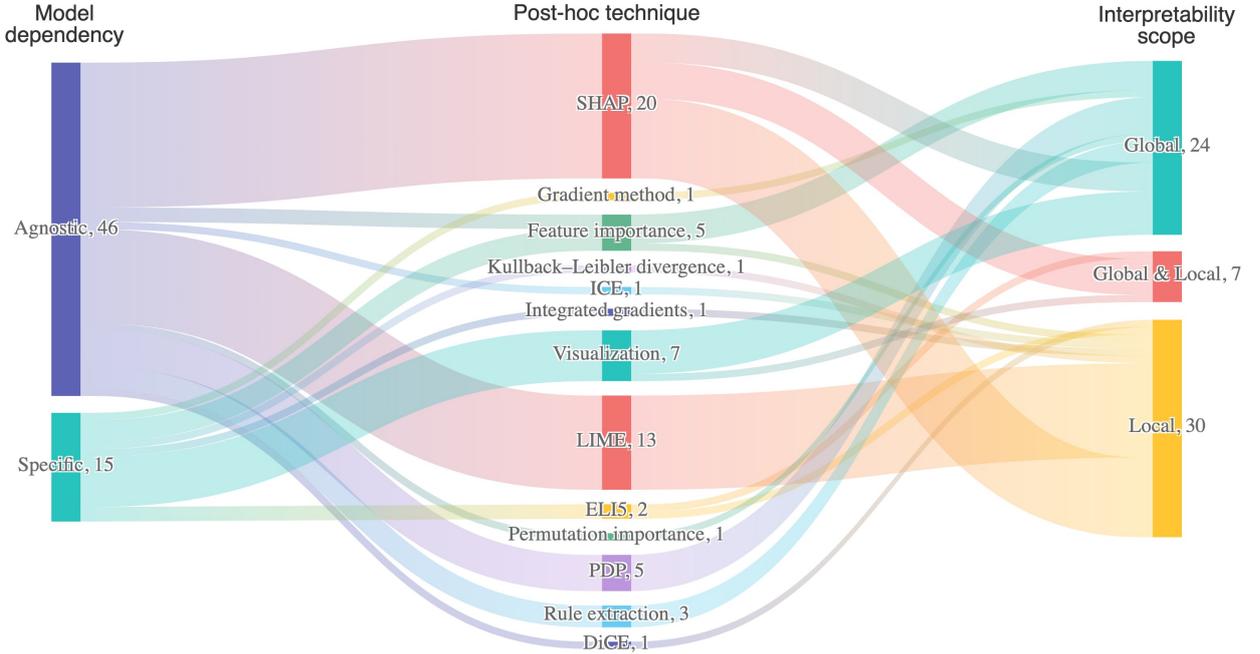


Figure 2.6. Sankey diagram depicting the connections of reviewed studies on different model dependencies, post-hoc techniques, and interpretability scopes.

2.1.3.1. Local interpretable model-agnostic explanations (LIME)

Local interpretable model-agnostic explanations (LIME) was proposed by Ribeiro et al. in 2016 as a model-agnostic approach to obtain local interpretation for individual predictions [60]. The local interpretation is obtained by training a local surrogate model to approximate the local characteristics of the black-box model in the region around the prediction sample. The interpretable model is obtained by optimizing the following objective $\xi(x)$:

$$\xi(x) = \operatorname{argmin}_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g) \quad (2.2)$$

where f is the black-box model, g is the local surrogate model from searching space G that defines the type of interpretable models such as linear or logistic models, π_x defines locality around data instance x , \mathcal{L} is a loss function that measures the fidelity of the surrogate model g to the black-box model f , and Ω measures the complexity of the surrogate model.

As LIME can give the contradict or support value of each input feature for a prediction sample, it is valuable to explain the prediction of classification problems. Wastensteiner et al. used LIME to interpret machine learning -based time-series classification for building energy consumption and analyzed the stability and reliability of the interpretation [61]. Madhikermi

et al. trained ANN and SVM for AHU fault diagnosis, and six samples were randomly selected to demonstrate the interpretability of LIME [62]. Srinivasan et al. experimented with interpreting three types of faults of chiller operation (i.e., scaling in condenser fins, sensor errors caused by pulsations in the flow, and false alarm) using LIME. On the one hand, decision-makers can know the foundation of the model output to support fault/normal decisions based on the contradict and support values given by LIME. On the other hand, LIME also provides information for the possible false alarms of the black-box model [63].

Apart from classification, LIME can also be applied to regression tasks. Fan et al. integrated the contradiction and support values into a single metric to evaluate the confidence level of a single prediction of chiller COP efficiency (i.e., low or high efficiency) [64]. Kotevska et al. used LIME to get the local linear approximation of the deep reinforcement learning (DRL) setpoint controller model. The results showed that the impact of zone temperature on setpoint recommendation varies in different ranges [65]. Zdravković et al. employed LIME to generate the local feature importance of prediction samples for heating demand prediction and anomaly detection in district heating systems [66,67]. Likewise, Arjunan et al. adopted LIME to improve the interpretability of the CatBoost model for building energy benchmarking [68]. In the case study, the authors gave an example of a building that consumed less energy than its peer group. According to the local interpretation provided by LIME, it was because the target building had a lower air-conditioned floor area. Jin et al. presented a LIME-based interpretable building energy benchmarking framework that could help evaluators understand the results [69]. For example, a building that consumes more energy than its peers would obtain a low score. Geyer et al. proposed a component-based methodology that predicted the heat flow of envelopes, heating/cooling demand, and final energy consumption by stages. DNN models were used for prediction in each stage, and LIME was employed to interpret the model output [70]. Besides, LIME was also used for other building management-related applications such as distributed PV power prediction [71], electricity demand prediction [61], and indoor CO₂ concentration prediction [72].

Although LIME is a model-agnostic technique suitable for any data-driven model, the interpretation obtained from LIME depends on data-driven models. Madhikermi et al. found that the interpretation of ANN and SVM models using LIME differs. For example, in FDD, the temperature of supply air after the heat recovery unit is the most influential feature for ANN, while the temperature of waste air is the most influential feature for SVM [62].

2.1.3.2. Shapley additive explanations (SHAP)

Shapley additive explanations (SHAP) is also a model-agnostic tool proposed by Lundberg and Lee in 2017 to interpret individual predictions [73]. SHAP computes Shapley values of each feature representing marginal contribution using a conditional expectation function. Although SHAP is designed as a local interpretability tool, the aggregation of Shapley values can be regarded as a global interpretation. For example, Carlsson et al. used the average of Shapley values as feature importance of the ANN model and found the most important features for energy consumption [74,75]. Ugwuanyi also used the average Shapley values for the global interpretation of CO₂ prediction [72].

Similar to LIME, SHAP is suitable for explaining the influential features of fault detection. In the study [76], SHAP generated the local and global interpretations of RF for FDD in district heating systems. Local interpretations revealed the influential features for individual prediction, while global interpretation showed the overall impact of each feature in the black-box model. Gao et al. used SHAP to interpret RF and LightGBM models for chiller FDD [77]. Santos et al. adopted XGBoost to detect fraud electricity consumption in the market, and SHAP was used to build interpretations for fraud activities afterward [78]. Additionally, SHAP can be used to interpret time-series classification for building energy consumption [61].

For building energy benchmarking, SHAP can determine the key features contributing to high or low energy usage intensity of individual buildings. In study [79], SHAP was used to interpret the XGBoost-based residential building energy benchmarking model in New York. According to SHAP values, unit density was the strongest predictor for energy use intensity of residential buildings in New York with the highest positive correlation, followed by property assessed value and number of floors. Arjunan et al. improved the interpretability of a conventional benchmarking method named EnergyStar by combining the XGBoost algorithm and SHAP interpretable machine learning framework [80].

Load/power prediction is the most popular application for SHAP. Chang et al. adopted SHAP to provide the interpretability analysis to reveal feature importance for PV power generation models (TS-SOM and XGBoost) [81]. Results showed that global horizontal irradiance for center value was the most influential feature, which was consistent with the Pearson correlation analysis. Movahedi and Derribea investigated the interpretation and interrelationship of three prediction models (electricity, water, and gas consumption) using SHAP [82], and results showed that the type of buildings (i.e., residential buildings or commercial buildings) and water

consumption were the most influential feature for electricity prediction. They also found that gas and water consumption were strongly interrelated because gas was used for water heating in target buildings. Bellahsen and Dagdougui used SHAP to rank the feature importance as global interpretation. The three most influential features were historical loads right ahead of the forecasting time, one day, and one week ahead of the forecasting time [83]. Results also showed that the RF model relied heavier on historical features instead of calendar features than other models. According to the SHAP values from the XGBoost model, Chakraborty et al. found that single-family homes were likely to have a more significant increase in building cooling energy consumption under global climate change [84]. Besides, buildings in hot-humid zones would consume more energy for cooling because of global warming. SHAP was adopted to interpret the performance-related indices (i.e., cooling capacity, COP, and wet/dew point efficiency) of a dew point cooler predicted by DNN in [85]. For example, a sample had a higher cooling capacity than the base value because of the relatively high intake air velocity. In [86], SHAP values showed that load and solar generation one-hour-ahead and the solar irradiance were the top three influential features for hourly ahead distributed PV power prediction. Similarly, Li and Wang summarized that day-ahead energy consumption was the most influential for daily load prediction [87].

2.1.3.3. Visualization and partial dependency plot

Visualization is a useful technique for users to build a better understanding of black-box models. The t-distributed stochastic neighbor embedding (t-SNE) creates two-dimensional projections for high-dimensional data using a non-linear transformation. Visualizing the embedding or hidden layer of neural networks using t-SNE has been widely adopted because it helps reveal the hidden mechanisms within neural networks. Kim and Cho added the state transition that can be visualized using t-SNE in the autoencoder model to improve the interpretability of electricity demand prediction results [88,89]. In [90], the authors visualized the latent states of autoencoders using t-SNE to explain the possible reasons for high or low energy consumption prediction. Singaravel et al. did similar research by visualizing the embedding layer of the CNN model to improve the understanding of building peak load prediction [91]. It was found that models with good generalization had higher separability than models with poor generalization when plotting using t-SNE.

Heatmap is another commonly used visualization tool that reveals the magnitude of a phenomenon in two dimensions. In [92], Kim and Cho analyzed class activation heatmaps to explore the influential features for load prediction. They found that one of the sub-metering

related to an electric water heater and an air conditioner was the most noteworthy feature. Based on a heatmap interpretation tool for DNN named Grad-CAM, Li et al. proposed a modified variant to obtain fault-discriminative information from the one-dimensional CNN for chiller fault diagnosis [93]. To improve the interpretability of LSTM-based electricity load prediction, Kim and Cho proposed a deep learning model that can visualize and analyze the correlation between latent variables and output. The results showed that the two latent variables had different time dependencies, i.e., short-term and long-term dependencies [94].

The partial dependency plot (PDP) is a visualization tool that generates global interpretations for black-box models. PDP measures the effect of a feature by averaging the marginal distribution of other features for the entire dataset. PDP shows the overall effect, whereas the individual conditional expectation (ICE) plot visualizes the impact of a feature for each sample. The limitation of PDP and ICE is that they assume that input features are uncorrelated. In the study [65], PDP indicating the global effect of input showed that indoor temperature was the most influential feature for setpoint recommendation in DRL. Additionally, the ICE plot revealed the feature variation impact of inputs by showing the control upper and lower bound. Overall, most interpretations from the DRL model were consistent with domain knowledge. Zhang et al. adopted PDP to assess the marginal impact of each input feature in the thermal comfort model [95]. They concluded that the marginal impact of each feature was different, and most features had a positive impact on PMV value. In the study [96], PDP was employed for feature importance, and the impact of floor area was much larger than building ID because the PDP curve of floor area had a larger variation. Mouakher et al. found that the dwelling type and the number of bedrooms were influential features for energy consumption prediction according to the PDP of the LSTM load prediction model [97].

2.2. Reliability of data-driven applications

2.2.1. Model generalization ability for fault detection and diagnosis

Over the past decade, FDD in HVAC systems has been widely studied, with three main approaches: expert rule-based, physical model-based, and machine learning-based approaches [98]. Expert rule-based FDD relies on domain knowledge to construct expert rules with predefined threshold values for decision-making [99,100]. This approach is easy to develop and implement but is less accurate due to oversimplicity [101]. Physical model-based FDD heavily relies on domain knowledge to develop physical models and fault indicators to measure the difference between the measurements and the indicators predicted by the physical models

[102]. Compared to the rule-based approach, the physical model-based approach is more reliable because it excels in dealing with dynamic operations. However, this approach requires detailed physical information about the system and is time-consuming to build and validate physical models, making it more information-demanding and labor-intensive. Machine learning-based FDD relies on historical data to develop supervised [103] or unsupervised models [104]. Compared to expert rule-based and physical model-based approaches, the machine learning-based approach is easier to develop because it utilizes historical operational data and does not require detailed information about the HVAC system. FDD is usually treated as a classification task in machine learning, which can be binary (for fault detection) or multi-class classification (for fault diagnosis) [105]. Binary classification detects whether a working condition is normal or faulty [106], while multi-class classification aims to determine whether the working condition is normal or suffers various faults [107,108]. The machine learning-based approach can address performance shifts of equipment during long-term operation by updating the model automatically, which is hardly achievable using expert rule-based or physical model-based approaches [109]. Previous studies have developed machine learning-based models using popular algorithms such as support vector machines [110,111], decision trees [112,113], artificial neural networks [114,115], and deep neural networks [116] for FDD in HVAC systems [98]. These studies demonstrated high accuracy and exemplified the great potential of the machine learning-based approach in FDD, given adequate labeled data, including both normal and faulty data.

However, the practical deployment of machine learning-based FDD encounters significant challenges because the labeled data is often limited in real-world scenarios. On the one hand, conventional machine learning algorithms usually require a large amount of labeled data to prevent overfitting due to the curse of dimensionality in machine learning (the demand for labeled data increases exponentially with dimensionality) [117]. On the other hand, faulty data usually come from limited maintenance records, which are insufficient to train and validate data-driven models. Therefore, it is important to develop a new FDD method with high generalization ability for HVAC systems in which only a few labeled samples are available. To improve the performance of machine learning-based FDD under limited labeled data, semi-supervised learning methods have recently been adopted for FDD in HVAC systems. Semi-supervised learning enhances the performance of supervised learning by leveraging large amounts of unlabeled data with high confidence scores [118]. The confidence score of an unlabeled sample is the probability associated with the output class from supervised models.

In [25], the support vector machine (SVM) was employed to calculate the confidence score. If the predicted class of an unlabeled sample has a higher probability than a predefined threshold, the sample will be assigned with the class label and added to the training set for the next round of training. By adopting the semi-supervised SVM, the results showed above 80% accuracy on the test data for AHU fault diagnosis. Similarly, a generative adversarial network (GAN)-based semi-supervised learning framework was proposed to improve the chiller fault diagnostic accuracy under limited and imbalanced labeled data [119]. The proposed framework could generate artificial fault samples to balance the training data, and the fault diagnostic accuracy reached 90% when each fault type had 30 samples. Li et al. compared GAN-based semi-supervised learning with a supervised baseline [120,121], and the proposed semi-supervised method showed a 3%–10% improvement in FDD accuracy. It was also found that when the number of labeled samples decreased, the proposed method gained more accuracy improvement. Fan et al. designed case studies to test the generalization ability of semi-supervised learning in detecting unseen faults in AHU operations [122,123]. The results showed that semi-supervised learning improved the fault detection rate by about 10% when the size of labeled data is small. The performance of semi-supervised learning gradually approached the baseline with the increase of labeled data.

Similarity learning is a novel supervised learning method for classification problems to enhance the generalization ability of classification problems, and it has been successfully adopted in a variety of FDD applications, such as bearing fault diagnosis [124,125], power system fault diagnosis [126,127], robot fault diagnosis [128,129], etc. Similarity learning measures the similarity between the new samples and labeled samples to make FDD predictions. Conventionally, the similarity of a pair of samples can be calculated in two ways. First, similarity can be calculated from the geometric differences, such as Euclidean distance [130,131], Manhattan distance [132], and cosine distance [133]. Second, correlation coefficients such as Pearson correlation coefficient [134] and Spearman's rank coefficient [135] can serve as similarity metrics. On the other hand, similarity learning learns the similarity by training neural networks in a supervised learning way. The distance-based and correlation-based similarity measures work well when used to compare univariate time-series samples. However, these similarity metrics cannot reasonably measure the similarity between two multivariate time-series samples because it is difficult to determine the weights of each variable [136]. Despite the success of similarity learning in many FDD applications, only a few studies have adopted similarity learning in building energy management. Tan et al. proposed a sensor

fusion framework for detecting occupancy in residential buildings [137]. In the proposed framework, similarity learning was adopted to classify indoor image data into two classes, i.e., occupied and vacant. The results proved the value of similarity learning, especially given limited labeled data. In view of the boom of deep learning and the success of similarity learning in various industries, similarity learning is a promising tool for detecting and diagnosing faults in building HVAC systems, particularly when labeled data is limited.

Another issue in previous studies on using machine learning algorithms to analyze time-series data is the neglect of the high autocorrelation in the time-series data, which often causes biased evaluation or overestimating model performance. Autocorrelation describes the degree of correlation of the same variables between two successive observations. For example, the supply air temperature of AHU at any time should be very close to the previous observation (e.g., the temperature a minute ago) [138]. When developing data-driven models for non-time-series data, training and test data are usually split randomly because data samples are usually independent of each other. However, when such a random split strategy is applied to time-series data, the actual performance of data-driven models might be exaggerated because of the high autocorrelation in the time-series data [139]. For example, if the random split ratio of training/test time-series data is 1:1, it means that every test sample can find similar neighbors in the training data. In this way, the test data have almost identical distribution with the training data, and the out-of-distribution generalization ability of the proposed methods may not be as good as that shown by test results. Therefore, the temporal train-test split is a preferable method to validate the out-of-distribution generalization ability of the proposed methods [140].

2.2.2. Robustness and stability for optimal control

2.2.2.1. Measurement uncertainty in measurement-based control problems

Due to the variability and seasonal patterns of cooling load, multiple-chiller systems are popular in large commercial buildings [141,142]. Chiller sequencing control plays a critical role in achieving the balance of cooling demand and supply because it determines the number of operating chillers (and associated pumps and cooling towers) and switches chillers on/off frequently in response to the measured cooling load. Chiller sequencing control strategies should provide buildings with sufficient cooling while minimizing energy consumption. The unnecessary chillers could meet the cooling load but consume extra electricity because of interlocked pumps and cooling towers. Insufficient operation of chillers cannot provide enough cooling and cause thermal discomfort. The significance of chiller sequencing control lies in its

ability to balance multiple conflicting objectives, such as energy consumption, chiller capacity utilization, and maintenance schedules, and to effectively handle real-world complexities, such as variable cooling loads, chiller failures, and maintenance schedules [143].

In the past few decades, various chiller sequencing control strategies have been developed and successfully implemented in chiller plants, employing either a direct or an indirect indicator of the building's real-time cooling load [143]. The direct approach is also called total cooling load-based (Q-based) sequencing control, which calculates the real-time cooling load from the measured chilled water supply and return temperatures and flow rate [144]. A chiller will be turned on if the measured cooling load exceeds the switching-on threshold, and one of the operating chillers will be turned off if the measured cooling load is below the switching-off threshold [145]. Usually, a dead band and minimum switching time are set to prevent frequent switching of chillers when the cooling load fluctuates around the threshold. The indirect approach uses an indicator of the operating chillers' loading ratio compared with their total cooling capacity, such as chiller water return temperature, electric power and current, and bypass water flow, to determine whether a chiller should be switched on or off.

Owing to its reliability, the direct approach is more commonly used in multiple-chiller systems [146,147]. However, the uncertainties associated with flow and temperature measurements in practical applications may result in a lack of robustness in control decisions [148]. Ensuring the accuracy of flow measurements in practical installations can be challenging due to sensor faults, data transmission problems, and other factors. This challenge is particularly pronounced after a certain period of operation, as it involves difficulties associated with site calibration and sensor replacement [149]. The problem of uncertainties with temperature measurements is particularly challenging and critical because the temperature difference between the supply and return temperatures of the chilled water is usually slight (around 5 K), which means a slight disturbance to the temperature measurements may cause a significant deviation of the measured cooling load. To deal with the flow measurement uncertainty, Sun et al. [150] developed a model of flow measurements using Bayesian inference and Markov chain Monte Carlo methods to make uncertainty-tolerant chiller sequencing decisions. Liu et al. [151] considered the uncertainty of maximum cooling capacity to develop a more robust chiller sequencing strategy.

A survey shows that *system reliability and stability* is the second most crucial criterion, following *total energy use* as a criterion when evaluating intelligent HVAC control systems [152]. Good system reliability and stability are crucial for reducing maintenance costs and

extending the life of service equipment. Frequently switching chillers may damage the chiller plant's service life and system performance [153]. Despite efforts to improve the robustness of chiller sequencing control by considering the uncertainties of measurements and chiller maximum capacity, these conventional strategies seldom consider cooling load fluctuation in the short term. Even though the measured cooling load is accurate and a dead band is set, the cooling load continuously varies with the changing operation conditions, including the outdoor air temperature, relative humidity, solar radiation, and indoor occupancy [154]. The load fluctuations in a short period, say 30 minutes, may cause unnecessary on/off switching. In such a case, the automatic chiller sequencing strategy may become unreliable and unstable [155].

In recent years, research has been conducted to enhance the robustness of chiller sequencing by utilizing cooling load predictions. Liao and Huang [13] proposed a hybrid predictive chiller sequencing control strategy by predicting the cooling load in the following hours using an autoregressive with exogenous (ARX) model. The case studies showed that the total switch number was reduced by 20%. However, this hybrid strategy ignores the uncertainty with cooling load prediction [156]. Uncertainty with data-driven models is a broad concern. Research showed that uncertainties within the data used to train the model would propagate and eventually significantly increase the uncertainty of the predictions [157]. Therefore, the uncertainty of cooling load predictions should be considered to make more robust sequencing actions.

Because both the direct and indirect approaches are based on real-time measurements, there are some major challenges in practical applications. Firstly, real-time measurements are ever-changing [158,159]. Therefore, even with a minimum control interval and a dead band, direct and indirect approaches may result in unnecessary on/off switching actions due to short-term measurement fluctuations. For example, when the terminals in a building, such as fan coil units (FCU) or variable air volume (VAV) terminals, are suddenly turned on in the morning to remove accumulated heat gain during the night, it can lead to a sudden increase in the cooling load of the entire chiller plant [160]. As a result, the cooling load may exceed the threshold for switching on an additional chiller, particularly if the chillers are already operating at a high PLR. Unlike the overall building cooling demand increase due to weather changes [161,162], this sudden increase in cooling load is a short-term fluctuation and usually does not require an additional chiller. Therefore, it is necessary to predict short-term cooling loads to help achieve more robust sequencing control [163]. Liao and Huang [13] employed an autoregressive with exogenous (ARX) prediction model and verified its contribution to chiller sequencing control,

aiming to minimize unnecessary sequencing actions. Huang et al. [144] also emphasized the value of cooling load prediction for optimal chiller sequencing control. Chen et al. [164] proposed a robust chiller sequencing control strategy, which evaluates the risks of possible chiller sequencing actions based on probabilistic cooling load predictions.

Another challenge for the applicability of chiller sequencing strategies based on real-time measurement is that both the measured values and the chiller switching thresholds can deviate from the actual values. For example, the most common Q-based strategy requires real-time measurements, including chilled water supply temperature, chilled water return temperature, and chilled water flow rate. There can be random uncertainty and systematic biases in flow and temperature sensors [150]. The random uncertainty can be handled by data preprocessing techniques, such as data smoothing and outlier detection [165]. Systematic biases are usually caused by sensor drift, requiring sensor recalibration to reduce it [166]. Furthermore, the switching on/off thresholds in the Q-based strategy are typically set to the rated maximum cooling capacity of chillers, assuming that the maximum cooling capacity of a chiller is constant. However, in practice, one of the main challenges of implementing such strategies is that the maximum cooling capacity of the chiller can vary with the chiller's operating conditions [167].

Due to the variation of the maximum cooling capacity, the chiller sequencing actions based on the fixed switching on/off thresholds can be inappropriate even if the cooling load is accurately measured. The existing research is mainly based on simulation platforms such as EnergyPlus [168] and TRNSYS [169], or validation using historical data. The assumption that the chiller's maximum cooling capacity is a known fixed value, typically the chiller's rated cooling capacity, is reasonable on simulation platforms. However, underestimating or overestimating the chiller's maximum cooling capacity may result in energy waste or affect end-user comfort. For example, underestimating the current operating chiller's maximum cooling capacity may lead to energy waste from running one or more surplus chillers [170]. On the other hand, overestimating the current operating chiller's maximum cooling capacity may result in insufficient cooling supply, affecting occupants' comfort. Liu et al. [151] propose a more accurate chiller model considering the variation of chiller maximum cooling capacity. The simulation results showed that the proposed chiller model had better robustness in chiller sequencing control.

2.2.2.2. Physical Inconsistency issue in data-driven equipment modeling: Variable speed pump as an example

Variable frequency drives (VFDs) convert sinusoidal power from the grid, which has a constant rated frequency and voltage, into power with variable frequency and pulse width modulation, which is then supplied to connected motors [171]. As the VFD changes the frequency of power supply to the motor, the pump spinning speed varies proportionally [172]. Theoretically, the energy saving of reducing pump speed is significant because power consumption varies with the cube of the pump speed [173]. With the increasing adoption of VFDs, more HVAC chilled water systems are incorporating variable speed pumps (VSPs) to save energy consumed by pumps when the chilled water system's flow rate falls below the design flow rate. Typically, conventional differential pressure control is used to ensure fixed differential pressure at both ends of the cooling coil in the chilled water loop (i.e., differential pressure setpoint) [174]. To further save pump energy and provide greater flexibility, many large chilled water systems employ variable speed parallel pumping [175]. The operation of multiple identical VSPs is controlled by conventional differential pressure control and rule-based pump sequencing. All VFDs controlling the secondary pump motors receive identical commands, which are adjusted to maintain the differential pressure setpoint. As for the sequencing control, an additional VSP is activated when the operating pumps' frequencies reach their maximum (e.g., 50 Hz), and one of the active VSPs is deactivated when the operating pumps' frequencies reach their minimum (e.g., 30 Hz) [176].

Although variable speed parallel pumping has better energy efficiency and flexibility, the conventional rule-based pump sequencing strategy only ensures the need for differential pressure setpoint without considering the optimum energy efficiency for VSPs. Therefore, huge energy savings can be achieved by optimizing the sequencing control of parallel pumps. Optimal sequencing control resolves the optimum number of operating pumps at any particular time and their corresponding speeds, which will ensure minimum energy consumption while satisfying both the thermal and hydraulic demands of the water system.

Wire-to-water efficiency, considered an effective tool for optimizing the sequencing control of parallel pumps, takes into account the entire energy conversion process from the electrical input (wire) to the hydraulic output (water) [177]. This comprehensive approach considers the combined efficiencies of the pump, motor, and VFD, allowing engineers and operators to evaluate and optimize the total pumping installation's performance [178]. When calculating the efficiency of a pump, manufacturers typically provide an efficiency curve at full speed. This

curve can be adjusted for partial speeds using the affinity laws. For motors and VFDs, efficiency curves from manufacturers or literature are commonly used. The optimal number of pumps for current working conditions can be determined by computing the wire-to-water efficiency under different pump operating conditions. The head-flow (H-Q) chart identifies the working condition area for each optimal number of operating pumps. Rishel et al. extensively researched optimizing the operation of parallel variable speed pumps in HVAC pumping systems using wire-to-water efficiency [177–179]. They aimed to sequence these pumps effectively based on wire-to-water efficiency or kW input to the pumping system. The core principle of their approach was to operate the pumps as closely as possible to their optimal efficiency curves.

The applicability of wire-to-water efficiency faces challenges in two aspects. Firstly, this method heavily relies on motor and VFD efficiency curves from manufacturer data or literature. Due to varying installation conditions, the actual pump efficiency curve may differ significantly from the theoretical curve. Moreover, motor and VFD efficiency curves can vary with manufacturer models, and for those without provided curves, reference curves from literature are typically used. Therefore, the combined efficiencies of the pump, motor, and VFD often struggle to accurately reflect the operation of parallel variable speed pumps, significantly limiting the practical application of Wire-to-water efficiency in engineering. Wang et al. proposed a motor equivalent circuit approach for calculating the motor efficiency under variable frequencies [180]. Secondly, the efficiency of parallel variable pumping systems deteriorates over time due to wear and tear, a factor that Wire-to-water efficiency cannot capture. The efficiency of motors could decrease by as much as 5%, while the efficiency of pumps could drop by up to 25% [181]. Bernier and Bourret analyzed how decreasing pump speed affects motor and VFD efficiencies over time [182]. It was found that the power needed at the inlet of a pump-motor-VFD system was notably higher, particularly for oversized motors, compared to the expected power using pump affinity laws.

The wide adoption of the Internet of things (IoT) and advanced measurement instruments technologies have led to an explosion of data available from HVAC systems. This data can be harnessed to develop data-driven control strategies that learn and adapt over time. Data-driven approaches offer the potential to overcome some of the limitations of the conventional pump sequencing methods by learning from real-time data. Wang developed data-driven models for modeling the efficiency of the motor and VFD together [183]. Results showed that the VFD input frequency and the motor shaft power can be used to regress the motor and VFD efficiency.

Jepsen et al. developed a data-driven pump model using the pump head and flow rate and validated the proposed model in a simple open-loop water system [184]. These data-driven methods require the accurate measurement of the pump head. However, the differential pressure between the discharge and suction pipes is the static head instead of the total head of the pump.

2.2.2.3. Unreliable control for deploying data-driven models in complex optimization problems

Optimization algorithms play an important role throughout the life cycle of buildings, especially in the design and operation stages. Optimization algorithms find the optimal solution for a given optimization objective with constraints, such as minimizing total energy consumption to maximize economic benefits [185]. Two types of variables need to be optimized: integer and continuous [186]. Integer variables represent the on/off status or the number of devices to be operated. Moreover, continuous variables can be device sizing and setpoints, depending on the specific optimization problems.

Optimization algorithms are widely used in building energy systems for optimal design, demand response, and optimal control [187]. Optimal design for building energy systems determines the optimal configuration of building energy systems, such as HVAC systems, energy storage, and distributed renewable energy based on the heating and cooling demand, occupancy patterns, utility prices, and renewable energy resources [188,189]. The optimal configuration of building energy systems ensures the ideal match between energy demand and supply. During the operation stage, optimization algorithms can take advantage of the energy flexibility from building energy systems to interact with the grid, such as peak load reduction [8,190]. Furthermore, optimization algorithms can be used for optimal control in building energy systems, such as non-predictive model-based control [191] and model-predictive control [15]. The input of models is control signals and system states. Moreover, the output can be energy use or carbon emissions for the building energy systems. White-box, grey-box, or black-box modeling methods can be used for optimal control, depending on the availability of operational data and detailed information about the building energy systems [154].

According to the processes to find the optimal solution, optimization algorithms can be classified into deterministic and metaheuristic algorithms [185]. Deterministic algorithms find the optimal solution through a specific set of mathematical rules [192]. For example, the branch and bound (B&B) algorithm works by exploring the space of possible solutions and pruning

away branches that cannot lead to the optimal solution [193]. Metaheuristic algorithms, on the other hand, are non-deterministic and stochastic [194]. Metaheuristic algorithms are often inspired by nature or physics to find optimal solutions. For example, particle swarm optimization (PSO) mimics the social behavior observed in birds flocking or fish schooling [195].

Table 2.1 presents a comparison of deterministic algorithms and metaheuristic algorithms. Deterministic algorithms are typically tailored for specific problems and can find the optimal solution when it exists. Unlike metaheuristic algorithms, deterministic algorithms are not influenced by randomness [196]. For a given optimization problem and constraints, the solution process is deterministic, and deterministic algorithms always produce the same output. However, using deterministic algorithms for online optimal control in building energy systems can cause stability issues. As the working conditions vary, the objective function and constraints of the optimization problem also vary over time. As a result, uncertainties in sensor measurements or fluctuations in building energy system operational measurements can lead to unstable control actions [150,164]. Fan et al. [197] applied the branch and bound algorithm to optimal chiller loading, and the results showed that the chillers switched on and off frequently due to fluctuations in cooling load. To enhance control stability, Fan et al. divided the cooling load into intervals of 10 and 20, respectively, and applied the branch and bound algorithm to only the maximum cooling load within each interval for optimal chiller loading. While the improved algorithm effectively enhanced control stability, it also increased energy consumption by 5%. Furthermore, as the number of intervals increased, control stability was compromised, indicating a tradeoff between control stability and energy efficiency.

Table 2.1. Comparison of deterministic algorithms and metaheuristic algorithms

Aspect	Deterministic algorithms	Metaheuristic algorithms
Solution quality	An optimal solution is guaranteed if one exists for specific problem types.	Near-optimal solution
Randomness	No randomness involved; the process is predictable.	Random processes are involved to escape local optima and explore the solution space.
Flexibility	Rigid and problem-specific	More flexible and adaptable to various types of problems
Examples	Branch and bound; Lagrange method	Particle swarm optimization; Genetic algorithms; Simulated annealing

Owing to the flexibility in handling various optimization problems, metaheuristic algorithms are increasingly being used to solve optimization problems in building energy systems. A recent literature review revealed that metaheuristics have been employed in 61% of studies related to the control of water-cooled central cooling systems [198]. Metaheuristic algorithms are generally divided into two categories based on their search processes: biology-based and physics-based. The most popular biology-based algorithms include particle swarm optimization (PSO) [199], genetic algorithm (GA) [200], and differential evolution (DE) [24], while simulated annealing (SA) [202] is the preferred physics-based algorithm. The solutions produced by metaheuristic algorithms vary due to random initialization and other random explorations, whether biology-based or physics-based, as shown in **Figure 2.7**. Metaheuristic algorithms do not guarantee an optimal solution but aim to find a satisfactory solution, i.e., a near-optimal solution. Therefore, metaheuristic algorithms often struggle to find only local optima within a limited search time. As shown in **Figure 2.8**, to maximize the well-known peaks function in MATLAB [203], there are two local optima and one global optimum. Even if the optimization problem remains unchanged, the solutions could shift from one local optimum to another local optimum if the random seed varies. Therefore, even when the working conditions of building energy systems are stable and the optimization problem remains unchanged, control actions derived from metaheuristic algorithms can still be unstable.

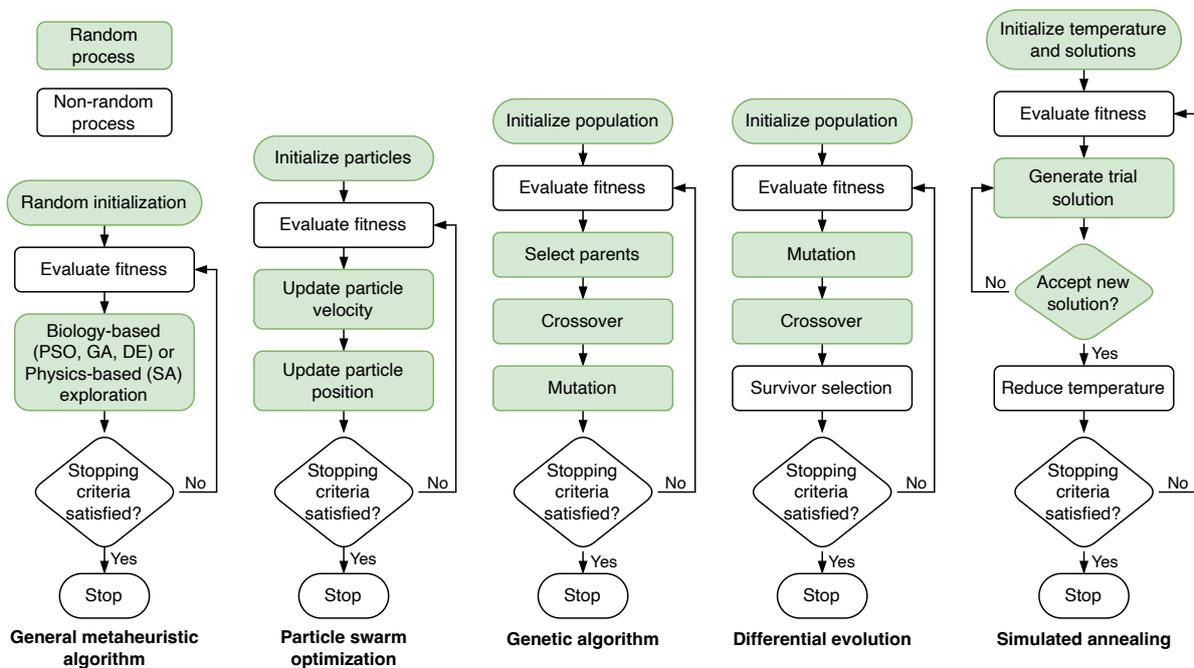


Figure 2.7. Random processes in metaheuristic algorithms

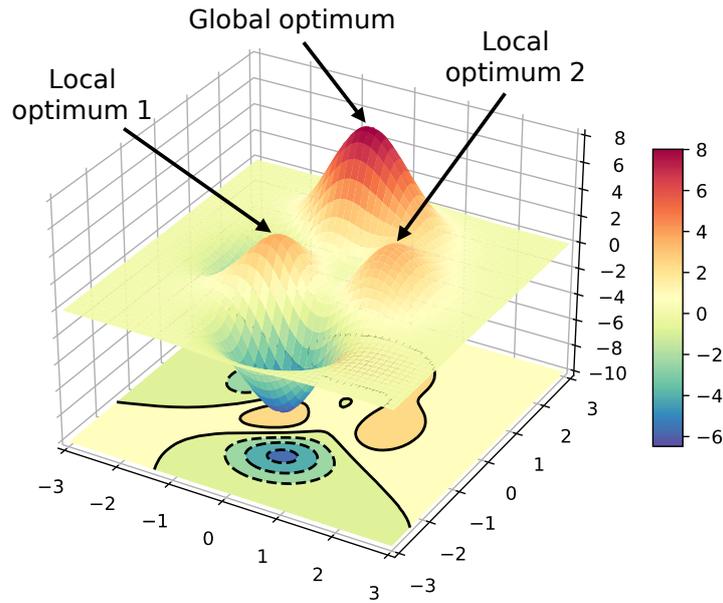


Figure 2.8. Local optima in optimization problems

To enhance the stability of GA for online optimal control, Ma and Wang [191] proposed a strategy that compares the energy consumption of the new optimal control action from GA to the adopted control action in the previous control period, updating the control action only if the energy savings from the new action exceeded 1%. Although this strategy improves the stability of control actions by somewhat sacrificing energy efficiency, it neglects the smoothness between successive control actions. Sun et al. [204] proposed a multiplexed optimization strategy that conducts an exhaustive search on each variable once a time within the control interval to avoid the dramatic variation in control actions. The results showed that multiplexed optimization improved control stability while achieving nearly the same energy efficiency as GA. However, the proposed multiplexed optimization is limited to optimizing continuous variables and does not extend to integer variables, precluding it from achieving optimal sequencing control for devices such as chillers, pumps, and cooling towers.

2.3. Summary of research gaps

Based on the literature review, the following research gaps have been identified for the four typical data-driven modeling applications: fault detection and diagnosis, cooling load prediction-based chiller sequence control, sequence control for parallel variable speed pumps, and model-based optimal control. Further research and development are needed to enhance the effectiveness and practical applicability of data-driven models and control strategies in building energy systems by improving interpretability and reliability:

1. **Poor generalization ability under data scarcity.** There is a lack of interpretability and generalization in conventional data-driven models due to the probability output and limited labeled data. Typical HVAC systems usually have limited labeled or unbalanced labeled data for training, which challenges the reliability of conventional data-driven models. In addition, existing data-driven models often do not address the high correlation between training and test data, leading to unreliable diagnostic results.
2. **Poor control reliability under measurement uncertainty.** Current point prediction models do not consider the uncertainty of measurement and output and, therefore, have poor interpretability and reliability for control problems such as chiller sequencing control. There is a need for robust chiller sequencing control strategies that can handle the uncertainty of cooling load predictions. Moreover, there is a lack of in-situ tests to validate the applicability in actual chiller plant operations.
3. **Physical Inconsistency issue in data-driven equipment modeling: Variable speed pump as an example.** Optimal pump sequencing is often overlooked despite the wide adoption of variable speed parallel pumping. There is a need for a control strategy that integrates physical knowledge with data-driven modeling to enhance interpretability and reliability. In addition, validating the proposed strategy with real building operational data is necessary to ensure its practical feasibility.
4. **Unreliable control for deploying data-driven models in complex optimization problems.** Existing optimization algorithms are not tailored for online optimal control, and their stability is rarely considered. Deterministic algorithms may suffer stability issues due to variations in the optimization problem during online control. Metaheuristic algorithms' solutions may be inconsistent due to their stochastic nature and exploration processes. There is a need for a novel approach to balance energy saving and stability in model-based online control.

CHAPTER 3 A SIMILARITY-LEARNING METHOD FOR ENHANCING MODEL GENERALIZATION ABILITY AND INTERPRETABILITY

This chapter proposes a similarity learning - based method to solve the problem of **data scarcity**, enhancing interpretability and generalization ability. Conventional data-driven models for FDD simply output the probability of fault, which lacks interpretability and generalization ability under limited labeled data. To this end, this study proposes a similarity learning-based fault detection and diagnosis method for building HVAC systems. The output is the similarity (ranging from 0 to 1) towards each fault, which provides better interpretability like the typical high interpretable k -nearest neighbors (k -NN) method. The temporal data-splitting method is adopted to tackle the issue of the high correlation of training data and test data when the random split is adopted. Two case studies were conducted to test the effectiveness of the proposed method, with insufficient labeled data and imbalanced data, respectively. The remainder of this chapter is organized as follows. Section 3.1 describes the research methodology developed to address the gaps in current research. Section 3.2 describes the two case studies conducted utilizing an open dataset for AHU FDD. The results of the two case studies are presented in Section 3.3, which includes in-depth discussions. Finally, Section 3.4 concludes the main findings of this study.

3.1. Proposed similarity learning-based fault diagnosis method

3.1.1. Overview of the proposed method

In conventional supervised multi-classification tasks, the input is a single sample X , and the desired output of the classification function $class(\cdot)$ is the class y to which X belongs, as shown in Eq. (3.1). If neural networks are used as the classification function, the raw output is the probabilities of the input belonging to each class given by a Softmax activation function. The final classification result is obtained by selecting the class with the maximum probability [205]. Cross-entropy is the most used loss function to train a classification task. Similarity learning differs from conventional supervised learning by transforming the multi-class classification problem into a binary one. Therefore, the loss function used in similarity learning is different. Binary cross-entropy and triplet loss are common loss functions for similarity learning [206]. The classification function of similarity learning $sim(\cdot)$ is formulated in Eq.

(3.2). The similarity function examines the similarity of a pair of input samples, i.e., X_1, X_2 to determine if the two samples belong to different classes. If yes, the desired output of the similarity function is 0; otherwise, the desired output is 1. When the similarity function is obtained, the similarity between new samples and labeled samples can be measured to make FDD predictions. In similarity learning, the labeled samples can be paired randomly to generate more input pairs for training than using the labeled samples alone. Therefore, the advantage of similarity learning over conventional supervised classification is the larger amount of training data to overcome the problem of limited labeled data and improve the FDD method's generalization ability. This makes similarity learning well-suited to tasks where labeled data are limited or challenging to collect.

$$class(X) = y, y \in \{0, 1, \dots, n\} \quad (3.1)$$

$$sim(X_1, X_2) = \begin{cases} 0, & X_1, X_2 \text{ belong to different classes} \\ 1, & X_1, X_2 \text{ belong to the same class} \end{cases} \quad (3.2)$$

Similarity learning can adopt various architectures and machine learning algorithms to formulate the similarity function, such as Siamese networks [136] and kernel-driven methods [207]. Siamese networks (also called twin neural networks) are widely adopted in similarity learning consisting of two identical neural networks as subnetworks that share the same structure and weights [208]. Siamese networks are typically used for learning the similarity between a pair of samples as input, such as two images or two speech records [209].

The workflow of the proposed similarity learning-based FDD method using Siamese networks is shown in **Figure 3.1**, which consists of two tasks: model training and fault diagnosis. During model training, input pairs are first randomly generated from training data. The various color in **Figure 3.1** denotes the raw label in the dataset (i.e., Normal, Fault 1, ...). The output of the Siamese networks is the similarity of the two samples of an input pair. If the input pair belongs to the same class, the target output is 1. Otherwise, the target output is 0. Finally, the loss function used to update the Siamese networks is binary cross-entropy loss. During fault diagnosis, the test sample is paired with a support set and outputs the most similar class using the trained Siamese networks. Note that the proposed method is applicable to fault detection when the data are only labeled as normal and faulty.

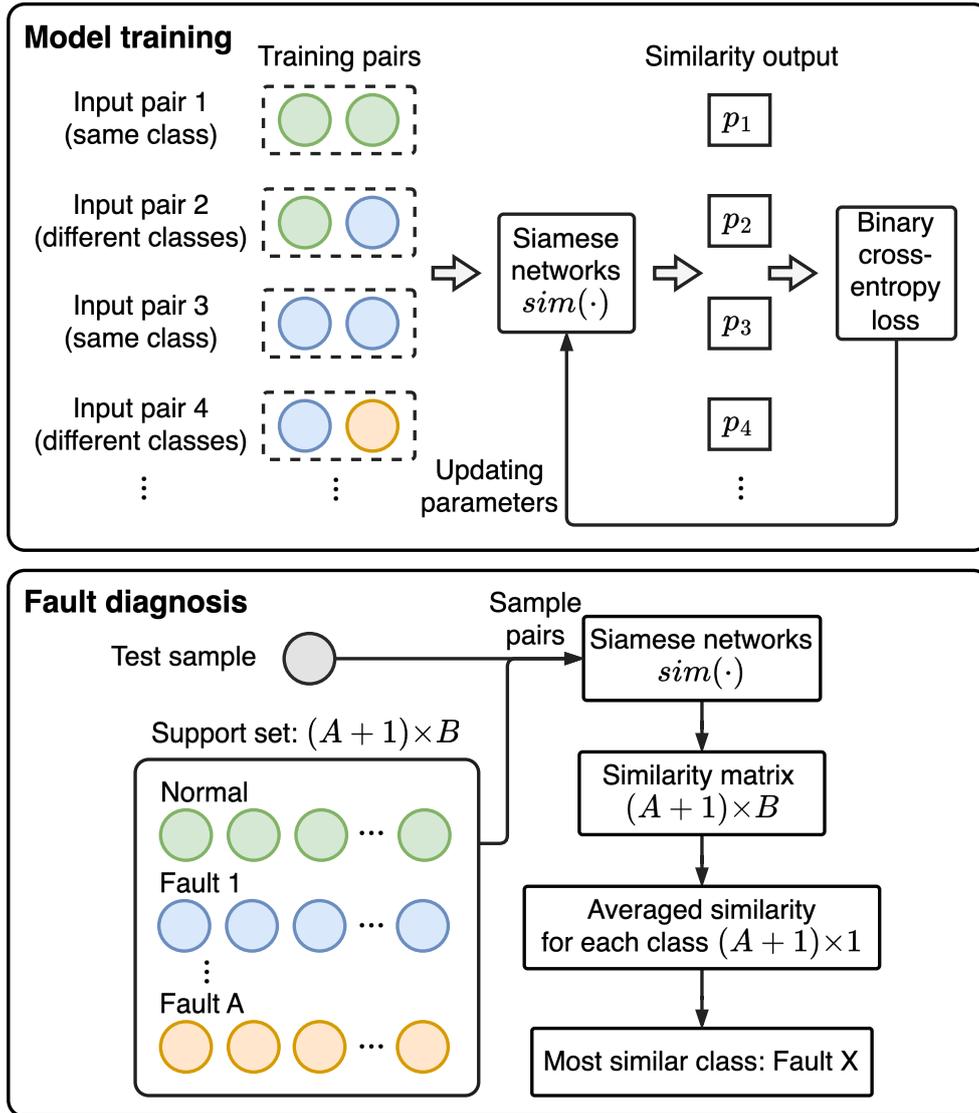


Figure 3.1. The proposed similarity learning-based FDD method using Siamese networks

3.1.2. Model training

The structure of the proposed Siamese network is shown in **Figure 3.2**, which contains two identical long short-term memory (LSTM) subnetworks. The input is a pair of multivariate time-series samples, and the output is the similarity of the input pair, ranging from 0 to 1.

