HOLISTIC OPTIMIZATION OF DATA CENTER COOLING SYSTEMS AND AIRFLOW MANAGEMENT

by

XU HAN

B.E., Harbin Institute of Technology, 2011

M.E., Tianjin University, 2014

A dissertation submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Civil, Environmental and Architectural Engineering

University of Colorado Boulder

2020

Committee Members:

Wangda Zuo

John Z. Zhai

Moncef Krarti

Debanjan Mukherjee

Wei Tian

ABSTRACT

Han, Xu (Ph.D., Architectural Engineering)

Holistic Optimization of Data Center Cooling Systems and Airflow Management

Dissertation directed by Professor Wangda Zuo

Data centers are often overcooled to address the reliability concerns, which may lead to a lower cooling efficiency. Model-based optimization in a holistic view can be utilized to improve the cooling efficiency while meeting the stringent thermal requirements of data centers, which usually involve non-uniform thermal environment. The co-simulation of building energy simulation (BES) and computational fluid dynamics (CFD) is promising to fulfill the needs. However, the existing BES-CFD co-simulation methods are not good enough for practical use in the real world because of either huge computational costs with BES-CFD coupled models or accuracy issues with BES-ROM-CFD coupled models that integrate reduced order models (ROMs). This dissertation therefore intends to improve the simulation methods from the following aspects to allow a computationally-practical and sufficiently accurate holistic optimization of cooling systems and airflow management towards energy efficient data centers with reliable operation.

First, as CFD is the bottleneck of the BES-CFD co-simulation speed, this dissertation develops an improved Fast Fluid Dynamics (FFD) model for fast and accurate simulation of data center airflow and thermal environment. FFD, which is a simplified CFD method, has been applied for fast airflow simulation. However, few research applied FFD for optimal design and operation of data center thermal management. This dissertation improves the FFD model for data centers and conducts a comprehensive evaluation and demonstration. The FFD model is improved by 1)

proposing new algorithms to solve equations for a better general quantity conservation than conventional FFD; and 2) adding new features and boundary conditions for data centers. The validation with a real data center shows that the new FFD model achieves a similar level of accuracy as CFD when compared to the experimental measurements, and is 61 times faster than CFD for the studied case.

Then, to improve the existing BES-CFD co-simulation methods, this dissertation proposes a new online BES-ROM-CFD co-simulation method. Among existing models, BES-CFD is accurate but time consuming, and BES-ROM-CFD is fast but may have accuracy issues. The proposed online BES-ROM-CFD method combines the advantages of both existing models, in which the ROM allows online learning and automatic error control. This methodology is realized by implementing a Modelica-ISAT-FFD model, which will be officially released in Modelica *Buildings* library to allow a broader range of applications. The new model is then comprehensively evaluated with a mixed convection case. The results show that the new model can generally control the prediction error within user-defined tolerances compared to an existing Modelica-FFD model that was validated by previous research. An annual simulation shows that the new model saves up to 95.7% of computing time against the existing Modelica-FFD model. A space heating case is also studied to demonstrate the capability of the new model to handle scenarios with feedback loop controls.

To further improve the online BES-ROM-CFD method, this dissertation proposes an adaptive online BES-ROM-CFD method, in which adaptive coupling frequencies are used to reduce the number of ROM-CFD calls during the co-simulation. It is further powered by distributed computing, which allows that BES-ROM runs on a CPU and CFD runs on a GPU in parallel during the co-simulation. Based on the new methods, an adaptive Modelica-ISAT-FFD model is developed,

which is verified against an existing Modelica-FFD model with a real middle-size data center. A holistic optimization platform is proposed based on the new model, which is then evaluated and demonstrated with optimization studies. It is found that a holistic optimization with annual co-simulations for a middle size data center in the real world can be finished within a day, which is estimated to take as much as many years if using the existing Modelica-FFD model. The results also show that the holistic optimization saves the annual energy consumption by up to 48.1% while meeting the data center thermal requirements.

In addition to the off-operation model-based optimization, this dissertation further proposes a novel machine learning assisted expert system (MLES) method towards a real-time optimal control for the studied data center cooling system. Model predictive control (MPC) has been widely studied for optimal control. Though it can achieve a good performance theoretically, formulating a precise and real-time model is often not easy especially for complex systems, such as data center cooling systems involving stratified airflow. To address this problem, this dissertation proposes a novel robust and easy-to-implement model-free MLES optimal control method. First, the optimization problem is simplified to allow formulating an expert system based on expert knowledge. Then, a machine learning model trained by more than a thousand of CFD simulations is used to assist the expert system for hot spots control towards reliable operation. A MPC is also implemented as benchmark. The case studies show that the MLES could achieve a similar performance as a well-designed ideal MPC, and is much faster than MPC. An annual simulation shows that with the MLES, the energy consumption is saved by up to 64.6% and meanwhile hot spots are generally well controlled.

To conclude, this dissertation concentrates on improving the data center cooling efficiency while maintaining reliable operation in a holistic view. The outcomes include 1) an open source

model for fast simulation of data center thermal environment, 2) an open source model for fast and accurate co-simulation of cooling systems and airflow management, 3) a computationally-practical and sufficiently accurate optimization platform for holistic optimization of cooling systems and airflow management, 4) a novel robust and easy-to-implement optimal control method that simultaneously considers cooling efficiency and reliable operation. The proposed models and methodologies may not be limited in data center applications, which can be expanded for studying other cooling systems involving non-uniform thermal environment in future research.

DEDICATION

To my wife, Xiangmei Li, and my daughter, Naomi Han

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Prof. Wangda Zuo for his strong support and help to me. Except for the patient guidance, his working attitude also inspired me and pushed me to become a better person. I would like to thank the rest of my thesis committee: Prof. John Zhai, Prof. Moncef Krarti, Prof. Debanjan Mukherjee and Dr. Wei Tian, for their insightful comments, which bring my research to the current depth.