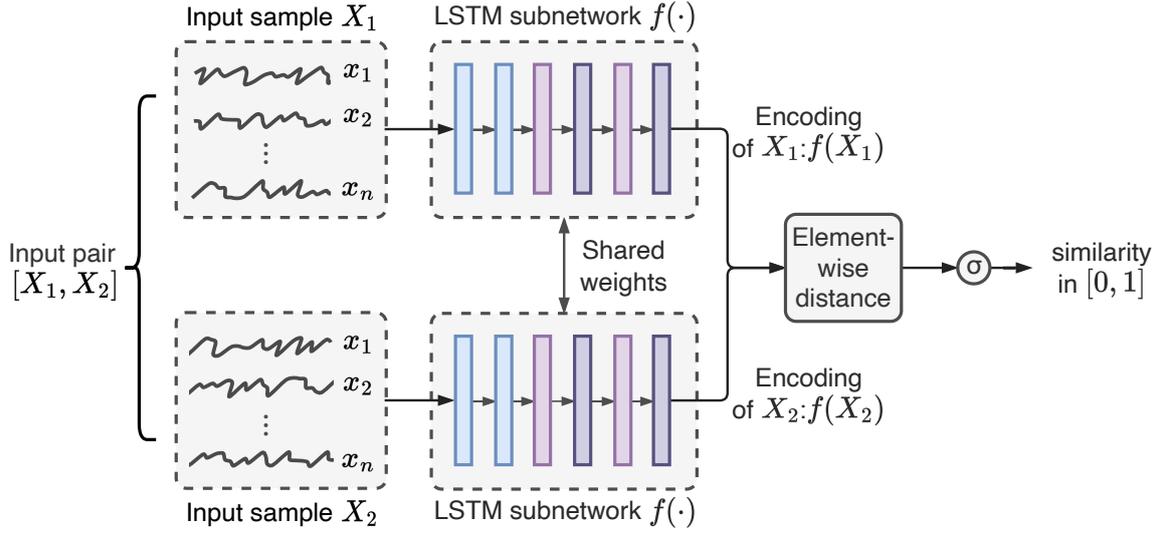


Figure 3.2. Structure of the proposed Siamese network

3.1.2.1. Input pair generation

The input samples X_1, X_2 are a pair of multivariate time-series sampled from historical operation data using a moving window. **Figure 3.3** exemplified how samples are obtained from a labeled multivariate time-series (e.g., labeled as Normal, Fault 1, Fault 2...). There are n features in the raw multivariate time-series, and the window samples data by moving itself from left to right. The window size m refers to the length of each sample, and the window stride is the number of time steps by which the window is shifted each time. The fault labels for the samples are identical to the label of the raw time-series. After all raw multivariate time-series are sampled, different samples with the shape of $m \times n$ are obtained.

Next, pairs of similar and dissimilar samples are generated by randomly pairing the above samples. As defined in Eq. (3.2), if both samples in the pair belong to the same class, the pair is labeled as similar ("1"). Otherwise, the pair is labeled as dissimilar ("0"). The number of similar and dissimilar pairs in the training dataset should be roughly the same to ensure that the distribution of similar and dissimilar pairs is balanced when creating sample pairs for training the Siamese network. If the distribution is imbalanced, the networks may not be able to learn the differences between the two classes of samples effectively. A binary random variable that can be 1 or 0 with equal probability generates similar or dissimilar pairs to achieve the balance.

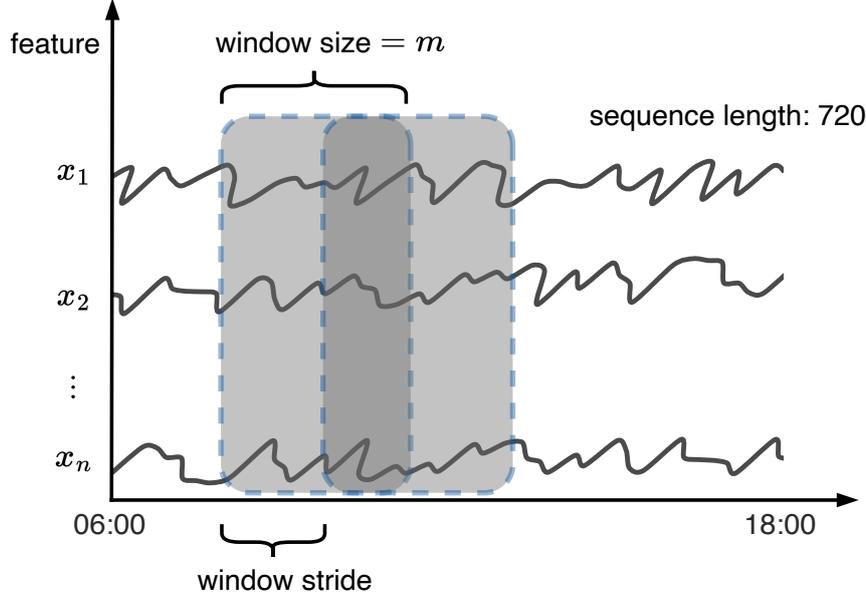


Figure 3.3. Moving window sampling for multivariate time-series

3.1.2.2. LSTM subnetworks for encoding multivariate time-series

Once a set of labeled pairs has been generated, Siamese networks can be trained by providing each pair to the LSTM subnetworks as input and using the labels to compute the loss. The Siamese networks learn to compare the two input samples and predict whether they are similar or dissimilar based on their features and learned relationships. The architecture of LSTM subnetworks is shown in **Table 3.1**, consisting of two LSTM layers, two one-dimensional batch normalization layers, and two fully connected layers.

Table 3.1. The architecture of LSTM subnetworks in the proposed Siamese network

Layer	Parameters	Output size
LSTM layer 1	hidden_size: 75	$m \times 75$
LSTM layer 2	hidden_size: 50	50
BatchNorm1d layer 1	num_features: 50	50
Fully connected layer 1 (ReLU)	out_features: 40	40
BatchNorm1d layer 2	num_features: 40	40
Fully connected layer 2 (ReLU)	out_features: 20	20

- **LSTM layers**

LSTM is a type of recurrent neural network that can capture long-term dependencies in data [210]. LSTM networks are well-suited to modeling sequences of data, such as time-series, natural language text, and audio data. An LSTM layer contains LSTM cells, which are composed of an input gate that determines what new information to store in the memory, a forget gate that determines what information to throw away from the cell's memory and an output gate that outputs the information stored in the memory, as shown in Eqs. (3.3)-(3.5). A

set of weights controls these gates learned during training. The cell state and hidden state are in Eqs. (3.6) and (3.7).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.3)$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.5)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3.6)$$

$$h_t = o_t \odot \tanh(c_t) \quad (3.7)$$

where x_t is the input to the LSTM cell at time step t , and b are the trainable weights and biases of the LSTM layer, \odot is the mathematical operator for the element-wise multiplication, and $\sigma(\cdot)$ and $\tanh(\cdot)$ are the sigmoid and hyperbolic tangent activation functions, respectively.

- **Batch normalization and fully connected layers**

The batch normalization layer normalizes each batch of data by shifting and scaling the previous layer's output, as shown in Eq. (3.8). The batch normalization can improve overall performance by stabilizing the learning process and accelerating convergence [211]. A fully connected layer is used after the batch normalization layer to generate a more abstract representation of the input data, as given by Eq. (3.9).

$$y = \frac{x - \mu}{\sqrt{\sigma + \epsilon}} \times \gamma + \beta \quad (3.8)$$

where μ and σ are the mean and standard deviation of the batch, respectively, ϵ is a small constant (1×10^{-5} in this study), added to prevent division by zero, and γ and β are trainable parameters.

$$y = \text{ReLU}(Wx + b) \quad (3.9)$$

where W and b are the weights and biases of the fully connected layer, respectively, and $\text{ReLU}(\cdot)$ is the rectified linear unit activation function.

After the Siamese subnetworks, a pair of 20-dimensional vectors $f(X_1), f(X_2)$ representing the encodings of the input pair X_1, X_2 are obtained, where $f(\cdot)$ denotes the LSTM subnetwork.

3.1.2.3. Similarity learning-based model training

During training, Siamese networks are fed with pairs of input samples with labels indicating whether the pair is similar or dissimilar. The networks compute the similarity between the two

time-series in the pair, and the loss is computed using the similarity and the pair's true label. The networks are then optimized using this loss, with the goal of minimizing the overall loss in the training data. This helps the networks learn to predict the similarity between pairs of input samples accurately. Siamese networks typically use binary cross-entropy loss or contrastive loss [212]. This study adopts the binary cross-entropy loss because the proposed similarity learning-based method treats fault diagnosis as a binary classification problem (i.e., either similar or dissimilar).

Before calculating binary cross-entropy, the encodings $f(X_1), f(X_2)$ are squashed into the range $[0,1]$ using a fully connected layer with the sigmoid activation function. First, the absolute elementwise difference between the encodings $f(X_1), f(X_2)$ is calculated. After that, the fully connected layer is applied, and the output of the layer is interpreted as the probability that the input pair belongs to the same class. The process is given as:

$$d = |f(X_1) - f(X_2)| \quad (3.10)$$

$$\hat{y} = \sigma(W \cdot d + b) \quad (3.11)$$

where \hat{y} is the output of the layer, W is the weight matrix, d is the absolute elementwise difference between encodings of input X_1, X_2 , and b is the bias vector.

The binary cross-entropy loss is computed as the average of the loss values for each sample in the dataset, as shown in Eq. (12). Like the cross-entropy loss, the binary cross-entropy loss penalizes confident but incorrect predictions, encouraging the model to output probabilities closer to the true labels. This helps improve the model's performance and accuracy.

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] \quad (3.12)$$

where N is the number of samples in the dataset, y_i is the true label of the i -th sample ("0" or "1" representing different and same labels, respectively), and \hat{y}_i is the predicted probability of the i -th sample belonging to the same class ("0" or "1").

3.1.3. Fault diagnosis

The next step involves predicting the labels of test samples using the Siamese network. A support set needs to be constructed, which provides a set of reference samples with labels. The support set can be used to pair up with one test sample as the inputs of the Siamese networks. As shown in **Figure 3.1**, if an FDD problem has A types of faults, the number of classes is $A + 1$, including the normal class for each class, B samples are randomly chosen from the training

data to construct the support set. Then, a similarity matrix is calculated by comparing the similarity of the input sample with each sample in the support set. Finally, the similarity of all test samples to the samples in each class is averaged to find the most similar class of the input sample.

3.2. Design of case studies using open dataset

In this chapter, two case studies are conducted to evaluate the proposed Similarity learning-based FDD method under two typical scenarios: insufficient labeled data and imbalanced labeled data. Insufficient labeled data means having a limited amount of labeled data for each fault in FDD for HVAC systems, making it hard to train and validate models effectively. Imbalanced labeled data occurs when a specific fault class has fewer labeled data than other classes during HVAC system operation, causing an uneven distribution of labeled data among fault classes.

3.2.1. Dataset description

AHU operational data from the ASHRAE project RP-1312 [213] are used in both case studies to verify the proposed method's performance. The schematic of AHU in the ASHRAE RP-1312 is shown in **Figure 3.4**. The outdoor air is initially combined with the return air, after which it undergoes treatment sequentially using a set of heating and cooling coils. Temperature, flowrate, and pressure sensors were installed, and the experimental data were collected under normal and various fault conditions. The faults were manually introduced in the AHU, such as the stuck air damper and the leaking heating valve. As listed in **Table 3.2**, five types of faults happened in different components of AHU. These faults were tested under various intensities and had 15 fault classes. The data in normal and fault classes are labeled as F_0, F_1, \dots, F_{15} , accordingly. The operation duration of AHU in this experiment was from 6:00 to 18:00, and data were collected with a one-minute interval. A 12-hour time-series was collected during a one-day experiment. Each faulty class contains 720 data points, while the normal class contains 720×4 data points. Each fault scenario was tested on different days, while the normal scenario was tested on four days. The 15 features used for fault diagnosis are listed in **Table 3.3**, mainly including temperatures, flow rates, and fan power signals.

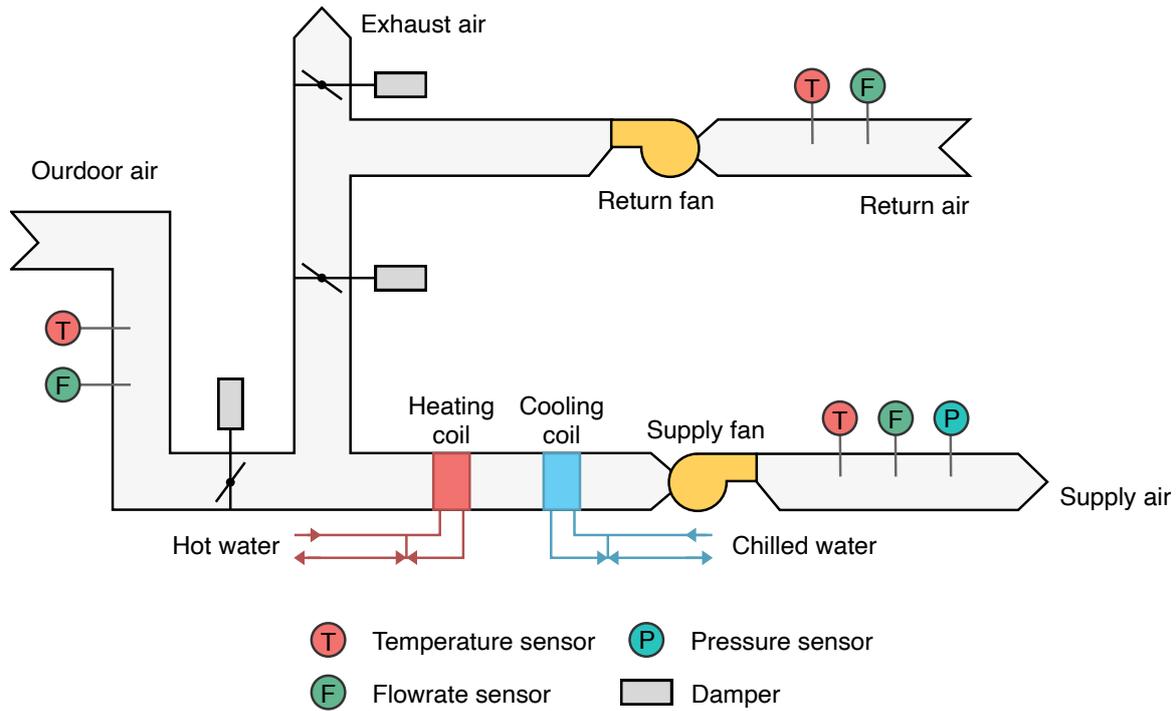


Figure 3.4. Schematic of AHU in the ASHRAE RP-1312 dataset

Table 3.2. Details of typical AHU faults in the ASHRAE RP-1312 dataset

Fault type	Fault (class) label	Fault detail	Experiment date
Normal	F_0	No fault	2007-08-19/25, 2007-09-04, 2007-09-10
Type A	F_1	Exhaust air damper stuck (100% open)	2007-08-20
	F_2	Exhaust air damper stuck (0% open)	2007-08-21
	F_3	Outdoor air damper stuck (0% open)	2007-08-26
Type B	F_4	Outdoor air damper leaking (45% open)	2007-09-05
	F_5	Outdoor air damper leaking (55% open)	2007-09-06
Type C	F_6	Heating coil valve leaking (0.4 GPM)	2007-08-28
	F_7	Heating coil valve leaking (1.0 GPM)	2007-08-29
	F_8	Heating coil valve leaking (2.0 GPM)	2007-08-30
Type D	F_9	AHU duct leaking (after supply fan)	2007-09-07
	F_{10}	AHU duct leaking (before supply fan)	2007-09-08
	F_{11}	AHU duct leaking (before supply fan)	2007-09-09
Type E	F_{12}	Cooling coil valve stuck (100% open)	2007-08-31
	F_{13}	Cooling coil valve stuck (0% open)	2007-08-27
	F_{14}	Cooling coil valve stuck (15% open)	2007-09-01
	F_{15}	Cooling coil valve stuck (65% open)	2007-09-02

Table 3.3. Features used for AHU fault diagnosis in the ASHRAE RP-1312

Feature name	Description	Unit
SA-TEMP	Supply air temperature	°F

OA-TEMP	Outdoor air temperature	°F
MA-TEMP	Mixed air temperature	°F
RA-TEMP	Return air temperature	°F
SA-CFM	Supplied air flow rates	CFM
RA-CFM	Return air flow rates	CFM
OA-CFM	Outdoor air flow rates	CFM
SF-WAT	Supply fan power	W
RF-WAT	Return fan power	W
SF-STS	Supply fan status	On = 1, Off = 0
RF-STS	Return fan status	On = 1, Off = 0
SF-DP	Differential pressure of supply fan	Pa
RF-DP	Differential pressure of return fan	Pa
CHWC-VLV	Cooling coil valve position	%
HWC-VLV	Heating coil valve position	%

3.2.2. Data-splitting

As discussed in Section 2.2.1, the temporal train-test split is preferable for time-series data. The temporal train-test split involves selecting a certain number of time steps from the beginning of the series as the training data and the rest as the test data. There is also a constraint that the test data must be chronologically after the training data. This constraint ensures that the model has not seen any data from the future when making predictions on the test data.

As shown in **Figure 3.5**, for each day's experimental data, the data from 6:00 to 14:00 was used as the training period, and the data from 14:00 to 18:00 was used as the test period. The data samples were generated using the method introduced in Section 3.1.2.1. The window stride size is set to 1 minute for the training period. In the following two cases, only a part of the training data was randomly selected to simulate the insufficient labeled and imbalanced data scenarios. The two scenarios will be introduced in Sections 3.2.4 and 3.2.5, respectively. For the test period, the window stride is 5 minutes. In both case studies, the window size m is 10.

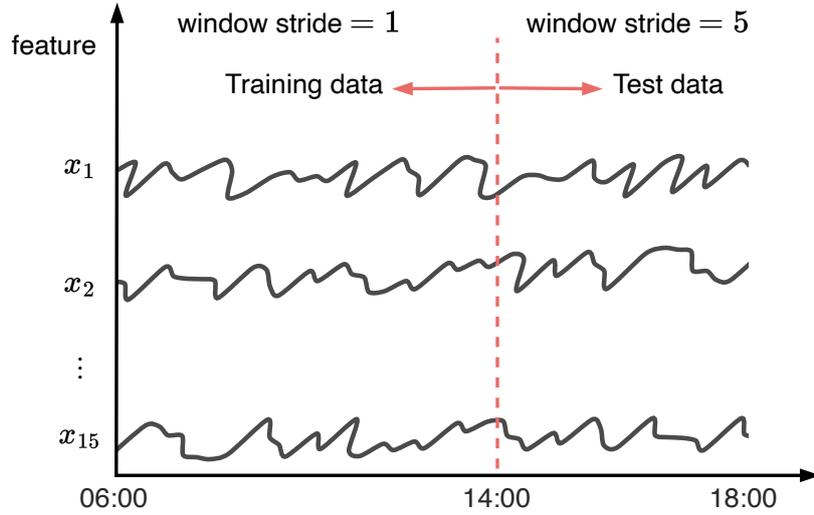


Figure 3.5. Data-splitting for case studies

3.2.3. Baseline model for comparison

To compare the proposed similarity learning-based method with the conventional data-driven approach, the same architecture of LSTM subnetworks is used as the baseline. The baseline model treats fault diagnosis as a multi-class classification problem. Therefore, the input and output of the baseline model differ from those of the Siamese networks. As shown in **Figure 3.6**, the input is a single sample, and the output is the probability of the sample belonging to each class, i.e., p_i . The probability vector is obtained using a fully connected layer with the Softmax activation function after the LSTM. Finally, the baseline model is optimized using the cross-entropy loss.

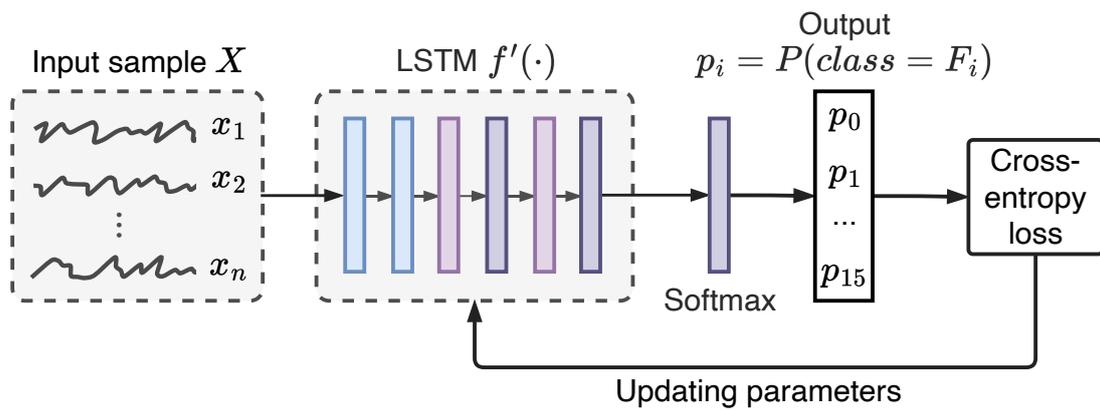


Figure 3.6. Baseline model for case studies

3.2.4. Scenario 1: Insufficient labeled data

As listed in **Table 3.2**, there are 16 normal/faulty working conditions in the dataset. The purpose of Scenario 1 is to test the performance of the proposed method when only a small amount of labeled data is available for each fault, which is a common situation when developing data-driven models for FDD in practical HVAC systems. The performance of the proposed method is tested by adjusting the number of samples per class in the training set N_t in the range of 5–100. In addition, the number of samples in each class in the support set N_s is also adjustable in the range of 1–10. When $N_s < N_t$, N_s samples are selected from each training class without replacement; when $N_s \geq N_t$, all N_t samples in each class used for training are selected. During the experiments, the specified number of samples were randomly selected from the training samples generated in Section 3.2.2. The experiments were repeated ten times using different random seeds to eliminate the effect of randomness on the results.

3.2.5. Scenario 2: Imbalanced labeled data

Scenario 2 is designed to test the performance of the proposed method under data imbalance, for example, when a particular fault class has less labeled data than other fault classes. This situation is also common as some faults occur more frequently than other faults during the operation of HVAC systems. In Scenario 2, each type of fault in **Table 3.2** was tested under data imbalance. The imbalance ratio of each imbalanced fault class and the remaining classes (faulty/normal conditions) is 1:10. The number of labeled samples for the imbalanced fault is 10, and the number of samples for the remaining faulty/normal conditions is 100. Similar to Scenario 1, the experiments were repeated ten times to avoid sampling bias.

3.2.6. Evaluation metric

The performance improvement rate (PIR) is used as an evaluation metric to measure the performance difference between the proposed similarity learning-based method and the baseline model, which is defined as:

$$PIR = \frac{Accuracy_{Siamese} - Accuracy_{Baseline}}{Accuracy_{Baseline}} \times 100\% \quad (3.13)$$

where $Accuracy_{Siamese}$ is the accuracy of the proposed Siamese networks on the test set; $Accuracy_{Baseline}$ is the accuracy of the baseline model on the test set.

3.3. Results and discussion

This section presents the results of Scenario 1 and Scenario 2. Scenario 1 analyzes the impacts of the number of samples per class in the training set and the support set on fault diagnostic

accuracy using the proposed Siamese network. Then the performance improvement rate versus the baseline model is evaluated. In Scenario 2, each imbalanced fault's diagnostic accuracy and the remaining classes is analyzed.

3.3.1. Scenario 1: Insufficient labeled data

3.3.1.1. Impact of the number of samples per class in the training set

Table 3.4 and **Figure 3.7** show how the fault diagnostic accuracy on the test data changes in different settings (the number of samples per class in the support set N_s and the number of samples per class in the test set N_t) are used for training Siamese network. Note that all experiments were repeated ten times using different random seeds, and **Figure 3.7** shows the average accuracy. Generally, there is a positive correlation between fault diagnostic accuracy and N_t under a given number of N_s , because more labeled data can provide extra information for the model and lead to better generalization performance. A significant increase in fault diagnostic accuracy can be observed when N_t increases from 5 to 10. The fault diagnostic accuracy becomes stable when N_t reaches a certain critical value. At this point, adding more labeled data to the training set does not improve the model's overall accuracy. Nevertheless, according to **Table 3.4**, a higher number of samples per class in the training set has a lower standard deviation of accuracy in the ten random experiments. Therefore, adding more labeled data to the training set can improve the model's stability and robustness, although the overall accuracy does not improve.

3.3.1.2. Impact of the number of samples per class in the support set

According to **Table 3.4** and **Figure 3.7**, there is a positive correlation between fault diagnostic accuracy and N_s under a given number of N_t . The more samples per class in the support set, the Siamese networks can make more reliable classification results based on the similarity matrix. Compared with the performance leap owing to the increase of N_t at the early stage, the performance improved with the increase of N_s is not significant, especially when $N_t \geq 20$. For example, when $N_t = 20$, the difference in fault diagnostic accuracy from $N_s = 1$ to $N_s = 10$ is only 4.1% (from 86.3% to 90.4%). It means that despite only one sample per class in the support set, Siamese networks can make accurate enough FDD results. Similarly, the correlation between the standard deviation of accuracy and N_s is also not as strong as the correlation between the standard deviation of accuracy and N_t . It means that the stability of Siamese networks is more sensitive to the number of samples per class in the training set than the number of samples per class in the support set.

Another finding is that when N_t is large, the improvement of fault diagnostic accuracy becomes small as N_s increases from 1 to 10. For example, the difference in model accuracy from $N_s = 1$ to $N_s = 10$ is 12.5% (from 60.6% to 73.1%) when N_t is 5, but is only 4.4% (from 87.6% to 92.0%) when $N_t = 100$. Thus, there is a trade-off between the selection of N_t and N_s . Fewer samples in the support set are needed when more samples are used to train the model.

Table 3.4. Accuracy of fault diagnosis on the test set using Siamese networks (%)

N_t	Number of samples per class in the support set N_s										Avg.
	1	2	3	4	5	6	7	8	9	10	
5	60.6 (7.6)	66.3 (12.5)	69.9 (6.3)	73.7 (6.5)	73.1 (7.4)	73.1 (7.4)	73.1 (7.4)	73.1 (7.4)	73.1 (7.4)	73.1 (7.4)	70.9 (7.7)
10	77.1 (11.4)	84.4 (7.9)	85.1 (5.8)	84.7 (5.0)	86.2 (5.7)	86.1 (5.4)	85.5 (5.5)	86.1 (6.1)	86.1 (6.2)	86.2 (6.1)	84.8 (6.5)
20	86.3 (5.4)	89.8 (6.6)	87.4 (5.2)	90.0 (5.4)	89.9 (5.0)	90.5 (6.1)	90.6 (4.8)	89.7 (5.7)	89.6 (6.1)	90.4 (5.8)	89.4 (5.6)
30	86.2 (6.0)	89.6 (5.6)	90.3 (4.8)	90.6 (4.6)	87.3 (6.0)	90.0 (4.3)	90.7 (3.9)	89.1 (5.1)	90.7 (4.1)	90.1 (4.2)	89.5 (4.9)
40	88.9 (7.1)	89.3 (4.7)	88.3 (3.8)	89.9 (4.1)	89.3 (5.0)	91.1 (4.3)	90.9 (5.2)	90.9 (4.1)	91.0 (4.5)	90.9 (4.7)	90.1 (4.8)
50	87.1 (6.9)	89.4 (7.2)	90.9 (3.9)	90.6 (6.7)	90.3 (5.7)	92.4 (4.2)	91.3 (4.8)	89.9 (6.1)	90.0 (4.5)	91.1 (4.2)	90.3 (5.4)
60	86.9 (5.3)	88.2 (5.2)	91.4 (5.1)	90.8 (5.9)	91.8 (4.0)	91.9 (4.1)	91.3 (5.1)	91.7 (5.0)	91.6 (4.9)	90.9 (4.8)	90.7 (4.9)
70	87.6 (7.1)	90.4 (4.6)	88.0 (6.6)	88.7 (5.1)	91.0 (4.8)	90.1 (4.4)	90.8 (3.3)	89.6 (5.0)	90.3 (5.5)	89.6 (4.9)	89.6 (5.1)
80	89.1 (4.3)	91.8 (4.7)	91.4 (3.5)	92.3 (4.7)	92.4 (3.6)	92.0 (3.2)	91.9 (5.1)	92.7 (3.5)	92.5 (3.7)	91.0 (5.1)	91.7 (4.1)
90	90.9 (3.9)	92.1 (3.6)	88.3 (5.6)	92.0 (3.4)	90.7 (4.1)	91.8 (3.4)	90.4 (4.2)	91.0 (4.1)	92.0 (3.4)	91.6 (4.2)	91.1 (4.0)
100	87.6 (5.2)	91.9 (4.0)	92.7 (2.7)	91.5 (3.0)	92.8 (2.0)	92.0 (2.3)	93.1 (2.6)	92.4 (2.5)	92.8 (2.4)	92.0 (2.5)	91.9 (2.9)
Avg.	84.4 (6.4)	87.6 (6.1)	87.6 (4.8)	88.6 (4.9)	88.6 (4.8)	89.2 (4.5)	89.1 (4.7)	88.7 (5.0)	89.1 (4.8)	88.8 (4.9)	88.2 (5.1)

Note: N_t refers to the number of samples per class in the training set; the unit of numbers is %; numbers in parentheses refer to the standard deviation of accuracy in ten repeated experiments using different random seeds.

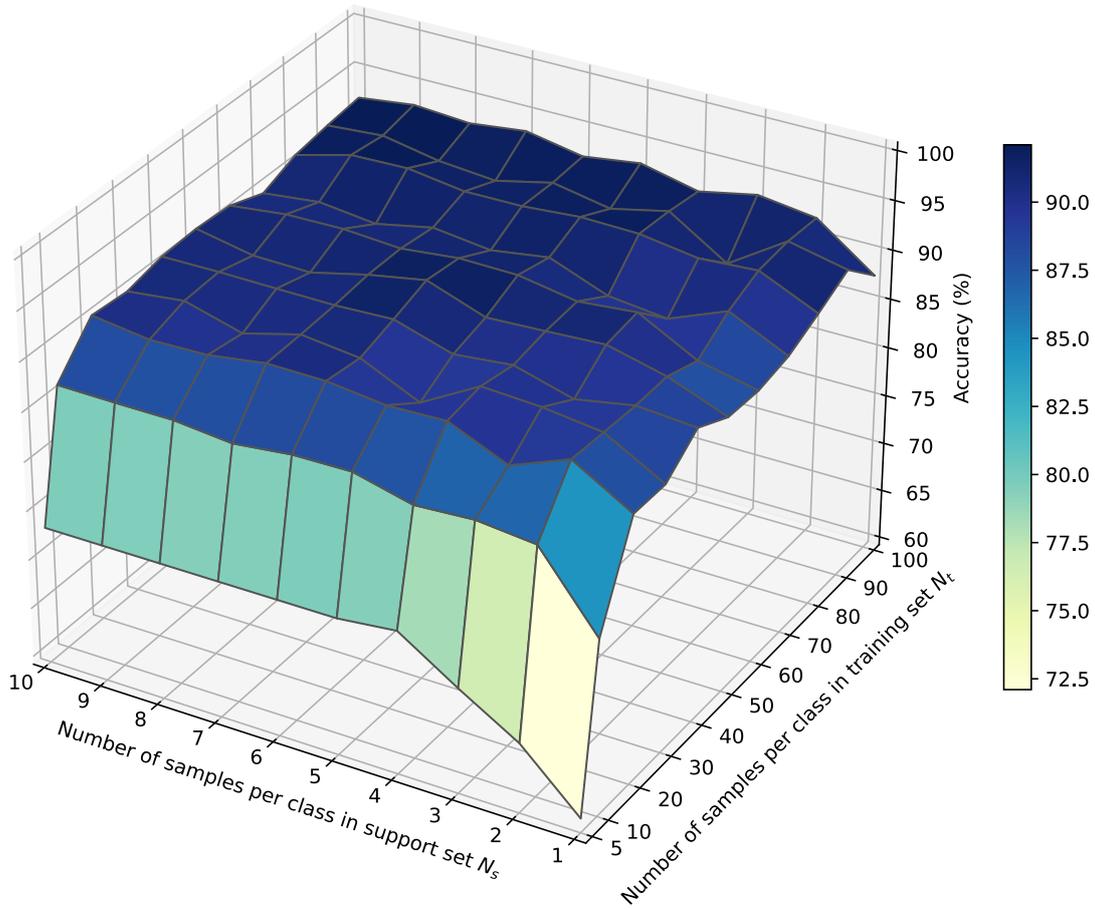


Figure 3.7. Average accuracy of fault diagnosis on the test set using the proposed method

3.3.1.3. Performance improvement rate

The performance improvement rate of the proposed Siamese networks compared with the baseline model is shown in **Figure 3.8**. The number of samples per class in the support set N_s is fixed at 10. Eq. (3.13) defines a positive PIR as the Siamese networks perform better than the baseline model. When the number of samples per class in the training set N_t is 5, an average of 45.7% PIR can be observed. It means that the proposed Siamese networks can boost fault detection accuracy when the number of labeled data is insufficient. With the increase of N_t , the PIR decreases fast ($N_t \leq 20$) and fluctuates in the range between 0%–10% ($N_t \geq 30$). In general, the proposed Siamese networks outperform the baseline model, especially when the number of labeled data is small ($N_t \leq 20$).

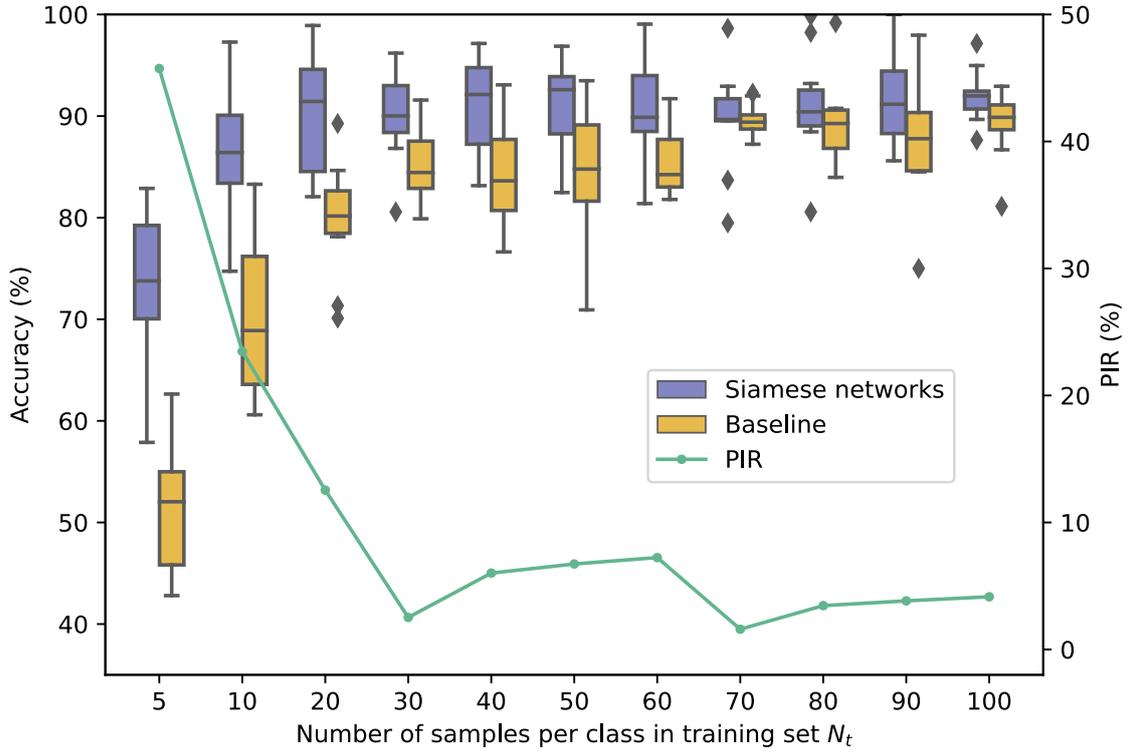


Figure 3.8. Performance improvement rate of the proposed Siamese networks ($N_s = 10$)

3.3.2. Scenario 2: Imbalanced labeled data

Figure 3.9 shows the testing accuracy of each imbalanced fault in Scenario 2. For example, when F_2 (exhaust air damper is stuck to 0% position) is the class with the imbalanced fault in the training data, the fault diagnostic accuracy of F_2 on the test data using the Siamese networks and baseline model is 82.6% and 73.0%, respectively. The proposed Siamese networks outperform the baseline model in all experiments except F_{13} , F_{14} , and F_{15} . Furthermore, the lowest accuracy of the imbalanced fault using Siamese networks is 56.5% compared with 13.9% using the baseline mode (when F_5 is the imbalanced fault in the training data).

Figure 3.10 shows the testing accuracy of the remaining classes. For example, when F_2 (exhaust air damper is stuck to 0% position) is the imbalanced fault in the training data, the fault diagnostic accuracy of F_2 on the test data using the Siamese networks is 82.6% and the diagnostic accuracy of the remaining classes is 88.7%. As for the baseline model which is trained using labeled samples instead of sample pairs, the fault diagnostic accuracy of F_2 and the remaining classes is 73.0% and 87.9%, respectively. As is seen in **Figure 3.10**, the proposed Siamese networks have higher diagnostic accuracy than the baseline model in all experiments. Compared with the results in **Table 3.4**, the data imbalance decreases the FDD accuracy in general.

Still, the proposed Siamese networks perform better than the baseline model in both imbalanced fault and the remaining classes.

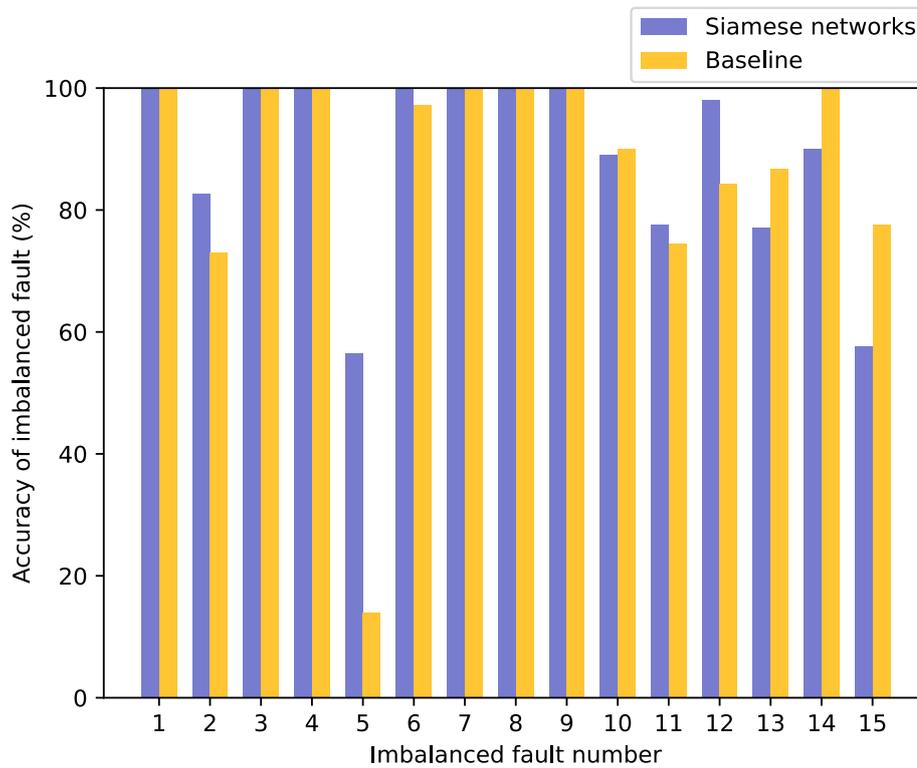


Figure 3.9. Fault diagnostic accuracy of each imbalanced fault

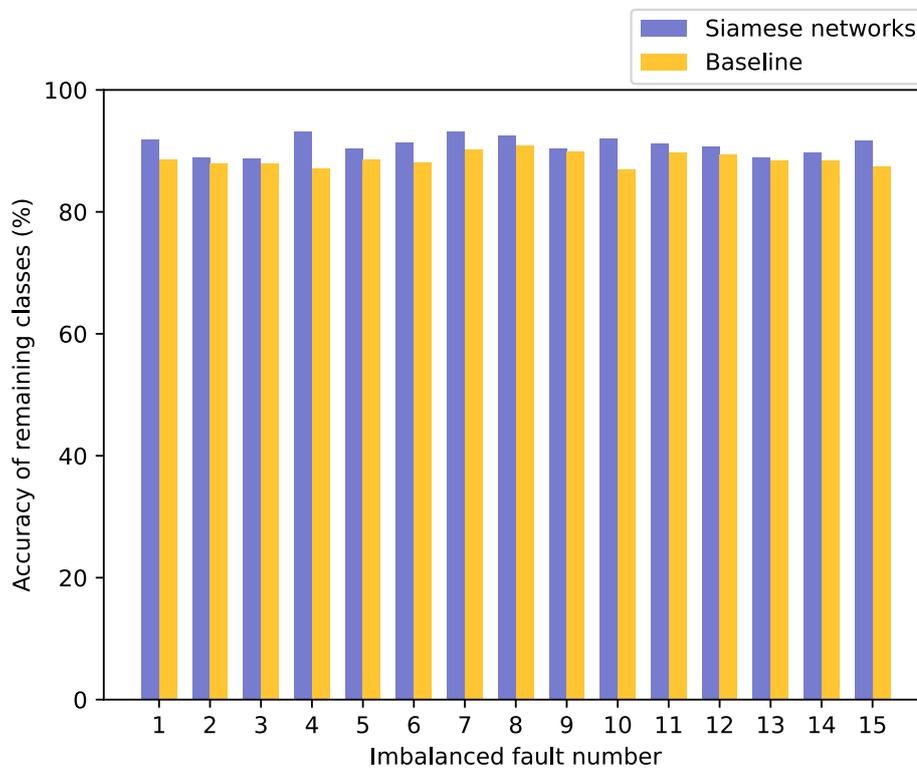


Figure 3.10. Fault diagnostic accuracy of remaining classes

3.4. Summary

Machine learning has been widely adopted for FDD in HVAC systems over the past decade due to the ever-increasing availability of massive building operational data. Machine learning-based FDD is flexible and accurate but heavily relies on the availability of sufficient labeled data to develop supervised or unsupervised models. However, collecting labeled data is usually labor-intensive for various types of faulty conditions, significantly limiting the practical implementation of machine learning-based FDD. Therefore, this study proposes a similarity learning-based method using Siamese networks to improve the interpretability and reliability of machine learning-based FDD in applications with limited labeled data. Unlike the conventional supervised approach, the proposed Siamese networks contain two identical long short-term memory subnetworks that take a pair of multivariate time-series samples from the building energy management system as input. The number of training samples can be significantly augmented by generating pairs randomly. In this way, the generalization ability of the machine learning-based FDD is significantly improved in practical applications. In addition, the comparison between the new test sample and the support set provides an interpretation similar to k-NN, which enhances interpretability. Two case studies were designed and conducted using experimental data when labeled data were limited and imbalanced to validate the proposed similarity learning-based method. Where the number of labeled data is limited, the proposed method improves the fault diagnostic accuracy by 45.7% than the baseline model. When the labeled data is imbalanced, the proposed method demonstrated better generalization ability.

CHAPTER 4 PROBABILISTIC MACHINE LEARNING FOR ENHANCING RELIABILITY AND INTERPRETABILITY ADDRESSING MEASUREMENT UNCERTAINTY

This chapter presents a probabilistic machine learning approach to address the issue of **measurement uncertainty** and improve interpretability and reliability. Cooling load prediction is important for the control of chiller plants, particularly for sequencing control for multiple-chiller systems. Conventional cooling load prediction provides point prediction without uncertainty information, which lacks interpretability and reliability for chiller sequencing control. Therefore, this study proposes a robust chiller sequencing control strategy that utilizes probabilistic cooling load predictions. One-hour-ahead probabilistic cooling load prediction in the form of a normal distribution is made using natural gradient boosting (NGBoost). A risk-based action evaluation scheme is designed to determine the optimal number of operating chillers by evaluating the risks associated with possible control actions. The risk interpretation can enhance the trust between the prediction model and the operators. Section 4.1 describes the research methodology utilizing probabilistic cooling load predictions for robust chiller sequencing. Section 4.2 details the test facilities and data collection from an actual chiller plant. Sections 4.3 and 4.4 present the chiller sequencing results based on historical operational data and an in-situ test, respectively. Lastly, Section 4.5 summarizes the main conclusions of this study.

4.1. Proposed risk-based chiller sequencing control strategy

4.1.1. Basic principle

A typical multiple-chiller plant is illustrated in **Figure 4.1** [150]. The chilled plant consists of two circulation loops: a chilled water loop and a condensing water loop. Each chiller is associated with a chilled water pump, a condensing water pump, and a cooling tower. Usually, switching on/off a chiller will automatically switch on/off the associated cooling tower and water pumps [214].

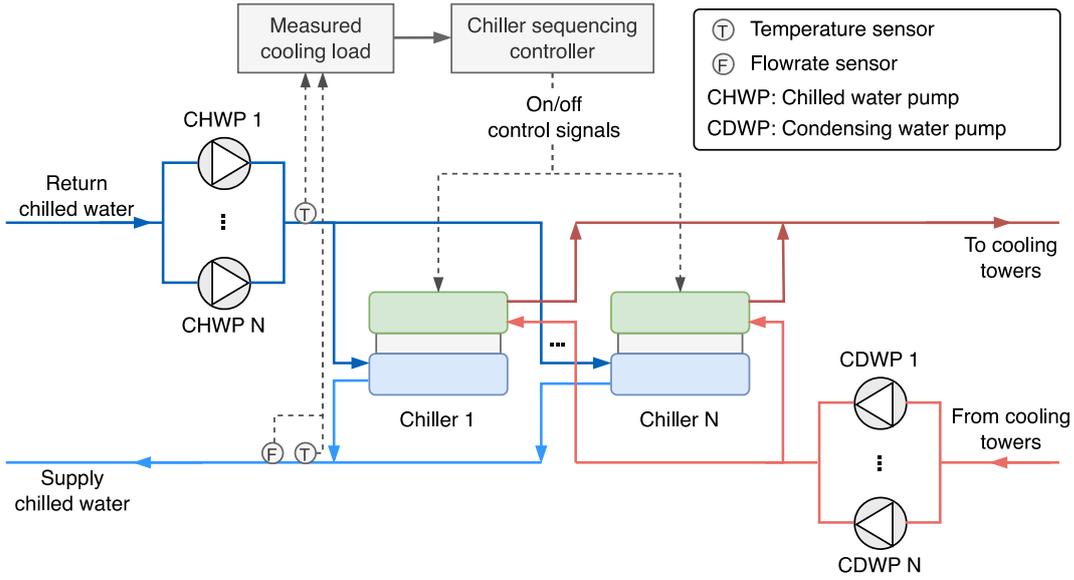


Figure 4.1. Schematic diagram of a typical multiple-chiller system and sequencing control

The conventional cooling load-based chiller sequencing control, denoted as conventional Q-based control in this study, uses the real-time cooling load calculated from chilled water's flow rate and supply and return temperatures. Then, the number of chillers to meet the cooling load is determined by dividing the calculated real-time cooling load by the nominal cooling capacity of the chiller [131,215]. The measured total cooling load is formulated as:

$$Q_t = c_w \cdot \dot{m}_w \cdot (T_{rtn} - T_{sup}) \quad (4.1)$$

where c_w represents the specific capacity of water ($\text{kJ}/\text{kg} \cdot ^\circ\text{C}$), \dot{m}_w is the chilled water flow rate in the header pipe (kg/s), and T_{rtn}, T_{sup} are the return and supply temperatures of chilled water.

When the chillers have the same capacity, the thresholds for switching off a chiller Q_n^{off} and switching on a chiller Q_n^{on} are defined as:

$$Q_n^{off} = (n - 1) \cdot Q_c \cdot (1 - d/2) \quad (4.2)$$

$$Q_n^{on} = n \cdot Q_c \cdot (1 + d/2) \quad (4.3)$$

where n is the current number of operating chillers, d is the dead band of cooling load to be set by the operators, and Q_c is the cooling capacity of one chiller.

The basic rule is that the cooling capacity provided by chillers should meet the actual cooling load. Therefore, the number of chillers to be switched on/off Δn is determined by:

$$\Delta n = \begin{cases} -1, & Q_t < Q_n^{off} \\ 0, & Q_n^{off} \leq Q_t \leq Q_n^{on} \\ 1, & Q_t > Q_n^{on} \end{cases} \quad (4.4)$$

where Q_t is the real-time measured building load at the current control period.

The main disadvantage of the conventional Q-based chiller sequencing strategies is that short-term load fluctuation may cause unnecessary on/off switching. Therefore, the proposed chiller sequencing control strategy combines the conventional Q-based and the probabilistic cooling load predictions to improve the robustness of sequencing control.

The proposed strategy's basic idea is illustrated using the daily load profile of an actual building as an example, as shown in **Figure 4.2**. The measured total cooling load Q_t is larger than the threshold Q_n^{on} at 10:30, the conventional Q-based sequencing strategy switches on a chiller without considering the future cooling load change. In contrast, the proposed strategy takes advantage of one-hour-ahead probabilistic cooling load predictions and determines whether a less risky action should be adopted. The risk of an action is defined to measure the probability that the cooling capacity cannot satisfy the cooling demand or that an unnecessary chiller is operated after the action is adopted [216]. Compared with conventional cooling load prediction [217], probabilistic cooling load prediction gives a distribution of cooling load for each time step. **Figure 4.2** shows the probabilistic cooling load prediction after 10:30 (25%, 50%, 75%, and 95% prediction intervals are illustrated), which is predicted based on the available information at 10:30.

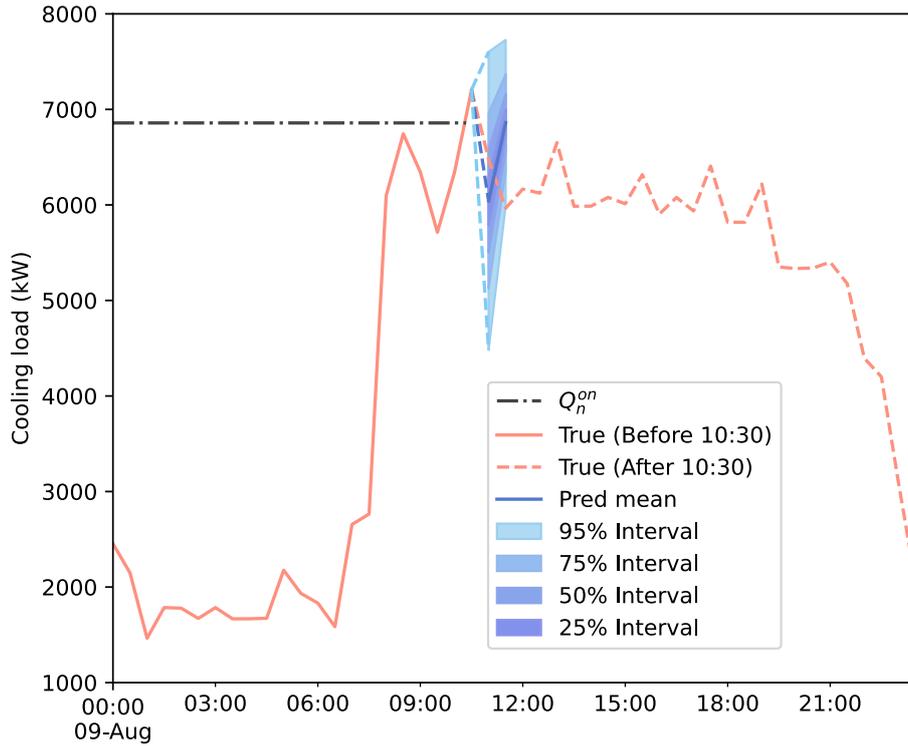


Figure 4.2. The schematic diagram illustrating the basic idea of the proposed probabilistic cooling load prediction-based sequencing control strategy

4.1.2. Overview of the proposed strategy

The framework of the proposed probabilistic cooling load prediction-based sequencing control strategy is illustrated in **Figure 4.3**. The proposed strategy comprises three modules: 1) probabilistic cooling load prediction, 2) Q-based sequencing action, and 3) risk-based action evaluation. The inputs and outputs of the modules are highlighted in yellow and purple, respectively. Inputs, which include weather data, cooling load, and the number of operational chillers, can be sourced from building management systems. The first module predicts the 1h-ahead probabilistic cooling load using the natural gradient boosting (NGBoost) algorithm. In the second module, the Q-based sequencing action is determined based on the current cooling load. The third module is bypassed if there is no need to switch any chiller on/off (meaning the supplied cooling capacity aligns with the current cooling demand). However, if one or more chillers need to be switched on/off, the third module evaluates the risk associated with the Q-based action based on the probabilistic cooling load predictions in the first module and decides if a compromised action is less risky. The magnitude of the compromised action is less than that of the Q-based action, so it generally maintains the existing number of chillers, thereby reducing the impact on the chiller plant. The third module evaluates the risks associated with

both actions by leveraging the probabilistic cooling load predictions. This module ensures that the sequencing optimally reduces the impact on the chiller plant while still fulfilling the cooling load demands.

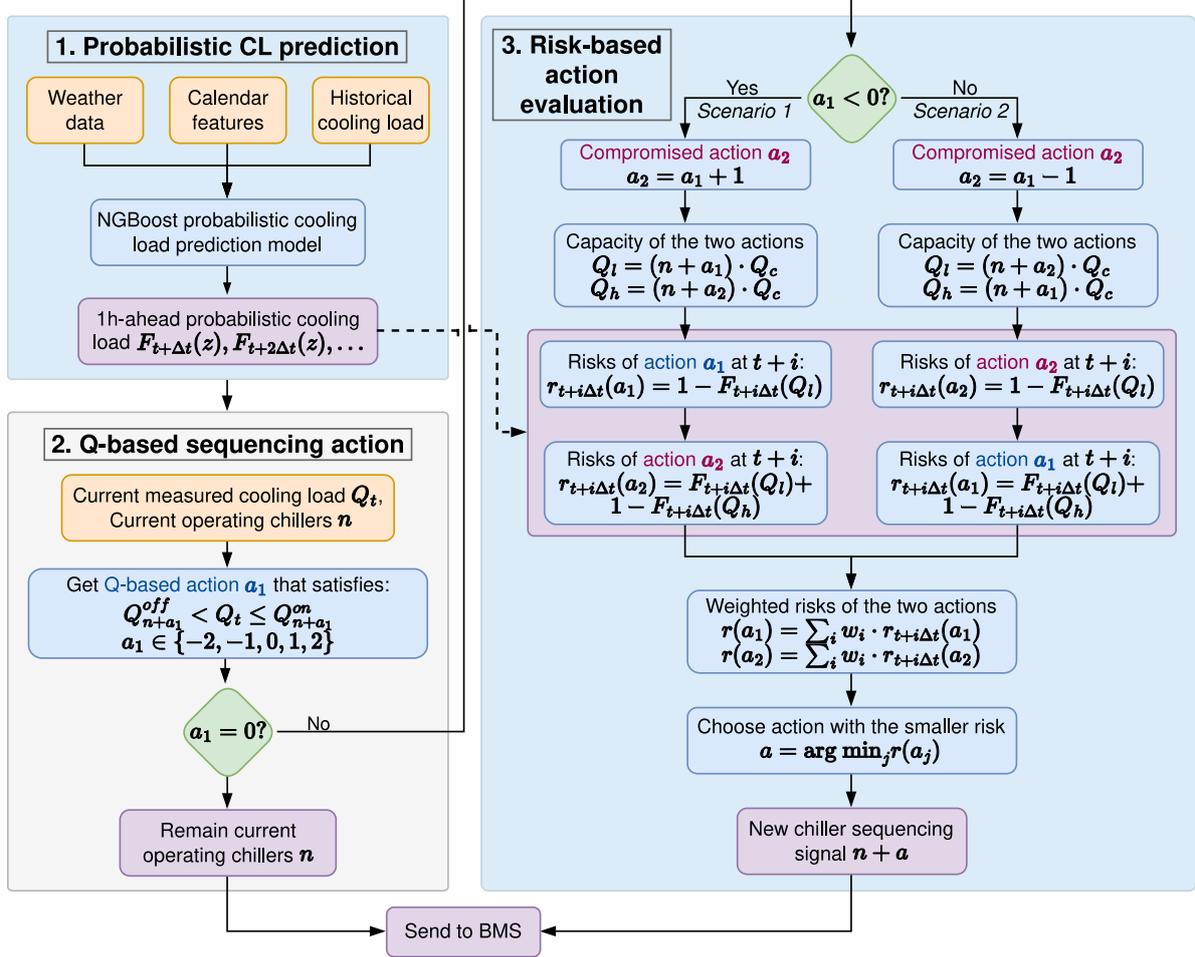


Figure 4.3. Framework of the proposed chiller sequencing control strategy

4.1.3. One-hour-ahead probabilistic cooling load prediction using natural gradient boosting

The first module predicts probabilistic cooling load using natural gradient boosting (NGBBoost). NGBBoost is a state-of-the-art machine learning algorithm that can make probabilistic predictions for various regression tasks [218]. Compared with conventional gradient boosting algorithms, NGBBoost has improvements for probabilistic prediction in three main aspects: prediction target, optimization function, and gradient used [219,220]. First, in conventional gradient boosting, the prediction target is a point estimate, typically the mean of the target variable given the input [108]. Instead of predicting a single value, NGBBoost aims to capture the uncertainty of predictions using a distribution denoted as $P_\theta(y | x)$. For example,

NGBoost can estimate a normal distribution's mean and standard deviation. The second improvement is the loss function used during model training. In conventional gradient boosting, loss functions like Mean Square Error are commonly used for regression tasks. The loss function measures the distance between the predicted and true values, and the boosting algorithm minimizes this distance iteratively. NGBoost introduces probabilistic loss functions for optimization, such as log-likelihood denoted. The log-likelihood measures how well the predicted distribution fits the observed data. The log-likelihood loss function is denoted as:

$$\mathcal{L}(\theta, y) = -\log P_{\theta}(y) \quad (4.5)$$

where θ represents the parameters of the predicted distribution and y represents the true values.

The third improvement is the gradient used. In conventional gradient boosting, the normal gradient is used to update the parameters. However, in probabilistic prediction, the "distance" between two parameter values of the distributions does not reflect an appropriate "distance" between the distributions. Therefore, in NGBoost, the natural gradient motivated by information geometry [221] is adopted, denoted as:

$$\tilde{\nabla}\mathcal{L}(\theta, y) = \mathcal{J}_{\mathcal{L}}(\theta)^{-1} \cdot \nabla\mathcal{L}(\theta, y) \quad (4.6)$$

where $\mathcal{J}_{\mathcal{L}}(\theta)$ is the Fisher Information that the distribution P_{θ} carries, denoted as:

$$\mathcal{J}_{\mathcal{L}}(\theta) = \mathbb{E}_{y \sim P_{\theta}}[\nabla_{\theta}\mathcal{L}(\theta, y) \cdot \nabla_{\theta}\mathcal{L}(\theta, y)^T] \quad (4.7)$$

Three categories of features are used to predict the one-hour-ahead cooling load [130]. First, calendar features include holiday or not (binary feature), hour of the day, day of the week, and month of the year. Second, weather data are used, including outdoor dry air temperature, outdoor dew-point temperature, and relative humidity. These features are lagged by one hour because the actual weather data for the cooling load to be predicted is unknown. For example, weather conditions at 13:00 predict the cooling load at 14:00. Last, lag features (1-hour, 24-hour, 48-hour, and 168-hour ahead) of cooling load are used.

The output of NGBoost is a probability distribution of cooling load prediction. In this study, the normal distribution is used as the output distribution [158]. Therefore, the probability distribution can be formulated as $\mathcal{N}(\mu, \sigma)$, where the mean value term μ is equivalent to the output of non-probabilistic methods (cooling load), and the standard deviation term σ represents the uncertainty of the prediction. The larger the standard deviation term σ is, the more uncertain the prediction is. Finally, the probability distribution is converted into the cumulative density function. The cumulative density function is a function that calculates the

probability of the cooling load being less than or equal to a value z , defined as Eq. (4.8). **Figure 4.4** shows the probabilistic prediction, and the purple area represents the probability that the cooling load is less than or equal to z .

$$F(z) = P(Q \leq z) \quad (4.8)$$

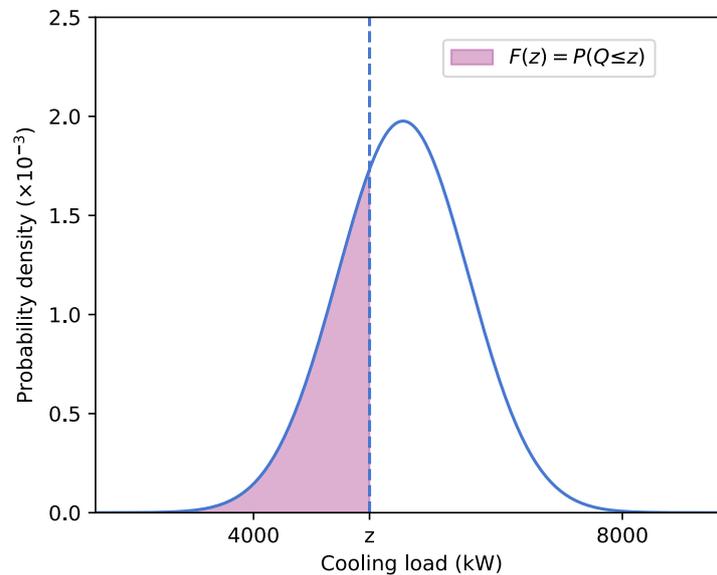


Figure 4.4. Probabilistic prediction of cooling load in the form of normal distribution

Therefore, the output of the probabilistic cooling load prediction module is formulated as: $F_{t+\Delta t}(z), F_{t+2\Delta t}(z), \dots$. Δt is the prediction step that can be adjusted according to the granularity of data collection. For example, if $\Delta t = 0.5h$, there are two probabilistic cooling load predictions in one hour: $F_{t+0.5h}(z), F_{t+1h}(z)$. By predicting the probability distribution of the cooling load, NGBoost provides insight into the uncertainty and variability of the cooling load, enabling more informed and effective chiller sequencing decisions.

4.1.4. Enhanced cooling load-based sequencing

The second module, the Q-based sequencing, determines switching actions according to the measured cooling load. The Q-based action a_1 is determined by Eq. (4.9), which is converted from Eqs. (4.2)-(4.4). The Q-based action a_1 denotes the number of chillers that should be switched on/off, ranging from -2 to 2 . Usually, switching one chiller on/off is enough to deal with the cooling load change. In some critical situations, when the cooling load changes dramatically due to indoor or outdoor disturbance [222], two chillers may need to be switched on/off.

$$Q_{n+a_1}^{off} < Q_t \leq Q_{n+a_1}^{on} \quad (4.9)$$

When the Q-based action $a_1 = 0$, the supplied cooling capacity aligns with the current measured cooling load. Therefore, the Q-based action a_1 is adopted without the risk-based action evaluation module. When the Q-based action $a_1 \neq 0$, indicating that one or more chillers need to be switched on or off, the risks associated with the Q-based action will be evaluated in the risk-based action evaluation module to determine whether a compromised action is more appropriate. Compared with the conventional strategy, the proposed strategy adds a risk-based action evaluation module based on the probabilistic cooling load prediction when the conventional strategy suggests switching on or off (i.e., $a_1 \neq 0$). Therefore, the control actions based on the data-driven model can be comprehended using the risks and the reliability is enhanced.

4.1.5. Risk-based action evaluation scheme

If Q-based action a_1 is adopted, Eq. (4.9) may not be satisfied in the following hour because the cooling load may change, leading to unnecessary on/off switching. The risk of an action is defined as the probability that the cooling capacity cannot satisfy the cooling demand or that an unnecessary chiller is operated after the action is adopted [216]. In this module, the risks of the Q-based action a_1 and the compromised action a_2 are evaluated according to one-hour-ahead probabilistic cooling load predictions $F_{t+\Delta t}(z), F_{t+2\Delta t}(z), \dots$. The mapping between Q-based action a_1 and compromised action a_2 is listed in **Table 4.1**, and the magnitude of the compromised action is less than that of the Q-based action.

When $a_1 = 0$ equals zero, the decision-making process of the conventional strategy, which refrains from executing switching on/off operations on the units, is preferred. Conversely, when $a_1 \neq 0$, the conventional strategy initiates switching on/off operations. At this juncture, risk-based action evaluation comes into play to assess actions a_1 and a_2 , aiming to identify the operation with lower risk. The strategy introduced herein integrates a risk assessment module into the conventional framework, ensuring that it does not resort to a strategy more aggressive than the conventional approach.

Table 4.1. Mapping between the Q-based action a_1 and compromised action a_2

Scenario	Q-based action	Compromised action
$a_1 < 0$ (Switching off)	$a_1 = -2$	$a_2 = -1$
	$a_1 = -1$	$a_2 = 0$
$a_1 > 0$ (Switching on)	$a_1 = 1$	$a_2 = 0$
	$a_1 = 2$	$a_2 = 1$

For each prediction step, the risks of Q-based action a_1 and compromised action a_2 are calculated according to the cooling capacity of adopting a_1 and a_2 . Then, the weighted risks of a_1 and a_2 are calculated based on the weights of each prediction step.

There are two scenarios: $a_1 < 0$ and $a_1 > 0$, representing switching off and switching on chillers, respectively.

4.1.5.1. Scenario 1: Switching off ($a_1 < 0$)

When $a_1 < 0$, the total chiller capacity that the compromised action a_2 can provide is higher than that of the Q-based action a_1 . Therefore, the lower chiller capacity limit Q_l and higher chiller capacity limit Q_h are given by Eqs. (4.10) and (4.11), respectively. **Figure 4.5** illustrates one step of the one-hour-ahead probabilistic cooling load predictions ($F_{t+i\Delta t}(z)$) and the two capacity limits. According to the definition of the cumulative density function, the white area under the distribution in **Figure 4.5** (a) represents the probability that the cooling demand matches the cooling supply when adopting Q-based action a_1 . Therefore, the risk of the Q-based action a_1 in the i -th prediction step is formulated as Eq. (4.12), as shown in the blue area in **Figure 4.5** (a). The risk of the compromised action a_2 in the i -th prediction step is given by Eq. (4.13), as shown in the green area in **Figure 4.5** (b). Finally, the weighted risks of a_1 and a_2 in the following hour are calculated as Eqs. (4.14) and (4.15), and the action with the minimum risk is adopted in the current control period.

$$Q_l = (n + a_1) \cdot Q_c \quad (4.10)$$

$$Q_h = (n + a_2) \cdot Q_c \quad (4.11)$$

$$r_{t+i\Delta t}(a_1) = 1 - F_{t+i\Delta t}(Q_l) \quad (4.12)$$

$$r_{t+i\Delta t}(a_2) = F_{t+i\Delta t}(Q_l) + 1 - F_{t+i\Delta t}(Q_h) \quad (4.13)$$

$$r(a_1) = \sum_i w_i \cdot r_{t+i\Delta t}(a_1) \quad (4.14)$$

$$r(a_2) = \sum_i w_i \cdot r_{t+i\Delta t}(a_2) \quad (4.15)$$

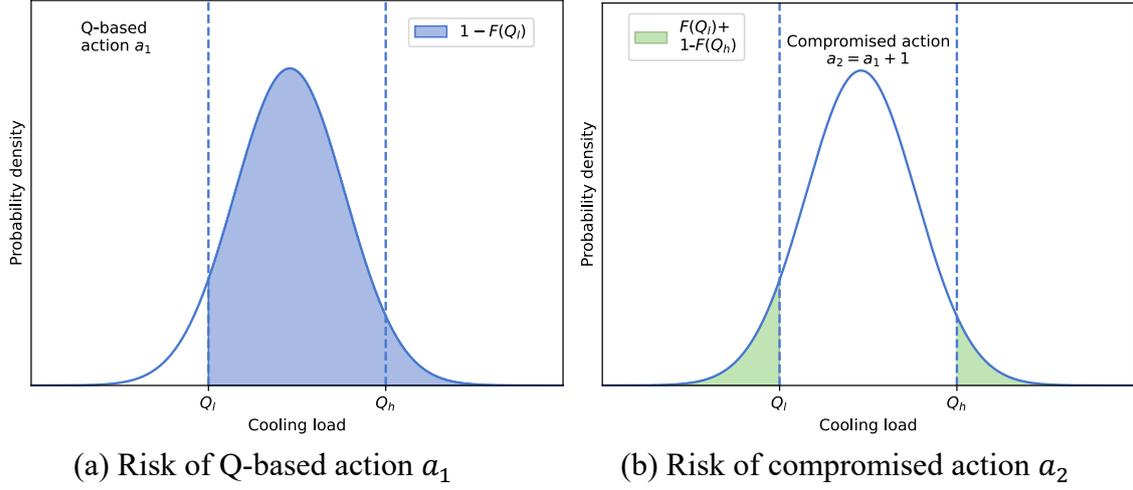


Figure 4.5. Probabilistic cooling load prediction and risks of actions when switching off

4.1.5.2. Scenario 2: Switching on ($a_1 > 0$)

When $a_1 > 0$, the total chiller capacity that Q-based action a_1 can provide is higher. Therefore, the lower chiller capacity limit Q_l and higher chiller capacity limit Q_h are given by Eqs. (4.16) and (4.17), respectively. Like in scenario 1, the risk of the compromised action a_2 in the i -th prediction step is formulated as Eq. (4.18), as shown in the blue area in **Figure 4.6** (a). Moreover, the risk of the Q-based action a_1 in the i -th prediction step is formulated as Eq. (4.19), as shown in the green area in **Figure 4.6** (b). Finally, the weighted risks of a_1 and a_2 are calculated as Eqs. (4.14) and (4.15), and the action with the minimum risk is adopted.

$$Q_l = (n + a_2) \cdot Q_c \quad (4.16)$$

$$Q_h = (n + a_1) \cdot Q_c \quad (4.17)$$

$$r_{t+i\Delta t}(a_2) = 1 - F_{t+i\Delta t}(Q_l) \quad (4.18)$$

$$r_{t+i\Delta t}(a_1) = F_{t+i\Delta t}(Q_l) + 1 - F_{t+i\Delta t}(Q_h) \quad (4.19)$$

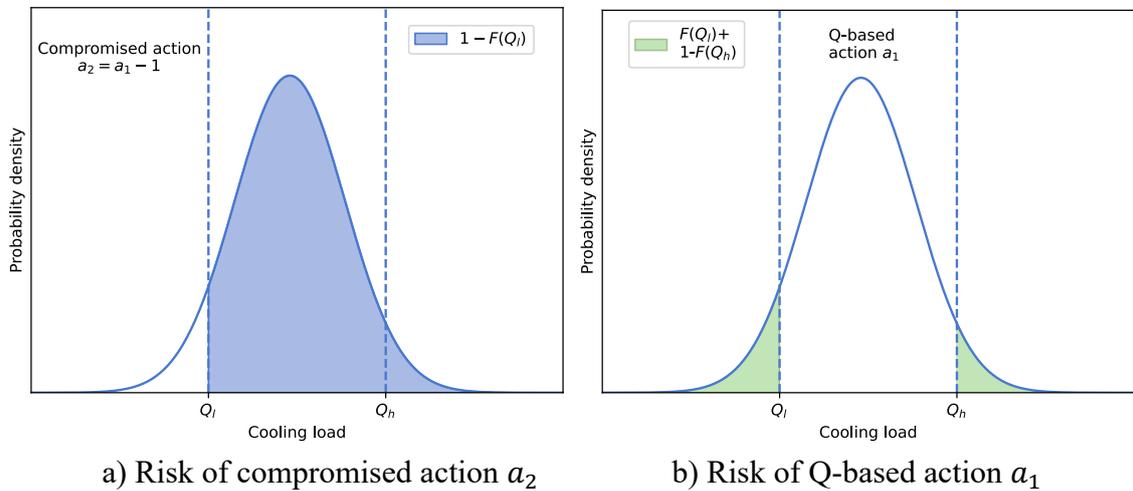


Figure 4.6. Probabilistic cooling load prediction and risks of actions when switching on

4.2. Test facilities and data collection

4.2.1. Chiller plant description

Figure 4.7 depicts the schematic of the chiller plant tested in this study, located at The Hong Kong Polytechnic University (PolyU). On the left side of the diagram, the chilled water system of the chiller plant is depicted, which is a typical primary-constant secondary-variable flow system. The supply main supply and return pipes of the primary loop are connected via a bypass pipe. The return chilled water shown in the diagram is returned from three secondary loops i.e., Phase 1, Phase 2, and library. The diagram also shows ten primary chilled water pumps (PCHWP) numbered 1-10. These ten pumps are connected in parallel and pump water to six water-cooled chillers (WCC) numbered 1-6 and two air-cooled chillers (ACC) numbered 1-2.

On the right side of the schematic, the condensing water loop of the chiller plant, which is used to cool the condenser water of WCC 1-6, is shown. The condensing water loop includes eight condensing water pumps (CDWP) numbered 1-8 and six cooling towers (CT) numbered 1-6. It is noteworthy that the chiller plant employs a combination configuration of large and small chillers, where WCC 1-5 are large chillers with a rated cooling capacity of 650 tons, and WCC 6 and ACC 1-2 are small chillers with a rated cooling capacity of 325 tons. The PCHWP, CDWP, and CT also have corresponding large and small models. The on and off of the PCHWP, CDWP, and CT are interlocked to the corresponding chillers. For example, when a large chiller is switched on, the corresponding large PCHWP, CDWP, and CT are also turned on. The number of PCHWP is two more than the total number of WCC and ACC, and the number of CDWP is two more than the total number of WCC due to the redundant configuration of both large and small PCHWP, each exceeding the number of chillers by one to handle pump failures or maintenance scenarios.

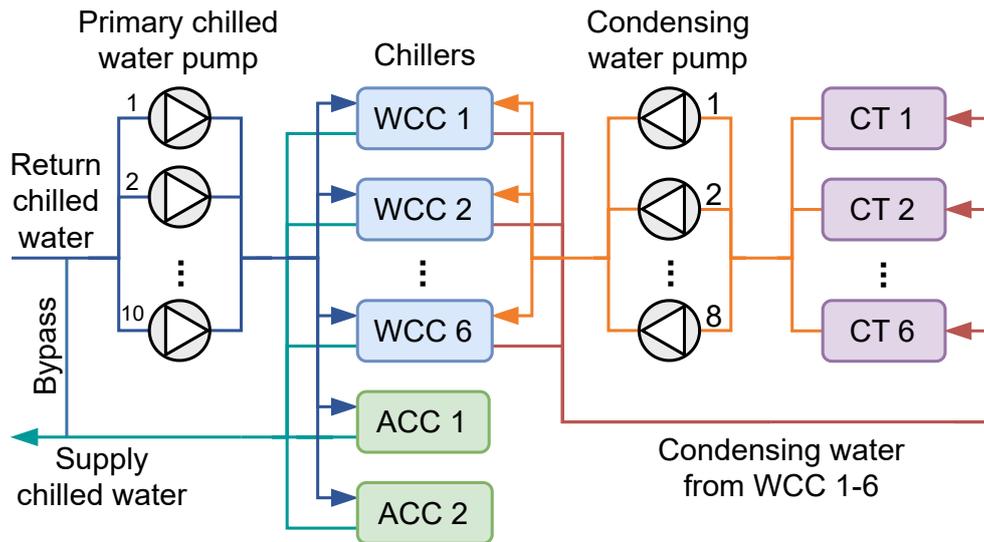


Figure 4.7. Diagram of the chiller plant in PolyU campus

Table 4.2. Specification of air-cooled/water-cooled chillers

Equipment	Model type	Speed	Capacity (RT)	Rated COP	Power (kW)	Flow rate (L/s)
WCC 1-2,5	Large	Variable	650	5.72	400	109.2
WCC 3-4	Large	Constant	650	5.73	399	109.2
WCC 6	Small	Variable	325	5.72	200	56.94
ACC 1-2	Small	Variable	325	3.63	328	56.94

Table 4.3. Specification of primary chilled water pumps and condensing water pumps

Equipment	Model type	Speed	Head (m)	Power (kW)	Flow rate (L/s)
PCHWP 1-3,10	Small	Constant	28	30	56.67
PCHWP 4-9	Large	Constant	33	55	109.2
CDWP 1-6	Large	Constant	32	75	132.0
CDWP 7-8	Small	Constant	40	45	69.0

Table 4.4. Specification of cooling towers

Equipment	Model type	Speed	Power (kW)	Flow rate (L/s)	Heat rejection capacity (kW)
CT 1-4	Large	Constant	30	128.6	2700
CT 5	Large	Constant	18.5	138	2900
CT 6	Small	Constant	15	69	1450

Table 4.2-Table 4.4 show the specifications of chillers, PCHWP, CDWP, and CT, respectively, where the model type indicates whether the equipment is a large or small model. For chillers, **Table 4.2** lists the speed, rated capacity, rated COP, power, and flow rate. **Table 4.3** gives the rated head, power, and flow rate for pumps. **Table 4.4** shows the rated power, flow rate, and heat capacity for cooling towers. The PCHWP, CDWP, and CT are all operated under constant

speed, so their electricity consumption can be estimated using their rated power and on/off signal. The real-time power data of chillers can be obtained from the BMS.

4.2.2. Data communication and processing

During the test, the proposed chiller sequencing control strategy is deployed in the AI workstation on the PolyU campus. **Figure 4.8** illustrates the data communication process between the AI workstation and the BMS. The BMS collects real-time measurements, including the water temperature of chilled and condensing water, flow rate of chilled and condensing water, on/off signals of pumps and cooling towers, chiller power, etc. These measurements are sent in real-time to the gateway via the BACnet protocol. The gateway, acting as an intermediary for data transmission, does not store data itself. Its real-time data is read and stored in 5-minute intervals in the time-series database InfluxDB [223]. The AI engine in the AI workstation uses the QUERY/INSERT command to read historical data and real-time data required by the chiller sequencing control strategy. At 00:00 each day, the NGBoost model will be retrained based on existing historical data and saved locally. At other times, the model will be directly loaded from the local file for prediction without retraining the model. Finally, the AI engine sends the AI recommendation value for chiller sequencing every 30 min via the BACnet protocol to the gateway. The BMS controls chillers by reading the AI recommendations values from the gateway. This achieves bidirectional communication between the AI workstation and BMS.

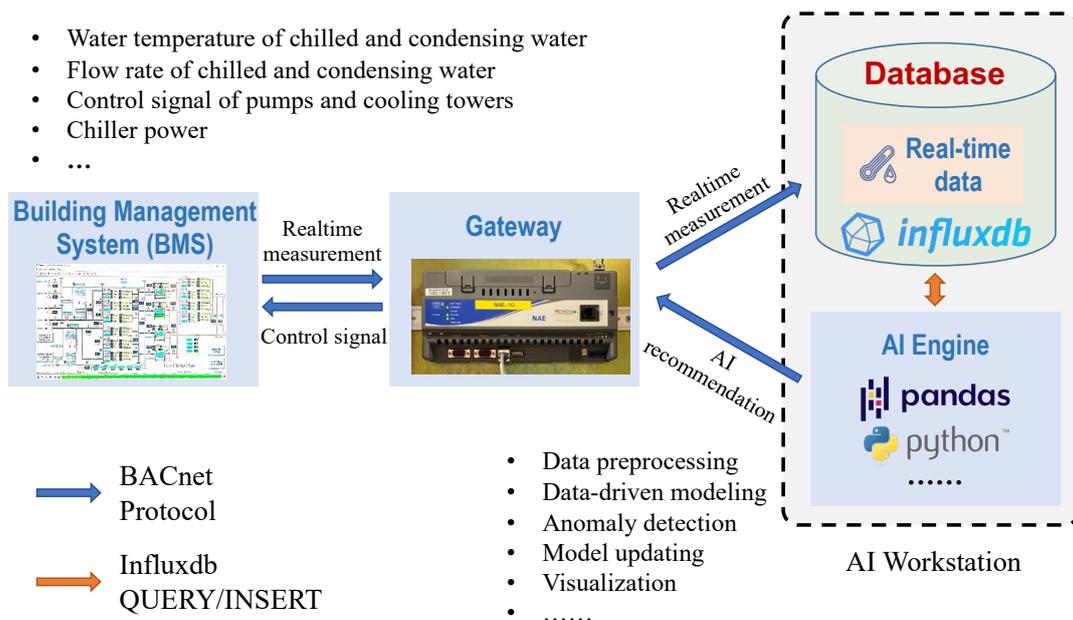


Figure 4.8. Data communication between AI workstation and BMS

4.3. Robustness validation using operational data

4.3.1. Real building operation data

The historical data from January 1 2022 to August 6 2022 was used to train the NGBoost model, and the test set was a one-week-long data set from August 7 2022 to August 13 2022. The test cooling load profile for the data experiment is shown in **Figure 4.9**. Occupancy schedules and weather conditions regularly influence the cooling load. The cooling load on Sunday (August 7 2022) is lower than on other days because of the occupancy schedule. In addition, weather conditions can influence the cooling load in a building by affecting the amount of sensible and latent heat that enters or leaves the building. The features used for one-hour-ahead probabilistic cooling load prediction include weather-related features, calendar features, and lagged cooling load, as presented in Section 4.1.3.

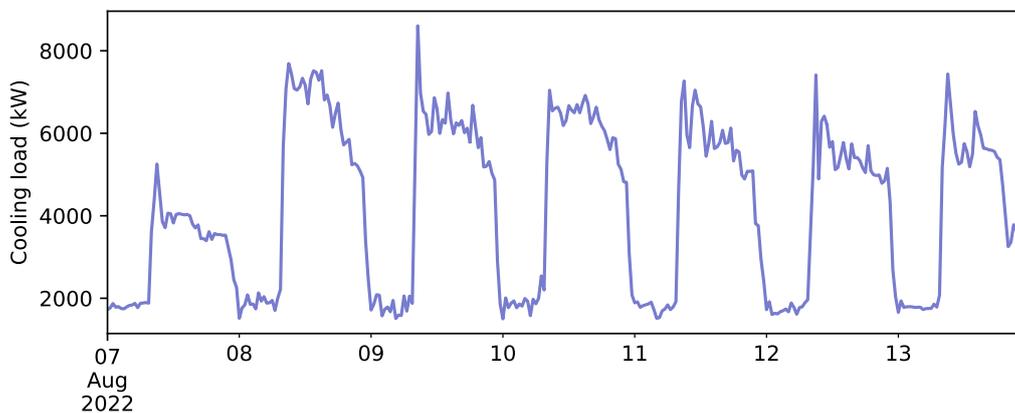


Figure 4.9. Test cooling load profile for the data experiment

4.3.2. Performance of one-hour-ahead probabilistic cooling load predictions

The results of the one-week test cooling load predictions are shown in **Figure 4.10**. The 25%, 50%, 75%, and 95% confidence intervals are filled with different shades of blue colors. The orange line represents the predicted mean values of each probabilistic cooling prediction. The detailed distribution of cooling load prediction at 12:00 on August 13 is shown in the right part of the figure. The mean value of the distribution is 5748.9 kW, and the standard deviation is 551.4 kW.

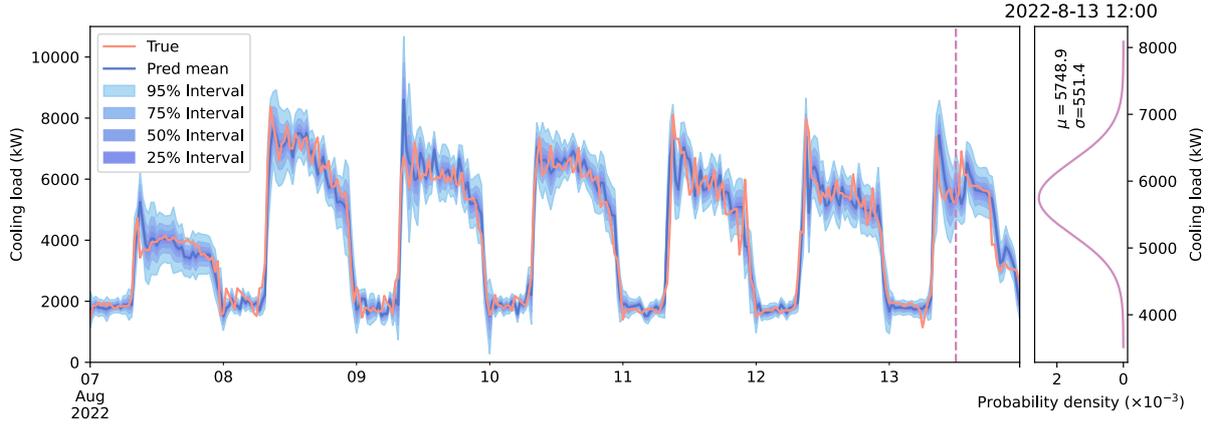


Figure 4.10. One-hour-ahead probabilistic cooling load predictions

The performance of probabilistic cooling load predictions is first evaluated based on mean values of probabilistic cooling load predictions. Root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of variation of root mean square error (CV-RMSE) are used as evaluation indices. The formulas for RMSE, MAPE, and CV-RMSE are given in Eqs. (4.20)-(4.22), respectively. **Table 4.5** shows the performance of probabilistic cooling load predictions based on the true values and predicted mean values. The RMSE, MAPE, and CV-RMSE are 553.3 kW, 9.7%, and 13.2%, respectively, which is acceptable according to a previous study on short-term load prediction [224]. Because of the uncertainty of the cooling load, the cooling load predictions are not accurate enough. Therefore, utilizing the uncertainty information of cooling load predictions is important to generate robust actions.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4.20)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (4.21)$$

$$\text{CV - RMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\frac{1}{n} \sum_{i=1}^n y_i} \times 100\% \quad (4.22)$$

Table 4.5. Performance of probabilistic cooling load predictions

RMSE	MAPE	CV-RMSE
553.3 kW	9.7%	13.2%

The probability integral transform (PIT) is a method that converts data values from any given continuous distribution to random variables having a standard uniform distribution. It can evaluate probabilistic predictions by comparing the observed values with the forecasted densities [225]. One way to apply PIT for evaluation is by inspecting the resulting PIT histogram, which shows the frequency of each transformed value. The PIT histogram should

be approximately uniformly distributed if the probabilistic predictions are consistent and stable. The PIT histogram of the probabilistic cooling load predictions on the whole dataset is shown in **Figure 4.11**, where the dashed red line represents the average frequency. The histogram is approximately uniformly distributed, meaning the consistency and steadiness of probabilistic cooling load predictions are reliable on the dataset.

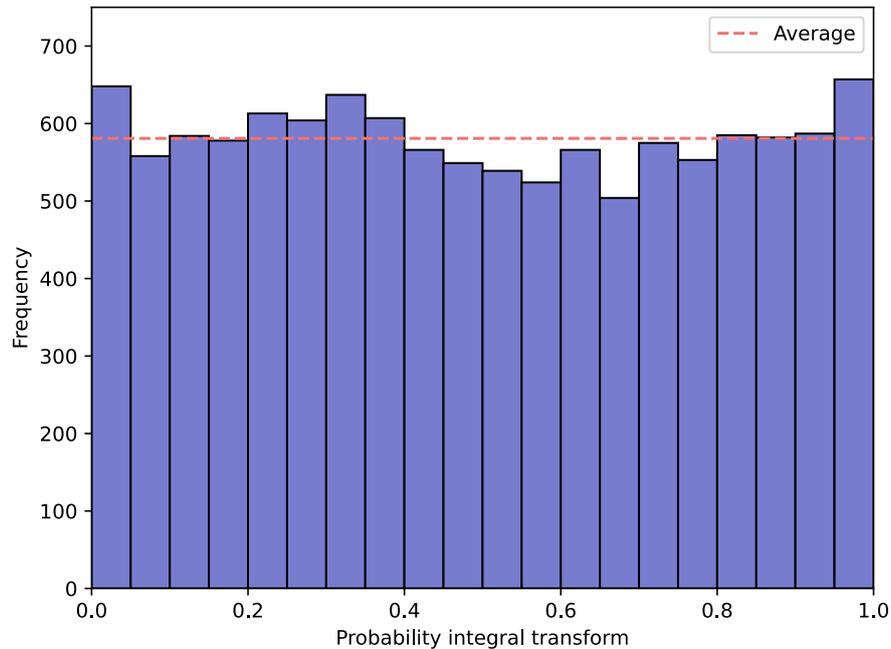


Figure 4.11. Histogram of probability integral transform

Prediction interval averaged width (PIAW) is used to evaluate the reliability of probabilistic predictions. PIAW is a metric that measures the average width between the lower and upper bounds of a prediction interval, as defined in Eq. (4.23), where α is the interval and \hat{y} is the predicted quantile at probability τ [226]. The narrower the PIAW is, the higher the reliability of the prediction is. **Figure 4.12** shows the PIAE at different times on the test set. When the interval range is 25%, the PIAW is mostly below 100 kW, and the average PIAW of all time is 79.8 kW. As for the 50% and 75% intervals, the average PIAWs of all time are 169.0 kW and 288.2 kW, respectively. The figure also shows that the peak of PIAW occurs at 08:30 because occupancy increases significantly, and additional chillers are turned on around this time.

$$PIAW(\alpha) = \frac{1}{N} \sum_{t=1}^N (\hat{y}_{t,\tau=1-\alpha/2} - \hat{y}_{t,\tau=\alpha/2}) \quad (4.23)$$

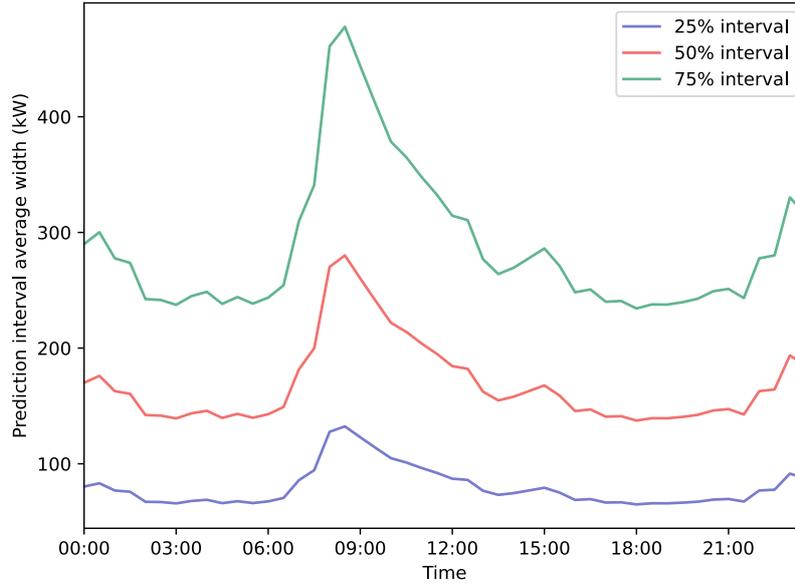


Figure 4.12. Prediction interval averaged width at different times

4.3.3. Chiller sequencing results

Figure 4.13 compares the number of operating chillers between the conventional Q-based and proposed chiller sequencing strategies (with a dead band of $d = 0.05$). In the data experiment, the weights w_1 and w_2 in Eqs (4.13) and (4.14) were set to 0.7 and 0.3, respectively. The results demonstrate that the conventional strategy leads to more unnecessary on/off switchings than the proposed strategy. These frequent on/off switches typically occurred under two conditions: firstly, during the morning start-up period, and secondly, when the measured cooling load fluctuated around the operational chillers' upper or lower cooling capacity boundaries. The morning cooling demand is usually high due to overnight heat accumulation and a sudden increase in cooling load caused by occupant arrival, equipment use, and changes in weather conditions. To meet this large cooling demand, more chillers than necessary often operate in the morning, removing the instantaneous building heat gain. During this period, the measured cooling load undergoes significant variations, typically increasing and then decreasing, leading to unnecessary on/off switching in the conventional Q-based sequencing strategy. The second condition may result from short-term indoor and outdoor disturbances and measurement uncertainties. Generally, eliminating unnecessary switching off does not affect end users' thermal comfort, as the supply cooling capacity exceeds the measured cooling load. The terminal control devices can maintain thermal comfort within the air-conditioned space. However, there is a risk of compromising end users' thermal comfort by neglecting to switch on an additional chiller based on the predicted decrease in cooling load for the following hour.