I would like to thank all the collaborators for providing support, data, help and advice for my work. They include Dr. Michael Wetter, Dr. Jianjun Hu and Dr. Michael Sohn from Lawrence Berkeley National Laboratory; Mr. Jim VanGilder, Dr. Wei Tian, Dr. Christopher M. Healey and Mr. Michael Condor from Schneider Electric; Prof. Wangda Zuo, Dr. Yangyang Fu and Mr. Cary Faulkner from University of Colorado Boulder; Mr. David Plamondon from University of Massachusetts Medical School; and Mr. Juan Quintero, Mr. F. James Doppelhammer from University of Miami. I also would like to thank Dr. Hongtao Qiao, Dr. Saleh Nabi and Dr. Christopher R Laughman for their help during my internship at Mitsubishi Electric Research Laboratories.

This dissertation is partially based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Award Number DE-0007688. This work emerged from the IBPSA Project 1, an international project conducted under the umbrella of the International Building Performance Simulation Association (IBPSA). Project 1 will develop and demonstrate a BIM/GIS and Modelica Framework for building and community energy system design and operation. The author of this dissertation was also sponsored by the National Science Foundation under Award No. IIS-1633338 and No. IIS1802017.

This dissertation was also supported by the Department of Civil, Environmental & Architectural Engineering at the University of Colorado Boulder through a Doctoral Assistantship for Completion of Dissertation. The author also would like to acknowledge the support from the Strauch Family Graduate Fellowship and ASHRAE College of Fellows Travel Award.

I would like to thank all of the former and current students in the Sustainable Buildings and Societies Laboratory. I usually spoke with them when I had questions. Among them are Dr. Sen Huang, Dr. Wei Tian, Thomas A. Sevilla, Dr. Yangyang Fu, Dr. Yunyang Ye, Jing Wang, Katy Hinkelman, Xing Lu, Danlin Hou, Yingli Lou, Cary Faulkner, Jes Stershic, Shifang Huang, Guang Zhou, Qiujian Wang, Peilin Yang, Angelique Fathy, Saranya Anbarasu.

Last but not the least, I would like to thank all of my family members for the strong support.

DISCLAIMER

This dissertation was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Background

Data centers house a large amount of mission-critical IT equipment, such as IT servers, network and communication servers, and data storage devices, which effectively dissipates its input power as heat during operation. The data centers usually require cooling system to remove the heat dissipation from IT equipment to ensure the IT equipment operate reliably. It was reported that data centers consumed approximately 1.8% of the total U.S. electricity [1], among which 24%-60% are consumed by the cooling system [2]. With the fast growing markets on cloud computing, data centers' share on global electricity was estimated to be as high as 3-13% in 2030 [3]. As the IT equipment becomes increasingly more power-intensive, it imposes more challenges on the data center cooling system [4]. As of today, air cooling is still the dominant method of primary heat removal, though liquid cooling is emerging as an energy-saving alternative.

Many data centers are significantly overcooled to ensure the reliable operation of IT equipment, which leads to a low cooling efficiency [5]. To improve the effectiveness and efficiency of data center cooling systems, simulation-based optimization [6] can be adopted. Different from other buildings, such as offices, hot spots are critical for reliable operation of IT equipment in data centers, which usually involve non-uniform airflow and temperature distribution. This brings a lot of challenges to conventional simulation-based optimization using a standalone building energy simulation (BES) program with a multi-zone room model, such as EnergyPlus [7], eQuest [8], TRNSYS [9] and Modelica [10]. To improve the cooling efficiency while maintaining the reliable operation, a simulation-based optimization that simultaneously considers

cooling systems and airflow management in a holistic view is promising. The co-simulation of BES and computational fluid dynamics (CFD) can be used to fulfill this need. However, CFD's huge computational demand makes the co-simulation very challenging to be applied in real applications, which usually requires hundreds of or even thousands of CFD simulations during co-simulation. Pre-trained reduced order models (ROMs) have been proposed to replace CFD to couple with BES. However, the traditional ROMs face challenges when the inputs are outside of the training domain. In addition, the ROMs commonly need to be carefully trained and tested to achieve a desired prediction accuracy. Therefore, this dissertation intends to proposed new models and methods to allow a computationally-practical and sufficiently accurate holistic optimization of cooling systems and airflow management towards energy efficient data centers with reliable operation.

1.2 Problem Statement

To support realizing energy efficient data centers with reliable operation, there are four research problems that need to be solved:

1) **Fast and accurate prediction model for data center airflow management**. CFD has been widely used to evaluate and improve the data center airflow management. However, CFD is computationally demanding, which may take several hours to simulate the airflow of a middle-size data center. This makes it difficult to support optimal design and operation of data center cooling systems. For example, a large amount of CFD simulations may be needed when multiple parameters are to be optimized in the early design stage.
2) **Fast and accurate co-simulation of cooling systems and airflow management.** CFD's huge computational demand makes the BES-CFD co-simulation very challenging to be

applied in real applications, which usually requires hundreds of or even thousands of CFD simulations during co-simulation. Pre-trained reduced order models (ROMs) have been proposed to replace CFD to couple with BES. However, the traditional ROMs face challenges when the inputs are outside of the training domain. In addition, the ROMs commonly need to be carefully trained and tested to achieve a desired prediction accuracy.

3) **Practical simulation-based optimization for cooling systems and airflow management in a holistic view.** BES-CFD co-simulation can be used to support the holistic optimization. However, to apply such a co-simulation based optimization in real applications, the BES-CFD co-simulation needs to be further improved to make the optimization process computationally-practical and sufficiently accurate.
4) **Optimal control that simultaneously optimizes cooling systems and airflow management in real-time.** Most studies focused on optimal control of cooling systems, and few of them simultaneously considered cooling system and airflow management. The challenges lie in two aspects. One the one hand, formulating an optimal control for the cooling system alone is already challenging because of its nonlinearity. Taking the non-uniform airflow into considerations will make it even harder. On the other hand, a more complex model may be able to capture the dynamics. However, the increasing complexity will make it very challenging to run the model in real-time.