Nevertheless, the results show that the interval between an unnecessary switch-on and switch-off action is relatively short, approximately 30 minutes. Given the thermal mass of the building and the chilled water system, the risk to end users' thermal comfort is considered low. The actual impact on end users' thermal comfort could not be quantified in the data experiment, representing a limitation of the current work.

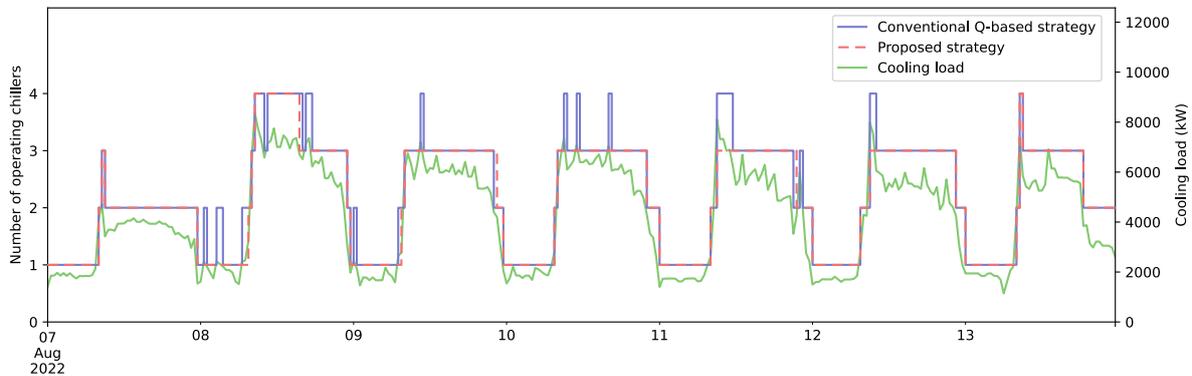


Figure 4.13. Control comparison of conventional and the proposed strategies

The total switching number of chillers is an important metric for evaluating the robustness of chiller sequencing strategies [150]. It is advisable to implement a chiller sequencing control strategy that minimizes the frequency of chiller on/off cycles. Reducing unnecessary switching helps prolong the lifespan of the chiller plant and reduce maintenance expenses. **Table 4.6** shows the control performance of conventional and proposed strategies. Compared to the conventional strategy, the proposed strategy reduces 43.6% of the total switch number (from 55 to 31). In addition to the total switching number, unnecessary on/off actions reflect the robustness of chiller sequencing strategies [13]. Unnecessary on/off actions refer to one chiller being switched on first and then off (or off first and then on) within 0.5 h or 1 h. **Table 4.6** also shows that by adopting the proposed strategy, the number of unnecessary on/off actions in 0.5 h is reduced by 83.3%, and the number of unnecessary on/off actions in 1 h is reduced by 88.2%. It could significantly improve the stability and reliability of the chiller plant.

Table 4.6. Control comparison of two strategies from August 7 to 13, 2022

Strategy	Total switching number	Unnecessary on/off in 0.5 h	Unnecessary on/off in 1 h
Conventional	55	12	17
Proposed	31	2	2
Reduction (%)	43.6	83.3	88.2

Table 4.7 shows the switching details of conventional and proposed strategies. The first two rows indicate the number of operating chillers before (N_t) and after a sequencing action

($N_{t+\Delta t}$), respectively. The third and fourth rows are the count of chiller plant switching from N_t to $N_{t+\Delta t}$ when adopting the conventional and proposed strategy, respectively. For example, when adopting the conventional strategy, the chiller plant switches from $N_t = 1$ to $N_{t+\Delta t} = 2$ for 10 times while the count is 7 when adopting the proposed strategy. According to **Table 4.7**, the proposed strategy mainly contributes to reducing switching on and off chillers when the cooling demand is high, i.e., reducing 85.7% from $N_t = 3$ to $N_{t+\Delta t} = 4$ and 80.0% from $N_t = 4$ to $N_{t+\Delta t} = 3$. When the demand is relatively low, the reduction is not as significant. This is because the cooling load during the daytime this week often exceeds the threshold for switching on the fourth chiller. The conventional strategy would result in unnecessary switching actions. The proposed strategy, on the other hand, avoids unnecessary switching through probabilistic cooling load prediction and risk-based action evaluation.

Table 4.7. Sequencing details from August 7 to 13, 2022

Switch from chiller number N_t	1	2	2	2	3	3	4
Switch to chiller number $N_{t+\Delta t}$	2	1	3	4	2	4	3
On/off count of conventional strategy	10	9	5	3	8	7	10
On/off count of proposed strategy	7	6	6	1	7	1	2
Reduction (%)	30.0	33.3	-20.0	66.7	12.5	85.7	80.0

Note: N_t and N_{t+1} denotes the number of operating chillers at time t and $t + \Delta t$, respectively.

4.3.4. Model interpretation: Risk analysis

The risks of Q-based actions and compromised actions are shown in **Figure 4.14** as the interpretation of the proposed strategy. In most sequencing actions (93.8%), risk-based action evaluation is bypassed, or the risk of Q-based action is less than that of the compromised action. Therefore, the conventional Q-based strategy is reliable under most circumstances. The proposed strategy combines the strength of conventional strategy that engineers can easily interpret. Meanwhile, the proposed strategy leverages probabilistic machine learning to make more robust sequencing control actions. The risk-based action evaluation explains when the proposed strategy recommends different control actions compared to the conventional strategy. **Figure 4.15** illustrates the chiller sequencing results of conventional and the proposed strategies and the risks on August 8. The figure reveals how risk-based action evaluation helps to produce robust sequencing results. At 00:30, according to the conventional Q-based strategy, the measured cooling load exceeds the threshold of switching on an additional chiller. However, according to the proposed strategy, the risk of the compromised action is lower than the Q-based action. It means that although the current cooling load is higher than the threshold, the

risk that one chiller cannot satisfy cooling demand in the following hour (adopting the compromised action) is lower than the risk that an unnecessary chiller is switched on (adopting the Q-based action) according to probabilistic cooling load predictions. At 10:00, the measured cooling load is lower than the threshold of switching off one chiller according to the conventional Q-based strategy. However, according to the proposed strategy, the risk of the compromised action is lower than the Q-based action. Although the current cooling load is lower than the threshold, the risk that an unnecessary chiller is kept running in the following hour (adopting the compromised action) is lower than the risk that three chillers cannot satisfy cooling demand (adopting the Q-based action) according to probabilistic cooling load prediction. At 15:30, the measured cooling load is still higher than the threshold of switching off one chiller according to the conventional Q-based strategy. However, the risk that three chillers cannot satisfy cooling demand in the following hour (adopting the compromised action) is lower than the risk that an unnecessary chiller is switched on (adopting the Q-based action) according to probabilistic cooling load predictions.

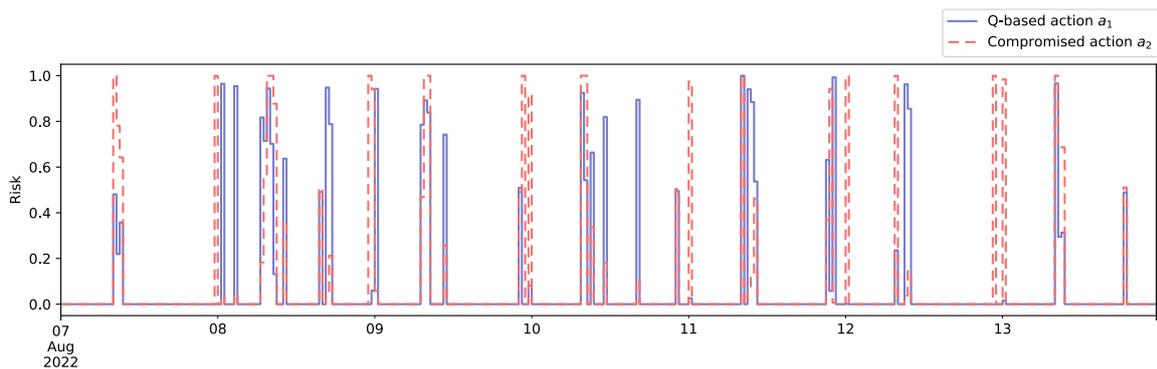


Figure 4.14. Risks of Q-based action and compromised actions on the test data

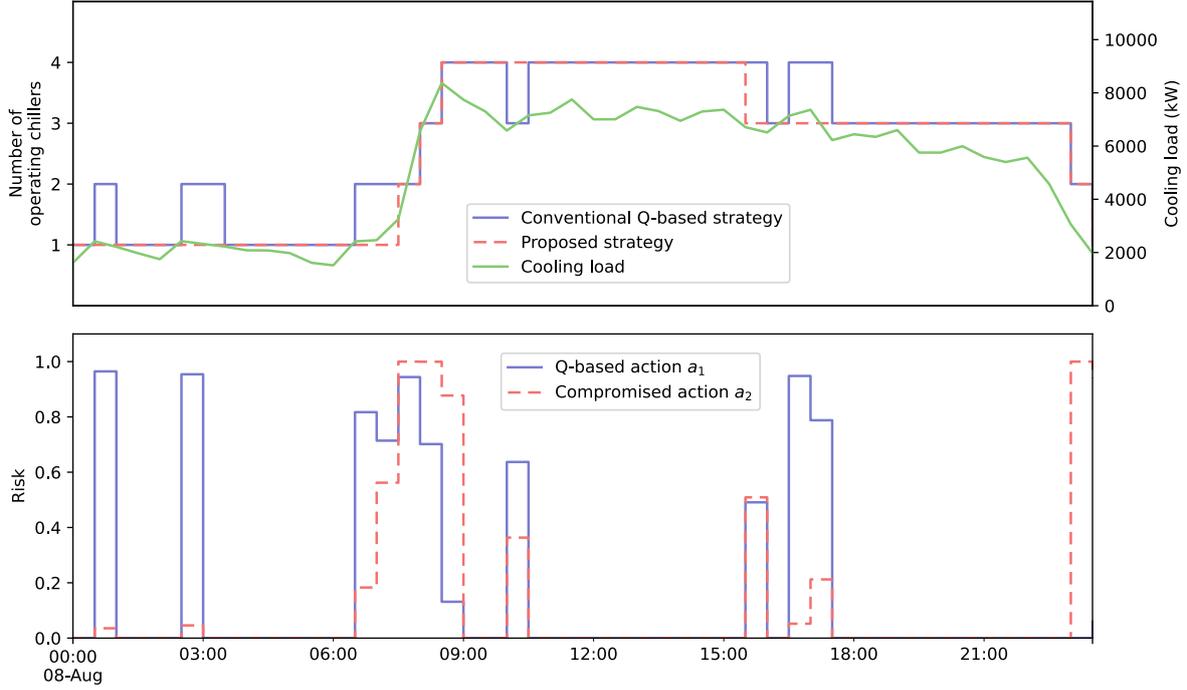


Figure 4.15. Sequencing results and risks evaluation on August 8

4.3.5. Robustness comparison with conventional strategy under different levels of cooling load uncertainty

In practical applications, measurement uncertainties, mainly arising from random uncertainty in flow sensors, can lead to variations in calculated cooling loads. To compare the proposed strategy with the conventional Q-based strategy under varying levels of uncertainty, this section assesses chiller sequencing results for both strategies in different degrees of cooling load measurement uncertainty. Previous studies have indicated that the random error of flow measurement can be considered to follow a normal distribution with a standard deviation [158]. Given that the chilled water pumps typically operate at a constant speed in multiple-chiller systems, the flow rate through one pump remains relatively stable. Depending on the number of chillers activated, it can be simplified as Eq. (4.24), where the measured cooling load Q'_t is the actual value Q_t multiplied by a random variable following a normal distribution, where σ represents the level of uncertainty.

$$Q'_t = Q_t \cdot \mathcal{N}(1, \sigma) \quad (4.24)$$

In this test, the measured cooling load is assumed to be the true value, and five different levels of uncertainty are introduced to the entire dataset: 0%, 5%, 10%, 15%, and 20%. The test result of 0% is the same as the results in Section 4.3 because no uncertainty is introduced. The introduced uncertainty affects chiller sequencing results for the conventional Q-based strategy

because it alters the real-time measurement. However, for the proposed strategy, the introduced uncertainty affects not only the real-time measurement but also the predictions of the NGBoost model, due to the uncertainty present in historical data.

In **Figure 4.16**, a comparison is made between the two strategies in terms of the total switching number at different uncertainty levels. As uncertainty increases, the total switching number for the conventional Q-based strategy significantly increases, with its average value going from 55 to 141.6, while the proposed strategy only increases from 31 to 41.7. This implies that even at an uncertainty level of 20%, the proposed strategy remains more robust than the conventional Q-based strategy. The green line representing the reduction in the total switching number for the proposed strategy compared to the conventional Q-based strategy shows that as uncertainty increases, the average reduction in total switching number increases from 43.6% to 70.4%.

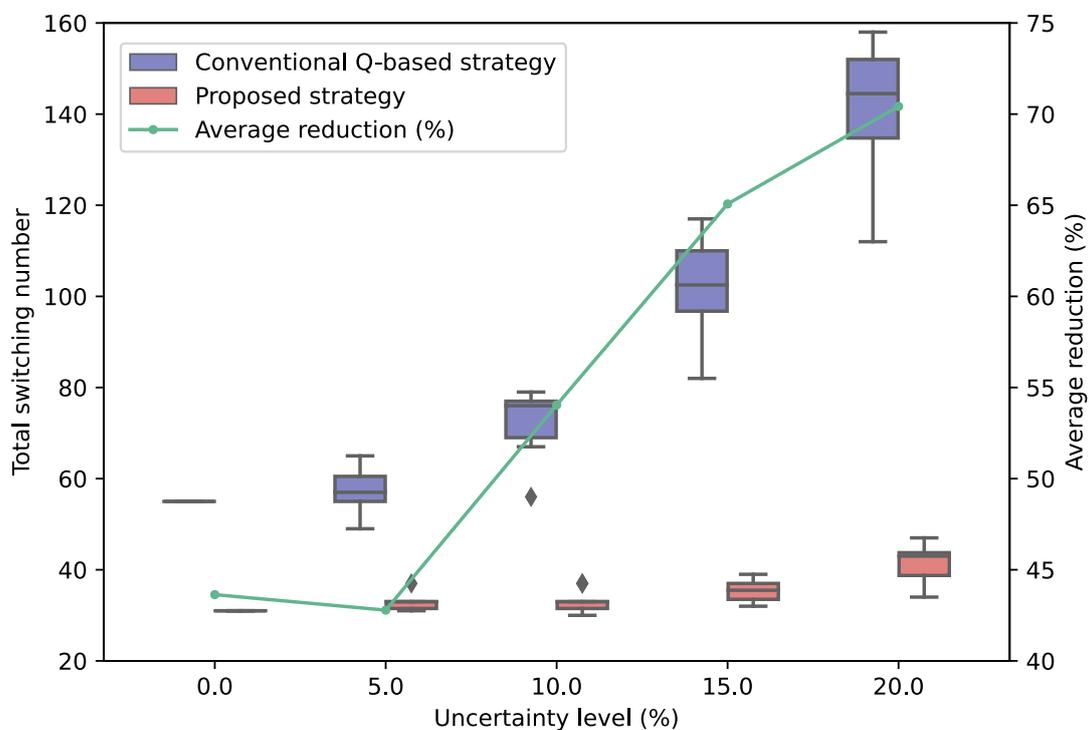


Figure 4.16. Impact of uncertainty level to the total switching numbers

4.4. In-situ test of the proposed strategy

After deploying the proposed strategy, testing was conducted on the chiller sequencing control strategy. Prior to deployment, a rule-based strategy from the building automation company was employed in the BMS, which switched on the chillers based on the time schedule, PLR of the chillers, and chilled water return temperature. Details of the rule-based strategy can be found in the Supplementary Materials.

For a fair comparison, it is necessary to select testing days and reference days with sufficiently similar weather conditions for both the proposed and rule-based strategies. Three sets of testing and reference days were selected based on two criteria. The first criterion required the testing day's weather conditions to be sufficiently similar to those of the reference day. The difference in maximum temperature, average temperature, and minimum temperature between the testing day and reference day should be within 1°C. The difference in average relative humidity should be within 10%. The second criterion required the type of working day to be consistent for the testing and reference days. As an educational building, Monday to Sunday was divided into three types according to the characteristics of cooling load: Monday, Tuesday to Friday, and Saturday and Sunday. Monday was singled out because many classrooms accumulated significant heat gain while closed during the weekend, resulting in greater cooling demand on Monday mornings compared to Tuesday to Friday under similar weather conditions. Finally, one set of testing and reference days was selected for each of these three types of working days. The control effects of the rule-based and proposed strategies were compared in three aspects: robustness of sequencing actions, thermal comfort, and energy efficiency. This section introduces the comparison of the first set of testing and reference days, and the comparison of the other two sets of testing and reference days can be found in Appendix A.

4.4.1. Testing day 1 and weather comparison with reference day 1

Reference day 1 (May 22, 2023) and testing day 1 (June 12, 2023) are both Mondays. The rule-based strategy was used in the BMS on reference day 1, while the proposed strategy was used on testing day 1. Both reference day 1 and testing day 1 were cloudy. A comparison of the weather between reference day 1 and testing day 1 is shown in **Figure 4.17**. The data was collected from the Hong Kong Observatory at 5-minute intervals. The outdoor temperature and relative humidity trends on these two days are very similar. **Table 4.8** compares the highest temperature, average temperature, lowest temperature, and average relative humidity on these two days. From these four values, it can be observed that the weather on these two days is very similar. The highest temperature on testing day 1 (33.6°C) is slightly higher than on reference day 1 (32.8°C).

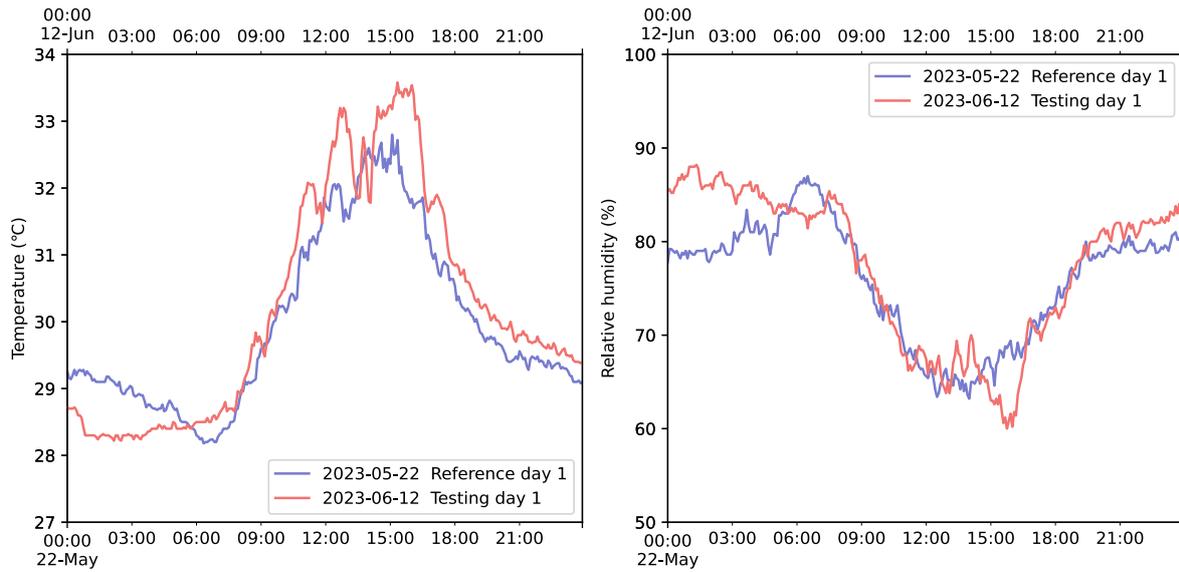


Figure 4.17. Weather conditions on reference day 1 and testing day 1

Table 4.8. Weather summary on reference day 1 and testing day 1

Date	Maximum temperature (°C)	Average temperature (°C)	Minimum temperature (°C)	Average relative humidity (%)
2023-05-22 Reference day 1	32.8	30.0	28.2	76
2023-06-12 Testing day 1	33.6	30.2	28.2	77

4.4.2. Probabilistic cooling load prediction

Figure 4.18 shows the probabilistic cooling load prediction on the testing day. The different shades of blue in the figure represent the 25%, 50%, 75%, and 95% confidence intervals. The width of the 90% confidence interval indicates that the probabilistic cooling load uncertainty is relatively small at night and increases significantly in the morning start-up period. This is because at night, the chillers only need to provide the basic cooling load for the 24-hour rooms (i.e., the 24-hour study center in the library), while in the morning, each room needs to remove the heat gain during the night, leading to larger load fluctuations. On the right side of **Figure 4.18**, the probabilistic cooling load probability distribution at 12:00 on the testing day is shown, with a predicted average value of 7425.7 kW and a standard deviation (i.e., uncertainty) of 495.5 kW. In testing day 1, the root mean square error (RMSE) for comparing the predicted average value of the probabilistic cooling load with the actual value is 407.0 kW, and the mean absolute percentage error (MAPE) is 9.5%.

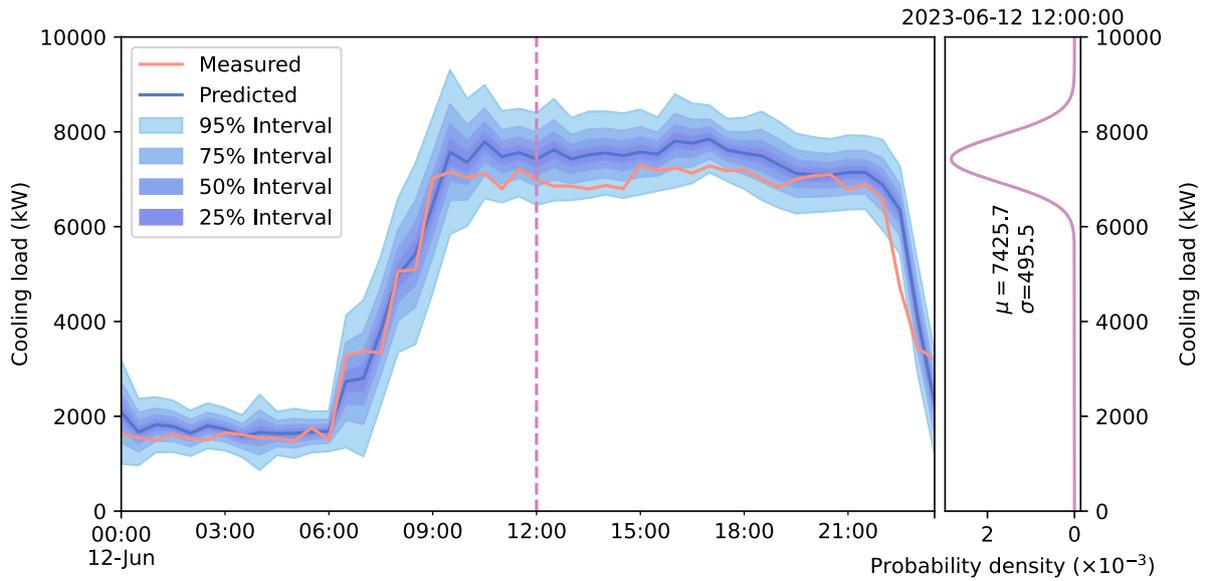


Figure 4.18. One-hour-ahead probabilistic cooling load prediction on testing day 1

4.4.3. Robustness of sequencing actions and thermal comfort

The chiller sequencing results for testing day 1 and reference day 2 are shown in **Figure 4.18**. The y-axis on the left represents the number of chillers in operation, while the y-axis on the right represents the chiller sequencing order. Therefore, by combining the two y-coordinates, the on/off status of each chiller at every moment can be determined on the reference day and testing day. Compared with the reference day under the rule-based strategy, the proposed strategy on the testing day significantly reduces the total switching number of chillers from 15 to 5, as shown in **Table 4.9**. Moreover, the maximum number of chillers turned on in the morning start-up period (6 a.m. to 12 a.m.) is reduced from 5 to 4. This comparison demonstrates that the adoption of the proposed strategy significantly improves the robustness of the chiller sequencing control.

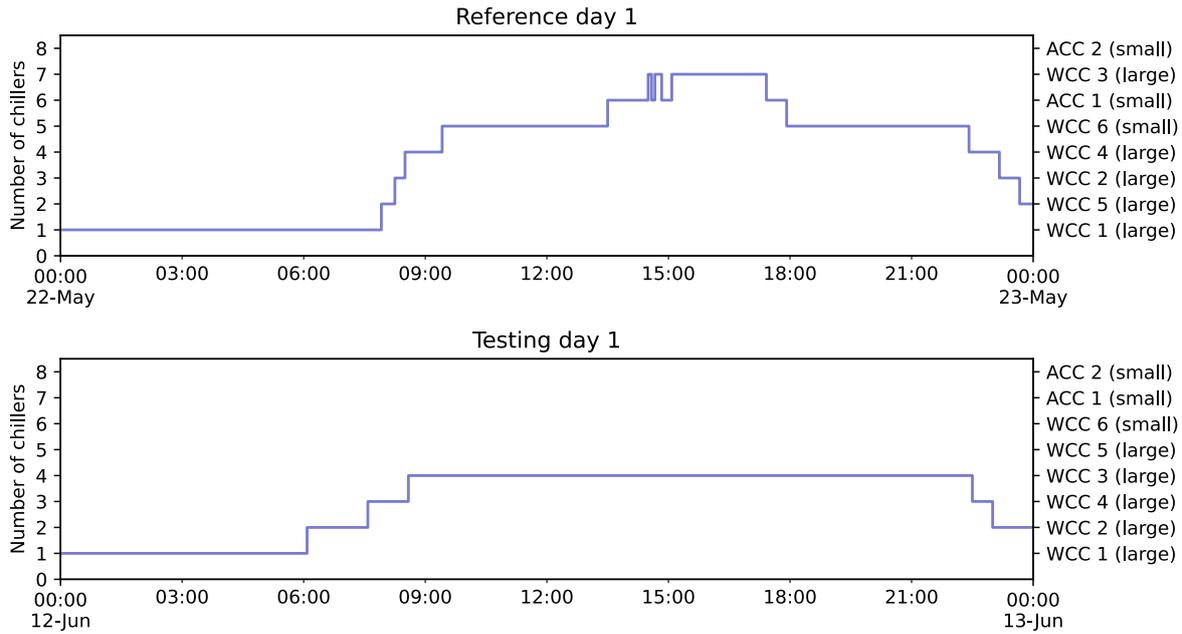


Figure 4.19. Chiller sequencing results on reference day 1 and testing day 1

Table 4.9. Chiller switching number on reference day 1 and testing day 1

Date	Total switching number	Maximum chiller during morning start-up
Reference day 1	15	5
Testing day 1	5	4

Figure 4.20 illustrates the chilled water return temperature of three loops (Phase 1, Phase 2, and Library) on reference day 1 and testing day 1. It can be seen from the figure that there are some problems with the rule-based strategy. Firstly, the highest chilled water return temperatures for Phases 1 and 2 both exceeded 19°C on the testing day due to a lack of cooling load prediction. According to the on-site operator's experience, when the chilled water return temperature of these two secondary loops remains above 16°C for more than half an hour, it can impact indoor thermal comfort. On reference day 1, the chilled water return temperature for Phase 1 was above 16°C from 7:15 to 10:25 (over 3 hours), and for Phase 2, it was above 16°C from 7:40 to 9:45 (over 2 hours), causing insufficient cooling in Phases 1 and 2 and resulting in many thermal complaints. Secondly, the chillers were switched on and off frequently, which increased the instability of the system operation.

When adopting the proposed strategy, the above two issues are significantly alleviated. First, the water temperature was more stable during the testing day. This indicates that the proposed strategy can make more reasonable on/off control of the chiller based on probabilistic cooling load predictions. Second, the control decisions based on probabilistic cooling load predictions

significantly avoid unnecessary on/off switching actions, which makes the chiller system operate more efficiently and effectively.

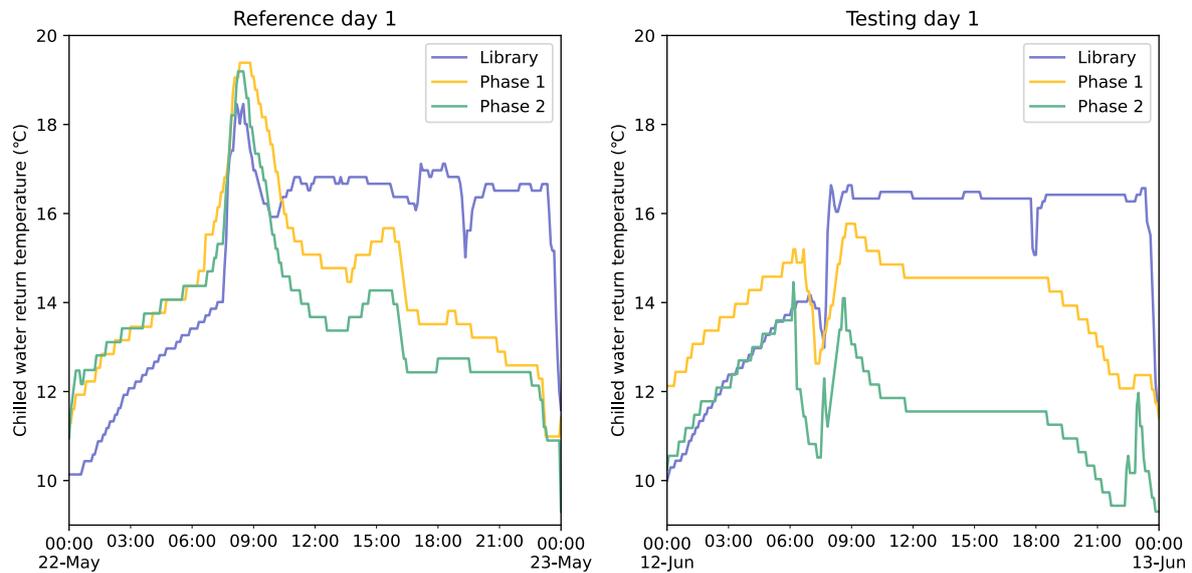


Figure 4.20. Chilled water return temperature of three loops

4.4.4. Energy efficiency

The energy consumption of chillers, primary chilled water pumps (PCHWP), condensing water pumps (CDWP), and cooling towers (CT) on reference day 1 and testing day 1 are shown in **Table 4.10**. Compared with the reference day which adopted the rule-based strategy, the proposed strategy significantly reduced the energy of various equipment on the testing day. The reduction of energy consumption mainly comes from two aspects. The first aspect is that the probabilistic cooling load prediction achieves timely and reasonable switching on, especially in the morning start-up period, which reduces the maximum number of chillers that need to be turned on during the peak period, and reduces the operation of the corresponding PCHWP, CDWP, and CT of the chillers. On the other hand, the timely chiller switching off is realized through the prediction of cooling load, which reduces the energy consumption of chillers and other equipment. Specifically, the energy consumption of chillers, PCHWP, CDWP, and CT are reduced by 1720.5 kWh, 546.7 kWh, 522.5 kWh, and 170.0 kWh, respectively. The total energy consumption is saved by 2959.7 kWh.

Table 4.10. Energy consumption on reference day 1 and testing day 1 (kWh)

Date	Chiller	PCHWP	CDWP	CT	Total
Reference day 1	29334.8	4474.6	5878.8	2035.5	41723.7
Testing day 1	27614.3	3927.9	5356.3	1865.5	38764.0
Difference	1720.5	546.7	522.5	170.0	2959.7

Table 4.11 provides a comparison of the energy efficiency of the entire chiller plant on reference day 1 and testing day 1. Owing to the proposed strategy, the system COP is improved by about 3.7%.

Table 4.11. Energy efficiency of chiller plant on reference day 1 and testing day 1

Date	Total energy consumption (kWh)	Total cooling load (kWh)	COP
Reference day 1	41723.7	125644	3.02
Testing day 1	38764.0	121067	3.13
Difference	-7.1%	-3.6%	3.7%

4.4.5. Testing day 2 and testing day 3

Given that the results of testing day 2 and testing day 3 are similar to those of testing day 1, the comparison of these two days with their respective reference day can be found in **Appendix A**.

4.4.6. Summary of the in-situ test

The chiller switching numbers on reference days 1-3 and testing days 1-3 are summarized in **Table 4.12**. In the three sets of reference and testing days, the daily chiller switching numbers were reduced by 66.7%, 62.5%, and 14.3%, respectively. The overall reduction of the three sets of days is 56.5%. The decrease in chiller switching numbers in the third set was not as significant because reference day 3 and testing day 3 are both Sundays, so the building cooling load is relatively low and stable. Therefore, the previous rule-based strategy can achieve a relatively robust chiller sequencing control. The reduction in chiller switching numbers demonstrates the superiority of the proposed strategy in terms of robustness over the rule-based strategy. Especially in dealing with rapidly changing cooling loads, the proposed strategy can make more timely switching-on actions based on probabilistic cooling load predictions, thereby avoiding switching on unnecessary chillers within a short period.

Table 4.12. Chiller switching number on reference and testing days 1-3

Date	Switching number on day 1	Switching number on day 2	Switching number on day 3	Switching number on days 1-3
Reference day	15	24	7	46
Testing day	5	9	6	20
Difference	-66.7%	-62.5%	-14.3%	-56.5%

Table 4.13 compares the energy consumption for the reference days 1-3 and testing days 1-3 of each type of equipment. The daily average energy consumption of chillers, PCHWP, CDWP, and CT were reduced by 2269.1 kWh, 685.0 kWh, 688.7 kWh, and 302.2 kWh. The energy savings for the chillers were the most significant. Because the chiller sequencing actions were

based on probabilistic cooling load prediction, the cooling supply matched the cooling demand to the greatest extent possible, thereby improving the PLR and the COP. For the PCHWP, CDWP, and CT, the energy savings resulted from reduced chillers during peak load periods and more timely switching-off control. Overall, the in-situ test demonstrated that the proposed strategy yielded an average daily energy savings of approximately 3945.1 kWh.

Table 4.13. Energy consumption on reference days 1-3 and testing days 1-3 (kWh)

Date	Chiller	PCHWP	CDWP	CT	Total
Reference days 1-3	81411.6	13453.8	17610.1	6293.2	118768.6
Testing days 1-3	74604.2	11398.8	15543.9	5386.5	106933.2
Difference	6807.4	2055	2066.2	906.7	11835.4
Daily average difference	2269.1	685.0	688.7	302.2	3945.1

Table 4.14 lists the energy efficiency of the chiller plant on the three sets of days reference and testing days. Owing to the proposed strategy, the system COP is improved by about 4.2% on average.

Table 4.14. Energy efficiency of chiller plant on reference days 1-3 and testing days 1-3

Date	Total energy consumption (kWh)	Total cooling load (kWh)	COP
Reference days 1-3	127347.3	380723	2.99
Testing days 1-3	116620.3	363369	3.12
Difference	-8.4%	-4.6%	4.2%

4.5. Summary

Multiple-chiller systems are widely adopted in large buildings due to their high flexibility and efficiency in providing cooling capacity. A reliable and robust chiller sequencing control strategy is crucial to ensure the energy efficiency and stability of the multiple-chiller systems. Therefore, this study proposes a robust chiller sequencing control strategy that utilizes probabilistic cooling load predictions. One-hour-ahead probabilistic cooling load prediction in the form of the normal distribution is made using NGBoost. Compared to conventional machine learning algorithms, NGBoost can predict not only the future cooling load but also the uncertainty of the predicted cooling load, which enables the load prediction to handle the uncertainties associated with the data/measurements adequately. A novel risk-based sequencing strategy is developed based on the probabilistic cooling load predictions. The risk interpretation can enhance the trust between the prediction model and the operators. The data experiment shows that the proposed strategy can significantly improve the stability and reliability of the chiller plant by reducing the total switching number by up to 43.6% (from 55

to 31 in a week). An in-situ test was conducted in an education building with a multiple-chiller system to validate the proposed strategy's superiority in robustness, thermal comfort, and energy efficiency over the conventional rule-based strategy. The test results show that, compared to rule-based strategies, the average daily chiller switching frequency decreased by 56.5%, resulting in an average daily energy savings of approximately 3945.1 kWh while maintaining thermal comfort. The chiller plant's coefficient of performance (COP) increased by an average of 4.2%.

CHAPTER 5 PHYSICALLY CONSISTENT DATA-DRIVEN MODELS FOR ENHANCING INTERPRETABILITY AND RELIABILITY

This chapter introduces a physically - consistent data - driven method to tackle the problem of **physical inconsistency**, enhancing interpretability and reliability. Although variable speed parallel pumping is widely adopted in HVAC water systems, optimal pump sequencing is usually overlooked. This study proposes a physically consistent data-driven optimal sequencing control, integrating physical rules with data-driven modeling, enhancing the interpretability and reliability for real applications. For different numbers of operating variable speed pumps (VSPs), features composed of frequency and total flow rate are used to train the interpretable ElasticNet data-driven model. The optimal sequencing is then determined using a physically consistent speed/frequency prediction method. Section 5.1 introduces the proposed physically consistent data-driven optimal sequencing control strategy. Section 5.2 details the validation of the proposed strategy using real building operational data. Finally, Section 5.3 summarizes the main conclusions of this study.

5.1. Proposed physically consistent data-driven optimal sequencing control strategy

Figure 5.1 illustrates the proposed physically consistent data-driven modeling and online optimal sequencing method. The left section describes the process of physically consistent data-driven modeling. It starts with historical operational data, including total water flow rate, pump speed or frequency, number of operating pumps, and total pump power. Based on the **operating point theorem** (see Section 5.1.1), interpretable ElasticNet models are developed, expressing total pump power as a polynomial function of water flow rate and pump speed for different numbers of operating pumps. Additionally, a physically consistent data-driven model is constructed to predict pump speed after switching while maintaining the same flow rate and head, where the pumps in parallel rule and affinity law are used. The right section illustrates the online optimal sequencing process. Real-time measurements provide the operating point, number of operating pumps, and total pump power. The predicted pump power for different pump numbers is calculated, and the optimal pump configuration is determined by selecting the number that minimizes power consumption.

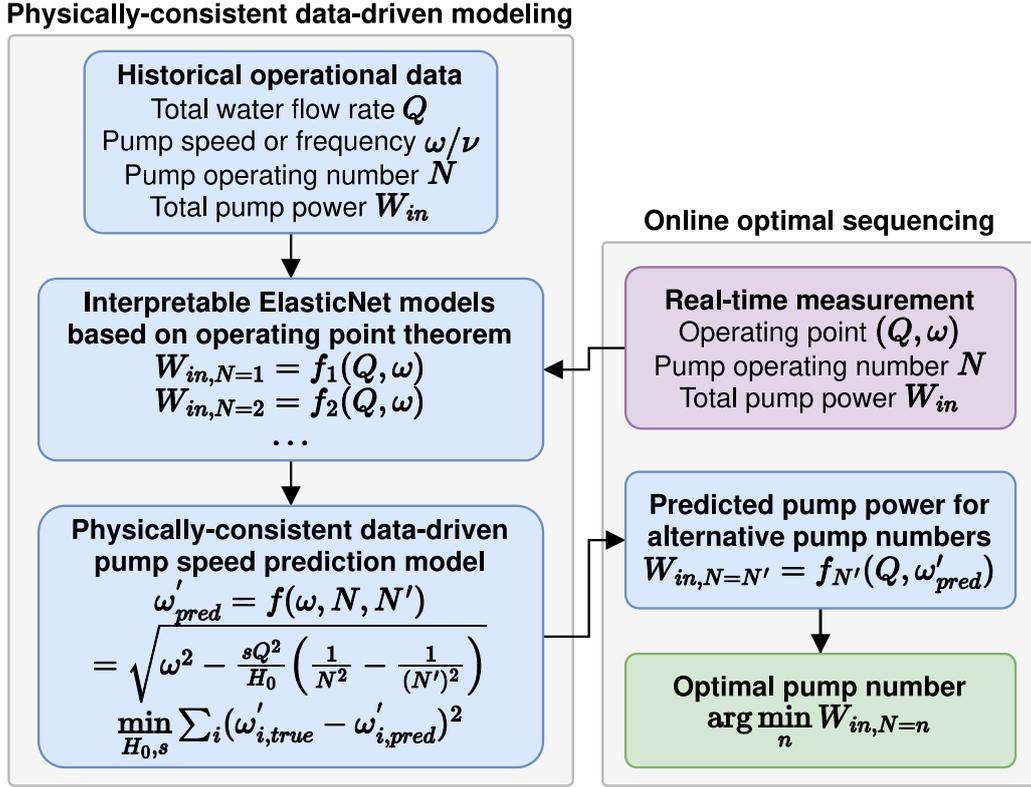


Figure 5.1. Overview of the proposed method

5.1.1. Derivation of operating point theorem

In this section, it is demonstrated that, similar to the total water flow rate Q and the pump head H , a combination of the total water flow rate Q and the speed of the operating pump ω is also adequate to represent the operating point.

For a parallel pump system, when the number of operating pumps N is determined, the operating point can be uniquely determined by the total water flow rate Q and the pump head H . Therefore, the speed (frequency) of operating pumps can be formulated as a function of the operating point (Q, H) :

$$\omega = f_\omega(Q, H) \quad (5.1)$$

To prove that a pair of total water flow rate Q and the speed of operating pump ω is also sufficient to represent the operating point, that is, there is an implicit function $f_{H,N}$ that satisfy:

$$H = f_H(Q, \omega) \quad (5.2)$$

First, an auxiliary function $F(Q, H, \omega)$ is defined as:

$$F(Q, H, \omega) = \omega - f_\omega(Q, H) = 0 \quad (5.3)$$

According to the physical characteristics of the water pump, when the flow rate remains constant, an increase in the speed of the water pump will also increase the pump's head:

$$\frac{\partial \omega}{\partial H} = \frac{\partial f_{\omega}(Q, H)}{\partial H} > 0 \quad (5.4)$$

Therefore, according to the implicit function theorem [227], if $\frac{\partial F}{\partial H}$ is non-zero, which here is:

$$\frac{\partial F}{\partial H} = \frac{\partial}{\partial H} (\omega - f_{\omega}(Q, H)) = -\frac{\partial f_{\omega}(Q, H)}{\partial H} \neq 0 \quad (5.5)$$

There exists a locally unique function $H = g_H(Q, \omega)$ satisfying Eqs. (5.1) and (5.3). This proves Eq. (52), which means that the pump head H can be uniquely determined given total water flow rate Q and the speed of operating pump ω . When the number of operating pumps N is considered as an independent variable, Eq. (5.2) becomes:

$$H = f_H(Q, \omega, N) \quad (5.6)$$

5.1.2. Derivation of pump power model

The pump efficiency of a pump is a function of the ratio of water flow rate Q to the square root of the pump head H :

$$\eta_{pump} = f\left(\frac{Q}{\sqrt{H}}\right) = f_{\eta, pump}(Q, H) \quad (5.7)$$

Therefore, the shaft power of the pump can also be formulated as the function of Q and H :

$$W_{shaft} = \frac{W_{water}}{\eta_{pump}} = \frac{Q \cdot H}{f_{\eta, pump}(Q, H)} = f_{W, shaft}(Q, H) \quad (5.8)$$

motor efficiency is a function of voltage V , frequency ν , and the pump shaft power W_{shaft} :

$$\eta_{motor} = f_{\eta, motor}(V, \nu, W_{shaft}) \quad (5.9)$$

Because the voltage is usually constant and the frequency ν is directly proportional to the pump speed ω , the motor efficiency can be represented using the pump speed ω , water flow rate Q , and the pump head H :

$$\eta_{motor} = f_{\eta, motor}(Q, H, \omega) \quad (5.10)$$

The efficiency of a VFD can be formulated as a function of the pump speed ω :

$$\eta_{vfd} = f_{\eta, vfd}(\omega) \quad (5.11)$$

Combining Eqs. (5.7), (5.10), and (5.11), the overall wire-to-water efficiency can be formulated as:

$$\eta_{wwe} = \eta_{pump} \cdot \eta_{motor} \cdot \eta_{vfd} = f_{\eta,wwe}(Q, H, \omega) \quad (5.12)$$

According to Section 3.2.1, pump head H can be uniquely determined given total water flow rate Q and the speed of operating pump ω . Therefore, the variable H in Eq. ((5.12) can be eliminated:

$$\eta_{wwe} = f_{\eta,wwe}(Q, \omega) \quad (5.13)$$

Therefore, the input power of a variable speed pump can be formulated as:

$$W_{in} = \frac{W_{water}}{\eta_{wwe}} = \frac{Q \cdot H}{f_{\eta,wwe}(Q, \omega)} \quad (5.14)$$

Similarly, the variable H in Eq. (5.14) can be eliminated:

$$W_{in} = f_{W,in}(Q, \omega) \quad (5.15)$$

Finally, if the number of operating pumps is considered as a variable, the wire-to-water efficiency and input power of parallel pumps can be formulated as:

$$\eta_{wwe} = f_{\eta,wwe}(Q, \omega, N) \quad (5.16)$$

$$W_{in} = f_{W,in}(Q, \omega, N) \quad (5.17)$$

In practical parallel pumping systems, using the pump speed as a basis for the pump's power model is more reliable than estimating the pump head through differential pressure between discharge and suction pipes.

5.1.3. Interpretable data-driven pump modeling using ElasticNet

ElasticNet is a popular interpretable machine learning algorithm that adopts regularization techniques to prevent overfitting and improve model performance of linear regression models, especially when dealing with highly correlated data or when the number of predictors (features) is much larger than the number of observations. It combines the properties of two popular methods: Ridge (L2 regularization) and Lasso (L1 regularization). Like Lasso, ElasticNet uses regularization to shrink the coefficients of less important features towards zero. This effectively performs feature selection, leading to simpler models with fewer variables. This makes it easier to understand which features are truly driving the predictions.

The cost function in ElasticNet is a combination of the cost function in linear regression and two regularization penalties:

$$\text{Cost}(\mathbf{w}) = \frac{1}{2N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \alpha \left(\lambda \|\mathbf{w}\|_1 + (1 - \lambda) \frac{1}{2} \|\mathbf{w}\|_2^2 \right) \quad (5.18)$$

where N is the number of observations, y_i and \hat{y}_i are the actual and predicted values respectively, w is the vector of coefficients, α is the regularization parameter controlling the overall strength of regularization, and λ is the mixing parameter controlling the balance between Lasso ($\lambda = 1$) and Ridge ($\lambda = 0$) penalties.

The first part, $\frac{1}{2n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$, represents the sum of squared residuals, which is the cost function used in ordinary least squares. It measures how close the model's predictions are to the actual values. The second part, $\alpha \left(\lambda \| \mathbf{w} \|_1 + (1 - \lambda) \frac{1}{2} \| \mathbf{w} \|_2^2 \right)$, is the penalty term. It combines Lasso's penalty ($\| \mathbf{w} \|_1$), which can shrink some coefficients completely to zero (thus performing variable selection), with Ridge's penalty ($\frac{1}{2} \| \mathbf{w} \|_2^2$), which shrinks the coefficients but typically does not set them exactly to zero. By adjusting α and λ , operators can control the balance between fitting the model accurately and keeping the model simple to avoid overfitting.

5.1.4. Physically consistent pump speed/frequency prediction after switching

If the number of operating pumps changes from N to N' , the total water flow rate remains the same, the pump speed of the parallel pump system is modulated from ω to ω' , and the input power is:

$$W'_{in} = f_{W,in}(Q, \omega', N') \quad (5.19)$$

In Eq. (5.19), the new pump speed ω' is unknown. Therefore, if the new pump speed ω' can be reasonably predicted, the pump operating number with the lowest power is adopted as the optimal pump operating number.

The head-flow curve of one pump is assumed as:

$$H = H_0 - sQ^2 \quad (5.20)$$

where H_0 and s are two unknown coefficients. Although the manufacture curve can be obtained, it could vary from the practical installation. Therefore, this study proposes a physically consistent pump speed/frequency prediction method.

Based on Eq. (5.20) and the affinity law for pumps, when N pumps operate in parallel and the pump speed is ω (the ratio of current speed to the maximum speed), the head-flow curve for each pump becomes:

$$\frac{H}{\omega^2} = H_0 - s \left(\frac{Q}{N\omega} \right)^2 \quad (5.21)$$

Therefore, the head-flow curve for each pump when N' pumps operate in parallel and the new pump speed is ω' becomes:

$$\frac{H}{(\omega')^2} = H_0 - s \left(\frac{Q}{N'\omega'} \right)^2 \quad (5.22)$$

where the pump head H and the total water flow in the parallel pump system remain the same as Eq. (5.21).

By solving Eqs. (5.21) and (5.22), simultaneously, the new pump speed ω' can be obtained:

$$\omega'_{pred} = \sqrt{\omega^2 - \frac{sQ^2}{H_0} \left(\frac{1}{N^2} - \frac{1}{(N')^2} \right)} \quad (5.23)$$

In a practical parallel pump system, when the number of operating pumps changes, Eq. (5.23) can be used to predict the frequency after switching. Because H_0 and s are unknown, the following optimization problem is used to find the optimal using operational data:

$$\min_{H_0, s} \sum_i (\omega'_{i,true} - \omega'_{i,pred})^2 \quad (5.24)$$

$$\text{s.t. } H_{0,min} \leq H_0 \leq H_{0,max}$$

$$s_{min} \leq s \leq s_{max}$$

where $\omega'_{i,true}$ is the true pump speed after switching, $\omega'_{i,pred}$ is the predicted pump speed after switching using Eq. (5.23), $H_{0,min}$, $H_{0,max}$, s_{min} , and s_{max} are the lower and upper bounds of H_0 and s , respectively.

5.2. Validation using real building operational data

5.2.1. Parallel pump system description

In the data experiment, operational data from the parallel pumping system at The Hong Kong Polytechnic University was collected. This parallel pumping system utilizes a primary-secondary chilled water system. Within the Secondary chilled water loop, there are three secondary chilled water pumps (SCHWPs 1-3) in parallel, model 250×150 CNJA, each driven by a 4-pole synchronous speed motor (50 Hz) and equipped with a variable speed drive (VFD). The design working conditions of the pumps are detailed in **Table 5.1**. The system allows for a maximum of two pumps to operate simultaneously, with the third pump serving as back up. The secondary chilled loop utilizes conventional differential pressure control, adjusting the differential pressure setpoint of the furthest cooling coil to meet the chilled water supply needs. This loop has two differential pressure setpoints, 60 kPa and 110 kPa, designed to satisfy the

demands during operational hours (7 am - 11 pm) and non-operational hours (11 pm - 7 am), respectively.

Table 5.1. Design working conditions of the variable speed pumps

Pump head (m)	Flow rate (L/s)	Motor power (kW)	Full speed (RPM)
57	125	110	1480

The operational data includes the total water flow rate of the parallel pumping system, as well as the frequency and power of each pump. As shown in **Figure 5.2**, when only one pump operates, and its frequency is fixed, the power of the pump varies approximately linearly with the total flow rate. The total water flow rate ranges from 50 L/s to 200 L/s. When the total water flow rate exceeds 130 L/s, the lowest frequency of 30 Hz cannot meet the differential pressure setpoint requirement, and the minimum operating frequency of the pump gradually increases. According to control area theory, the minimum head of the control area increases with the increase in the system flow rate. Therefore, when only one pump operates, a frequency greater than 30 Hz is needed to meet the minimum head requirement, and the required minimum operating frequency gradually increases. The lower boundary of the control area remains relatively flat only when the friction loss ratio r approaches 0. When two pumps operate and their frequency is fixed, the power of the pumps also varies approximately linearly with the total flow rate. Unlike operating a single pump, with the increase of the total water flow rate, the lowest frequency of 30 Hz can still meet the minimum head requirement in the control area.

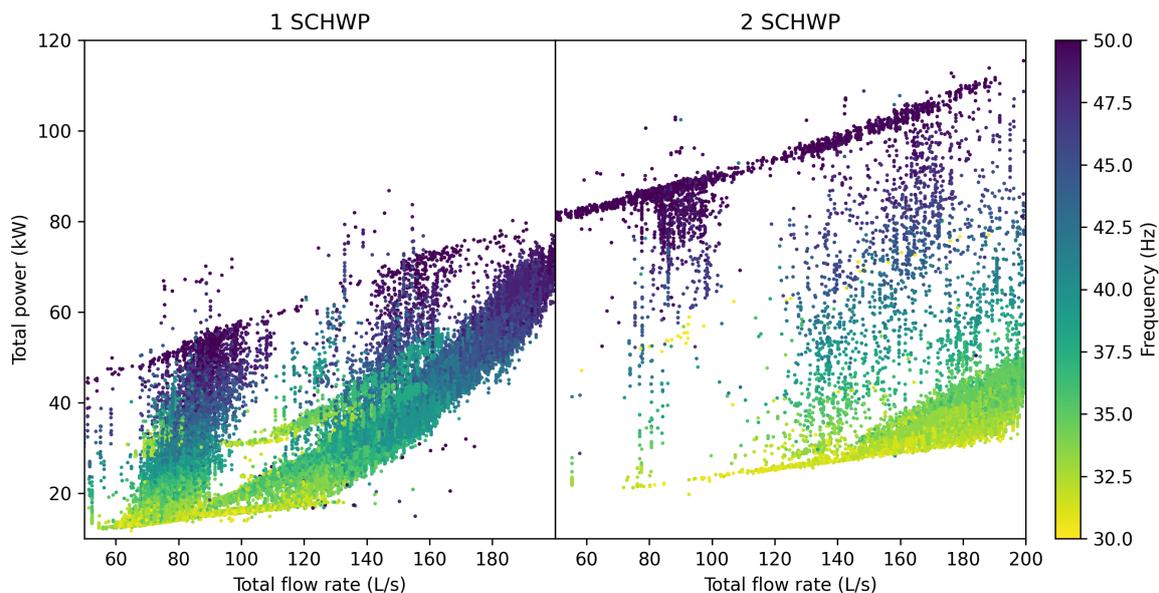


Figure 5.2. Total pump power operating 1 and 2 SCHWPs

5.2.2. Pump power models

Firstly, historical data operating one SCHWP ($N = 1$) were selected, and polynomial terms up to the power of 2 were created for the pump frequency ν and flow rate Q . Then, each term was processed with a standard scaler, and ElasticNet was used for training, resulting in the equation as shown in Eq. (5.25). After normalization, the first-order term of the flow rate Q was discarded, making the equation more concise and improving the interpretability of the model [109]. From the coefficients of the model, it can be seen that the square of the frequency has the greatest influence on the pump power, while the influence of other terms is relatively close. **Figure 5.3** shows the comparison between the predicted power and the actual power of the model. The coefficient of determination R^2 is 0.938, achieving accurate and interpretable predictions of pump power consumption through frequency and flow rate.

$$W_{in,N=1} = 3.3745\nu + 5.8252\nu^2 + 3.2236\nu Q + 3.7266Q^2 + 29.4660 \quad (5.25)$$

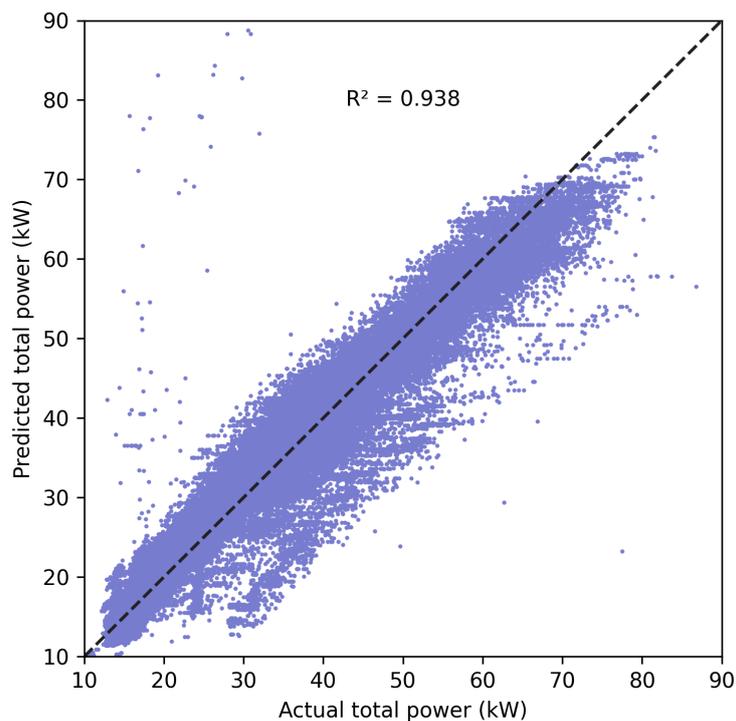


Figure 5.3. Power predictions of 1 SCHWP

Similarly, historical data operating two SCHWPs ($N = 2$) were filtered, and the ElasticNet model was trained using the same steps as above, resulting in the equation as shown in Eq. (5.26). After normalization, the first-order term of the flow rate Q was also discarded like Eq. (5.25). The square of the frequency has the greatest influence on the pump power, followed by the first power of the frequency. **Figure 5.4** shows the comparison between the predicted power

and the actual power of the model. The R^2 is 0.953, indicating slightly higher accuracy than when the $N = 1$.

$$W_{in,N=2} = 8.1660v + 12.5046v^2 + 4.4835vQ - 1.2526Q^2 + 50.5333 \quad (5.26)$$

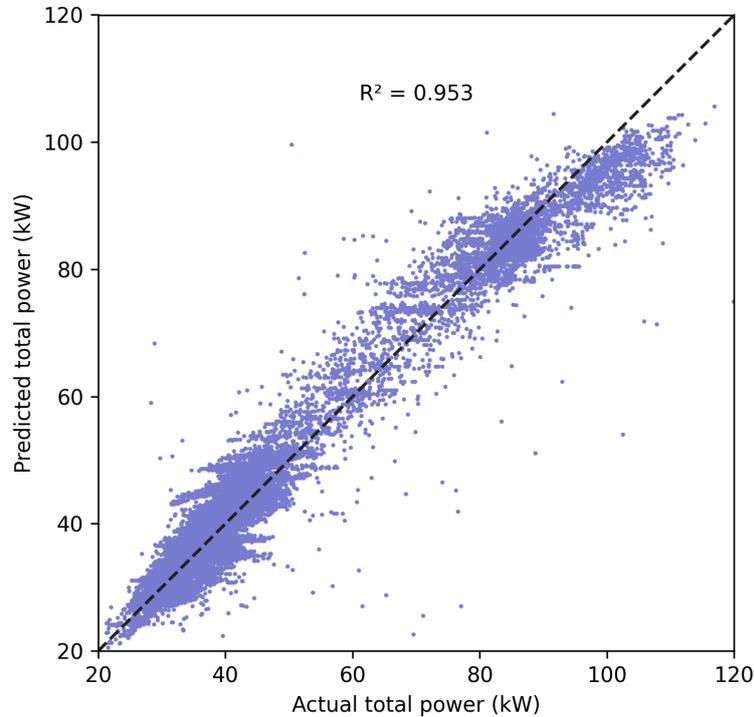


Figure 5.4. Power predictions of 2 SCHWPs

5.2.3. Frequency prediction after sequencing

In this section, the frequency after switching from 2 to 1 SCHWP is predicted. The historical data collected at 5-minute intervals selected data with a relatively stable total flow rate before and after the switch (the total system flow change does not exceed 10 L/s). **Figure 5.5** shows the prediction results of the frequency, with an R^2 of 0.820 and a mean absolute error of 0.773Hz. This prediction error is acceptable because the maintenance of the differential pressure setpoint is a dynamic process after the pump switch.

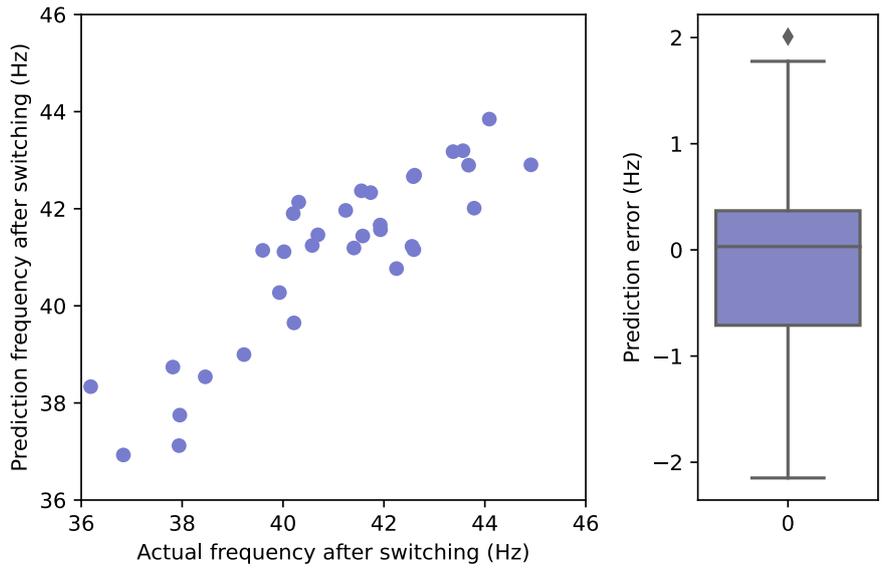


Figure 5.5. Frequency prediction accuracy

Based on the frequency prediction model, the contour map shown in **Figure 5.6** presents the frequency prediction after switching from 1 to 2 SCHWPs, where the minimum operating frequency of the pump is 30 Hz. To maintain the same differential pressure as when 1 SCHWP operates, the frequency decreases significantly after the pump switch as the flow rate increases. Therefore, switching from 1 pump to 2 pumps has more energy-saving potential when the flow rate is high.

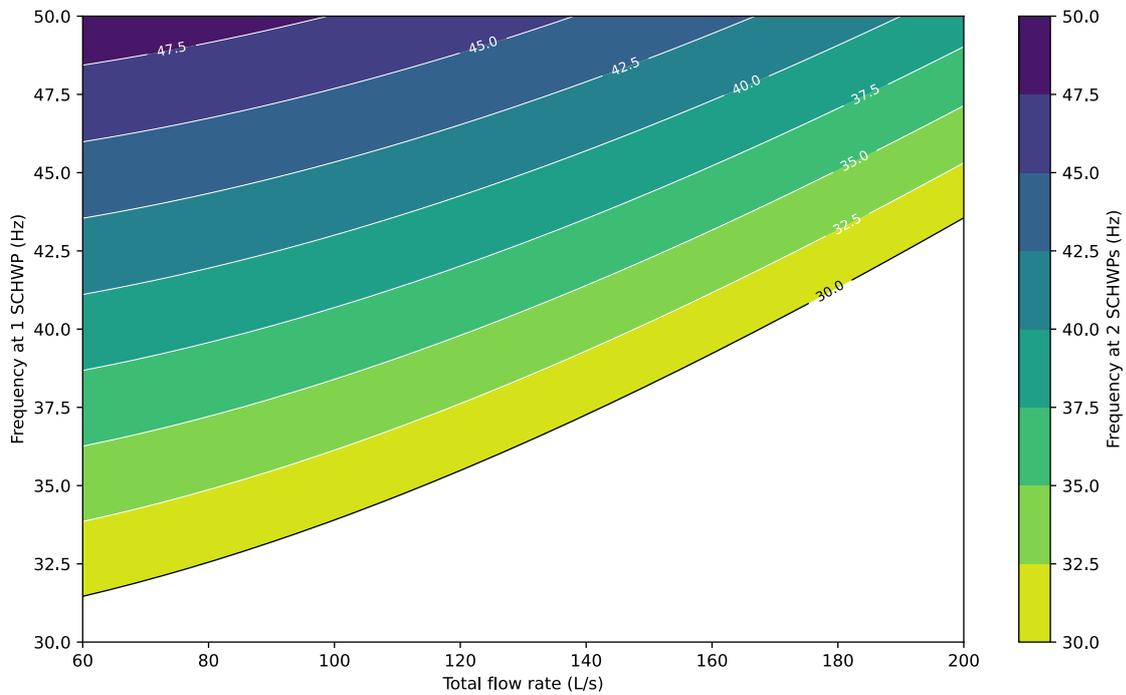


Figure 5.6. Frequency prediction after switching from 1 to 2 SCHWPs

Figure 5.7 displays the frequency prediction after switching from 2 to 1 SCHWP, where the maximum operating frequency of the pump is 50 Hz. The frequency increase becomes insignificant after switching from 2 to 1 SCHWP as the flow rate decreases. It implies that switching 2 to 1 SCHWP has more energy-saving potential when the flow rate is low.

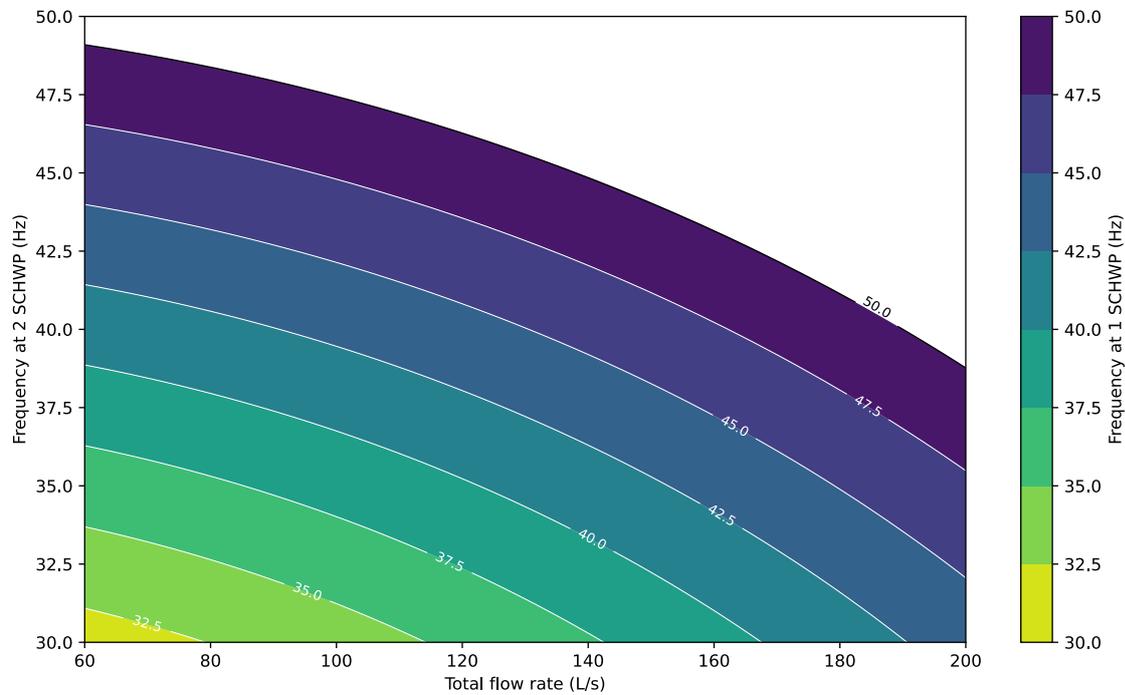


Figure 5.7. Frequency prediction after switching from 2 to 1 SCHWP

5.2.4. Energy saving analysis

5.2.4.1. Energy saving from 1 to 2 SCHWPs

By combining the pump power and frequency prediction models, the optimal sequencing of the pump can be analyzed, and energy savings can be evaluated. **Figure 5.8** shows the energy saving from 1 to 2 SCHWPs, where the minimum frequency requirement is 30 Hz. The pink points in the figure represent the data points of operating 1 SCHWP collected from June 2023 to June 2024. The data points have two distinct clusters, corresponding to the two-stage differential pressure setpoints set during the day and night. The lower differential pressure setpoint corresponds to the low-flow cluster, with flow rates distributed between 60-100 L/s. The higher differential pressure setpoint corresponds to the high-flow cluster, with flow rates distributed between 85-200 L/s. The purple line in the figure corresponds to the contour line at a frequency of 30 Hz in **Figure 5.6**. The contour lines shown in **Figure 5.8** bend when intersecting with the purple line because of the minimum frequency requirement. The white line on the contour map represents energy savings greater than 0. A total of 47.0% of the points

fall in the area where the energy savings are greater than 0, with energy savings ranging from 0-25 kW. This indicates that the conventional differential pressure-based control cannot guarantee that parallel SCHWPs operate at the lowest energy consumption, and there is a considerable potential for energy savings. If the allowable minimum frequency of the SCHWPs can be lower, the potential for energy saving can be further enhanced.

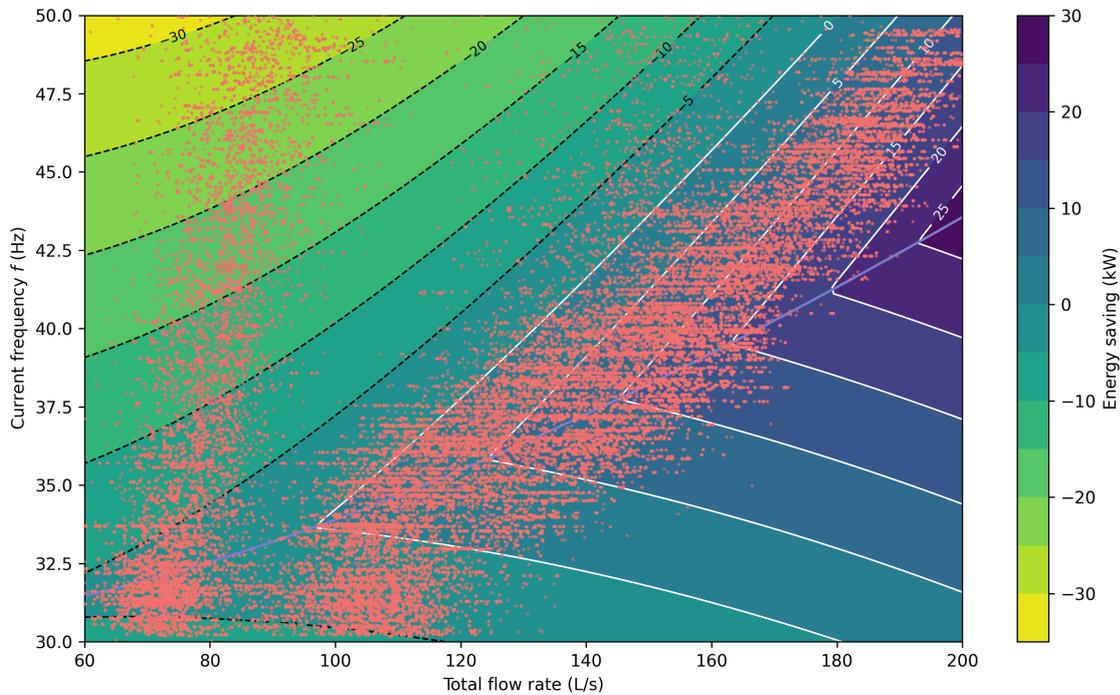


Figure 5.8. Energy saving from 1 to 2 SCHWPs (Min. frequency: 30 Hz)

Figure 5.9 shows the contour map of energy saving percentage from 1 to 2 SCHWPs, where the minimum frequency requirement is 30 Hz. The contour line marked 0 in the figure is consistent with the contour line marked 0 in **Figure 5.8**, and the remaining contour lines are different. It can be seen from the figure that the highest energy savings is more than 40%. Therefore, operating different numbers of SCHWPs to maintain the same head and flow rate can result in substantial differences in energy consumption, especially when the system flow rate is high.

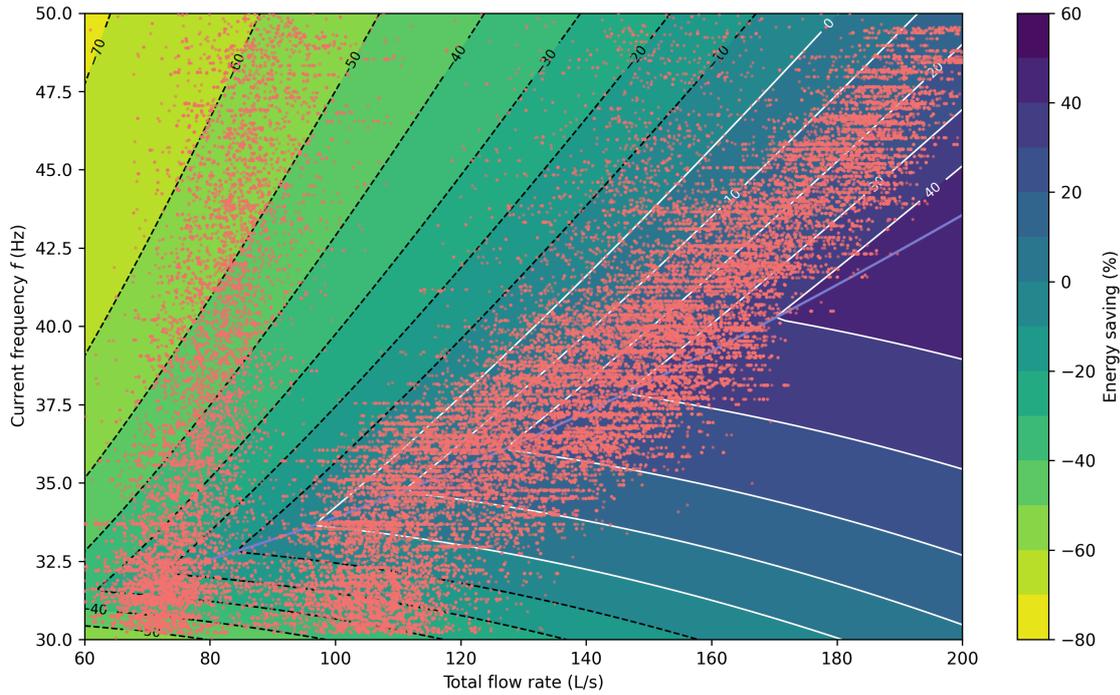


Figure 5.9. Energy saving percentage from 1 to 2 SCHWPs (Min. frequency: 30 Hz)

5.2.4.2. Energy saving from 2 to 1 SCHWP

Figure 5.10 shows the contour map of energy saving from 2 to 1 SCHWP, where the maximum frequency requirement is 50 Hz. The white area in the figure corresponds to the white part above the contour line at a frequency of 50 Hz in **Figure 5.7**, indicating the impracticability to switch from 2 SCHWPs to 1 SCHWP. The pink points in the figure represent the data points of operating 1 SCHWP collected from June 2023 to June 2024. 7.0% of the points fall in the area where the energy savings are greater than 0, with energy savings ranging from 0-35kW, which is smaller than the energy-saving potential of operating 1 SCHWP.

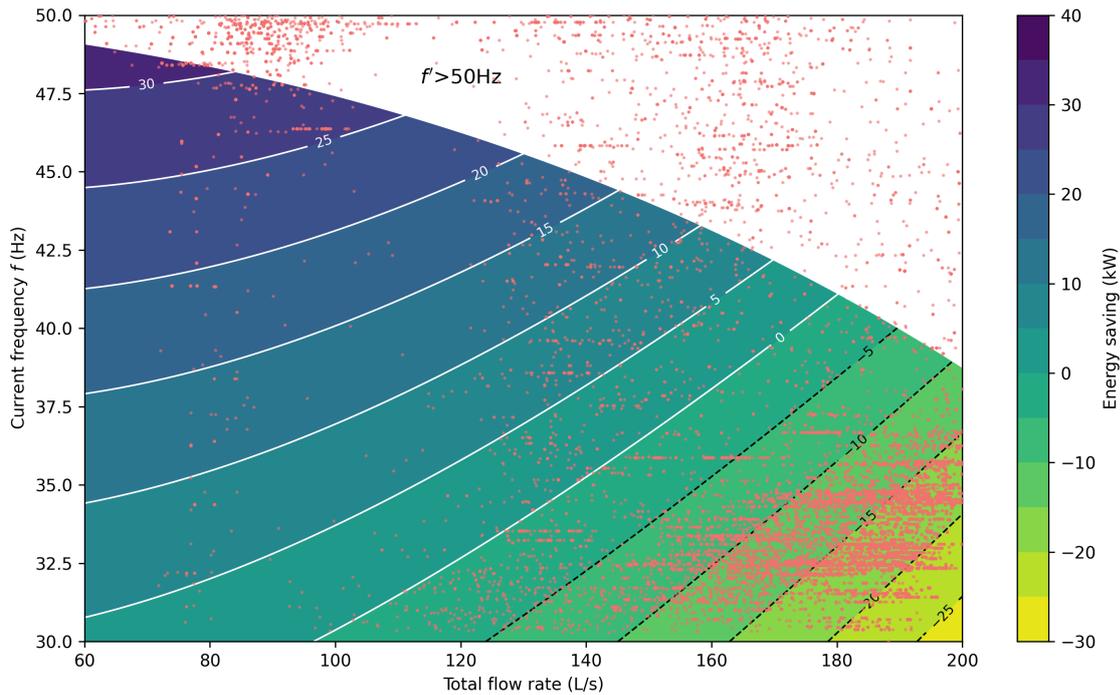


Figure 5.10. Energy saving from 2 to 1 SCHWP (Max. frequency: 50 Hz)

Figure 5.11 shows the contour map of the Energy saving percentage from 2 to 1 SCHWPs, where the maximum frequency requirement is 50 Hz. The contour line marked 0 in the figure is consistent with the line marked 0 in **Figure 5.10**. The highest energy savings is more than 40% when the system flow rate is low, and the pump frequency is high.

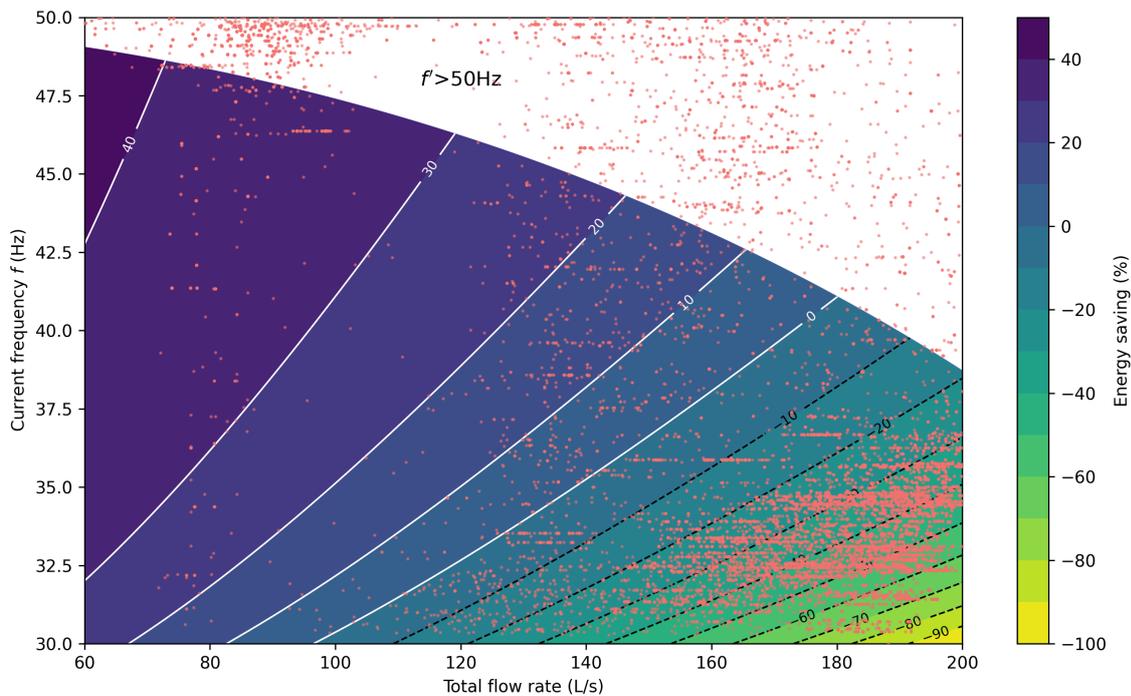


Figure 5.11. Energy saving percentage from 2 to 1 SCHWP (Max. frequency: 50 Hz)

5.2.4.3. Energy saving in historical data

The monthly energy savings from June 2023 to June 2024 using the proposed strategy is shown in **Figure 5.12**. The purple and yellow bars represent the energy consumption before and after using the proposed optimal pump sequencing strategy, and the green and red bars with white shadow represent the Energy saving from 1 to 2 SCHWPs and Energy saving from 2 to 1 SCHWP, respectively. The potential for energy savings in winter is minimal, with only a 2.5% reduction observed in January 2024, whereas substantial savings of 20.4% are evident in June 2024 during the summer. On average, the energy consumption of SCHWPs is reduced by approximately 10%. Most energy savings come from switching from two to one SCHWPs because the SCHWPs are oversized, which is common in practice. This means running one SCHWP is sufficient on most days.

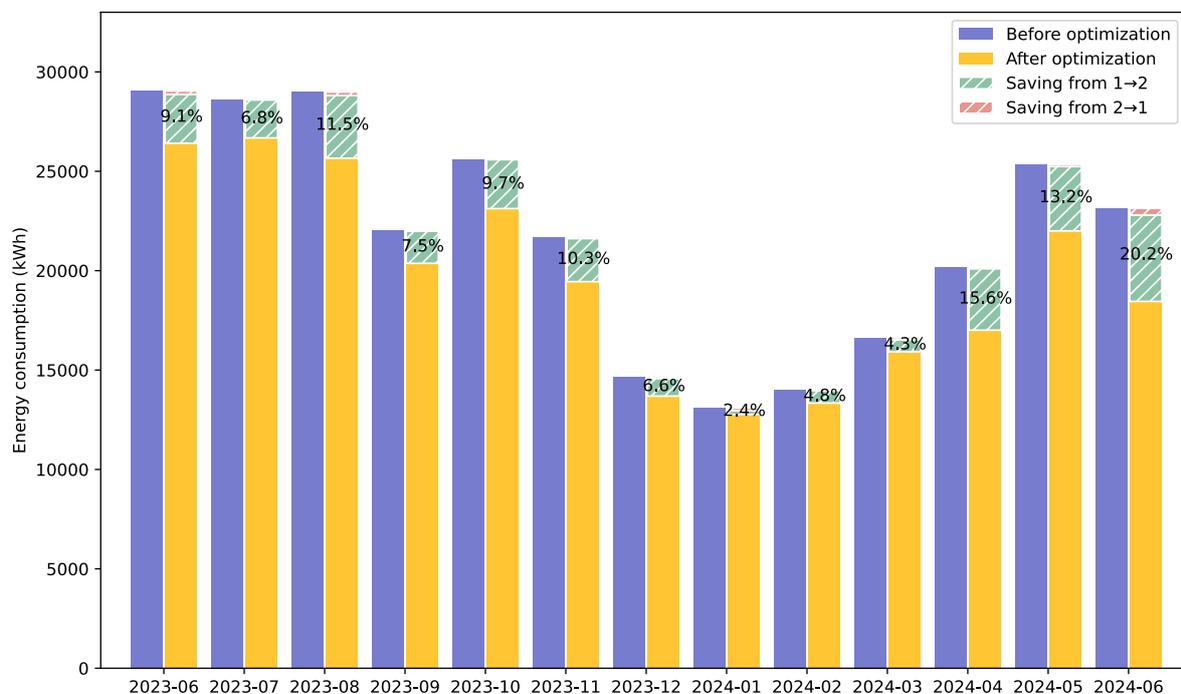


Figure 5.12. Monthly energy saving

Figure 5.13 shows an energy-saving analysis of a typical working day. Before using the proposed optimal pump sequencing strategy, only SCHWP-2 operated after 7am; its frequency is shown as the blue solid line. Although one SCHWP can maintain the differential pressure setpoint, the power is large due to the large system total flow rate and the pump needs to operate at a high frequency. The orange dashed line represents the predicted frequency after switching from 1 to 2 SCHWPs. After the switch, although one pump is added, the total power decreases

by 10%-35%, as shown in the middle of the figure, with an average power reduction of about 25.7%.

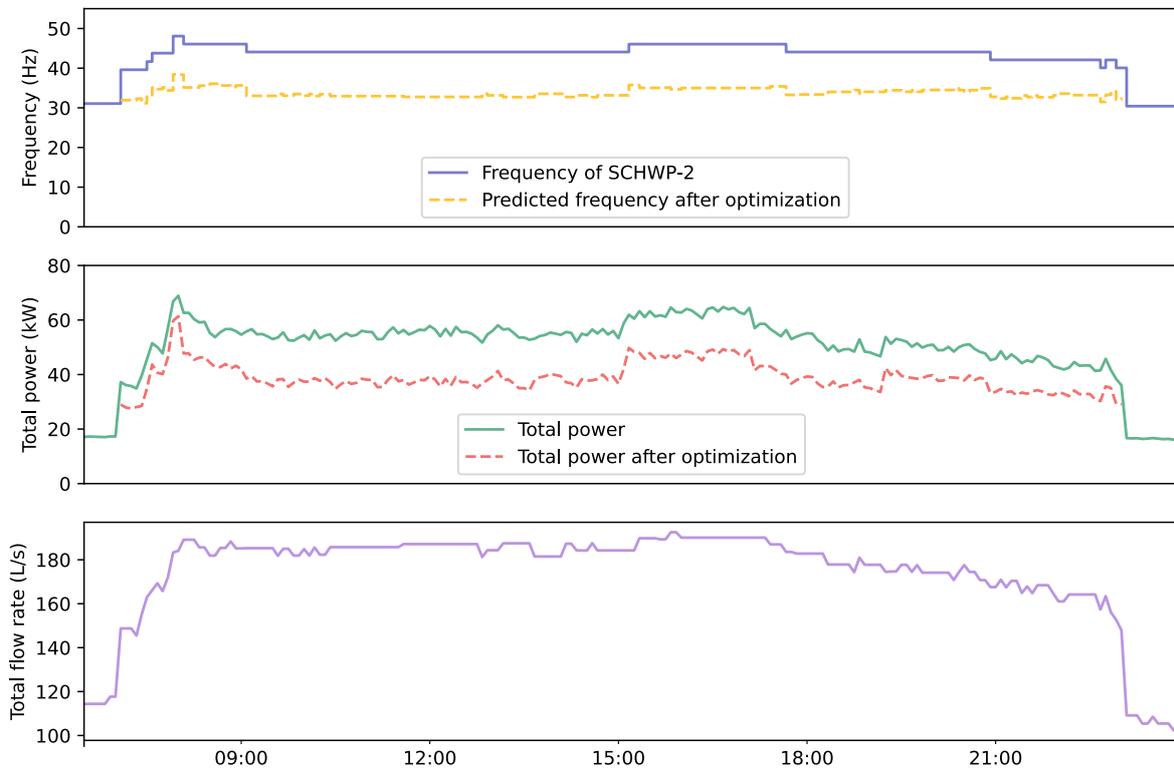


Figure 5.13. Detailed energy saving on a typical working day (21-July 2023)

5.3. Summary

HVAC systems consume roughly 40% of the total energy used in buildings, where about 15% of this system's total energy is used by pumps. Hence, optimal control of the pumping system provides substantial energy savings. The rule-based strategy currently employed in pump sequencing does not guarantee optimal energy efficiency for the VSPs. This research proposes a novel physically consistent data-driven optimal sequencing control strategy for variable speed pumps used in HVAC systems. For different numbers of operating VSPs, features composed of frequency and total flow rate are used to train the ElasticNet data-driven model. The optimal sequencing is then determined using a physically consistent frequency prediction method. The proposed strategy is validated using historical data from a real parallel variable pumping system with three VSPs. The results show that operating different numbers of VSPs to maintain the same head and flow rate can result in substantial differences in energy consumption, especially when the system flow rate is high. And the average energy saving is about 10% compared to the conventional rule-based strategy. The proposed strategy is promising to be applied to various variable speed parallel pumping systems that adopt differential pressure control.

CHAPTER 6 GENERAL OPTIMIZATION FRAMEWORK FOR DATA-DRIVEN MODEL-BASED ONLINE OPTIMIZATION WITH ENHANCED CONTROL SMOOTHNESS

This chapter develops a multi-objective optimization-based framework to achieve **reliable online control** with enhanced smoothness. Optimization algorithms, including deterministic algorithms and metaheuristic algorithms (biology-based and physics-based), are widely used in building energy systems for system design optimization, demand response and grid interaction, and energy efficiency operation. However, these optimization algorithms are not designed for online optimal control, and their stability is usually overlooked. The stability of deterministic algorithms could suffer from variations in the optimization problem (e.g., variations in objection function and constraints) during online optimal control. The solutions produced by metaheuristic algorithms can vary with random seed due to the stochastic nature of their initialization and exploration processes. Furthermore, successive control actions derived from these algorithms may differ significantly due to the shifts across local optima. Therefore, this study proposes a multi-objective optimization (MOO)-based approach for enhancing the stability in model-based online control. Besides energy saving, the distance between two successive control actions is used as one of the optimization objectives. Section 6.1 describes the proposed MOO-based approach. Section 6.2 presents the setup of the case studies. Section 6.3 analyzes the results of the case studies, and Section 6.4 concludes the main findings of the study.

6.1. Proposed multi-objective optimization-based approach for online optimal control

6.1.1. Overview of the proposed approach

Figure 6.1 illustrates the overview of the proposed multi-objective optimization (MOO)-based approach for online optimal control. The MOO-based approach includes three main parts: multi-objective optimization, modeling, and distance measurement. The multi-objective optimization algorithm adopted in this study is the non-dominated sorting genetic algorithm II (NSGA-II) [131], ensuring the flexibility of the proposed approach to various model-based control and model predictive control problems and optimization objectives. The modeling and

distance measurement parts are employed to compute two optimization objectives within the multi-objective optimization. Objective 1 is the same as conventional optimization approaches, such as to minimize total power, carbon emissions, or to maximize economic benefits. On the other hand, Objective 2 is to minimize the distance between the population of the NSGA-II and the previous control action. The L1-norm and L2-norm are used to calculate the distances of integer and continuous variables, respectively. After obtaining the non-dominated optimal solutions (Pareto front), an appropriate solution is selected from those solutions as the new control action based on the tolerance level of Objective 1 as the selection criterion.

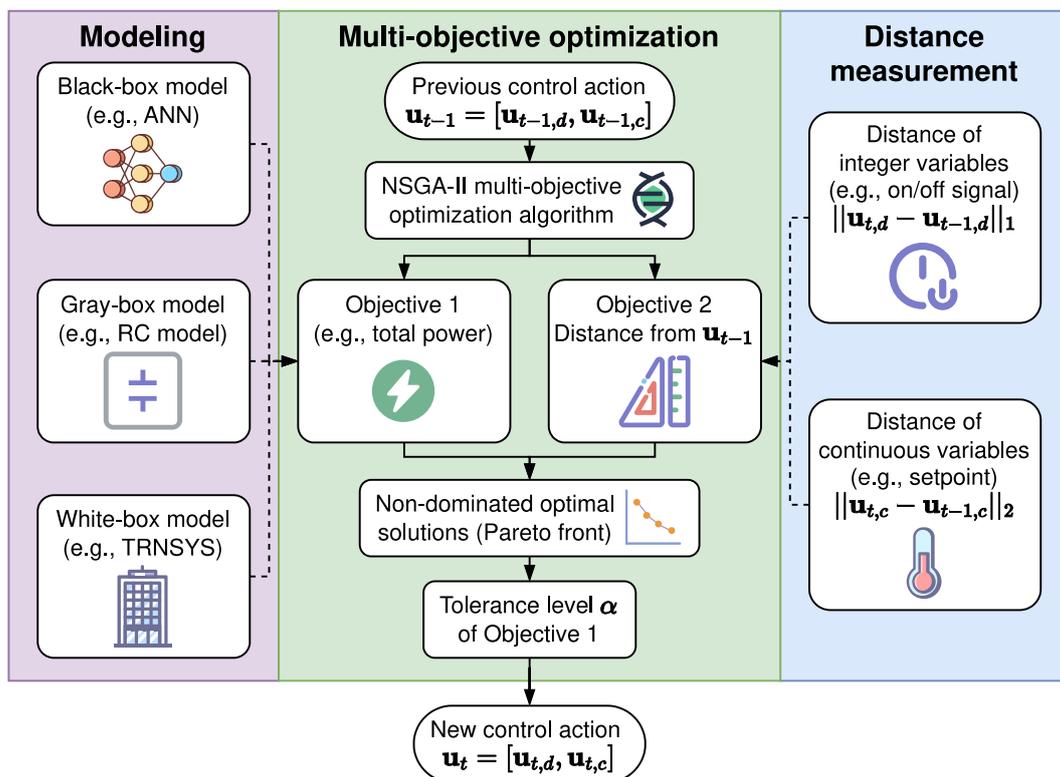


Figure 6.1. Overview of the proposed multi-objective optimization-based approach

6.1.2. Optimization Objective 1: Energy efficiency

Objective 1 aims to minimize total power and carbon emissions or maximize economic benefits based on models of building energy systems. The controllable inputs of the models are control actions such as the number of devices to be operated or the setpoints of devices. The models can be white-box, grey-box, or black-box models depending on the availability of operational data and detailed information about the building energy systems. White-box models, such as EnergyPlus, TRNSYS, and Modelica [38], can be co-simulated and optimized using Python. Grey-box models, like thermal resistance-capacitance (RC) models, are typically employed as simplified representations to understand and predict the thermal behavior of a building [228].