1.3 Objectives

There are five objectives for this dissertation to solve the research problems proposed in Section 1.2. The five objectives are shown as follows:

1) **Develop a fast and accurate prediction model for data center airflow management**. This model is based on an existing Fast Fluid Dynamics (FFD), which is a simplified CFD model and was reported to be 50 times faster than CFD [11]. Compared to the existing FFD model, the improvements of this study include 1) solving the advection and diffusion equations together using an upwind scheme instead of a semi-Lagrangian advection solver in the conventional FFD to improve its performance on local mass and energy conservation; 2) implementing new features and boundary conditions to allow modeling and simulations for data center airflow, such as a pressure correction method to simulate plenum airflow and dynamic boundary conditions for IT racks; 3) demonstrating the capability of FFD on handling realistic cases with complex boundary conditions, which has not been fully understood based on existing research.
2) **Propose a new online BES-ROM-CFD co-simulation method to allow fast and accurate co-simulation of cooling systems and airflow management**. This new online BES-ROM-CFD co-simulation method integrates an online learning ROM into BES-CFD co-simulation. In comparison with the conventional BES-ROM-CFD co-simulation, the improvements of this new method include 1) providing accurate predictions using CFD simulations when the inputs of the ROM are outside of the training domain; 2) automatically estimating the error between the ROM and CFD predictions and determines the training needs; 3) efficient and automatic training of the ROM during the co-simulation without expert knowledge and pre-training.
3) **Propose a novel co-simulation-based optimization platform for practical holistic optimization of data center cooling systems and airflow management**. This further

improves the online BES-ROM-CFD co-simulation to allow its applications in the real world by 1) using adaptive coupling frequencies to reduce the number of ROM-CFD calls during the co-simulation; 2) adopting distributed computing technology to maximize the capacity of computing hardware of a computer, in which BES-ROM runs on CPU and CFD runs on GPU in parallel. A holistic optimization platform is then proposed based on the new model.

4) **Propose a robust and easy-to-implement machine learning assisted expert system (MLES) optimal control method that simultaneously optimizes cooling systems and airflow management in real-time**. To realize such a MLES optimal control, the optimization problem is first simplified to allow formulating an expert system based on expert knowledge. Then, a machine learning model trained by more than a thousand of CFD simulations is used to assist the expert system to determine optimal actions in real-time for hot spots control towards reliable operation.
5) **Publicly release models developed in this study**. To allow a broader range of applications of models developed in this dissertation, the new FFD model for the first objective has been open source released, and the Modelica-ISAT-FFD model for the second objective will be officially released in Modelica Buildings library [12].

1.4 Structure of the Dissertation

This dissertation consists of seven chapters. Chapter 1 is introduction. Chapters 2 presents a literature review. Chapters 3-6 introduce the work in details: Chapter 8 summarizes the conclusions and proposes future research.

1) **Chapter 1: Introduction**. This chapter provides a general introduction of the background, existing problems, research objectives, and scope of the dissertation.
2) **Chapter 2: Literature Review**. This chapter reviews the overview of data center cooling technologies and modeling methods to improve the cooling efficiency and airflow management.
3) **Chapter 3: An Open Source Fast Fluid Dynamics Model for Data Center Thermal Management**. This chapter introduces a new FFD model for fast prediction of data center airflow and thermal environment. It is validated and then demonstrated to support optimal data center thermal management with a real middle-size data center.
4) **Chapter 4: An Online BES-ROM-CFD Co-Simulation Method for Fast Simulation of HVAC Systems with Non-Uniform Thermal Environment**. This chapter presents the new online BES-ROM-CFD co-simulation method. A Modelica-ISAT-FFD is implemented to demonstrate this new method, in which a Modelica model is for HVAC systems and ISAT-FFD model is used for thermal environment.
5) **Chapter 5: Towards Holistic Optimization of Data Center Cooling Systems and Airflow Management: Adaptive Online BES-ROM-CFD Co-Simulation Powered by Distributed Computing**. This chapter further improves the Modelica-ISAT-FFD model from using adaptive coupling frequencies and distributed computing technology. An optimization platform is proposed based on the improved model. A real middle size data center is adopted to demonstrate the capability of the optimization platform.
6) **Chapter 7: Machine Learning Assisted Expert System for Optimal Control of Data Center Cooling Systems with Air Side Economizer**. This chapter presents a MLES

method to achieve a real-time optimal control for data center cooling systems. A standard model predictive control (MPC) approach is also implemented to evaluate the performance of the proposed method.

7) **Chapter 8: Conclusions and Future Research**. This chapter summarizes the conclusions and proposes future research.

Chapter 2 Literature Review

This chapter reviews the overview of data center cooling configurations and modeling methods to improve the cooling efficiency and airflow management.

2.1 Overview of Data Center Cooling Infrastructure and Airflow Pattern

Data center thermal management is crucial for reliable operation of IT equipment and energy efficiency of the cooling system in the data center, which typically involves non-uniform and complex airflow dynamics [13]. For example, medium and large-size data centers are usually configured with alternating cold aisles and hot aisles to ensure the IT rack inlet temperatures are within a safe threshold by eliminating or reducing the mixing of the cooling system supply cold air and IT rack exhaust hot air. However, many data centers are over-cooled, such as more cold air is supplied than needed, just to dilute a few local hot spots (locations at the intake of IT equipment where the measured temperature is greater than the recommended value), which leads to a low energy efficiency of the cooling system. Model-based methods can be adopted to improve the thermal management in data centers by evaluating what-if scenarios or performing a model based optimization. For example, in the design phase, the model can help determine the layout of IT racks, the open-area-ratio (ratio of open area to total area) and location of the perforated tiles, and the depth of the raised-floor plenum. In the operation phase, the model can help determine the optimal setpoints of the supply air flow rate and temperature to achieve the best energy performance while still meeting the thermal requirements of data centers.