With the increasing development of metering technologies and IoT sensors, black-box models (i.e., data-driven models) are becoming more popular in building energy modeling. Since most data-driven models are complex and non-linear, deterministic algorithms usually cannot solve them. Therefore, metaheuristic algorithms are the only choice for black-box model-based optimization. The NSGA-II used in this study is a metaheuristic algorithm, making it flexible in addressing various model-based control and model predictive control problems.

6.1.3. Optimization Objective 2: Control stability

Objective 2 aims to minimize the distance between the population of the NSGA-II and the previous control action, i.e., to improve the stability between two control actions. The previous control action is formulated as the following action vector:

$$\mathbf{u}_{t-1} = [\mathbf{u}_{t-1,d}, \mathbf{u}_{t-1,c}] \quad (6.1)$$

where $\mathbf{u}_{t-1,d}$ is the vector of integer control variables (discrete variables), and $\mathbf{u}_{t-1,c}$ is the vector of continuous control variables normalized between 0 and 1.

In the optimal control of building energy systems, integer control variables are typically the on/off signals (labeled as 1/0) of devices or the number of devices to be operated in the following control period. Continuous control variables usually refer to the setpoints of devices or local controllers, for example, the chilled water supply temperature, the frequency of variable speed pumps, or the indoor air temperature setpoint. For each continuous variable, the range is normalized between 0 and 1 based on the minimum and maximum allowable operation range.

To measure the distance between two vectors with integer and continuous elements, the L1-norm and L2-norm are used to calculate the distance of integer and continuous variables, respectively:

$$\text{dist}(\mathbf{u}_t, \mathbf{u}_{t-1}) = \|\mathbf{u}_{t,d} - \mathbf{u}_{t-1,d}\|_1 + \|\mathbf{u}_{t,c} - \mathbf{u}_{t-1,c}\|_2 \quad (6.2)$$

where L1-norm and L2-norm are defined as:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i| \quad (6.3)$$

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2} \quad (6.4)$$

where \mathbf{x} is a n -dimensional vector and x_i refers to the i -th element of the vector.

The L1-norm of a vector, also known as the Manhattan distance, is the sum of the absolute values of each element in a vector. In two-dimensional space, for example, the L1-norm represents the sum of horizontal and vertical distances from the origin to the point where the vector lies. For example, if $\mathbf{u}_{t,d} = [1,0]$ and $\mathbf{u}_{t-1,d} = [0,1]$, this means that only the first device is on and only the second device is on at time t and $t - 1$, respectively. The distance between $\mathbf{u}_{t,d}$ and $\mathbf{u}_{t-1,d}$ is $|1 - 0| + |0 - 1| = 2$. The L1-norm distance measures the number of on/off actions between the two control actions and is preferable for measuring the distance between integer variables

The L2-norm of a vector, also called Euclidean distance, is the square root of the sum of squares of each element in the vector. For example, if $\mathbf{u}_{t,c} = [0.4,0.8]$ and $\mathbf{u}_{t-1,c} = [0.7,0.4]$, it means that only the first continuous variable becomes 0.4 from 0.7, and the second continuous variable becomes 0.8 from 0.4. The distance between $\mathbf{u}_{t,c}$ and $\mathbf{u}_{t-1,c}$ is $\sqrt{(0.4 - 0.7)^2 + (0.8 - 0.4)^2} = 0.5$. Compared to the L1-norm, the L2-norm is more sensitive to the outliers because of the square operator. Therefore, L2-norm of $\mathbf{u}_{t,c} - \mathbf{u}_{t-1,c}$ is more suitable for measuring the distance between continuous variables.

6.1.4. Tolerance level-based solution selection from the Pareto front

In NSGA-II, the Pareto front refers to a group of solutions representing the optimal tradeoff between conflicting objectives, i.e., objectives 1 and 2 in this study. NSGA-II employs the concept of Pareto dominance, which defines the quality of solutions. Solution A is dominated by solution B if solution B is equal to or better than solution A in all optimization objectives and better than solution A in at least one objective. The Pareto front is the set of all non-dominated solutions in the population, which offers more profound insights into the tradeoff between conflicting objectives.

Selecting an optimal solution from the Pareto front is also important in practice. Techniques for order of preference by similarity to ideal solution (TOPSIS) [229] and simple additive weighting (SAW) [230] have been widely adopted in previous studies. In this study, the tolerance-based selection method is proposed based on the characteristic of objective 2.

Assume that the objective matrix is a matrix with m rows (representing each solution in the Pareto front found by NSGA-II) and two columns (representing Objectives 1 and 2). The steps to find the solution for control actions are as follows:

Step 1. Sort the objective matrix based on the values of Objective 2, i.e., the distance between the previous control action and each solution in the Pareto front in ascending order.

Step 2. Group the solutions based on the floor value (the largest integer that the current number is less than or equal to) of Objective 2.

Step 3. Select the solution with optimal Objective 1 from each group and drop other solutions.

Step 4. Start selecting from the first remaining solution until the value of Objective 1 of the following solution differs from the current solution by less than the tolerance level α . In this study, the tolerance level α is a parameter to balance the energy efficiency and control stability. In the following case studies, the tolerance level α is set from 1% to 10%. Finally, the new control action $\mathbf{u}_t = [\mathbf{u}_{t,d}, \mathbf{u}_{t,c}]$ is obtained from the Pareto front.

6.2. Design of case studies

To validate the stability of the proposed multi-objective optimization (MOO)-based approach for online optimal control, a typical optimization problem in building energy systems is studied in the case studies: optimal chiller loading. The problem of optimal chiller loading is an important research issue as it involves maintaining an optimal balance between cooling load, chiller efficiency, and system stability in large commercial buildings. In previous studies, the performance of optimal chiller loading methods was compared under individual cooling load conditions instead of actual cooling load curves. The variation of cooling load and randomness could affect the stability of online optimal chiller loading. There have been limited studies comparing these algorithms and analyzing the control stability. The case studies present a stability comparison of several major optimization algorithms for online optimal chiller loading. These algorithms are tested and evaluated using the measured cooling load curve.

6.2.1. Typical optimization problem: optimal chiller loading

The case studies use a chiller plant with four 1280 RT and two 1250 RT (total 7620 RT) chillers adopted by numerous studies [231]. The power of each chiller is formulated as a quadratic function of the partial load ratio (PLR), as shown in Eq. (6.5) and **Table 6.1**. The curves were fitted based on data when the entering cooling water temperature of the chillers was 24.5 °C. In actual application, the cooling water temperature should also be an independent variable for the chiller model. Since these simple quadratic chiller power curves have been widely used in previous optimal chiller loading studies, the cooling water temperature water is regarded as a

fixed value at 24.5 °C in this study because the variation of cooling load affects the optimal chiller loading results most.

$$P_{\text{chiller},i} = a_i + b_i \times PLR_i + c_i \times PLR_i^2 \quad (6.5)$$

where $P_{\text{chiller},i}$ is the power (kW) of the i -th chiller, PLR_i is the partial load ratio of the chiller, and a_i, b_i, c_i are coefficients.

Table 6.1. Chiller performance coefficients

Chiller	a_i	b_i	c_i	Capacity (RT)
1	399.345	-122.12	770.46	1280
2	287.116	80.04	700.48	1280
3	-120.505	1525.99	-502.14	1280
4	-19.121	898.76	-98.15	1280
5	-95.029	1202.39	-352.16	1250
6	191.750	224.86	524.04	1250

Figure 6.2 shows the quadratic power curves of the six chillers. The performance of the six chillers is different because of the differences in model type, performance decay, and site conditions. The power curves of chillers 1, 2, and 6 are convex functions, while those of chillers 3, 4, and 5 are non-convex. It means that chillers 3, 4, and 5 remain highly efficient when the cooling load is high. Such performance differences in multiple-chiller systems are common, making optimal chiller loading a promising way to improve energy efficiency compared to the equally loading strategy.

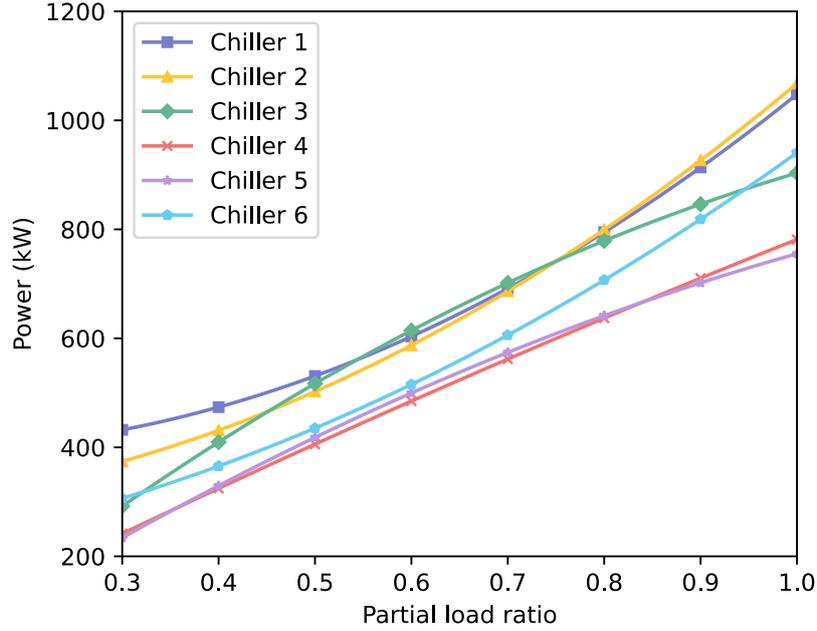


Figure 6.2. Power curve against partial load ratio of chillers

Objective 1 aligns with previous studies on optimal chiller loading, aiming to minimize the total power of the chiller plant under a given cooling load, as formulated in Eq.(6.6). The first constraint ensures that the sum the cooling load for each chiller should equal the total cooling load, as shown in Eq. (6.7). The second constraint sets the minimum and maximum values for the chiller partial load ratio (PLR), as shown in Eq. (6.8). The third constraint related the PLR and the on/off status of each chiller. Therefore, the control variables for the optimization problem are X_i and PLR_i , which are integer and continuous variables, respectively: $\mathbf{u}_t = [\mathbf{u}_{t,d}, \mathbf{u}_{t,c}]$.

$$\min \sum_i P_{\text{chiller},i} \quad (6.6)$$

$$\sum_{i=1}^6 PLR_i \times Cap_i = CL \quad (6.7)$$

$$0.3 \leq PLR_i \leq 1, \text{ or } PLR_i = 0 \quad (6.8)$$

$$PLR_i = X_i \times PLR_i \quad (6.9)$$

$$X_i \in \{0,1\} \quad (6.10)$$

6.2.2. Data collection

In the case studies, four weeks of cooling load data were collected from the chiller plant at an educational building with a 30-minute interval [164]. Since the cooling load of the building is not as large as the chiller configurations in **Table 6.1**, the actual cooling load was multiplied

by 2, as shown in **Figure 6.3**. At night, the cooling load is about 1000 RT. The peak load can vary between 2000 RT and 5000 RT during the daytime depending on weather conditions and occupancy schedule.

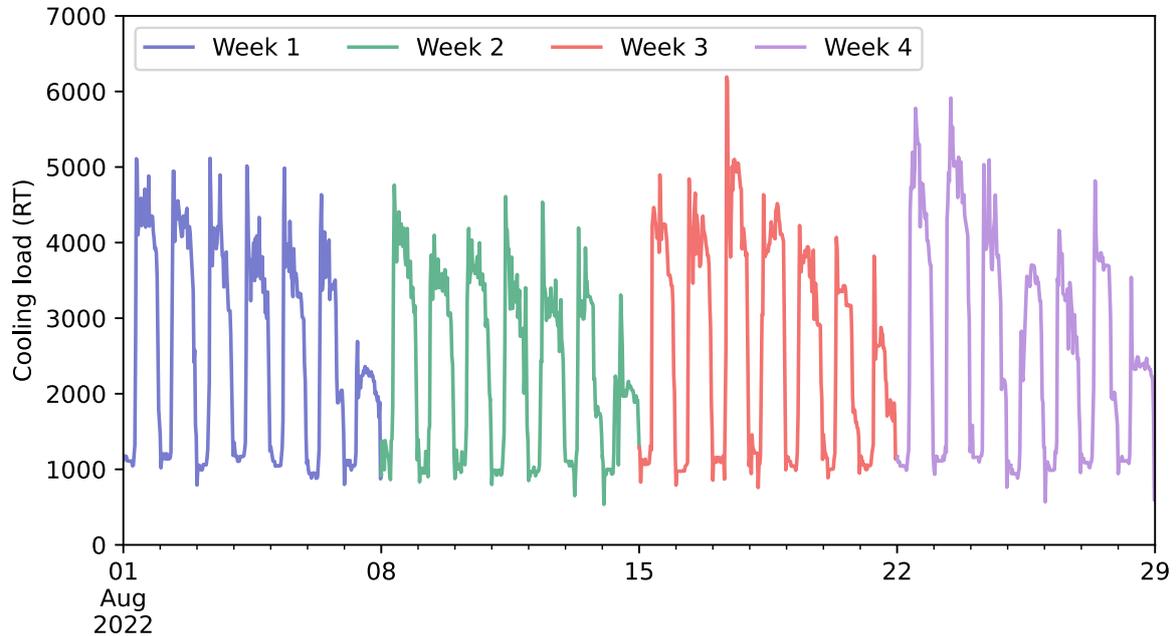


Figure 6.3. Actual 4-week cooling load from the chiller plant at an educational building

Figure 6.4 illustrates the frequency distribution of the chiller plant PLR for the tested cooling load curve. The partial load ratio was calculated by dividing the total cooling capacity of the six chillers, i.e., 7620 RT. There are two peaks in the distribution. The first peak occurs when the PLR is between 0.1 and 0.2, corresponding to the night period when the cooling load is relatively low and stable. The second peak occurs when the PLR is between 0.4 and 0.6, which is typical in actual multiple-chiller systems. In this PLR range, at least more than half of the chillers (three chillers) should be operated to provide cooling capacity, and optimal chiller loading algorithms should determine the optimal number of chillers and their loading.

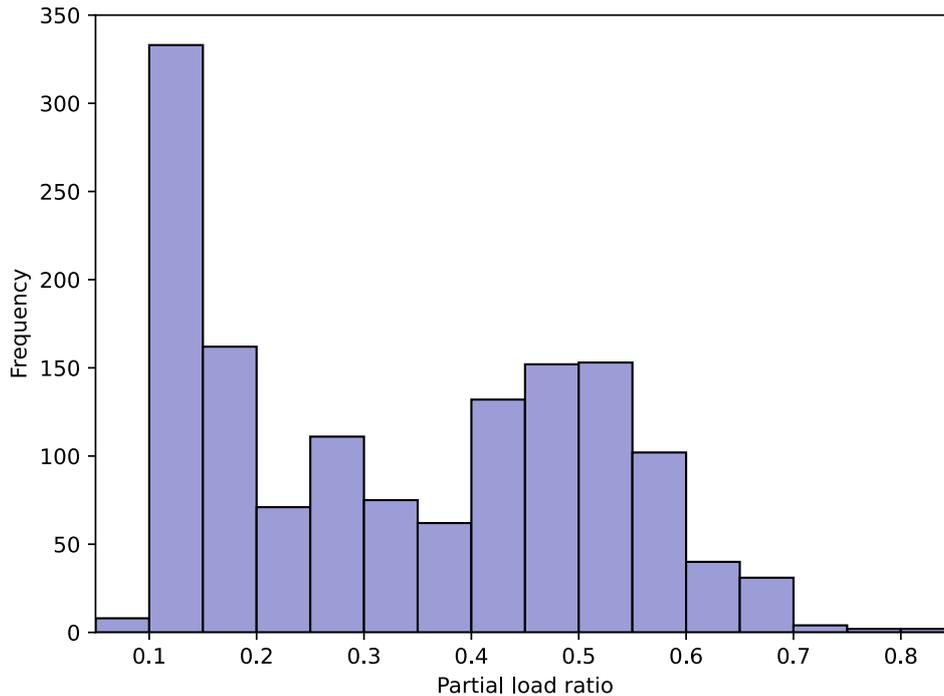


Figure 6.4. Frequency distribution of chiller plant PLR for the tested cooling load curve

6.2.3. Baseline algorithms for comparison

6.2.3.1. Metaheuristic algorithms

The four most popular were used as baseline metaheuristic algorithms: particle swarm optimization (PSO) [232], genetic algorithm (GA) [233], differential evolution (DE) [234], and simulated annealing (SA) [235]. These metaheuristic algorithms were implemented using the scikit-opt Python package with their default hyperparameter settings. For each metaheuristic algorithm, ten repeated experiments were conducted, with random seeds ranging from 0 to 9. In Section 6.3.1, the impact of random seeds on the control stability is analyzed.

6.2.3.2. Deterministic algorithm: Gurobi solver

The optimization problem formulated in Section 6.2.1 is a non-convex mixed integer quadratically constrained programming (MIQCP) problem. Therefore, it can be solved using the Gurobi solver [236]. Gurobi solver employs techniques such as branch and cut and cutting planes to find the global optimal solutions for non-convex MIQCP problems. In the case studies, the tolerance gap of the Gurobi solver is set to 0.01%.

6.2.4. Stability metrics

Two metrics are adopted in the case studies to quantify the control stability of optimization algorithms: chiller switching number and smoothness of control signals. These two metrics correspond to the stability of integer and continuous control variables in the optimization problem, i.e., the on/off status and the PLR of chillers in the case studies.

6.2.4.1. Chiller switching number

The chiller switching number is a widely adopted metric for measuring the control stability and robustness of chiller sequencing control [164,237]. The chiller switching number is the count of one chiller (or chillers) being switched on or off over a specific period:

$$\text{Chiller switching number} = \sum_i \sum_t s_{i,t} \quad (6.11)$$

$$s_{i,t} = \begin{cases} 0 & \text{if } X_{i,t} = X_{i,t-1} \\ 1 & \text{if } X_{i,t} \neq X_{i,t-1} \end{cases} \quad (6.12)$$

where $s_{i,t}$ is a binary variable that indicates the change of on/off status of the i -th chiller at time t , as defined in Eq. (6.12).

A lower chiller switching number indicates better control stability in chiller sequencing control, meaning the optimization algorithm is better at handling load variations without frequent on/off switching, which can wear down devices and damage their lifespan.

6.2.4.2. Smoothness of control signals: Average absolute change in PLR

Average absolute change is usually used to measure the smoothness of a sequence [238] by calculating the average changes between successive values in the sequence:

$$\text{Average absolute change} = \frac{1}{T-1} \sum_t |PLR_{t+1} - PLR_t| \quad (6.13)$$

where T refers to the length of the control signal sequence.

A lower average absolute change indicates better smoothness in the control signals, meaning the local control loop can track the control signals (setpoints) more responsively, and the tracking error for the setpoints can be reduced [204].

6.3. Results and discussion

6.3.1. Control stability

6.3.1.1. Chiller switching number

Figure 6.5 illustrates the total chiller switching number of the proposed MOO-based approach in four weeks. Each box represents the 25th, 50th (median), and 75th percentiles of the total

chiller switching number under a certain tolerance level α with the random seed ranging from 0 to 9. When the tolerance level $\alpha = 1\%$, the median total switching number is 195. As the tolerance level α increases from 1% to 8%, the median total switching number decreases roughly linearly to 125. With each 1% increase in the tolerance level α , the median total switching number decreases by approximately 10. As the tolerance level α continues to increase to 9% and then to 10%, the median total switching number decreases to 121 and 120.5, respectively. Therefore, beyond a certain point, as the tolerance level α increases, the benefits to control stability offered by the tolerance level diminish.

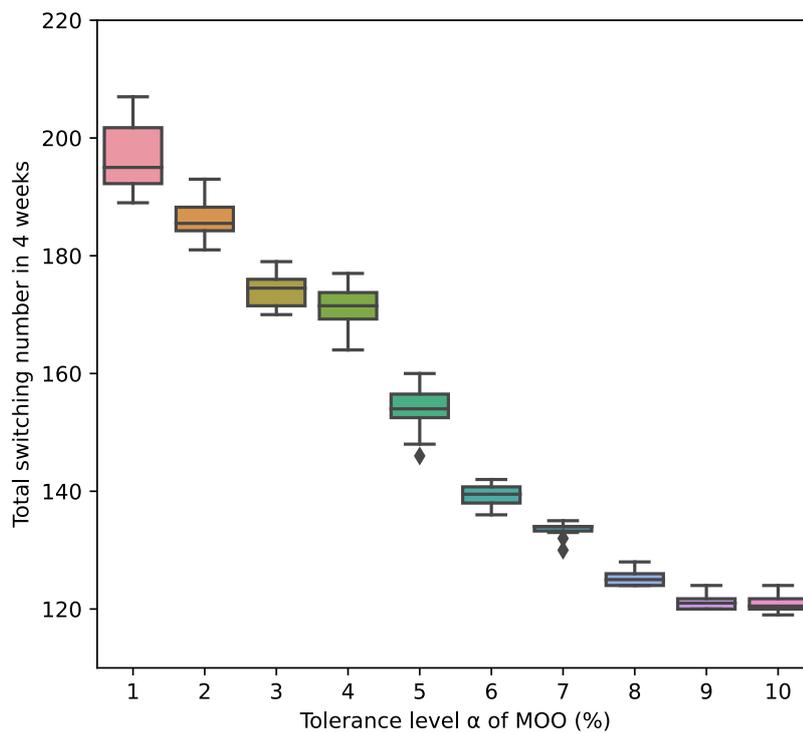


Figure 6.5. Total chiller switching number of the proposed MOO-based approach

Figure 6.6 shows the difference compared to baseline algorithms in total switching numbers in four weeks when the tolerance level α ranges from 1% to 10%. The total switching numbers of metaheuristic algorithms under random seeds from 0 to 9 are averaged. When the tolerance level α is 1%, the differences in total switching number compared to PSO, GA, SA, DE, and Gurobi are -67.0%, -47.0%, -29.5, -18.9%, and -24.8%, respectively. This means DE is the most stable conventional optimization algorithm for online optimal chiller loading, even more stable than Gurobi, which provides global optimal solutions. The proposed MOO-based approach can reduce the total chiller switching number by 18.9% compared to the most stable optimization algorithm. When the tolerance level α increases from 1% to 10%, the differences in the total switching number become even larger: -79.7%, -67.5%, -56.7%, -50.2%, and -

53.9%, respectively. This indicates that the total chiller switching number can be reduced by half, even compared to the most stable conventional optimization algorithm.

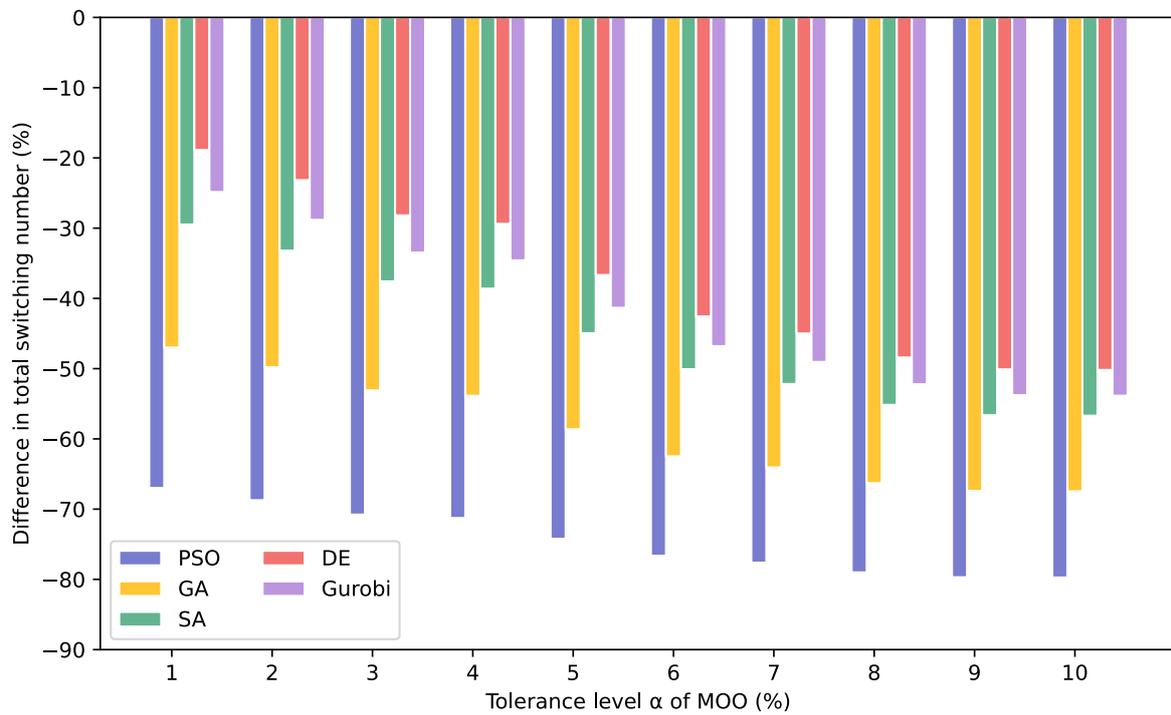


Figure 6.6. The difference compared to baseline algorithms in total switching numbers

Figure 6.7 shows the daily switching number in four weeks, with a tolerance level α of 5% for the MOO-based approach. Each curve in the figure represents the average daily switching number under the corresponding algorithm for random seeds 0-9. The proposed MOO-based approach consistently exhibits lower chiller switching numbers. It demonstrates that the proposed MOO-based approach offers better stability than conventional algorithms under different daily cooling load curves. The shaded regions above and below the lines represent the 75% percentile interval. PSO has the widest interval, indicating that the random seed most easily affects its optimization results. Gurobi, as a deterministic algorithm, is not influenced by random seeds, so its interval width is 0.

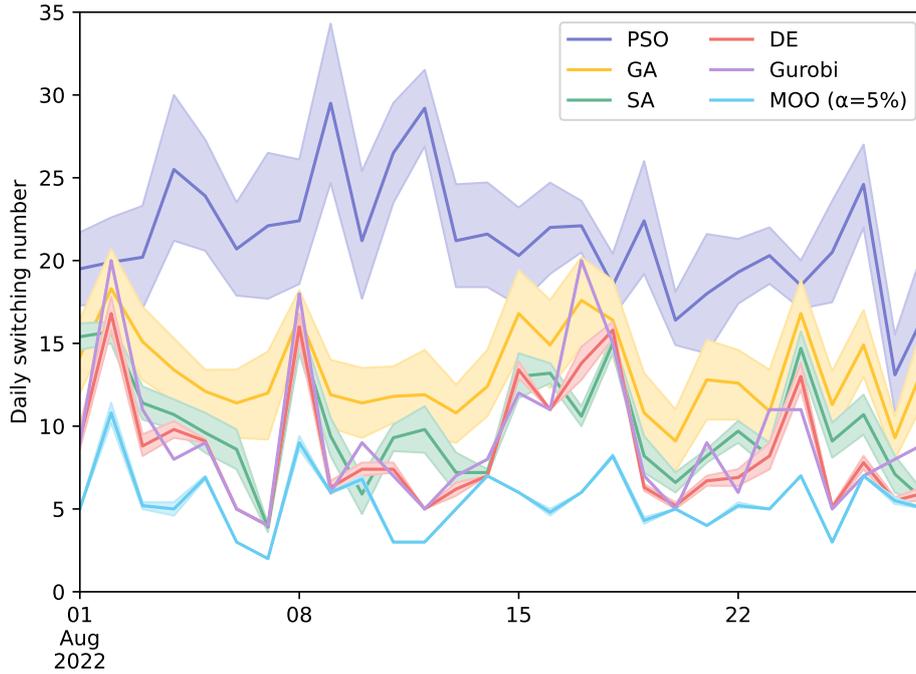


Figure 6.7. Daily chiller switching number (random seed 0-9)

Table 6.2 shows the total chiller switching number in four weeks under random seed from 0 to 9 for each algorithm. PSO shows the highest chiller switching numbers on most days, followed by GA, SA, Gurobi, DE, and the proposed MOO-based approach. It is important to note that the Gurobi solver that provides global optimum for optimization problems does not guarantee greater stability for online optimal control. Although DE does not provide global optimum, it has even better stability than Gurobi. An in-depth analysis for this phenomenon is provided in Section 6.3.4. The 'Max.-Min.' column represents the difference between the maximum total chiller switching number and minimum total chiller switching number under seed from 0-9. For example, GA's maximum total switching number in four weeks is 561 when the random seed is 0, while PSO's minimum total chiller switching number is only 223 when the random seed is 8. This huge difference indicates that the random seed can easily influence the solutions of conventional metaheuristic algorithms. DE is the metaheuristic algorithm that is least affected by the random seed. The proposed MOO-based approach reduced the difference caused by the random seed by 60%.

Table 6.2. Total chiller switching number in four weeks (random seed 0-9)

Algorithm	Average	Maximum	Minimum	Max.-Min.
PSO	596.4	928	425	503
GA	371.5	561	223	338
SA	279.2	307	267	40

DE	242.7	258	223	35
Gurobi	262	262	262	0
MOO ($\alpha = 5\%$)	153.7	160	146	14

The proposed MOO-based approach is hardly affected by random seeds compared to other metaheuristic algorithms. The random seed does not affect the chiller switching numbers over 17 days (61%). **Figure 6.8** shows the difference between maximum and minimum switching numbers in the day under random seeds from 0 to 9. The difference of the proposed MOO-based approach ranges from 0 to 4. In contrast, the differences of DE, SA, GA, and PSO range from 0 to 10, 2 to 12, 12 to 26, and 13 to 40, respectively, which are significantly larger than that of the proposed MOO-based approach. Therefore, the random seed can easily influence the stability of conventional metaheuristic algorithms. The proposed MOO-based approach can largely eliminate the instability caused by randomness and is close to deterministic algorithms.

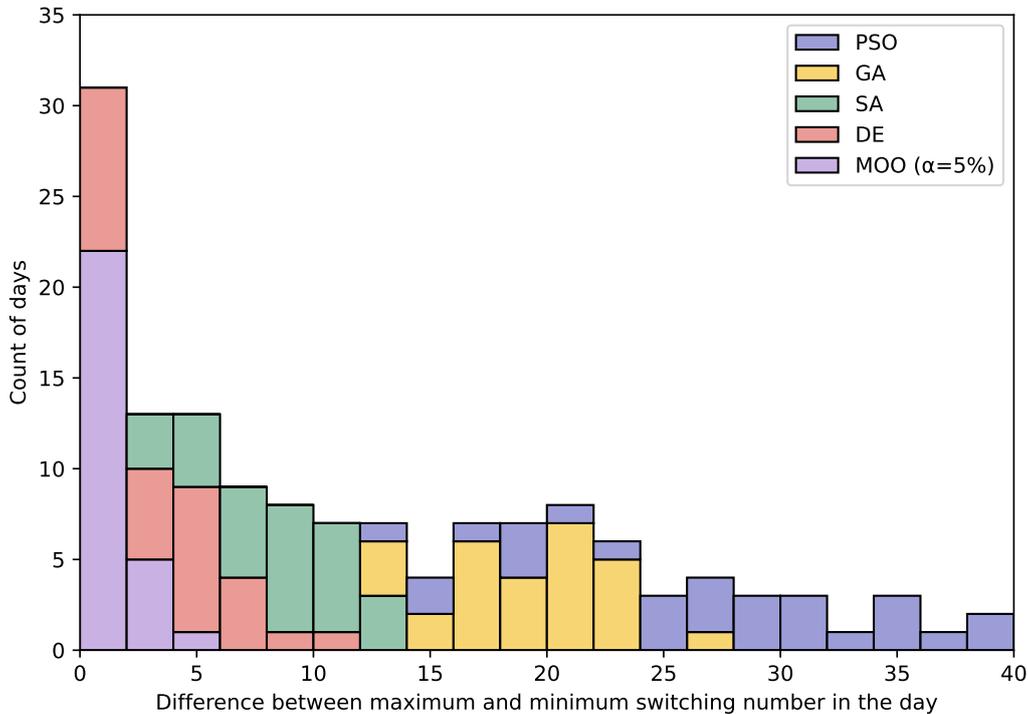


Figure 6.8. Difference between maximum and minimum switching number in the day (random seed 0-9)

6.3.1.2. Average absolute change in PLR

Figure 6.9 shows the daily average absolute change in PLR (%) in four weeks, with a tolerance level $\alpha = 5\%$ for the MOO-based approach. Each curve in the graph represents the average absolute change in PLR (%) under the corresponding algorithm for random seeds 0-9. The

proposed MOO-based approach consistently shows lower average absolute change, demonstrating enhanced stability compared to conventional algorithms.

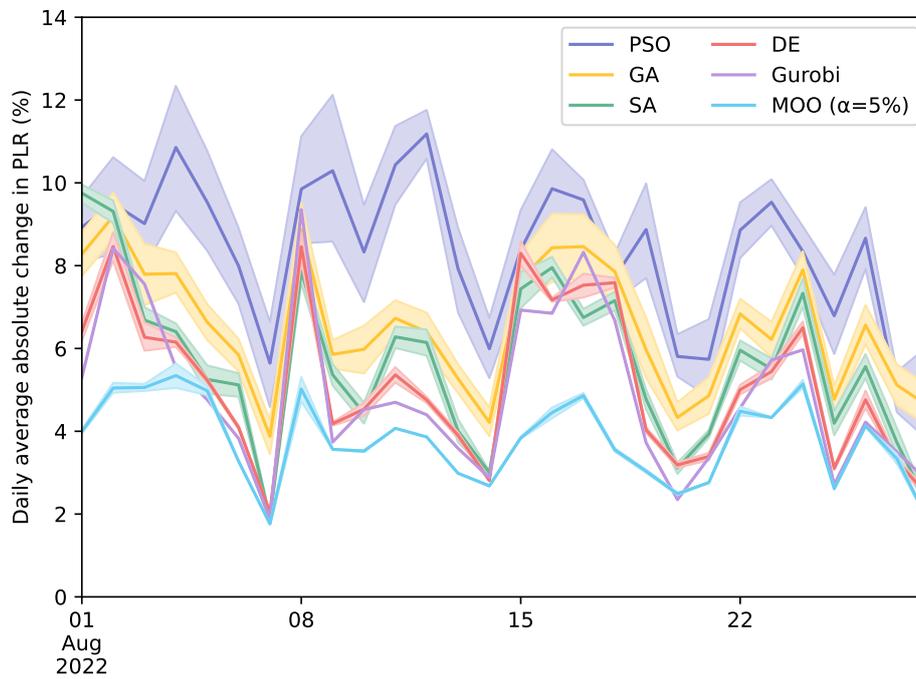


Figure 6.9. Daily average absolute change in PLR (%) (random seed 0-9)

Table 6.3 presents the average absolute change of PLR (%) in four weeks under random seeds from 0 to 9 (tolerance level $\alpha = 5\%$). Overall, PSO shows the highest average absolute change in PLR on most days, followed by GA, SA, DE, Gurobi, and the proposed MOO-based approach. It is important to note that, unlike the results for daily chiller switching numbers, the Gurobi solver provides slightly more stable PLR signals than DE. The reason is that global optimal solutions provide stable control actions when the cooling load is stable. However, for example, when the cooling load exceeds the threshold at which the deterministic algorithm to change the number of operating chillers, metaheuristic algorithms might maintain the current chiller number. As a result, the sudden variations in load fluctuation could lead to unstable decisions on integer variables. The total switching number is a more sensitive distance metric than the average absolute change in measuring such unstable control actions.

Table 6.3. Average absolute change of PLR (%) in four weeks (random seed 0-9)

Algorithm	Average	Maximum	Minimum	Max.-Min.
PSO	8.34	13.19	4.95	8.24
GA	6.51	8.92	4.61	4.31
SA	5.61	7.34	4.38	2.96
DE	5.16	5.8	4.64	1.16

Gurobi	4.93	4.93	4.93	0
MOO ($\alpha = 5\%$)	3.79	4.06	3.55	0.51

6.3.2. Total energy consumption

Figure 6.10 shows the energy consumption of the proposed MOO-based approach in four weeks. Each box represents the 25th, 50th (median), and 75th percentiles of the total chiller switching number under a certain tolerance level α when the random seed varies from 0 to 9. The dashed horizontal lines represent the average total energy consumption of five baseline algorithms: 578.0×10^3 , 579.3×10^3 , 580.0×10^3 , 582.7×10^3 , and 593.5×10^3 kWh for Gurobi, DE, SA, GA, and PSO, respectively. Gurobi solver, as the baseline for global optimal solutions, has the lowest energy consumption. However, the differences in energy consumption among the various algorithms are far less significant compared to the differences in control stability. DE, SA, and GA only consume 0.22%, 0.35%, and 0.81% more energy, respectively, than the global optimal solutions from the Gurobi solver. Even the least energy-efficient algorithm, PSO, only consumes 2.68% more energy than Gurobi. This conclusion aligns with previous research, indicating that various metaheuristic algorithms can find near-optimal solutions, with differences of around 1% [239]. When the tolerance level α is 1%, the average total energy consumption is 579.1×10^3 kWh, which is lower than all baseline metaheuristic algorithms. This means that the proposed MOO-based approach has better energy efficiency and control stability when the tolerance level is set at a low percentage. As the tolerance level α increases, the average total energy consumption rises first and becomes stable. When α is 10%, the average total energy consumption is 584.2×10^3 kWh, which is only 0.88% higher when α is 1%.

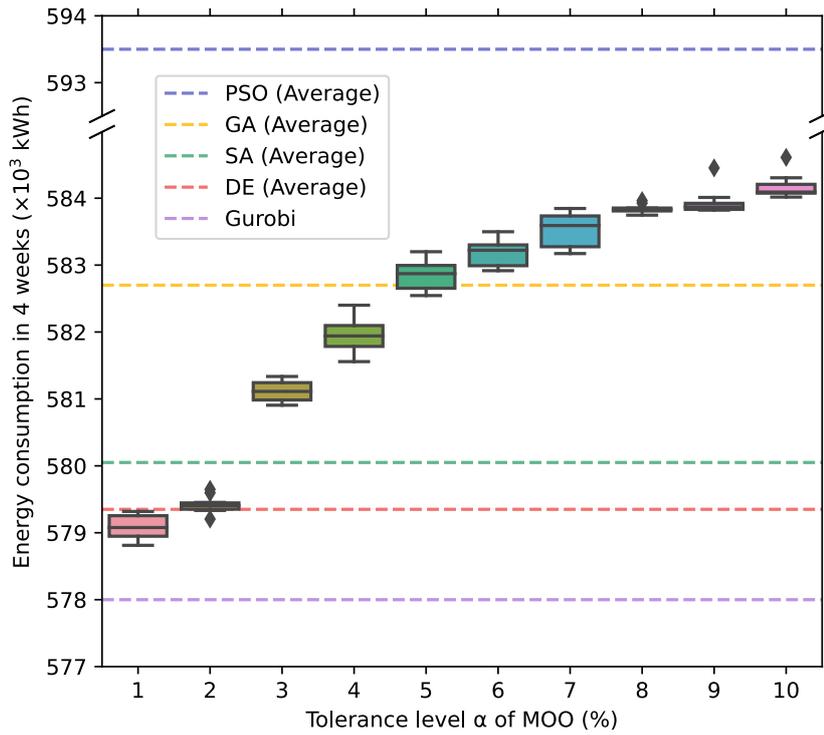


Figure 6.10. Energy consumption of the proposed MOO-based approach in four weeks

6.3.3. Chiller switching details

The heatmap in **Figure 6.11** shows the daily switching number of each chiller at the tolerance level α of 5% and with a random seed of 0. Each cell in the heatmap contains the switching number for the chiller on that day, with darker cell colors indicating higher switching numbers. These algorithms exhibit different tendencies when selecting preferable chillers. As shown in **Figure 6.11** (a), Chillers 4 and 5 are switched on and off frequently using PSO, while Chillers 1 and 6 are frequently operated using GA, as shown in **Figure 6.11** (b). PSO and GA are usually stuck in local optima and fail to find the most energy-efficient chillers as the Gurobi solver. The total switching numbers for SA, DE, and Gurobi are similar, primarily operating Chillers 4-6. In contrast, the proposed MOO-based approach does not operate Chiller 3 during the four weeks. On August 17, 2022, all baseline algorithms switch on Chiller 3 during the peak period (see daily average cooling load in **Figure 6.11** (f)). Due to the 5% tolerance level setting, the proposed MOO-based approach does not operate Chiller 3.

Figure 6.12 shows the PLRs of each chiller on August 1, 2022. The numbers on the right side of each heatmap represent the total daily switching number of the chiller, corresponding to the sequencing details in the first column of the heatmap in **Figure 6.11**. Due to the minimum PLR constraint of 0.3 in the optimization, the deep blue area in the heatmap indicates the off status of chillers. The remaining gradient from lighter blue to red indicates the PLRs of the operating chillers. **Figure 6.12 (a)** shows that Chiller 4 undergoes frequent on-and-off actions using PSO. On the other hand, Chiller 5 undergoes frequent on-and-off actions using GA and SA, as depicted in **Figure 6.12 (b)** and **(c)**, respectively. As observed in **Figure 6.12 (g)**, fluctuations in the cooling load cause this phenomenon. Furthermore, due to a significant drop in the cooling load after 9:00 AM, DE and Gurobi unnecessarily switched Chiller 5 off and on at 9:30 and 10:00, respectively, which harms the stability of the chiller plant. Overall, the proposed MOO-based approach effectively handles variations in cooling load, making energy-efficient, stable, and robust control actions for online optimal chiller loading.

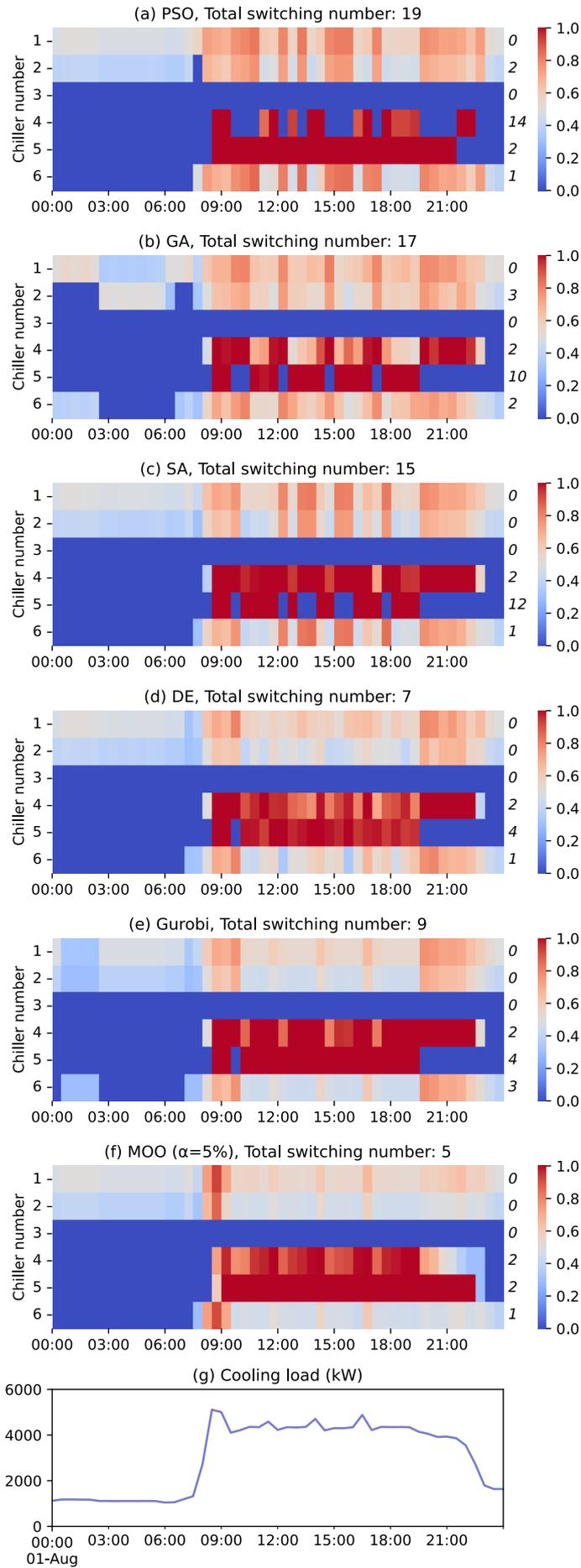


Figure 6.12. Partial load ratio of each chiller on August 1, 2022 (Monday)

Figure 6.13 illustrates the average absolute change in PLR (%) for each chiller on August 1, 2022. For each chiller, the proposed MOO-based approach demonstrates the lowest average absolute change. This indicates that the chillers are better able to follow control signals, resulting in smoother load increases and decreases to minimize control tracking errors. Chillers 1 and 2, which operate almost continuously throughout the day, exhibit relatively minor differences in average absolute change. In contrast, Chillers 4, 5, and 6, which utilize on/off control, show a more significant improvement owing to the proposed MOO-based approach.

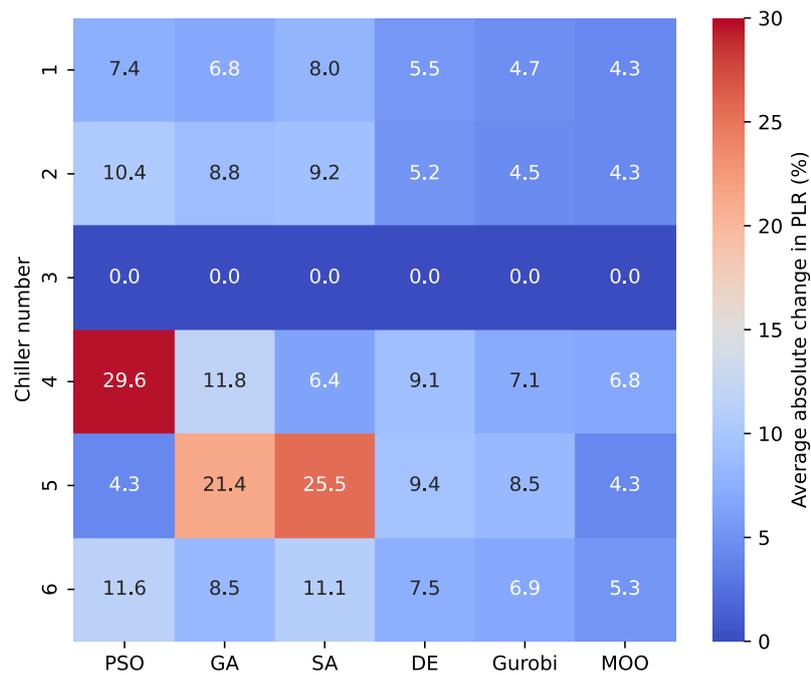


Figure 6.13. Average absolute change in PLR (%) of each chiller on August 1, 2022

6.3.4. The tradeoff between control stability and energy efficiency

Figure 6.14 compares optimization algorithms in terms of energy efficiency and control stability. Higher energy efficiency indicates a better ability to find global optimal solutions. The Gurobi solver provides global optimization in this experimental data, making it the most energy-efficient. In actual optimization problems, models are typically complex, and many can only be solved using metaheuristic algorithms, especially with the development of data-driven modeling methods. If energy efficiency is the only concern, metaheuristic algorithms such as DE and SA are adequate for finding satisfactory near-optimal solutions. However, these conventional optimization algorithms are not designed for online optimization problems, leaving their control stability unsolved. The proposed MOO-based approach can significantly

improve control stability by reducing the total chiller switching frequency by 18.9% ($\alpha = 1\%$) to 50.2% ($\alpha = 10\%$) compared to the most stable baseline algorithm, DE.