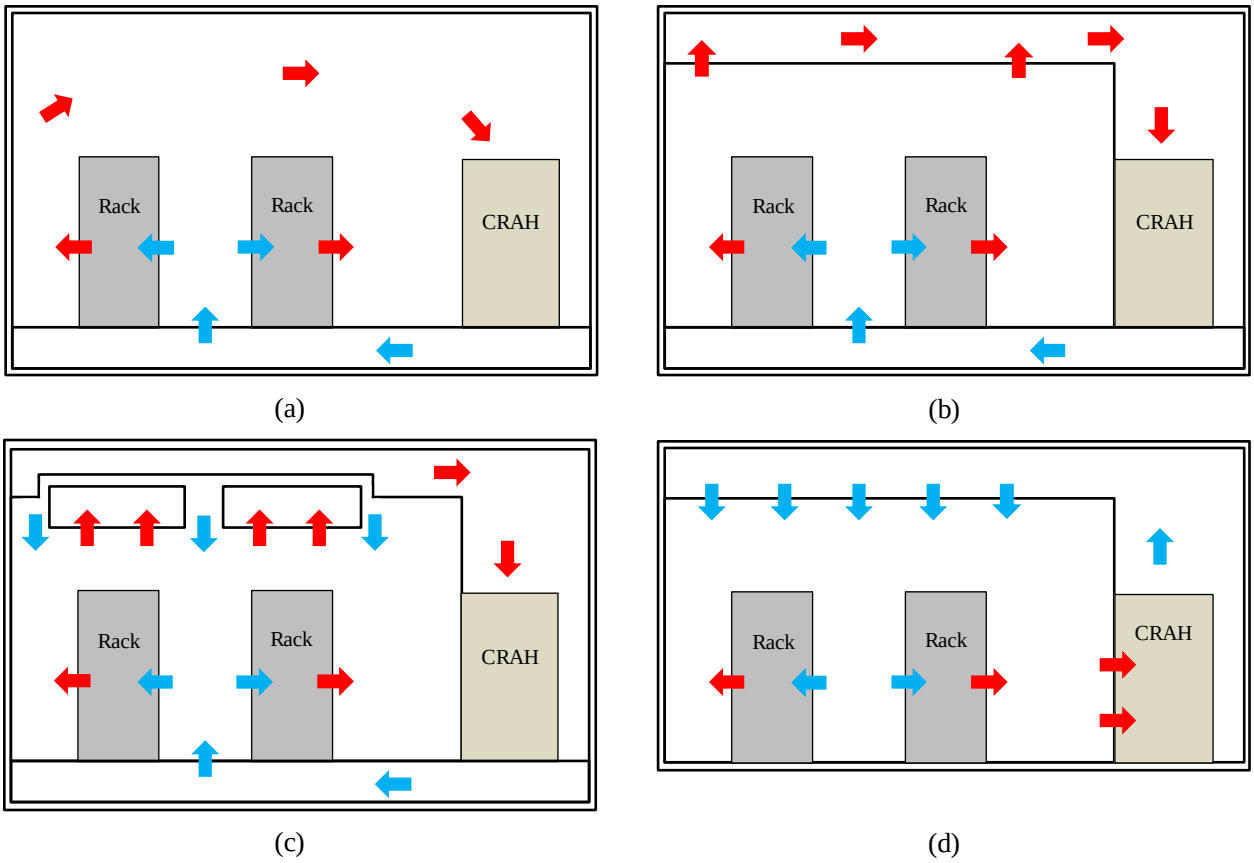


Figure 2-1 Typical airflow configurations (a) raised-floor supply; (b) raised-floor supply/ceiling return; (c) raised-floor supply/ceiling supply; (d) non-raised floor/ceiling supply [14].

The typical configurations and intended airflow directions of data centers are shown in *Figure 2-1*. The above-floor white space (allocated for IT equipment) in the data center is partitioned into cold aisles and hot aisles by rows of racks. For the airflow configurations in *Figure 2-1* (a-c), the raised-floor plenum and perforated floor tiles are designed to uniformly distribute the cold air to the cold aisles. The cold air from the cooling system, i.e. Computer Room Air Handlers (CRAH) or Computer Room Air Conditioners (CRAC), is supplied to the cold aisle, then flows through the IT racks and carries heat dissipated from IT equipment and then exhausts to the

hot aisle and returns to the cooling system through the room in *Figure 2-1* (a) or the ceiling plenum in *Figure 2-1* (b-c). The cold is supplied from both the floor and the ceiling for the configuration in *Figure 2-1* (c). The cold air is supplied form the ceiling and goes back to the cooling system from the side in *Figure 2-1* (d).

2.1.1 Thermal Guidelines and Performance Metrics

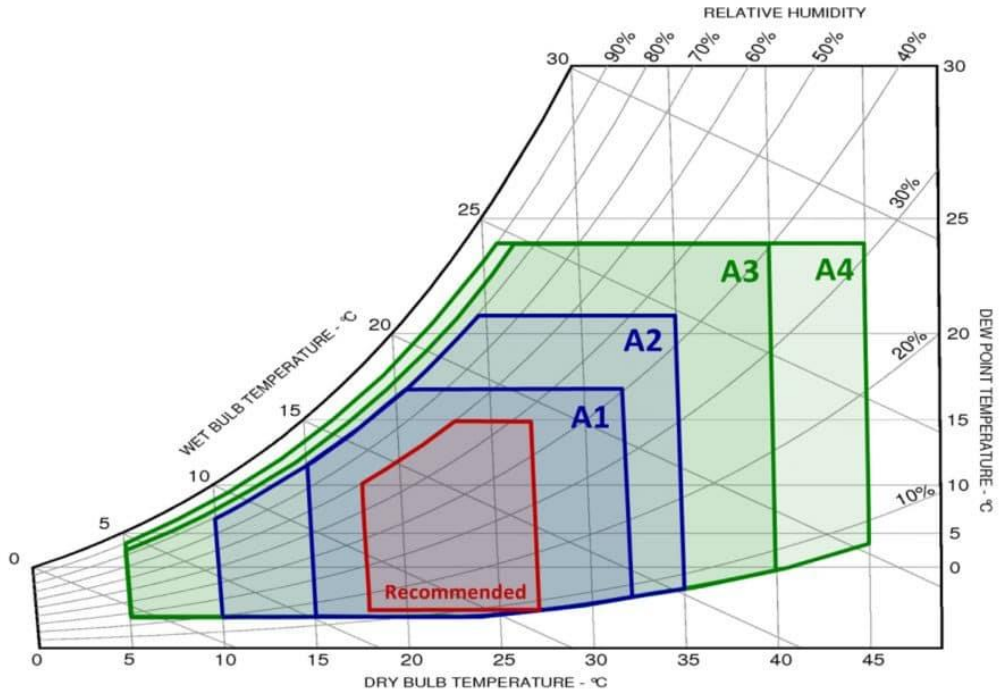


Figure 2-2 ASHRAE guidelines for data center thermal environment [15]