With the increase in the tolerance level α , control stability and energy efficiency are tradeoffs. When the tolerance level α is increased from 1% to 10%, the average total energy consumption rises from 579.1×10^3 kWh to 584.2×10^3 kWh (a subtle 0.88% increase). On the other hand, the average of the total switching number in four weeks decreases from 196.9 to 120.9 (a significant 38.60% decrease). In real-world applications, the tolerance should be determined based on preferences regarding energy efficiency and control stability.

Compared to previous work to enhance control stability, the proposed MOO-based approach offers two advantages. Firstly, energy efficiency is not compromised. When the tolerance level α is 1%, the energy efficiency of the proposed approach is better than all baseline metaheuristic algorithms and close to global optimal solutions from the Gurobi solver. Even when the tolerance level α is 5%, the energy efficiency is close to GA and better than PSO. Secondly, the proposed approach is not designed for a specific optimization problem and is more general in various model-based optimization problems in building energy systems.

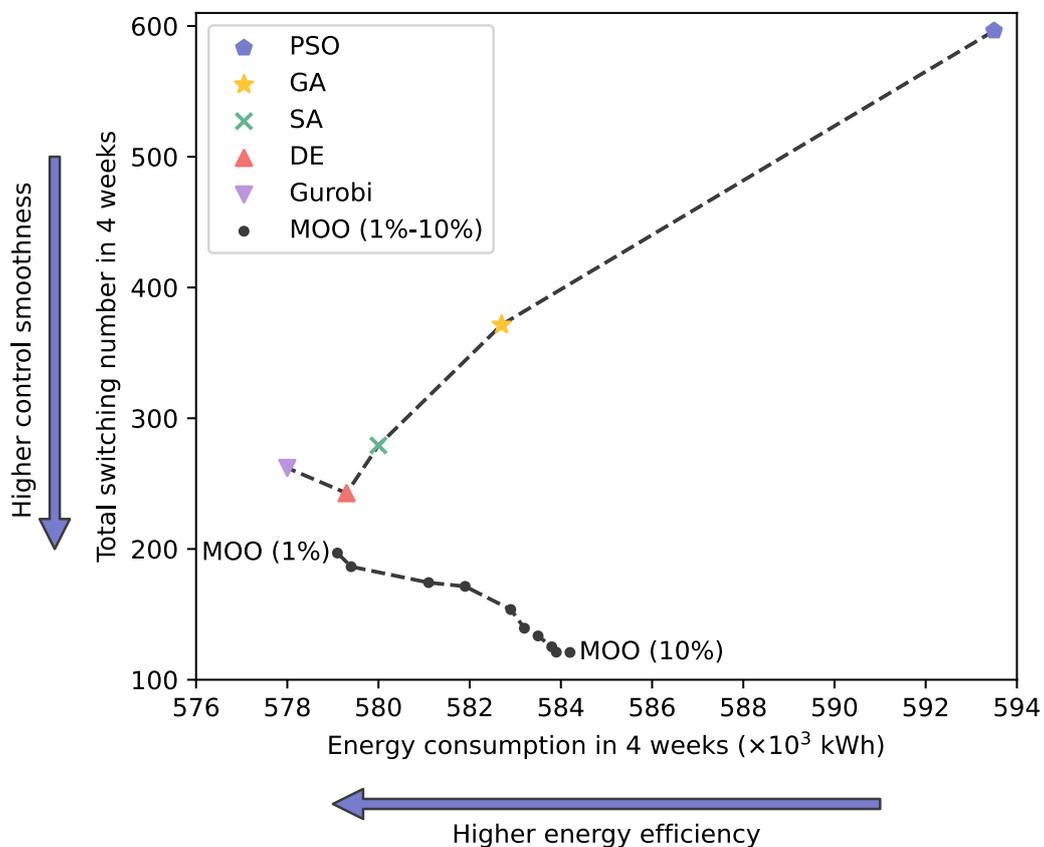


Figure 6.14. Comparison in energy efficiency and control stability

In **Figure 6.14**, the control stability provided by Gurobi, which offers a global optimal solution, is not superior to that of DE. The reason is that the disturbances in online control come from two aspects: shifts across local optima and variations in the optimization problem, as shown in **Figure 6.15**. Shifts across local optima occur because metaheuristic algorithms fail to find the global optimum and become trapped in local optima. Therefore, even if the optimization problem varies little from one control period to the next (for example, if the cooling load varies slightly), the randomness of the search process means that metaheuristic algorithms may find a near-optimal solution that is very different from the previous control action. This type of disturbance lessens as the optimization algorithm's ability to find the global optimum improves. Variations in the optimization problem refer to how, during the online optimization process, the objective function and constraints vary over time. For instance, in the optimal chiller optimization problem, constraints correspond to the real-time cooling load; the optimization algorithm needs to find a new solution under a new given cooling load to control the chillers. This disturbance intensifies as the optimization algorithm's ability to find the global optimum increases. This is because when the cooling load is relatively stable, a stronger ability to find the global optimum results in more stable control actions. However, when the cooling load exceeds or drops below a certain threshold, a stronger optimization algorithm can find a better solution within limited iterations than a weaker one, but at the cost of changing integer variables in the control actions (such as the number of chillers operating). Therefore, sudden load fluctuation could lead to unstable decisions regarding integer variables. Combining the disturbances from the shifts across local optima and the variations in optimization problems, DE exhibits superior control stability among the five baseline algorithms in the case studies. On the other hand, the proposed MOO-based online method reduces the disturbances from variations in optimization problems by measuring the distance of two successive control actions.

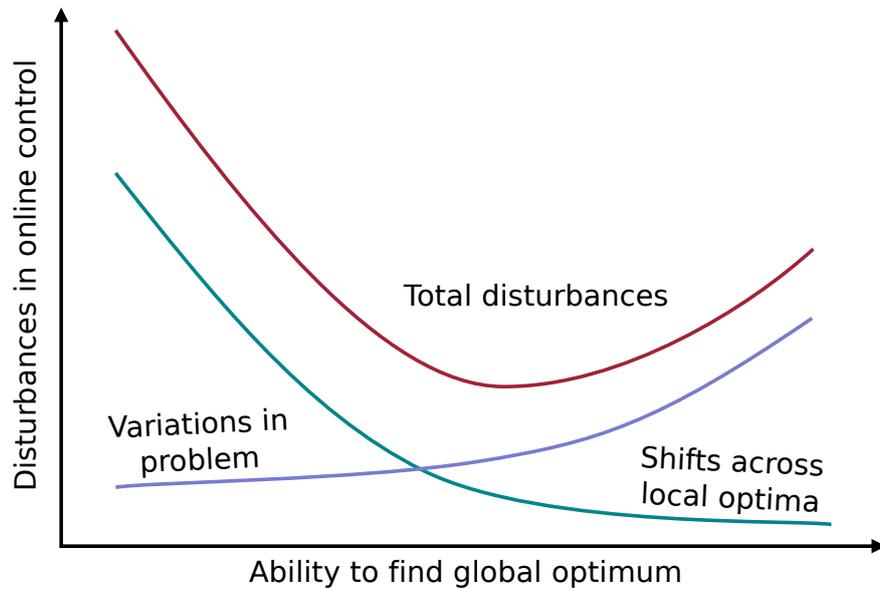


Figure 6.15. Disturbances in online control problems

6.4. Summary

Conventional deterministic and metaheuristic algorithms are not tailored for online optimization problems, and their control stability faces challenges in real-world applications. Therefore, this study introduces a multi-objective optimization (MOO)-based approach to enhance the stability in model-based online control for building energy systems. By considering the distance between successive control actions as one of the optimization objectives, the proposed approach can effectively search for non-dominated optimal solutions that are stable and energy-efficient, i.e., the Pareto front. The proposed MOO-based approach is validated using the optimal chiller loading problem with a 4-week cooling load curve. The approach is compared with the Gurobi solver, which provides global optimal solutions and four popular metaheuristic algorithms: particle swarm optimization (PSO), genetic algorithm (GA), simulated annealing (SA), and differential evolution (DE). The results demonstrate that the proposed approach substantially improves control stability while maintaining energy efficiency. Compared to DE, the most stable baseline algorithm, the MOO-based approach reduces the total chiller switching number by 18.87% at a tolerance level of $\alpha = 1\%$ and consumes 0.03% less energy. At a tolerance level of $\alpha = 10\%$, the MOO-based approach significantly reduces the total chiller switching number by 50.19% while only consuming 0.85% more energy than DE. The proposed MOO-based approach is promising for future applications to achieve energy-efficient, stable, and robust control in building energy systems.

CHAPTER 7 CONCLUSIONS

Machine learning and data-driven modeling have effectively facilitated building energy management in various typical applications in the past decade, such as cooling load prediction, fault detection and diagnosis. However, the data-driven approach faces two major challenges regarding practical applications: a lack of interpretability and reliability. Therefore, this study aims to develop data-driven methods of enhanced interpretability and reliability for optimal control and diagnosis of air-conditioning systems.

7.1. Summary of main contributions

The main contributions of this PhD study are summarized as follows, as illustrated in **Figure 7.1**. The new methods address general issues in the development of data-driven models (**Q1** to **Q3**) and the application of data-driven models in complex optimization problems (**Q4**), enhancing the applicability for deploying data-driven models.

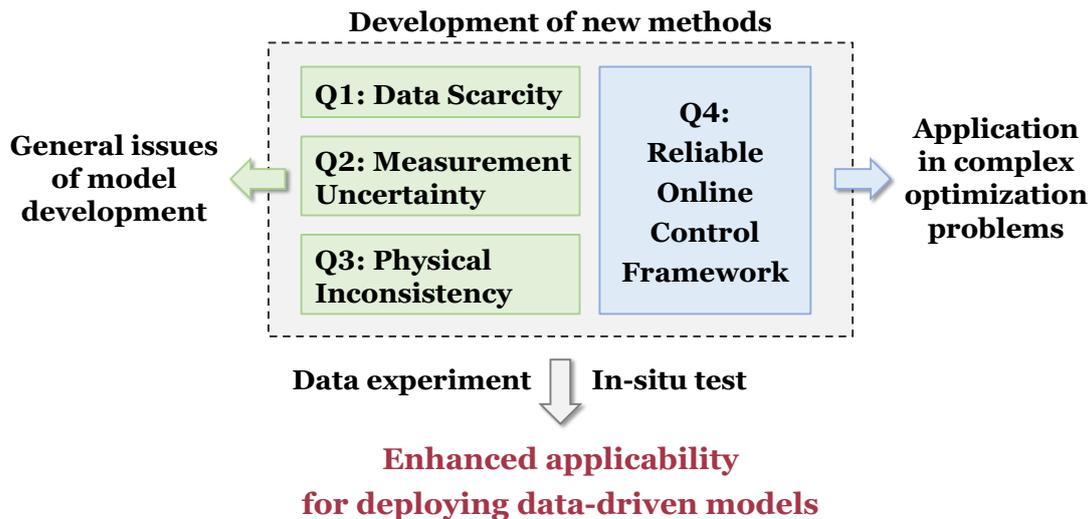


Figure 7.1. Contributions of this PhD study

1. To address the **data scarcity** issue in developing data-driven models in building HVAC systems, a novel similarity learning-based FDD method is proposed. The output is the similarity towards each fault, which provides better interpretability. The proposed Siamese network with LSTM subnetworks strengthens the model generalization ability by creating a large number of input pairs from limited labeled data. The temporal data-splitting method is adopted to tackle the issue of the high correlation of training data and test data using random split. Two case studies utilizing an open dataset were conducted to test the

effectiveness of the proposed method, with insufficient labeled data and imbalanced data, respectively.

2. To overcome the **measurement uncertainty** issue, a novel probabilistic machine learning method and risk evaluation scheme are proposed. One-hour-ahead probabilistic cooling load prediction in the form of a normal distribution is made using natural gradient boosting. A novel risk-based scheme then evaluates the risks of sequencing actions based on probabilistic predictions. This approach aims to make more informed decisions to better handle the uncertainties in cooling load predictions. The proposed strategy was deployed to a chiller plant, and an in-site experiment was conducted.
3. To tackle the **physical inconsistency** issue in HVAC pump modeling, a physically consistent data-driven optimal sequencing control integrating physical knowledge with data-driven modeling, enhancing the interpretability and reliability is proposed for real application. For different numbers of operating variable speed pumps, features composed of frequency and total flow rate are used to train the interpretable ElasticNet data-driven model. The optimal sequencing is then determined using a physically consistent frequency prediction method.
4. To achieve **reliable online control**, a multi-objective optimization (MOO)-based approach for enhancing the stability in model-based online control is proposed. The distance between two successive control actions is used as one of the optimization objectives besides energy saving. The proposed MOO-based approach significantly improves online control stability without compromising energy efficiency compared to conventional metaheuristic algorithms. The tolerance level of energy efficiency can be adjusted based on the requirements of energy efficiency and control stability in real-world applications. Previous strategies to improve control stability are usually tailored for specific problems. The proposed MOO-based approach is designed to be applicable to various model-based optimization problems with enhanced stability.

7.2. Conclusions

1. **A similarity-learning method for enhancing model generalization ability and interpretability**
 - The proposed Siamese network for fault diagnosis contains two identical LSTM subnetworks. The Siamese network is trained by pairs of multivariate time-series samples from the building energy management system and learns a similarity function.

- The case studies evaluate the proposed similarity learning-based FDD method under two typical scenarios during the operation of HVAC systems: when labeled data are limited (scenario 1) and imbalanced (scenario 2). The normal and faulty AHU operation data collected from the ASHRAE RP-1043 are used in the case studies. The data were treated using a temporal train-test split to ensure the generalization ability of the proposed method is properly evaluated.
- Scenario 1 tests the performance of the proposed method when all faults have a limited amount of labeled data. Compared to the baseline model, the proposed similarity learning-based method improved fault diagnostic accuracy by at most 45.7% when only a limited number of labeled data is available. Scenario 2 validates the performance of the proposed method when labeled data were imbalanced. The method demonstrates superior generalization ability when dealing with imbalanced labeled data, improving fault diagnosis accuracy from 73.0% to 82.6%.

2. Probabilistic machine learning for enhancing reliability and interpretability addressing measurement uncertainty

- This study proposes a novel probabilistic cooling load prediction-based chiller sequencing control strategy that considers prediction uncertainty to make sequencing actions more robust. An online risk-based action evaluation module is designed to determine the number of operating chillers based on probabilistic predictions, improving the interpretability and reliability of control decisions.
- A data experiment based on measured operational data is carried out to validate the proposed strategy. Results show that the total switching number is reduced by 43.6%, significantly decreasing unnecessary switching actions within 1 hour by 88.2%, thereby improving the reliability and lifespan of the chiller plant. In addition, the proposed strategy demonstrates better robustness when managing uncertainty from measurements.
- An in-situ test in an education building was carried out to validate the proposed strategy in robustness, thermal comfort, and energy efficiency. Three sets of test and reference days were selected based on two criteria: similar weather conditions and consistent types of working days. The results show that, first, the daily chiller switching number was reduced by 56.5%. Second, compared to reference days, the proposed strategy achieved an average daily energy saving of about 3945.1 kWh. Last, the COP of the chiller plant was increased by 4.2% on average.

3. Physically consistent data-driven models for enhancing interpretability and reliability

- This study proposes a physically consistent data-driven optimal sequencing control, integrating physical knowledge with data-driven modeling, enhancing the interpretability and reliability for real applications. For different numbers of operating VSPs, features composed of frequency and total flow rate are used to train the interpretable ElasticNet data-driven model. The optimal sequencing is then determined using a physically consistent frequency prediction method.
- The proposed strategy is validated using historical data from a real parallel variable pumping system. Results show that a total of 47.0% of the points that operating one secondary chilled water pump have energy saving potential, ranging from 0-25 kW (up to 40%). This indicates that the conventional differential pressure-based control cannot guarantee that parallel SCHWPs operate at the lowest energy consumption, and there is considerable potential for energy savings. 7.0% of the points fall that operating two secondary chilled water pumps have energy saving potential, ranging from 0-35 kW (up to 40%).
- The monthly energy savings analysis shows that the potential for energy savings in winter is minimal, with only a 2.5% reduction observed in January 2024, whereas substantial savings of 20.4% are evident in June 2024 during the summer. On average, the energy consumption of SCHWPs is reduced by approximately 10%.

4. General optimization framework for data-driven model-based online optimization with enhanced control smoothness

- This study proposes a MOO-based general optimization framework for enhancing the smoothness in model-based online control. The distance between two successive control actions is used as one of the optimization objectives besides energy saving. An adjustable tolerance level is used to select the appropriate solution from the Pareto front.
- The proposed approach is validated using a common optimization problem, i.e., optimal chiller loading for chillers. The proposed approach is compared with the Gurobi solver as the deterministic approach to find the global optimum, and the four most popular metaheuristic algorithms to find the near-optimal solution. Case studies show that when the tolerance level is set to 1%, the proposed MOO-based approach can save 24.8%, 18.9%, 29.5%, 47.0%, and 67.0% total chiller switching number compared to Gurobi, DE, SA,

GA, and PSO, respectively. The total energy saving of the proposed approach is also saved from 0.0% to 2.4% compared to metaheuristic algorithms. And the proposed consumes 0.2% more energy consumption compared to Gurobi. When the tolerance level is set to 10%, the proposed approach can reduce 53.9%, 50.2%, 56.7%, 67.5%, and 79.7% of the total switching number compared to Gurobi, DE, SA, GA, and PSO, respectively. And the total energy consumption is only 1.1%, 0.8%, 0.7, and 0.2% more than Gurobi, DE, SA, and GA.

- Compared to the previous work on enhancing control stability, the proposed MOO-based approach has two advantages. First, the energy efficiency is not compromised. Second, the proposed approach is not designed for a specific optimization problem and is more general in various model-based optimization problems. Therefore, the proposed MOO-based approach is a promising tool for real-world applications in achieving energy-efficient, stable, and robust control in building energy systems.

7.3. Recommendations for future work

This PhD study has made serious efforts on developing data-driven methods of enhanced interpretability and reliability for optimal control and diagnosis of air-conditioning systems. Case studies using open datasets and real operational data, as well as in-situ tests instead of simulation, have been carried out to validate the applicability of the proposed methods. In the future, more efforts can be made on the following aspects further to enhance the interpretability and reliability of data-driven applications:

1. Although the proposed similarity learning-based FDD method significantly improves the fault diagnosis accuracy when only limited labeled data is available, deploying the proposed method is still challenging. This is because only a few faults occur and are logged in a standalone HVAC system. In the future, techniques such as transfer learning can further enhance the reliability of the proposed method by sharing the labeled data across buildings.
2. Although the proposed probabilistic cooling load prediction-based robust chiller sequencing strategy proves to be robust and energy efficient based on in-situ tests, the operators predetermine the chiller sequencing order. In the future, the proposed strategy can be further improved by utilizing the day-ahead cooling load prediction and chiller COP prediction model to determine the optimal chiller sequencing order. In this way, energy savings can be further improved.

3. Although the proposed optimal pump sequencing strategy proves to be interpretable, reliable, and energy efficient, its deployment to the primary-secondary chilled water system is still on-going. To further improve the reliability for deployment, the robustness of sequencing control should be considered. Like the problem of chiller sequencing control, the control robustness of pump sequencing control can be improved by utilizing a risk-based scheme leveraging probabilistic prediction.
4. Although the proposed multi-objective optimization (MOO)-based approach proves its superior stability over conventional deterministic algorithms and metaheuristic algorithms in model-based online control using a case study based on historical data, it has not been deployed and validated in the actual HVAC system. More practical problems, such as model updating and raw data preprocessing, should be considered to enhance its practicability in the future.

APPENDICES

Appendix A: Results of testing day 2 and testing day 3 in Chapter 4

A.1 Testing day 2

A.1.1 Weather comparison with reference day 2

Reference day 2 (June 2, 2023) and testing day 2 (July 13, 2023) are Friday and Thursday, respectively. The rule-based strategy was used in the BMS on reference day 2, while the proposed strategy was used on testing day 2. Both reference day 2 and testing day 2 were cloudy. A comparison of the weather between reference day 2 and testing day 2 is shown in **Figure A.1**. The outdoor temperature and relative humidity trends on these two days are very similar. **Table A.1** also presents a comparison of the highest temperature, average temperature, lowest temperature, and average relative humidity on these two days. From these four values, it can be observed that the weather on these two days is very similar. The highest temperature on testing day 2 (34.8°C) is slightly lower than on reference day 2 (35.0°C). And the average relative humidity on testing day 2 (71%) is slightly lower than on reference day 2 (76%).

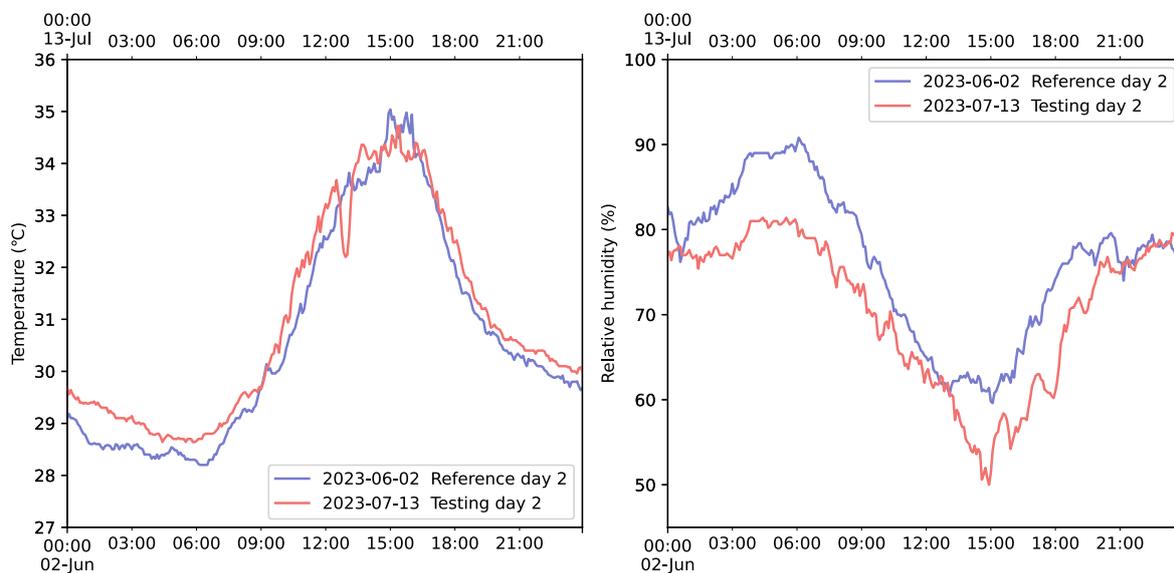


Figure A.1. Weather conditions on reference day 2 and testing day 2

Table A.1. Weather summary on reference day 2 and testing day 2

Date	Maximum temperature (°C)	Average temperature (°C)	Minimum temperature (°C)	Average relative humidity (%)
2023-06-02 Reference day 2	35.0	30.6	28.2	76
2023-07-13 Testing day 2	34.8	31.0	28.6	71

A.1.2 Probabilistic cooling load prediction

Figure A.2 shows the probabilistic cooling load prediction on the testing day. The different shades of blue in the figure represent the 25%, 50%, 75%, and 95% confidence intervals. The width of the 90% confidence interval indicates that the probabilistic cooling load uncertainty is relatively small at night and increases significantly in the morning start-up period. This is because at night, the chillers only need to provide the basic cooling load for the 24-hour rooms (i.e., the 24-hour study center in the library), while in the morning, each room needs to remove the heat gain from the night, leading to larger load fluctuations. On the right side of **Figure A.2**, the probabilistic cooling load probability distribution at 12:00 on the testing day is shown, with a predicted average value of 7474.0 kW and a standard deviation (i.e., uncertainty) of 452.4 kW. In testing day 2, the root mean square error (RMSE) for comparing the predicted average value of the probabilistic cooling load with the actual value is 393.0 kW, and the mean absolute percentage error (MAPE) is 8.8%.

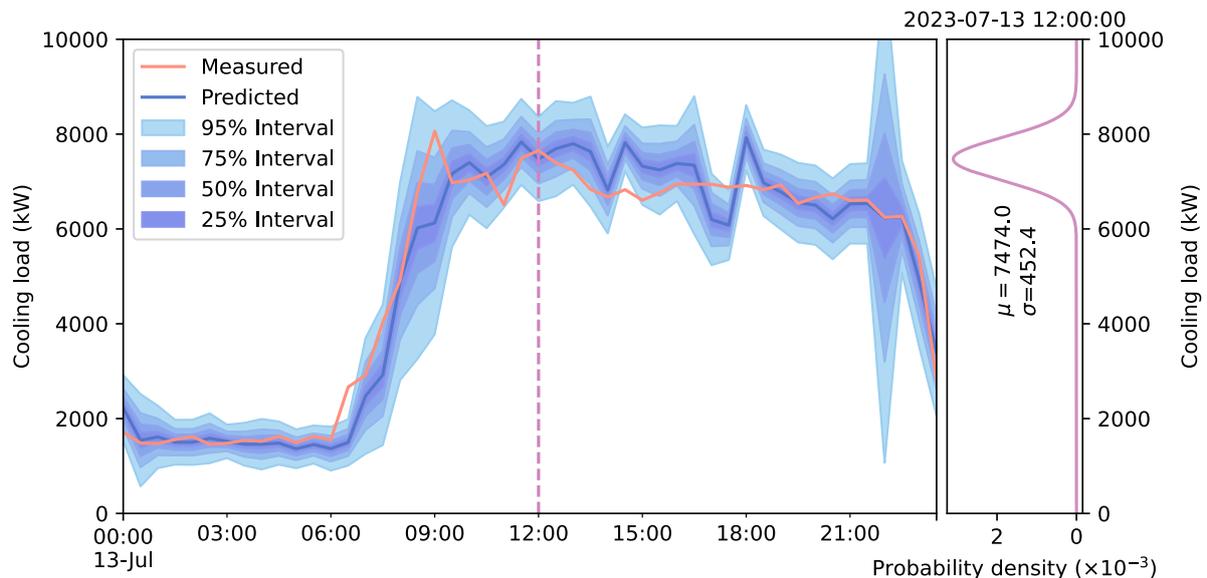


Figure A.2. 1 hour ahead probabilistic cooling load prediction on testing day 2

A.1.3 Robustness of sequencing actions and thermal comfort

The chiller sequencing results for testing day 2 and reference day 2 are shown in **Figure A.3**. The y-axis on the left represents the number of chillers in operation, while the y-axis on the right represents the chiller sequencing order. Therefore, by combining the two y-coordinates, the on/off status of each chiller at every moment can be determined on the reference day and testing day. Compared with the reference day under the rule-based strategy, the proposed strategy on the testing day significantly reduces the total switching number of chillers from 24 to 9, as shown in **Table A.2**. Moreover, the maximum number of chillers turned on in the morning start-up period (6 a.m. to 12 a.m.) is reduced from 8 to 5. This comparison demonstrates that the adoption of the proposed strategy significantly improves the robustness of the chiller sequencing control.

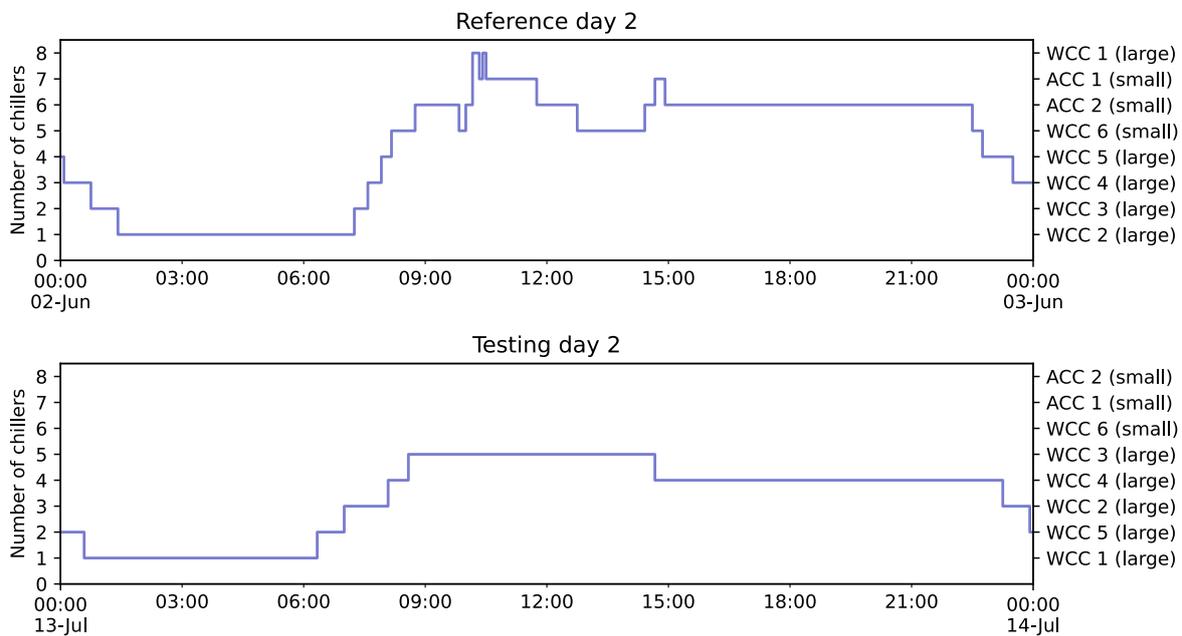


Figure A.3. Chiller sequencing results of reference day 2 and testing day 2

Table A.2. Chiller switching number on reference day 2 and testing day 2

Date	Total switching number	Maximum chiller during morning start-up
Reference day 2	24	8
Testing day 2	9	5

The chilled water return temperature of three loops (Phase 1, Phase 2, and Library) on reference day 2 and testing day 2 are shown in **Figure A.4**. It can be seen from the figure that there are some problems with the rule-based strategy. The chillers were switched on and off frequently, which increased the instability of chilled water temperature. When adopting the proposed strategy, the issue is significantly alleviated. The control decisions based on

probabilistic cooling load predictions significantly avoid unnecessary on/off switching actions, which makes the chiller system operate more efficiently and effectively.

Additionally, at 14:30 on testing day 2, the proposed strategy recommended a switching-off control action based on probabilistic cooling load prediction. Following the implementation of this control action, the chilled water return temperature did not exhibit the same abrupt changes as when using the rule-based strategy.

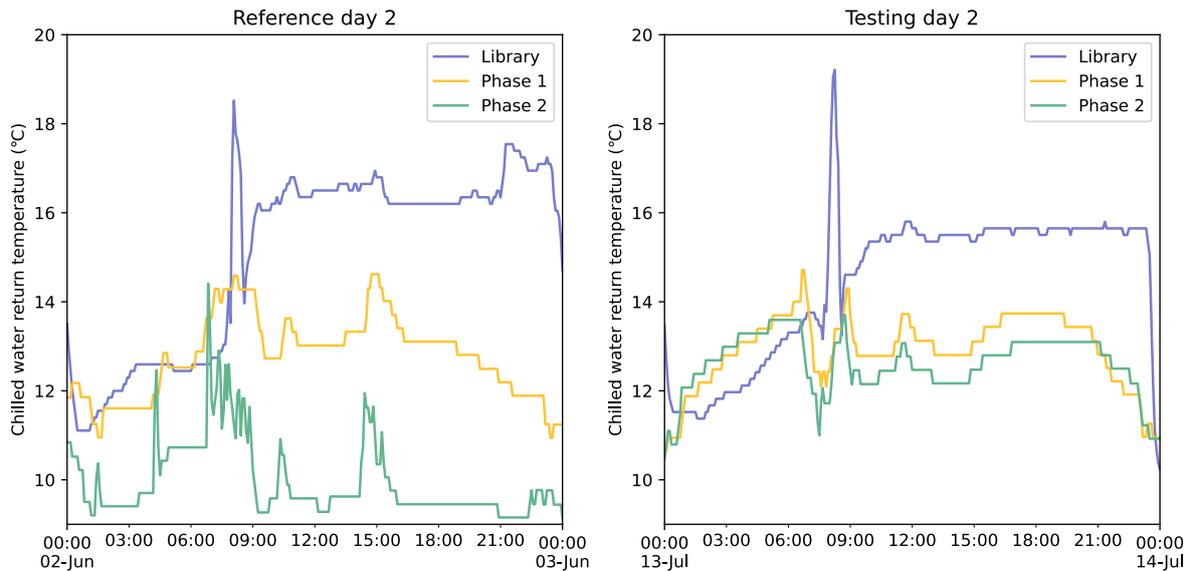


Figure A.4. Chilled water return temperature of three loops

A.1.4 Energy efficiency

The energy consumption of chillers, primary chilled water pumps (PCHWP), condensing water pumps (CDWP), and cooling towers (CT) on reference day 2 and testing day 2 are shown in **Table A.3**. Compared with the reference day which adopted the rule-based strategy, the proposed strategy significantly reduced the energy of various equipment on the testing day. The reduction of energy consumption mainly comes from two aspects. The first aspect is that the probabilistic cooling load prediction achieves timely and reasonable switching on, especially in the morning start-up period, which reduces the maximum number of chillers that need to be turned on during the peak period, and also reduces the operation of the corresponding PCHWP, CDWP, and CT of the chillers. On the other hand, the timely chiller switching off is realized through the prediction of cooling load, which reduces the energy consumption of chillers and other equipment. Specifically, the energy consumption of chillers, PCHWP, CDWP, and CT are reduced by 3765.5 kWh, 564.2 kWh, 256.2 kWh, and 221.7 kWh, respectively. The total energy consumption is saved by 4807.6 kWh.

Table A.3. Energy consumption on reference day 2 and testing day 2 (kWh)

Date	Chiller	PCHWP	CDWP	CT	Total
Reference day 2	30241.5	4996.3	6300.0	2362.2	43899.9
Testing day 2	26476.0	4432.1	6043.8	2140.5	39092.3
Difference	3765.5	564.2	256.2	221.7	4807.6

Table A.4 provides a comparison of the energy efficiency of the entire chiller plant on reference day 2 and testing day 2. Owing to the proposed strategy, the system COP is improved by about 5.2%.

Table A.4. Energy efficiency of chiller plant on reference day 2 and testing day 2

Date	Total energy consumption (kWh)	Total cooling load (kWh)	COP
Reference day 2	43899.9	129435	2.95
Testing day 2	39092.3	121235	3.10
Difference	-11.0%	-6.3%	5.2%

A.2 Testing day 3

A.2.1 Weather comparison with reference day 3

Reference day 3 (June 4, 2023) and testing day 3 (June 11, 2023) are both Sunday. The rule-based strategy was used in the BMS on reference day 3, while the proposed strategy was used on testing day 3. Both reference day 3 and testing day 3 were cloudy. A comparison of the weather between reference day 3 and testing day 3 is shown in **Figure A.6**. The outdoor temperature and relative humidity trends on these two days are very similar. **Table A.5** also presents a comparison of the highest temperature, average temperature, lowest temperature, and average relative humidity on these two days. From these four values, it can be observed that the weather on these two days is very similar. The highest temperature on testing day 3 (32.4°C) is slightly lower than on reference day 3 (32.6°C). And the average relative humidity on testing day 3 (83%) is slightly higher than on reference day 3 (81%).

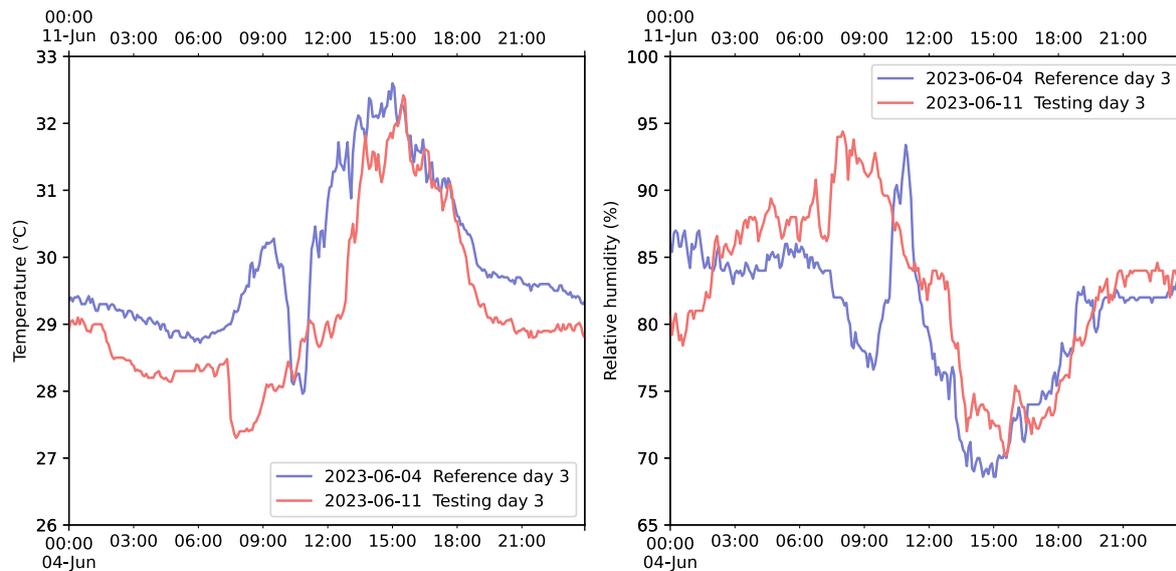


Figure A.6. Weather conditions on reference day 3 and testing day 3

Table A.5. Weather summary on reference day 3 and testing day 3

Date	Maximum temperature (°C)	Average temperature (°C)	Minimum temperature (°C)	Average relative humidity (%)
2023-06-04 Reference day 3	32.6	30.0	28.0	81
2023-06-11 Testing day 3	32.4	29.2	27.3	83

A.2.2 Probabilistic cooling load prediction

Figure A.7 shows the probabilistic cooling load prediction on the testing day. The different shades of blue in the figure represent the 25%, 50%, 75%, and 95% confidence intervals. The width of the 90% confidence interval indicates that the probabilistic cooling load uncertainty is relatively small at night and increases significantly in the morning start-up period. This is because at night, the chillers only need to provide the basic cooling load for the 24-hour rooms (i.e., the 24-hour study center in the library), while in the morning, each room needs to remove the heat gain from the night, leading to larger load fluctuations. On the right side of **Figure A.7**, the probabilistic cooling load probability distribution at 12:00 on the testing day is shown, with a predicted average value of 4101.1 kW and a standard deviation (i.e., uncertainty) of 550.4 kW. In testing day 3, the root mean square error (RMSE) for comparing the predicted average value of the probabilistic cooling load with the actual value is 334.0 kW, and the mean absolute percentage error (MAPE) is 11.3%.

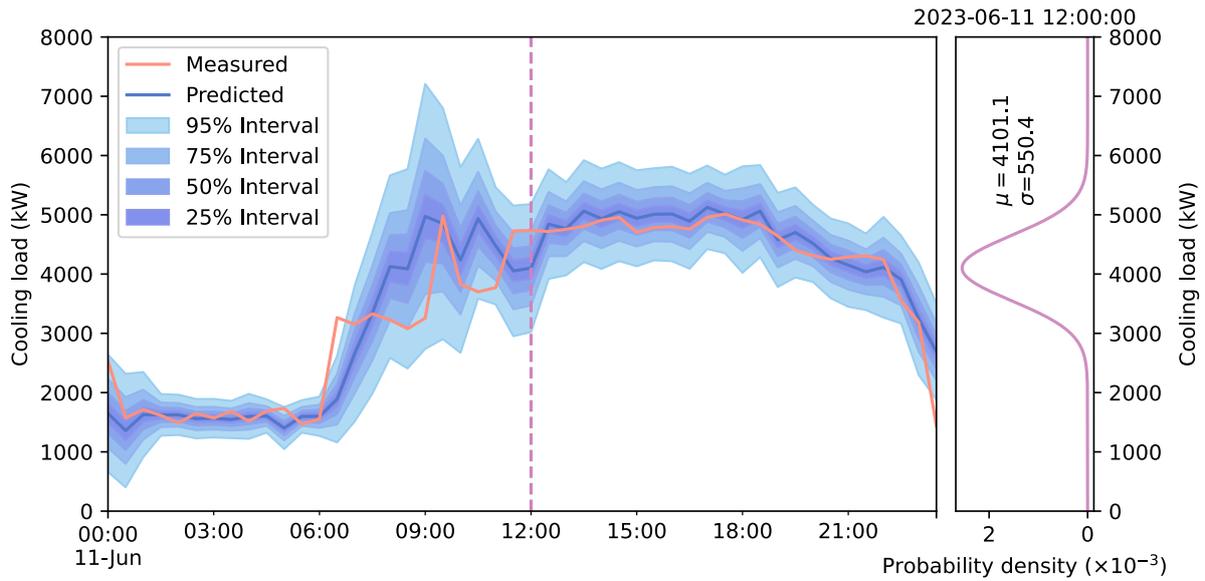


Figure A.7. 1 hour ahead probabilistic cooling load prediction on testing day 3

A.2.3 Robustness of sequencing actions and thermal comfort

The chiller sequencing results for testing day 3 and reference day 3 are shown in **Figure A.8**. The y-axis on the left represents the number of chillers in operation, while the y-axis on the right represents the chiller sequencing order. Therefore, by combining the two y-coordinates, the on/off status of each chiller at every moment can be determined on the reference day and testing day. Compared with the reference day under the rule-based strategy, the proposed strategy on the testing day significantly reduces the total switching number of chillers from 7 to 6, as shown in **Table A.6**. Moreover, the maximum number of chillers turned on in the morning start-up period (6 a.m. to 12 a.m.) is reduced from 4 to 3. This comparison demonstrates that the adoption of the proposed strategy significantly improves the robustness of the chiller sequencing control.

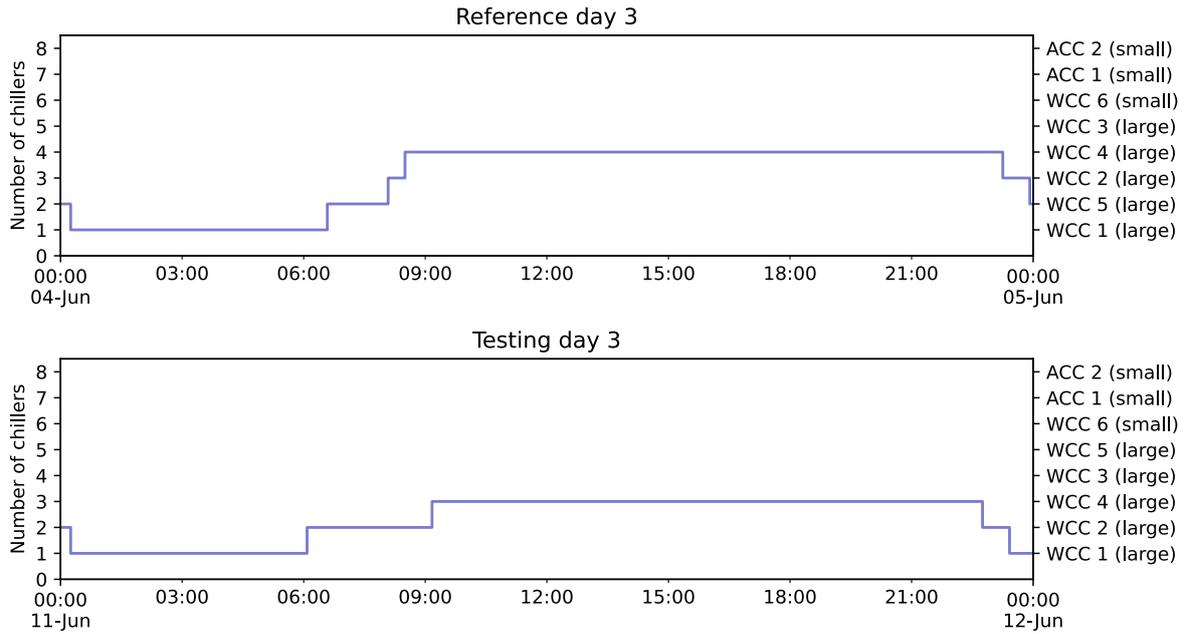


Figure A.8. Chiller sequencing results on reference day 3 and testing day 3

Table A.6. Chiller switching number on reference day 3 and testing day 3

Date	Total switching number	Maximum chiller during morning start-up
Reference day 3	7	4
Testing day 3	6	3

The chilled water return temperature of three loops (Phase 1, Phase 2, and Library) on reference day 3 and testing day 3 are shown in **Figure A.9**. When adopting the proposed strategy, the water temperature was more stable during the testing day. This indicates that the proposed strategy can make more reasonable on/off control of the chiller based on probabilistic cooling load predictions. Furthermore, the control decisions based on probabilistic cooling load predictions significantly avoid unnecessary on/off switching actions, which makes the chiller system operate more efficiently and effectively.

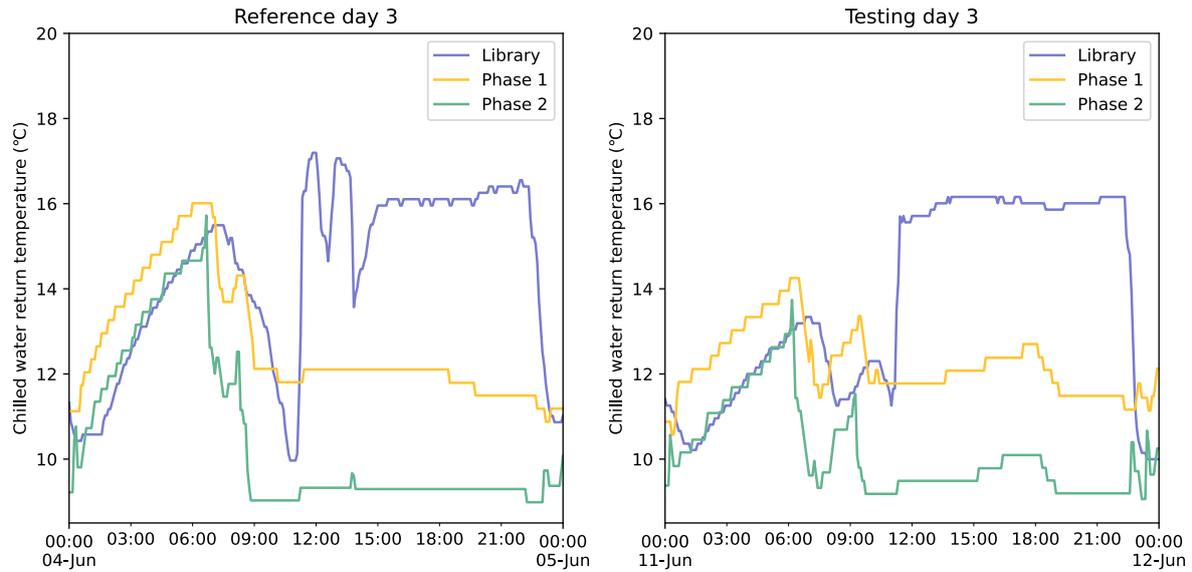


Figure A.9. Chilled water return temperature of three loops

A.2.4 Energy efficiency

The energy consumption of chillers, primary chilled water pumps (PCHWP), condensing water pumps (CDWP), and cooling towers (CT) on reference day 3 and testing day 3 are shown in **Table A.7**. Compared with the reference day which adopted the rule-based strategy, the proposed strategy significantly reduced the energy of various equipment on the testing day. The reduction of energy consumption mainly comes from two aspects. The first aspect is that the probabilistic cooling load prediction achieves timely and reasonable switching on, especially in the morning start-up period, which reduces the maximum number of chillers that need to be turned on during the peak period, and also reduces the operation of the corresponding PCHWP, CDWP, and CT of the chillers. On the other hand, the timely chiller switching off is realized through the prediction of cooling load, which reduces the energy consumption of chillers and other equipment. Specifically, the energy consumption of chillers, PCHWP, CDWP, and CT are reduced by 1321.4 kWh, 944.2 kWh, 1287.5 kWh, and 515.0 kWh, respectively. The total energy consumption is saved by 4068.0 kWh.