The thermal requirements of data centers are shown in *Figure 2-2* [15]. The recommended and allowable envelopes are presented for ASHRAE Classes A1, A2, A3 and A4. It is noteworthy that all these envelopes pertain to air entering the IT equipment. This is different from other types of buildings, such as residential buildings or offices, which control the averaged room temperature.

The recommended ranges for dry bulb temperature are from 18 ^{o}C to 27 ^{o}C. The recommended envelopes for humidity are from -9 ^{o}C DP (dewpoint temperature) to 15 ^{o}C DP and

60% RH (relative humidity). The allowable ranges are different for different classes. For example, the dry bulb temperature ranges from 15 ºC to 32 ºC and the humidity ranges from -12 ºC DP and 8% RH to 27 ºC DP and 80% RH.

2.1.2 Scope of the Dissertation

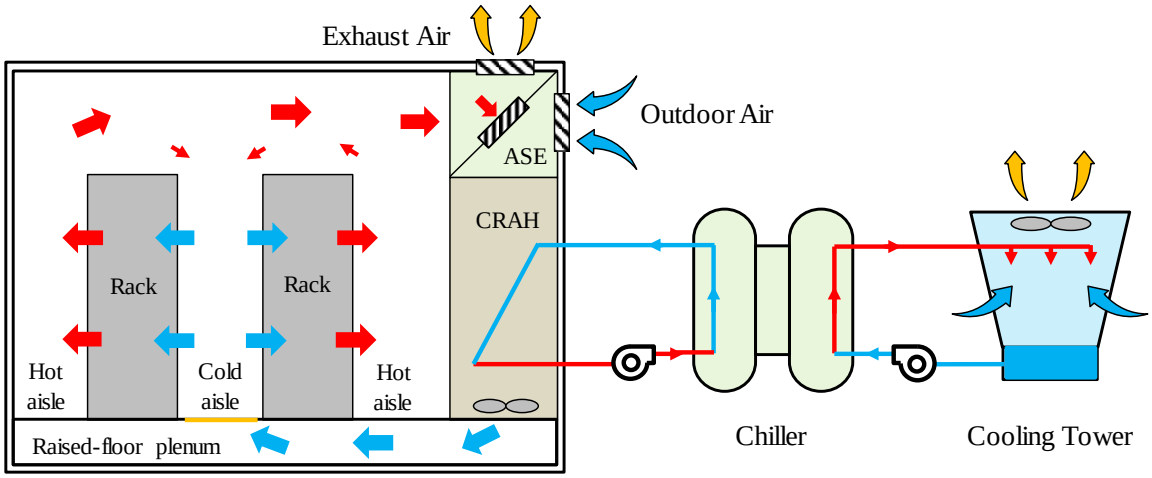


Figure 2-3 A typical data center cooling system with air-side economizer

Figure 2-3 shows a typical data center cooling system and airflow pattern. This dissertation focuses on the cooling systems with air-side-economizer (ASE), in which outdoor air is induced to the room to cool the IT equipment when the weather condition meets requirements. The cooling system operates on free cooling (FC), partial mechanical cooling (PMC) or fully mechanical cooling (FMC) according to a cooling mode controller. The airflow configuration studied in this dissertation is "raised-floor supply" that supplies cold air to the cold aisle through a raised floor plenum, though the models and methods proposed in this dissertation may also be applied for other configurations.

2.2 Modeling of Data Center Cooling Systems

Simulation can be an effective way to assist the design and operation of data centers. Many tools have been developed in academia and industry to perform computer simulation of cooling systems in data centers. For example, eQuest [16, 8], EnergyPlus [7, 17], TRNSYS [18], and some customized simulation tool such as Energy Modeling Protocol [19] have been widely used to study the cooling systems with waterside economizers (WSEs) and airside economizers (ASEs) in data centers.

The abovementioned tools utilize imperative programming languages such as FORTRAN, C/C++, which makes the tools less extensible. In such programs, models usually tightly couple physical equations, input/output routines with numerical solution methods, by making the numerical solution procedure part of the actual model equations [20]. This intertwinement makes it difficult to extend these programs to support various use cases [21], co-simulations with each other [22, 23] and effective optimization [24]. What's more, some energy simulation tools are not suitable for evaluating the system dynamics and the semantics of their control has little in common of how actual control works. For example, in EnergyPlus, the commonly used PI control loop is assumed to be ideal, i.e., there will be no overshoot [25]. EnergyPlus also idealizes dead band or waiting time, which are frequently used in the building control process. Moreover, many equipment models have built-in idealized control that requests flow rates, and flow rates are ideally distributed within a system rather than the results of friction-based flow distribution for a given valve or pump control signal. This makes it impossible to model, test and verify actual control.

One way to address these problems is that physical equations and their numerical solutions should be separated where possible. Instead of imperative programming language such as C/C++,

the equation-based language Modelica can be used to realize the objective. Details about Modelica and how it can benefit system modeling and optimization can be found in previous publications [21, 26].

The Modelica Buildings library (MBL) has been developed by Lawrence Berkeley National Laboratory (LBNL) to support various use cases regarding to HVAC systems in buildings [12]. MBL is an open-source, free library with component and system models for building energy and control systems. Besides the conventional energy analysis, this library can also provide support for rapid prototyping [27], modeling of arbitrary HVAC system topologies [27], evaluation of the stabilization of feedback control and Fault Detection and Diagnostics at the whole building system level [28, 20, 29]. It can also be used in the design and operation of cooling systems for data centers [12].