Table A.7. Energy consumption on reference day 3 and testing day 3 (kWh)

Date	Chiller	PCHWP	CDWP	CT	Total
Reference day 3	21835.3	3982.9	5431.3	1895.5	33145.0
Testing day 3	20513.9	3038.8	4143.8	1380.5	29076.9
Difference	1321.4	944.2	1287.5	515.0	4068.0

Table A.8 provides a comparison of the energy efficiency of the entire chiller plant on reference day 3 and testing day 3. Owing to the proposed strategy, the system COP is improved by about 4.8%.

Table A.8. Energy efficiency of chiller plant on reference day 3 and testing day 3

Date	Total energy consumption (kWh)	Total cooling load (kWh)	COP
Reference day 3	33145.0	91000.6	2.75
Testing day 3	29076.9	83628.5	2.88
Difference	-12.3%	-8.1%	4.8%

REFERENCES

- [1] IEA. 2021 Global status report for buildings and construction. United Nations Environment Programme 2021.
- [2] Ramesh T, Prakash R, Shukla KK. Life cycle energy analysis of buildings: An overview. *Energy and Buildings* 2010;42:1592–600.
- [3] Zhou N, Khanna N, Feng W. Policy roadmap to 50% energy reduction in Chinese buildings by 2050. *Procs ACEEE Summer Study on Energy Efficiency in Buildings* 2016:9–1.
- [4] Environmental Bureau HKSAR. Hong Kong's Climate Action Plan 2050. 2021.
- [5] Langevin J, Harris CB, Reyna JL. Assessing the Potential to Reduce U.S. Building CO2 Emissions 80% by 2050. *Joule* 2019;3:2403–24.
- [6] European Environment Agency. Greenhouse gas emissions from energy use in buildings in Europe n.d.
- [7] Hurtado LA, Mocanu E, Nguyen PH, Gibescu M, Kamphuis RIG. Enabling Cooperative Behavior for Building Demand Response Based on Extended Joint Action Learning. *IEEE Transactions on Industrial Informatics* 2018;14:127–36.
- [8] Tang H, Wang S, Li H. Flexibility categorization, sources, capabilities and technologies for energy-flexible and grid-responsive buildings: State-of-the-art and future perspective. *Energy* 2021;219:119598.
- [9] Chen Y, Chen Z, Yuan X, Su L, Li K. Optimal Control Strategies for Demand Response in Buildings under Penetration of Renewable Energy. *Buildings* 2022;12:371.
- [10] Capozzoli A, Cerquitelli T, Piscitelli MS. Chapter 11 - Enhancing energy efficiency in buildings through innovative data analytics technologies. In: Dobre C, Xhafa F, editors. *Pervasive Computing*, Boston: Academic Press; 2016, p. 353–89.
- [11] IBM. Building artificial intelligence into buildings n.d.
- [12] Zhang L, Wen J, Li Y, Chen J, Ye Y, Fu Y, et al. A review of machine learning in building load prediction. *Applied Energy* 2021;285:116452.
- [13] Liao Y, Huang G. A hybrid predictive sequencing control for multi-chiller plant with considerations of indoor environment control, energy conservation and economical operation cost. *Sustainable Cities and Society* 2019;49:101616.
- [14] Afroz Z, Shafiullah G, Urme T, Higgins G. Modeling techniques used in building HVAC control systems: A review. *Renewable and Sustainable Energy Reviews* 2018;83:64–84.
- [15] Jorissen F, Boydens W, Helsen L. TACO, an automated toolchain for model predictive control of building systems: implementation and verification. *Journal of Building Performance Simulation* 2019;12:180–92.
- [16] Mariano-Hernández D, Hernández-Callejo L, Zorita-Lamadrid A, Duque-Pérez O, Santos García F. A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis. *Journal of Building Engineering* 2021;33:101692.

- [17] Gaur M, Makonin S, Bajić IV, Majumdar A. Performance Evaluation of Techniques for Identifying Abnormal Energy Consumption in Buildings. *IEEE Access* 2019;7:62721–33.
- [18] Molnar C. *Interpretable Machine Learning*. Morisville, North Carolina: Lulu.com; 2019.
- [19] Vollert S, Atzmueller M, Theissler A. Interpretable Machine Learning: A brief survey from the predictive maintenance perspective. 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 2021, p. 01–8.
- [20] Naser MZ. An engineer’s guide to eXplainable Artificial Intelligence and Interpretable Machine Learning: Navigating causality, forced goodness, and the false perception of inference. *Automation in Construction* 2021;129:103821.
- [21] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 2020;58:82–115.
- [22] Thampi A. *Interpretable AI*. Manning; 2022.
- [23] Fei J, Chen Y, Liu L, Fang Y. Fuzzy Multiple Hidden Layer Recurrent Neural Control of Nonlinear System Using Terminal Sliding-Mode Controller. *IEEE Trans Cybern* 2022;52:9519–34.
- [24] Burkart N, Huber MF. A Survey on the Explainability of Supervised Machine Learning. *Jair* 2021;70:245–317.
- [25] Yan K, Zhong C, Ji Z, Huang J. Semi-supervised learning for early detection and diagnosis of various air handling unit faults. *Energy and Buildings* 2018;181:75–83.
- [26] Kamath U, Liu J. *Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning*. Cham: Springer International Publishing; 2021.
- [27] Chen Y. Physics-informed neural networks for building thermal modeling and demand response control. *Building and Environment* 2023.
- [28] Sha H, Xu P, Lin M, Peng C, Dou Q. Development of a multi-granularity energy forecasting toolkit for demand response baseline calculation. *Applied Energy* 2021;289:116652.
- [29] Zhuang H, Wang X, Bendersky M, Grushetsky A, Wu Y, Mitrichev P, et al. Interpretable Ranking with Generalized Additive Models. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, New York, NY, USA: Association for Computing Machinery; 2021, p. 499–507.
- [30] Toubreau J-F, Bottieau J, Wang Y, Vallee F, Shan S, Cao B, et al. Forecasting the Short-Term Electricity Consumption of Building Using a Novel Ensemble Model. *IEEE Access* 2019;7:88093–106.
- [31] Wang H, Cai R, Zhou B, Aziz S, Qin B, Voropai N, et al. Solar irradiance forecasting based on direct explainable neural network. *Energy Conversion and Management* 2020;226:113487.
- [32] Zhang X, Chung F-L, Wang S. An Interpretable Fuzzy DBN-Based Classifier for Indoor User Movement Prediction in Ambient Assisted Living Applications. *IEEE Trans Ind Inf* 2020;16:42–53.
- [33] Kim E. Interpretable and Accurate Convolutional Neural Networks for Human Activity Recognition. *IEEE Trans Ind Inf* 2020;16:7190–8.

- [34] Chen Y, Zhang D. Theory-guided deep-learning for electrical load forecasting (TgDLF) via ensemble long short-term memory. *Advances in Applied Energy* 2021;1:100004.
- [35] Chen Z, Xiao F, Shi J, Li A. Dynamic model development for vehicle air conditioners based on physics-guided deep learning. *International Journal of Refrigeration* 2022;134:126–38.
- [36] Yu Z, Yang X, Gao F, Huang J, Tu R, Cui J. A Knowledge-based reinforcement learning control approach using deep Q network for cooling tower in HVAC systems. 2020 Chinese Automation Congress (CAC), 2020, p. 1721–6.
- [37] Drgoňa J, Tuor AR, Chandan V, Vrabie DL. Physics-constrained deep learning of multi-zone building thermal dynamics. *Energy and Buildings* 2021;243:110992.
- [38] Di Natale L, Svetozarevic B, Heer P, Jones CN. Physically Consistent Neural Networks for building thermal modeling: theory and analysis. arXiv:211203212 [Cs, Eess] 2022.
- [39] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate 2022.
- [40] Li C, Dong Z, Ding L, Petersen H, Qiu Z, Chen G, et al. Interpretable Memristive LSTM Network Design for Probabilistic Residential Load Forecasting. *IEEE Trans Circuits Syst I* 2022;69:2297–310.
- [41] Rojat T, Puget R, Filliat D, Del Ser J, Gelin R, Díaz-Rodríguez N. Explainable Artificial Intelligence (XAI) on TimeSeries Data: A Survey. arXiv:210400950 [Cs] 2021.
- [42] Li D, Li D, Li C, Li L, Gao L. A novel data-temporal attention network based strategy for fault diagnosis of chiller sensors. *Energy and Buildings* 2019;198:377–94.
- [43] Gao Y, Ruan Y. Interpretable deep learning model for building energy consumption prediction based on attention mechanism. *Energy and Buildings* 2021;252:111379.
- [44] Azam MF, Younis MS. Multi-Horizon Electricity Load and Price Forecasting Using an Interpretable Multi-Head Self-Attention and EEMD-Based Framework. *IEEE Access* 2021;9:85918–32.
- [45] Li A, Xiao F, Zhang C, Fan C. Attention-based interpretable neural network for building cooling load prediction. *Applied Energy* 2021;299:117238.
- [46] Gangopadhyay T, Tan SY, Jiang Z, Sarkar S. Interpretable Deep Attention Model for Multivariate Time Series Prediction in Building Energy Systems. In: Darema F, Blasch E, Ravela S, Aved A, editors. *Lecture Notes in Computer Science*, Cham: Springer International Publishing; 2020, p. 93–101.
- [47] Gangopadhyay T, Tan SY, Jiang Z, Meng R, Sarkar S. Spatiotemporal Attention for Multivariate Time Series Prediction and Interpretation. ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, p. 3560–4.
- [48] Guo T, Lin T, Lu Y. An interpretable LSTM neural network for autoregressive exogenous model. arXiv:180405251 [Cs, Stat] 2018.
- [49] Bhatia A, Garg V, Haves P, Pudi V. Explainable Clustering Using Hyper-Rectangles for Building Energy Simulation Data. *IOP Conf Ser: Earth Environ Sci* 2019;238:012068.
- [50] Grimaldo A, Novak J. Explainable Needn't Be (Much) Less Accurate: Evaluating an Explainable AI Dashboard for Energy Forecasting. In: Maglogiannis I, Macintyre J, Iliadis L, editors. *Artificial Intelligence Applications and Innovations. AIAI 2021 IFIP*

- WG 12.5 International Workshops, Cham: Springer International Publishing; 2021, p. 340–51.
- [51] Grimaldo AI, Novak J. User-Centered Visual Analytics Approach for Interactive and Explainable Energy Demand Analysis in Prosumer Scenarios. In: Tzovaras D, Giakoumis D, Vincze M, Argyros A, editors. *Computer Vision Systems*, vol. 11754, Cham: Springer International Publishing; 2019, p. 700–10.
- [52] Grimaldo AI, Novak J. Combining Machine Learning with Visual Analytics for Explainable Forecasting of Energy Demand in Prosumer Scenarios. *Procedia Computer Science* 2020;175:525–32.
- [53] Amato U, Antoniadis A, De Feis I, Goude Y, Lagache A. Forecasting high resolution electricity demand data with additive models including smooth and jagged components. *International Journal of Forecasting* 2021;37:171–85.
- [54] Bujalski M, Madejski P. Forecasting of Heat Production in Combined Heat and Power Plants Using Generalized Additive Models. *Energies* 2021;14:2331.
- [55] Pathak N, Ba A, Ploennigs J, Roy N. Forecasting Gas Usage for Big Buildings Using Generalized Additive Models and Deep Learning. 2018 IEEE International Conference on Smart Computing (SMARTCOMP), 2018, p. 203–10.
- [56] Charalampopoulos I. A comparative sensitivity analysis of human thermal comfort indices with generalized additive models. *Theor Appl Climatol* 2019;137:1605–22.
- [57] Voss M, Heinekamp JF, Krutzsch S, Sick F, Albayrak S, Strunz K. Generalized Additive Modeling of Building Inertia Thermal Energy Storage for Integration Into Smart Grid Control. *IEEE Access* 2021;9:71699–711.
- [58] Sundararajan A, Ollis B. Regression and Generalized Additive Model to Enhance the Performance of Photovoltaic Power Ensemble Predictors. *IEEE Access* 2021;9:111899–914.
- [59] Khamma TR, Zhang Y, Guerrier S, Boubekri M. Generalized additive models: An efficient method for short-term energy prediction in office buildings. *Energy* 2020;213:118834.
- [60] Ribeiro MT, Singh S, Guestrin C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. arXiv:160204938 [Cs, Stat] 2016.
- [61] Wastensteiner J. Explainable AI for Tailored Electricity Consumption Feedback – An Experimental Evaluation of Visualizations 2021:19.
- [62] Madhikermi M, Malhi AK, Främling K. Explainable Artificial Intelligence Based Heat Recycler Fault Detection in Air Handling Unit. In: Calvaresi D, Najjar A, Schumacher M, Främling K, editors. *Explainable, Transparent Autonomous Agents and Multi-Agent Systems*, Cham: Springer International Publishing; 2019, p. 110–25.
- [63] Srinivasan S, Arjunan P, Jin B, Sangiovanni-Vincentelli AL, Sultan Z, Poolla K. Explainable AI for Chiller Fault-Detection Systems: Gaining Human Trust. *Computer* 2021;54:60–8.
- [64] Fan C, Xiao F, Yan C, Liu C, Li Z, Wang J. A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning. *Applied Energy* 2019;235:1551–60.

- [65] Kotevska O, Munk J, Kurte K, Du Y, Amasyali K, Smith RW, et al. Methodology for Interpretable Reinforcement Learning Model for HVAC Energy Control. 2020 IEEE International Conference on Big Data (Big Data), IEEE; 2020, p. 1555–64.
- [66] Zdravković M, Ćirić I, Ignjatović M. Towards explainable AI-assisted operations in District Heating Systems. *IFAC-PapersOnLine* 2021;54:390–5.
- [67] Zdravković M, Ćirić I, Ignjatović M. Explainable heat demand forecasting for the novel control strategies of district heating systems. *Annual Reviews in Control* 2022.
- [68] Arjunan P, Poolla K, Miller C. BEEM: Data-driven building energy benchmarking for Singapore. *Energy and Buildings* 2022;260:111869.
- [69] Jin X, Xiao F, Zhang C, Li A. GEIN: An interpretable benchmarking framework towards all building types based on machine learning. *Energy and Buildings* 2022;260:111909.
- [70] Geyer P, Singh MM, Chen X. Explainable AI for engineering design: A unified approach of systems engineering and component-based deep learning. arXiv:210813836 [Cs, Eess] 2022.
- [71] Kuzlu M, Cali U, Sharma V, Guler O. Gaining Insight Into Solar Photovoltaic Power Generation Forecasting Utilizing Explainable Artificial Intelligence Tools. *IEEE Access* 2020;8:187814–23.
- [72] Ugwuanyi C. Using interpretable machine learning for indoor CO₂ level prediction and occupancy estimation. PhD Thesis. University of Strathclyde, 2021.
- [73] Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc.; 2017.
- [74] Carlsson L, Samuelsson P, Jönsson P. Using Interpretable Machine Learning to Predict the Electrical Energy Consumption of an Electric Arc Furnace. *Stahl Und Eisen (1881)* 2019;139:24–9.
- [75] Carlsson LS, Samuelsson PB, Jönsson PG. Interpretable Machine Learning—Tools to Interpret the Predictions of a Machine Learning Model Predicting the Electrical Energy Consumption of an Electric Arc Furnace. *Steel Research International* 2020;91:2000053.
- [76] Park S, Moon J, Hwang E. Explainable Anomaly Detection for District Heating Based on Shapley Additive Explanations. 2020 International Conference on Data Mining Workshops (ICDMW), 2020, p. 762–5.
- [77] Gao Y, Han H, Lu H, Jiang S, Zhang Y, Luo M. Knowledge mining for chiller faults based on explanation of data-driven diagnosis. *Applied Thermal Engineering* 2022;205:118032.
- [78] Santos RN, Yamouni S, Albiero B, Vicente R, Silva J, Souza T, et al. Gradient boosting and Shapley additive explanations for fraud detection in electricity distribution grids. *Int Trans Electr Energy Syst* 2021;31:e13046.
- [79] Papadopoulos S, Kontokosta CE. Grading buildings on energy performance using city benchmarking data. *Applied Energy* 2019;233–234:244–53.
- [80] Arjunan P, Poolla K, Miller C. EnergyStar++: Towards more accurate and explanatory building energy benchmarking. *Applied Energy* 2020;276:115413.
- [81] Chang X, Li W, Ma J, Yang T, Zomaya AY. Interpretable Machine Learning In Sustainable Edge Computing: A Case Study of Short-Term Photovoltaic Power Output

- Prediction. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, p. 8981–5.
- [82] Movahedi A, Derrible S. Interrelated Patterns of Electricity, Gas, and Water Consumption in Large-Scale Buildings 2020.
- [83] Bellahsen A, Dagdougui H. Aggregated short-term load forecasting for heterogeneous buildings using machine learning with peak estimation. *Energy and Buildings* 2021;237:110742.
- [84] Chakraborty D, Alam A, Chaudhuri S, Başığaoğlu H, Sulbaran T, Langar S. Scenario-based prediction of climate change impacts on building cooling energy consumption with explainable artificial intelligence. *Applied Energy* 2021;291:116807.
- [85] Golizadeh Akhlaghi Y, Aslansefat K, Zhao X, Sadati S, Badiei A, Xiao X, et al. Hourly performance forecast of a dew point cooler using explainable Artificial Intelligence and evolutionary optimisations by 2050. *Applied Energy* 2021;281:116062.
- [86] Lu Y, Murzakhanov I, Chatzivasileiadis S. Neural network interpretability for forecasting of aggregated renewable generation. 2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 2021, p. 282–8.
- [87] Li M, Wang Y. Power load forecasting and interpretable models based on GS_XGBoost and SHAP. *J Phys: Conf Ser* 2022;2195:012028.
- [88] Kim J-Y, Cho S-B. Electric Energy Consumption Prediction by Deep Learning with State Explainable Autoencoder. *Energies* 2019;12:739.
- [89] Kim J-Y, Cho S-B. Explainable prediction of electric energy demand using a deep autoencoder with interpretable latent space. *Expert Systems with Applications* 2021;186:115842.
- [90] Kim J-Y, Cho S-B. Interpretable Deep Learning with Hybrid Autoencoders to Predict Electric Energy Consumption. In: Herrero Á, Cambra C, Urda D, Sedano J, Quintián H, Corchado E, editors. *Advances in Intelligent Systems and Computing*, Cham: Springer International Publishing; 2021, p. 133–43.
- [91] Singaravel S, Suykens J, Janssen H, Geyer P. Explainable deep convolutional learning for intuitive model development by non-machine learning domain experts. *Design Science* 2020;6.
- [92] Kim T-Y, Cho S-B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* 2019;182:72–81.
- [93] Li G, Yao Q, Fan C, Zhou C, Wu G, Zhou Z, et al. An explainable one-dimensional convolutional neural networks based fault diagnosis method for building heating, ventilation and air conditioning systems. *Building and Environment* 2021;203:108057.
- [94] Kim J-Y, Cho S-B. Electric Energy Demand Forecasting with Explainable Time-series Modeling. 2020 International Conference on Data Mining Workshops (ICDMW), 2020, p. 711–6.
- [95] Zhang W, Wen Y, Tseng KJ, Jin G. Demystifying Thermal Comfort in Smart Buildings: An Interpretable Machine Learning Approach. *IEEE Internet Things J* 2021;8:8021–31.
- [96] Zhang W, Liu F, Wen Y, Nee B. Toward explainable and interpretable building energy modelling. *Proceedings of the 8th ACM International Conference on Systems for*

- Energy-Efficient Buildings, Cities, and Transportation, New York, NY, USA: Association for Computing Machinery; 2021, p. 255–8.
- [97] Mouakher A, Inoubli W, Ounoughi C, Ko A. Expect: EXplainable Prediction Model for Energy Consumption. *Mathematics* 2022;10:248.
- [98] Zhao Y, Li T, Zhang X, Zhang C. Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. *Renewable and Sustainable Energy Reviews* 2019;109:85–101.
- [99] Wang H, Feng D, Liu K. Fault detection and diagnosis for multiple faults of VAV terminals using self-adaptive model and layered random forest. *Building and Environment* 2021;193:107667.
- [100] Wang H, Chen Y, Chan CWH, Qin J. An online fault diagnosis tool of VAV terminals for building management and control systems. *Automation in Construction* 2012;22:203–11.
- [101] Capozzoli A, Piscitelli MS, Brandi S, Grassi D, Chicco G. Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings. *Energy* 2018;157:336–52.
- [102] Deshmukh S, Samouhos S, Glicksman L, Norford L. Fault detection in commercial building VAV AHU: A case study of an academic building. *Energy and Buildings* 2019;201:163–73.
- [103] Chakraborty D, Elzarka H. Early detection of faults in HVAC systems using an XGBoost model with a dynamic threshold. *Energy and Buildings* 2019;185:326–44.
- [104] Yan K, Huang J, Shen W, Ji Z. Unsupervised learning for fault detection and diagnosis of air handling units. *Energy and Buildings* 2020;210:109689.
- [105] Li G, Zheng Y, Liu J, Zhou Z, Xu C, Fang X, et al. An improved stacking ensemble learning-based sensor fault detection method for building energy systems using fault-discrimination information. *Journal of Building Engineering* 2021;43:102812.
- [106] Turner WJN, Staino A, Basu B. Residential HVAC fault detection using a system identification approach. *Energy and Buildings* 2017;151:1–17.
- [107] Schein J, Bushby S. A Hierarchical Rule-Based Fault Detection and Diagnostic Method for HVAC Systems. *HVAC&R Res* 2006;12:111–25.
- [108] Chen Z, Xu P, Feng F, Qiao Y, Luo W. Data mining algorithm and framework for identifying HVAC control strategies in large commercial buildings. *Build Simul* 2021;14:63–74.
- [109] Chen Z, Xiao F, Guo F, Yan J. Interpretable machine learning for building energy management: A state-of-the-art review. *Advances in Applied Energy* 2023;9:100123.
- [110] Han H, Gu B, Hong Y, Kang J. Automated FDD of multiple-simultaneous faults (MSF) and the application to building chillers. *Energy and Buildings* 2011;43:2524–32.
- [111] Capozzoli A, Lauro F, Khan I. Fault detection analysis using data mining techniques for a cluster of smart office buildings. *Expert Systems with Applications* 2015;42:4324–38.
- [112] Li D, Zhou Y, Hu G, Spanos CJ. Fault detection and diagnosis for building cooling system with a tree-structured learning method. *Energy and Buildings* 2016;127:540–51.

- [113] Yan R, Ma Z, Kokogiannakis G, Zhao Y. A sensor fault detection strategy for air handling units using cluster analysis. *Automation in Construction* 2016;70:77–88.
- [114] Tassou SA, Grace IN. Fault diagnosis and refrigerant leak detection in vapour compression refrigeration systems. *International Journal of Refrigeration* 2005;28:680–8.
- [115] Zhou Q, Wang S, Xiao F. A Novel Strategy for the Fault Detection and Diagnosis of Centrifugal Chiller Systems. *HVAC&R Res* 2009;15:57–75.
- [116] Shahnazari H, Mhaskar P, House JM, Salsbury TI. Modeling and fault diagnosis design for HVAC systems using recurrent neural networks. *Computers & Chemical Engineering* 2019;126:189–203.
- [117] Verleysen M, François D. The Curse of Dimensionality in Data Mining and Time Series Prediction. In: Cabestany J, Prieto A, Sandoval F, editors. *Computational Intelligence and Bioinspired Systems*, Berlin, Heidelberg: Springer; 2005, p. 758–70.
- [118] van Engelen JE, Hoos HH. A survey on semi-supervised learning. *Mach Learn* 2020;109:373–440.
- [119] Yan K, Chong A, Mo Y. Generative adversarial network for fault detection diagnosis of chillers. *Building and Environment* 2020;172:106698.
- [120] Li B, Cheng F, Cai H, Zhang X, Cai W. A semi-supervised approach to fault detection and diagnosis for building HVAC systems based on the modified generative adversarial network. *Energy and Buildings* 2021;246:111044.
- [121] Li B, Cheng F, Zhang X, Cui C, Cai W. A novel semi-supervised data-driven method for chiller fault diagnosis with unlabeled data. *Applied Energy* 2021;285:116459.
- [122] Fan C, Liu Y, Liu X, Sun Y, Wang J. A study on semi-supervised learning in enhancing performance of AHU unseen fault detection with limited labeled data. *Sustainable Cities and Society* 2021;70:102874.
- [123] Fan C, Liu X, Xue P, Wang J. Statistical characterization of semi-supervised neural networks for fault detection and diagnosis of air handling units. *Energy and Buildings* 2021;234:110733.
- [124] Fang Q, Wu D. ANS-net: anti-noise Siamese network for bearing fault diagnosis with a few data. *Nonlinear Dyn* 2021;104:2497–514.
- [125] Yang Y, Wang H, Liu Z, Yang Z. Few-shot Learning for Rolling Bearing Fault Diagnosis Via Siamese Two-dimensional Convolutional Neural Network. 2020 11th International Conference on Prognostics and System Health Management (PHM-2020 Jinan), 2020, p. 373–8.
- [126] Zhang J, Wang Y, Zhu K, Zhang Y, Li Y. Diagnosis of Interturn Short-Circuit Faults in Permanent Magnet Synchronous Motors Based on Few-Shot Learning Under a Federated Learning Framework. *IEEE Trans Ind Inf* 2021;17:8495–504.
- [127] de Oliveira LM, Mendes da Silva FE, da Cunha Pereira GM, Oliveira DS, Marcelo Antunes FL. Siamese Neural Network Architecture for Fault Detection in a Voltage Source Inverter. 2021 Brazilian Power Electronics Conference (COBEP), 2021, p. 1–4.
- [128] Utkin LV, Zaborovsky VS, Popov SG. Siamese neural network for intelligent information security control in multi-robot systems. *Aut Control Comp Sci* 2017;51:881–7.

- [129] Oh Y, Kim Y, Na K, Youn BD. A deep transferable motion-adaptive fault detection method for industrial robots using a residual–convolutional neural network. *ISA Transactions* 2022;128:521–34.
- [130] Chen Z, Chen Y, Xiao T, Wang H, Hou P. A novel short-term load forecasting framework based on time-series clustering and early classification algorithm. *Energy and Buildings* 2021;251:111375.
- [131] Chen Z, Chen Y, He R, Liu J, Gao M, Zhang L. Multi-objective residential load scheduling approach for demand response in smart grid. *Sustainable Cities and Society* 2022;76:103530.
- [132] Kar P, Shareef A, Kumar A, Harn KT, Kalluri B, Panda SK. ReViCEE: A recommendation based approach for personalized control, visual comfort & energy efficiency in buildings. *Building and Environment* 2019;152:135–44.
- [133] Lin G, Claridge D. Two Similarity Measure Approaches to Whole Building Fault Diagnosis 2012.
- [134] Sha H, Xu P, Hu C, Li Z, Chen Y, Chen Z. A simplified HVAC energy prediction method based on degree-day. *Sustainable Cities and Society* 2019;51:101698.
- [135] Miller C. What’s in the box?! Towards explainable machine learning applied to non-residential building smart meter classification. *Energy and Buildings* 2019;199:523–36.
- [136] Chicco D. Siamese Neural Networks: An Overview. In: Cartwright H, editor. *Artificial Neural Networks*, New York, NY: Springer US; 2021, p. 73–94.
- [137] Tan SY, Jacoby M, Saha H, Florita A, Henze G, Sarkar S. Multimodal sensor fusion framework for residential building occupancy detection. *Energy and Buildings* 2022;258:111828.
- [138] Granderson J, Lin G, Harding A, Im P, Chen Y. Building fault detection data to aid diagnostic algorithm creation and performance testing. *Sci Data* 2020;7:65.
- [139] Splitting data randomly can ruin your model | Data Science 2021.
- [140] Ma R, Boubrahimi SF, Hamdi SM, Angryk RA. Solar flare prediction using multivariate time series decision trees. 2017 IEEE International Conference on Big Data (Big Data), 2017, p. 2569–78.
- [141] Salari E, Askarzadeh A. A new solution for loading optimization of multi-chiller systems by general algebraic modeling system. *Applied Thermal Engineering* 2015;84:429–36.
- [142] Daixin T, Hongwei X, Huijuan Y, Hao Y, Wen H. Optimization of group control strategy and analysis of energy saving in refrigeration plant. *Energy and Built Environment* 2022;3:525–35.
- [143] Liao Y, Sun Y, Huang G. Robustness analysis of chiller sequencing control. *Energy Conversion and Management* 2015;103:180–90.
- [144] Huang S, Zuo W, Sohn MD. Amelioration of the cooling load based chiller sequencing control. *Applied Energy* 2016;168:204–15.
- [145] Sun Y, Wang S, Xiao F. In situ performance comparison and evaluation of three chiller sequencing control strategies in a super high-rise building. *Energy and Buildings* 2013;61:333–43.

- [146] Wang S. *Intelligent buildings and building automation*. London ; New York: Spon Press; 2010.
- [147] Niu J, Zhou R, Tian Z, Zhu J, Lu Y, Ma J. Energy-saving potential analysis for a 24-hour operating chiller plant using the model-based global optimization method. *Journal of Building Engineering* 2023;69:106213.
- [148] Yan B, Li X, Malkawi AM, Augenbroe G. Quantifying uncertainty in outdoor air flow control and its impacts on building performance simulation and fault detection. *Energy and Buildings* 2017;134:115–28.
- [149] Huang G, Sun Y, Li P. Fusion of redundant measurements for enhancing the reliability of total cooling load based chiller sequencing control. *Automation in Construction* 2011;20:789–98.
- [150] Sun S, Shan K, Wang S. An online robust sequencing control strategy for identical chillers using a probabilistic approach concerning flow measurement uncertainties. *Applied Energy* 2022;317:119198.
- [151] Liu Z, Tan H, Luo D, Yu G, Li J, Li Z. Optimal chiller sequencing control in an office building considering the variation of chiller maximum cooling capacity. *Energy and Buildings* 2017;140:430–42.
- [152] Wong JKW, Li H. Construction, application and validation of selection evaluation model (SEM) for intelligent HVAC control system. *Automation in Construction* 2010;19:261–9.
- [153] Chang Y-C. An Outstanding Method for Saving Energy—Optimal Chiller Operation. *IEEE Trans On Energy Conversion* 2006;21:527–32.
- [154] Chen Y, Guo M, Chen Z, Chen Z, Ji Y. Physical energy and data-driven models in building energy prediction: A review. *Energy Reports* 2022;8:2656–71.
- [155] Jiang W, Reddy TA. General Methodology Combining Engineering Optimization of Primary HVAC&R Plants with Decision Analysis Methods—Part I: Deterministic Analysis. *HVAC&R Res* 2007;13:93–117.
- [156] Chen Y, Chen Z. Short-term load forecasting for multiple buildings: A length sensitivity-based approach. *Energy Reports* 2022;8:14274–88.
- [157] Einbinder B-S, Romano Y, Sesia M, Zhou Y. Training Uncertainty-Aware Classifiers with Conformalized Deep Learning 2022.
- [158] Sun S, Wang S, Shan K. Flow measurement uncertainty quantification for building central cooling systems with multiple water-cooled chillers using a Bayesian approach. *Applied Thermal Engineering* 2022;202:117857.
- [159] Li G, Xiong J, Sun S, Chen J. Validation of virtual sensor-assisted Bayesian inference-based in-situ sensor calibration strategy for building HVAC systems. *Build Simul* 2023;16:185–203.
- [160] Li B, Wang S. Multi-objective optimal control of multi-zone VAV systems for adaptive switching between normal and pandemic modes. *Building and Environment* 2023;243:110626.
- [161] Amonkar Y, Doss-Gollin J, Farnham DJ, Modi V, Lall U. Differential effects of climate change on average and peak demand for heating and cooling across the contiguous USA. *Commun Earth Environ* 2023;4:1–9.

- [162] Huo X, Yang L, Li DHW, Lun I, Lou S, Shi Y. Impact of climate change on outdoor design conditions and implications to peak loads. *Build Simul* 2022;15:2051–65.
- [163] Sha H, Xu P, Yan C, Ji Y, Zhou K, Chen F. Development of a key-variable-based parallel HVAC energy predictive model. *Build Simul* 2022;15:1193–208.
- [164] Chen Z, Zhang J, Xiao F, Madsen H, Xu K. Probabilistic Machine Learning for Enhanced Chiller Sequencing: A Risk-Based Control Strategy. *Energy and Built Environment* 2024.
- [165] Zhang M, Li X, Wang L. An Adaptive Outlier Detection and Processing Approach Towards Time Series Sensor Data. *IEEE Access* 2019;7:175192–212.
- [166] Teh HY, Kempa-Liehr AW, Wang KI-K. Sensor data quality: a systematic review. *J Big Data* 2020;7:11.
- [167] Lam JC. Energy analysis of commercial buildings in subtropical climates. *Building and Environment* 2000;35:19–26.
- [168] Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO, et al. EnergyPlus: creating a new-generation building energy simulation program. *Energy and Buildings* 2001;33:319–31.
- [169] Fiksel A, Thornton JW, Klein SA, Beckman WA. Developments to the TRNSYS Simulation Program. *Journal of Solar Energy Engineering* 1995;117:123–7.
- [170] Sun Y, Wang S, Huang G. Chiller sequencing control with enhanced robustness for energy efficient operation. *Energy and Buildings* 2009;41:1246–55.
- [171] Agamloh EB. Power and Efficiency Measurement of Motor-Variable-Frequency Drive Systems. *IEEE Trans on Ind Applicat* 2017;53:766–73.
- [172] Hughes A, Drury B. *Electric motors and drives: fundamentals, types and applications*. Newnes; 2019.
- [173] Jeter SM, Wepfer WJ, Fadel GM, Cowden NE, Dymek AA. Variable speed drive heat pump performance. *Energy* 1987;12:1289–98.
- [174] Zhang X, Li Z, Li Z, Qiu S, Wang H. Differential pressure reset strategy based on reinforcement learning for chilled water systems. *Build Simul* 2022;15:233–48.
- [175] Cai J, Yang H, Lai T, Xu K. Parallel pump and chiller system optimization method for minimizing energy consumption based on a novel multi-objective gorilla troops optimizer. *Journal of Building Engineering* 2023;76:107366.
- [176] Xiong Q, Li Z, Cai W, Wang Z. Model free optimization of building cooling water systems with refined action space. *Build Simul* 2023;16:615–27.
- [177] Rishel JB. Control of Variable Speed Pumps for HVAC Water Systems. *ASHRAE Transactions* 2003;109:380–9.
- [178] Rishel JB. Wire-to-water efficiency of pumping systems. *Ashrae Journal* 2001;43:40+42+44-46.
- [179] Rishel JB. *Control of Variable-Speed Pumps on Hot- and Chilled-Water Systems* 1991.
- [180] Wang G, Zheng X, Kiamehr K. Sequencing Control of Parallel Pumps in Variable-Flow Systems Using Wire-to-Water Efficiency. *ASHRAE Transactions* 2017;123:291–301.

- [181] Wang G. Efficiency degradation detection for VFD-motor-pump systems. *Science and Technology for the Built Environment* 2018;24:974–81.
- [182] Barnier MA, Bourret B. Pumping energy and variable frequency drives. *ASHRAE Journal* 1999;41:37.
- [183] Wang G. Data-driven energy models for existing VFD-motor-pump systems. *Science and Technology for the Built Environment* 2019;25:732–42.
- [184] Jepsen KL, Hansen L, Mai C, Yang Z. Power consumption optimization for multiple parallel centrifugal pumps. 2017 IEEE Conference on Control Technology and Applications (CCTA), 2017, p. 806–11.
- [185] Rajwar K, Deep K, Das S. An exhaustive review of the metaheuristic algorithms for search and optimization: taxonomy, applications, and open challenges. *Artif Intell Rev* 2023;56:13187–257.
- [186] Dullinger C, Struckl W, Kozek M. A general approach for mixed-integer predictive control of HVAC systems using MILP. *Applied Thermal Engineering* 2018;128:1646–59.
- [187] Barber KA, Krarti M. A review of optimization based tools for design and control of building energy systems. *Renewable and Sustainable Energy Reviews* 2022;160:112359.
- [188] Zang X, Li H, Wang S. Optimal design of energy-flexible distributed energy systems and the impacts of energy storage specifications under evolving time-of-use tariff in cooling-dominated regions. *Journal of Energy Storage* 2023;72:108462.
- [189] Tang H, Wang S. Life-cycle economic analysis of thermal energy storage, new and second-life batteries in buildings for providing multiple flexibility services in electricity markets. *Energy* 2023;264:126270.
- [190] Tang H, Wang S. Game-theoretic optimization of demand-side flexibility engagement considering the perspectives of different stakeholders and multiple flexibility services. *Applied Energy* 2023;332:120550.
- [191] Ma Z, Wang S. Supervisory and optimal control of central chiller plants using simplified adaptive models and genetic algorithm. *Applied Energy* 2011;88:198–211.
- [192] Wetter M, Wright J. A comparison of deterministic and probabilistic optimization algorithms for nonsmooth simulation-based optimization. *Building and Environment* 2004;39:989–99.
- [193] Mayer B, Killian M, Kozek M. A branch and bound approach for building cooling supply control with hybrid model predictive control. *Energy and Buildings* 2016;128:553–66.
- [194] Halim AH, Ismail I, Das S. Performance assessment of the metaheuristic optimization algorithms: an exhaustive review. *Artif Intell Rev* 2021;54:2323–409.
- [195] Afroz Z, Shafiullah GM, Urmee T, Shoeb MA, Higgins G. Predictive modelling and optimization of HVAC systems using neural network and particle swarm optimization algorithm. *Building and Environment* 2022;209:108681.
- [196] Beloglazov A, Buyya R. Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in Cloud data centers. *Concurrency and Computation* 2012;24:1397–420.

- [197] Fan B, Jin X, Du Z. Optimal control strategies for multi-chiller system based on probability density distribution of cooling load ratio. *Energy and Buildings* 2011;43:2813–21.
- [198] Jia L, Wei S, Liu J. A review of optimization approaches for controlling water-cooled central cooling systems. *Building and Environment* 2021;203:108100.
- [199] Kusiak A, Xu G, Tang F. Optimization of an HVAC system with a strength multi-objective particle-swarm algorithm. *Energy* 2011;36:5935–43.
- [200] Su L, Nie T, On Ho C, Yang Z, Calvez P, Jain RK, et al. Optimizing pipe network design and central plant positioning of district heating and cooling System: A Graph-Based Multi-Objective genetic algorithm approach. *Applied Energy* 2022;325:119844.
- [201] Wang X, Cai W, Yin X. A global optimized operation strategy for energy savings in liquid desiccant air conditioning using self-adaptive differential evolutionary algorithm. *Applied Energy* 2017;187:410–23.
- [202] Junghans L, Darde N. Hybrid single objective genetic algorithm coupled with the simulated annealing optimization method for building optimization. *Energy and Buildings* 2015;86:651–62.
- [203] Peaks function - MATLAB peaks n.d.
- [204] Sun Y, Huang G, Li Z, Wang S. Multiplexed optimization for complex air conditioning systems. *Building and Environment* 2013;65:99–108.
- [205] ElBedwehy MN, Behery GM, Elbarougy R. Face Recognition Based on Relative Gradient Magnitude Strength. *Arab J Sci Eng* 2020;45:9925–37.
- [206] Beuve N, Hamidouche W, Deforges O. DmyT. Proceedings of the 1st Workshop on Synthetic Multimedia - Audiovisual Deepfake Generation and Detection, New York, NY, USA: Association for Computing Machinery; 2021, p. 17–24.
- [207] Kang Z, Peng C, Cheng Q. Kernel-driven similarity learning. *Neurocomputing* 2017;267:210–9.
- [208] Chakladar DD, Kumar P, Roy PP, Dogra DP, Scheme E, Chang V. A multimodal-Siamese Neural Network (mSNN) for person verification using signatures and EEG. *Information Fusion* 2021;71:17–27.
- [209] Li Y, Chen CLP, Zhang T. A Survey on Siamese Network: Methodologies, Applications, and Opportunities. *IEEE Trans Artif Intell* 2022;3:994–1014.
- [210] Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Computation* 1997;9:1735–80.
- [211] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32nd International Conference on Machine Learning, PMLR; 2015, p. 448–56.
- [212] Lian Z, Li Y, Tao J, Huang J. Speech Emotion Recognition via Contrastive Loss under Siamese Networks. Proceedings of the Joint Workshop of the 4th Workshop on Affective Social Multimedia Computing and first Multi-Modal Affective Computing of Large-Scale Multimedia Data, New York, NY, USA: Association for Computing Machinery; 2018, p. 21–6.

- [213] Wen J, Li S. Tools for evaluating fault detection and diagnostic methods for air-handling units, ASHRAE RP-1312 Final Report. American Society of Heating, Refrigerating and Air Conditioning Engineers Inc: Atlanta, GA, USA 2011.
- [214] Yongjun Sun, Shengwei Wang, Gongsheng Huang. Model-based optimal start control strategy for multi-chiller plants in commercial buildings. *Building Services Engineering Research and Technology* 2010;31:113–29.
- [215] Gao D, Sun Y, Lu Y. A robust demand response control of commercial buildings for smart grid under load prediction uncertainty. *Energy* 2015;93:275–83.
- [216] Zhuang C, Wang S, Shan K. A risk-based robust optimal chiller sequencing control strategy for energy-efficient operation considering measurement uncertainties. *Applied Energy* 2020;280:115983.
- [217] Zhao Y, Zhang C, Zhang Y, Wang Z, Li J. A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis. *Energy and Built Environment* 2020;1:149–64.
- [218] Duan T, Anand A, Ding DY, Thai KK, Basu S, Ng A, et al. NGBoost: Natural Gradient Boosting for Probabilistic Prediction. *Proceedings of the 37th International Conference on Machine Learning, PMLR*; 2020, p. 2690–700.
- [219] Mitrentsis G, Lens H. An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting. *Applied Energy* 2022;309:118473.
- [220] Mei Y, Sun Y, Li F, Xu X, Zhang A, Shen J. Probabilistic prediction model of steel to concrete bond failure under high temperature by machine learning. *Engineering Failure Analysis* 2022;142:106786.
- [221] Natural Gradient Works Efficiently in Learning n.d.
- [222] Obuka NS, Utazi DN, Onyechi PC, Okoli NC, Eneh CO. Determination Of A Nonresidential Space Cooling Load: Vb Program Application. *International Journal of Scientific and Technology Research* 2015;4.
- [223] Zhang J, Ma T, Xu K, Chen Z, Xiao F, Ho J, et al. Smart Data-Driven Building Management Framework and Demonstration. In: Jørgensen BN, Da Silva LCP, Ma Z, editors. *Energy Informatics*, vol. 14467, Cham: Springer Nature Switzerland; 2024, p. 168–78.
- [224] Ding Y, Zhang Q, Yuan T. Research on short-term and ultra-short-term cooling load prediction models for office buildings. *Energy and Buildings* 2017;154:254–67.
- [225] Lindberg O, Lingfors D, Arnqvist J, van der Meer D, Munkhammar J. Day-ahead probabilistic forecasting at a co-located wind and solar power park in Sweden: Trading and forecast verification. *Advances in Applied Energy* 2023;9:100120.
- [226] Liu H. Interval Prediction of Electric Load Time Series. In: Liu H, editor. *Non-intrusive Load Monitoring: Theory, Technologies and Applications*, Singapore: Springer; 2020, p. 247–77.
- [227] Polak E. *Optimization: Algorithms and consistent approximations*. *Applied Mathematical Sciences* 1997;124.
- [228] Zhou P, Wang S, Zhou J, Hussain SA, Liu X, Gao J, et al. A modelling method for large-scale open spaces orientated toward coordinated control of multiple air-terminal units. *Build Simul* 2023;16:225–41.

- [229] Tzeng G-H, Huang J-J. Multiple attribute decision making: methods and applications. CRC press; 2011.
- [230] Afshari A, Mojahed M, Yusuff RM. Simple Additive Weighting approach to Personnel Selection problem. *International Journal of Innovation* 2010;1.
- [231] Chang Y-C. Genetic algorithm based optimal chiller loading for energy conservation. *Applied Thermal Engineering* 2005;25:2800–15.
- [232] Ardakani AJ, Ardakani FF, Hosseinian SH. A novel approach for optimal chiller loading using particle swarm optimization. *Energy and Buildings* 2008;40:2177–87.
- [233] Chang Y-C, Lin J-K, Chuang M-H. Optimal chiller loading by genetic algorithm for reducing energy consumption. *Energy and Buildings* 2005;37:147–55.
- [234] Lee W-S, Chen Y-T, Kao Y. Optimal chiller loading by differential evolution algorithm for reducing energy consumption. *Energy and Buildings* 2011;43:599–604.
- [235] Chang Y-C, Chen W-H, Lee C-Y, Huang C-N. Simulated annealing based optimal chiller loading for saving energy. *Energy Conversion and Management* 2006;47:2044–58.
- [236] The Leader in Decision Intelligence Technology. Gurobi Optimization n.d.
- [237] Zou W, Sun Y, Gao D, Cui Z, You Z, Ma X. Robust enhancement of chiller sequencing control for tolerating sensor measurement uncertainties through controlling small-scale thermal energy storage. *Energy* 2023;280:128152.
- [238] Perrone N, Kao R. A general finite difference method for arbitrary meshes. *Computers & Structures* 1975;5:45–57.
- [239] Cai J, Yang H, Lai T, Xu K. A new approach for optimal chiller loading using an improved imperialist competitive algorithm. *Energy and Buildings* 2023;284:112835.