2.3 Modeling of Airflow Management

To model airflow management in data centers, many available thermal management models have been utilized with varying levels of complexity and accuracy. CFD was used to predict detailed airflow and temperature fields in the data center [30, 31]. Potential flow models [32] and linear abstract heat flow models [33] were also employed as fast prediction techniques. To further accelerate the speed of predictions, artificial neural networks [34] and reduced order models (e.g. Proper Orthogonal Decomposition) [35] were utilized to obtain some critical information of the data center thermal environment. Among these methods, CFD has been widely used as a more sophisticated method than others. A recent study showed that, with careful calibration, CFD can make reliable predictions of perforated-tile air flow rates and rack-inlet temperatures [36]. However, while versatile and generally accurate, CFD is computationally

expensive (especially when the size of the data center is large). This makes its use in optimal design and operation of data centers difficult, especially when multiple parameters are optimized, which requires a large number of simulations.

Conventional CFD commonly employs the Semi-Implicit Method for Pressure Linked Equations (SIMPLE) [37] to solve the governing equations for fluid flow. For simplicity, the conventional CFD is referred to as CFD in the rest of the paper. To accelerate CFD, an alternative CFD model called fast fluid dynamics (FFD) was proposed and applied to predict indoor environment. FFD solves the same set of governing equations as CFD, but with a time-split method and semi-Lagrangian advection solver. FFD was reported to be 50 times faster than CFD and it can additionally speedup 30 times by running in parallel on graphics processing unit (GPU) [38, 11]. This level of speed improvement has great potential to significantly accelerate the process of model-based design and operation for data centers. FFD was first introduced and verified in building simulation applications by Zuo and Chen [39]. FFD was then improved to simulate different cases such as indoor airflow [11, 38, 40], airflow around buildings [41], cabin environment [42], and urban-scale airflows [43]. However, the airflow and thermal dynamics in data centers are quite different compared to these applications [13]. One is that the high thermal load and airflow rate make the airflow pattern in the data center more complex than that in regular indoor environment. The other is that special treatments for boundary conditions unique for data centers are needed, such as perforated floor tiles and IT racks with server fans inside. Some previous studies [44, 45, 36] tried to use FFD for data center airflow and thermal modeling and showed potential, but these studies lack a comprehensive evaluation of using FFD for optimal

design and operation of data center thermal management. In addition, the FFD programs implemented in those studies are proprietary.

2.4 Holistic Optimization for Cooling Systems and Airflow Management

2.4.1 State of the Art of Related Simulation Techniques

Ventilation systems involving stratified airflow and temperature distributions have been widely used since they can achieve required thermal comfort more effectively and efficiently. The typical applications are stratum ventilation in large spaces, ventilation in spaces with intense heat generation, displacement ventilation and natural ventilation [46, 45, 47]. The data center, which is the focus of this dissertation, is one of the typical applications. Model-based methods can be employed to evaluate and improve the design of such systems. However, it can be challenging to study such systems with the widely used Building Energy Simulation (BES) programs, such as EnergyPlus [48], TRNSYS [49], ESP-r [50], IDA-ICE [51], and BSim [52], which usually adopt a multizone room model. The multizone model (also called nodal model) assumes the air is well mixed in each zone, which can be represented by a node. This assumption in the multizone model is not valid for large spaces or ventilation systems with stratified airflow and temperature distributions [46].

To model the stratified indoor airflow and temperature distributions, multi-node nodal model, zonal model, and computational fluid dynamics (CFD) were proposed with increasing fidelity, which can be coupled with BES [46]. The multi-node nodal model was derived from the conventional multizone model by adopting multiple nodes to represent the stratified temperature distribution in a single zone [53]. The limitations of this model are 1) requirement of priori and empirical knowledge of various coefficients, such as convective heat transfer coefficient [54], and

2) requirement of in-depth understanding and detailed descriptions of the airflow pattern [55]. The zonal model is an intermediate model between the nodal model and CFD model, which is capable of capturing detailed stratified airflow and temperature distributions by dividing a room into a limited number of subzones and solving the equations for mass and energy balance [56]. However, the applications of the zonal model might be limited for flows with significant momentum because it does not solve the momentum equations. Even though special treatments can be added to handle airflow with significant momentum, such as the jet region of an inlet, the additional complexity and computing demand make the zonal model show few advantages against CFD [46]. Compared to the nodal model and zonal model, CFD can get the most accurate simulation results as it divides the space into the highest resolution of control volume and solves a set of governing equations for each control volume. Despite requiring high computational demand, CFD has become more and more popular on evaluating ventilation performance because of its great capability to model complex airflow and temperature distributions [46]. Given the complexity for data center applications, e.g. predicting hot spots, coupling CFD with BES is more suitable compared to the nodal model and zonal model, which cannot predict the hot spots even though they can recognize the stratified airflow and temperature distribution to some extent.

Many attempts have been made to couple CFD with BES for different purposes. First, CFD can improve the accuracy of heating or cooling load calculation by providing more accurate convective heat transfer coefficients [57], non-uniform airflow and temperature distributions [58, 59], infiltration and exfiltration rates [60], and local microclimates [59]. Second, CFD was adopted to evaluate the performance of advanced ventilation systems involving stratified airflow, such as displacement ventilation [61], natural ventilation [54] and double-skin facades/roofs [62, 63].

Moreover, CFD was coupled with BES to study the control dynamics of ventilation systems, such as control performance evaluation for a VAV system with different thermostat locations [64, 65].

However, as pointed out in a recent review paper [66], one of the major limitations of coupling BES with CFD is the high computational demand. As a result, most research that dynamically couples BES with CFD only simulate typical days instead of a whole year. For example, Srebric, Chen, et al. [59] studied a house (6 m × 4 m × 2.5 m) and an atrium (7 m × 4.3 m × 4.5 m), and found that it took about three hours to simulate the two cases for a typical day. If a more realistic case is studied, for example Zhang, Lam, et al. [54] coupled EnergyPlus with Fluent to study a naturally ventilated building, it was found that the computing time of one CFD simulation is 27 minutes, which results in about 11 hours to simulate the case for a day [54]. The application of coupling BES with CFD becomes even more challenging when studying the control of ventilation systems, which usually require a much shorter time interval to exchange data between BES and CFD than applications that target energy prediction to capture short term thermal dynamics. For example, Sun and Wang [67] embedded ventilation and control systems into a CFD program to evaluate control dynamics of a VAV system and found it took about 10 hours to simulate a ventilated room (3 m × 3 m × 2.5 m) for 1 hour. Even with Fast Fluid Dynamics (FFD) [11], which is an alternative to CFD and was reported to be 50 times faster than traditional CFD [ref: indoor air paper], it took 430 s to simulate a ventilated room (2.44 m × 2.44 m × 2.44 m) for 800 s using a coupled Modelica-FFD model [47]. This is still not fast enough for a long-term evaluation, which provides a more comprehensive performance evaluation by considering varying weather conditions, schedule profiles, and partial loads compared to a typical design day.

To reduce the computational cost of the BES-CFD co-simulation, the BES-ROM-CFD approach (also called bin coupling scheme or virtual dynamic coupling [58]) was proposed to couple BES with a reduced order model (ROM) trained by pre-computed CFD simulation results [68]. The conventional BES-ROM-CFD coupled model has the similar data synchronization scheme as the BES-CFD coupled model on the online stage. The only difference is that the BES-CFD coupled model calls CFD to make predictions for airflow but the conventional BES-ROM-CFD coupled model calls a ROM pre-trained by CFD instead. Consequently, the BES-CFD coupled model is accurate but time-consuming and the conventional BES-ROM-CFD coupled model is fast but may be less accurate. This conventional BES-ROM-CFD approach can be done with different forms of ROMs. Chen and Van der Kooi [58] coupled BES with curve-fitted functions generated by CFD results to conduct building energy analysis. Kim, Braun, et al. [69] developed a linear time-invariant (LTI) model to predict the indoor environment information to support the control study of rooftop units. As a widely-used ROM of CFD, the proper orthogonal decomposition (POD) constructed by interpolating from a set of CFD samples was coupled with BES to study the load placement in an open-aisle air-cooled data center [70]. An artificial neural network trained by CFD results was integrated into BES to predict the energy consumption and thermal environment of a large-space atrium [71]. Compared to BES-CFD, these BES-ROM-CFD studies can reduce the simulations time from hours to seconds and are generally accurate when the inputs of ROMs are within the training domain. However, the major limitation of these BES-ROM-CFD approaches is that the prediction accuracy might drastically drop when the model inputs for ROM are outside of the training domain. Consequently, inaccurate predictions might be returned from ROMs if the training domain is too small to cover all possibilities. On the other hand, the

ROM training process might be challenging, sometimes even impractical, due to high computational costs if the ROM is trained in a sufficiently large domain, especially for multivariate problems. Moreover, even a sufficiently large training dataset that covers the range of each variable is created from CFD simulations, it does not necessarily warrant the prediction accuracy of the ROMs. In fact, the ROMs commonly need to be carefully trained and tested to achieve a required prediction accuracy. For example, [72] mentioned that the representativeness of the data samples (i.e. training dataset) from CFD simulations is critical for the accuracy of the POD method. Even with the methodology proposed in [72], multiple iterations may be needed to identify sufficient data samples. In addition to the conventional BES-ROM-CFD approach, Zhai and Chen [68] mentioned a dynamic bin coupling method, which first predicted the airflow details in typical days by running BES-CFD co-simulation and then used those results (i.e. bins) in BES for the days with similar conditions. However, it is not clear what environmental and building operating conditions can be identified to be similar as "typical days". Also, the simulation errors introduced by using bins have not been fully investigated and understood.

2.4.2 Applications in Data Centers

ASHRAE [15] recommends 27 ^{o}C as the maximum inlet temperature for the IT equipment. In reality, computer room air conditioners (CRACs) in many data centers are operated at a significantly lower supply temperature to avoid hot spots. While providing redundant cooling, this leads to reduced cooling efficiency [5]. Therefore, it has been a research topic on how to improve the energy efficiency while still providing enough cooling to the data center whitespace, where the IT equipment is placed. This can be done through a holistic optimization framework as shown in

Figure 2-4, in which the simulation of HVAC systems and controls are dynamically coupled with simulation of non-uniform thermal environment in data centers.

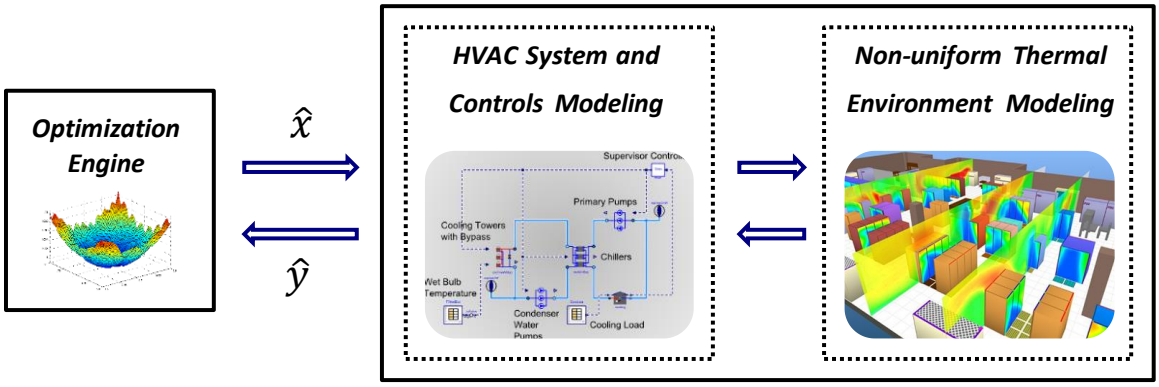


Figure 2-4 Holistic optimization framework for the data center cooling systems and airflow management

Moore, Chase, et al. [73] proposed several algorithms for workload placement to maximize the supply air temperature while ensuring the server inlet temperature not exceeding the threshold. Tang, Gupta, et al. [74] investigated several load placement strategies to minimize heat recirculation. Among these research, the CFD simulations were conducted to obtain the thermal map of the data center with regard to workload distribution, which required huge computational efforts to cover a large number of scenarios.

The bottleneck of the holistic optimization of cooling systems and airflow management is the high computational costs of CFD simulations. To reduce the computational cost, simplified airflow models were proposed for fast prediction of indoor environment. Zhang, VanGilder, et al. [75] developed a lumped model to predict rack-inlet temperatures under various supply airflow and rack-load scenarios. Since only one effective aggregated rack and cooling unit were considered based on a well-mixed assumption, the lumped model may fail to capture local airflow pattern and

temperature distribution. Potential flow models were used to predict the airflow and temperature distribution [76]. However, the potential flow model is an approximate and simplified model, which may not be as accurate as CFD, particularly when simulating the jet flows from the perforated tiles.

Other than physic-based models, statistics-based approaches, sometimes referred to as reduced order models (ROMs) were also employed. Proper orthogonal decomposition (POD) was used to predict the velocity and temperature distribution in an office [35] and rack-inlet temperature distribution in a raise-floor data center [77]. Artificial neural network (ANN) models were used to predict thermal map of data centers [78], perforated-tile flowrates and rack-inlet temperatures [34]. However, the predictions of such statistics-based models may become inaccurate when the queries lie beyond the training domain.

Chapter 3 A New Open Source Fast Fluid Dynamics Model for Data Center Thermal Management

This chapter introduces a new FFD model for fast prediction of data center airflow and thermal environment. The FFD model is validated and then demonstrated to support optimal data center thermal management with a real middle-size data center.

3.1 Introduction

To support model-based design and operation for data centers, many available thermal management models have been utilized with varying levels of complexity and accuracy. CFD was used to predict detailed airflow and temperature fields in the data center [30, 31]. Potential flow models [32] and linear abstract heat flow models [33] were also employed as fast prediction techniques. To further accelerate the speed of predictions, artificial neural networks [34] and reduced order models (e.g. Proper Orthogonal Decomposition) [35] were utilized to obtain some critical information of the data center thermal environment. Among these methods, CFD has been widely used as a more sophisticated method than others. A recent study showed that, with careful calibration, CFD can make reliable predictions of perforated-tile air flow rates and rack-inlet temperatures [36]. However, while versatile and generally accurate, CFD is computationally expensive (especially when the size of the data center is large). This makes its use in optimal design and operation of data centers difficult, especially when multiple parameters are optimized, which requires a large number of simulations.

Conventional CFD commonly employs the Semi-Implicit Method for Pressure Linked Equations (SIMPLE) [37] to solve the governing equations for fluid flow. For simplicity, the conventional CFD is referred to as CFD in the rest of the paper. To accelerate CFD, an alternative CFD model called fast fluid dynamics (FFD) was proposed and applied to predict indoor environment. FFD solves the same set of governing equations as CFD, but with a time-split method and semi-Lagrangian advection solver. FFD was reported to be 50 times faster than CFD and it can additionally speedup 30 times by running in parallel on graphics processing unit (GPU) [38, 11]. This level of speed improvement has great potential to significantly accelerate the process of model-based design and operation for data centers. FFD was first introduced and verified in building simulation applications by Zuo and Chen [39]. FFD was then improved to simulate different cases such as indoor airflow [11, 38, 40], airflow around buildings [41], cabin environment [42], and urban-scale airflows [43].

However, the airflow and thermal dynamics in data centers are quite different compared to these applications [13]. One is that the high thermal load and airflow rate make the airflow pattern in the data center more complex than that in regular indoor environment. The other is that special treatments for boundary conditions unique for data centers are needed, such as perforated floor tiles and IT racks with server fans inside. Some previous studies [44, 45, 36] tried to use FFD for data center airflow and thermal modeling and showed potential, but these studies lack a comprehensive evaluation of using FFD for optimal design and operation of data center thermal management. In addition, the FFD programs implemented in those studies are proprietary. This research develops the necessary FFD modules for data center and conducts a comprehensive evaluation of FFD for that regarding prediction accuracy, computing speed and performance for

model-based design and operation in a real data center. The new FFD model for data center thermal management is implemented using OpenCL, which is a cross-platform parallel computing language. To enable a broader range of applications, the FFD code has been publicly released under a free open-source license.

This paper is organized as follows. First, a comprehensive introduction of the new FFD model is provided including governing equations, methods to solve these equations, treatments of special boundary conditions in data centers and their implementation. The new FFD model is validated with two classical cases for indoor environment modeling [79] in Section 3.1 and a real data center case located in Massachusetts, U.S.A in Section 3.2. Subsequently, the application of the FFD model is demonstrated using three case studies based on the aforementioned data center.

3.2 Methodology

3.2.1 Data Center Airflow and Thermal Dynamics

The configuration and intended airflow directions of a typical raised-floor data center are shown in *Figure 3-1*. The above-floor white space (i.e. allocated for IT equipment) in the data center is partitioned into cold aisles and hot aisles by rows of racks. The raised-floor plenum and perforated floor tiles are designed to uniformly distribute the cold air to the cold aisles. The cold air from the cooling system is supplied to the cold aisle, then flows through the IT racks and carries heat dissipated from IT equipment, and then exhausts to the hot aisle and returns to the cooling system through the ceiling plenum. ASHRAE [15] defines the requirements of the data center airflow and thermal management. One key parameter is to ensure the rack inlet temperatures are within a safe threshold (e.g. not exceeding 27 ^{o}C). The segregated cold and hot aisles help lower

the IT rack inlet temperature by preventing or mitigating the mixing of cold and hot air. Sometimes, physical containment of the hot or cold aisle is employed to avoid such mixing.

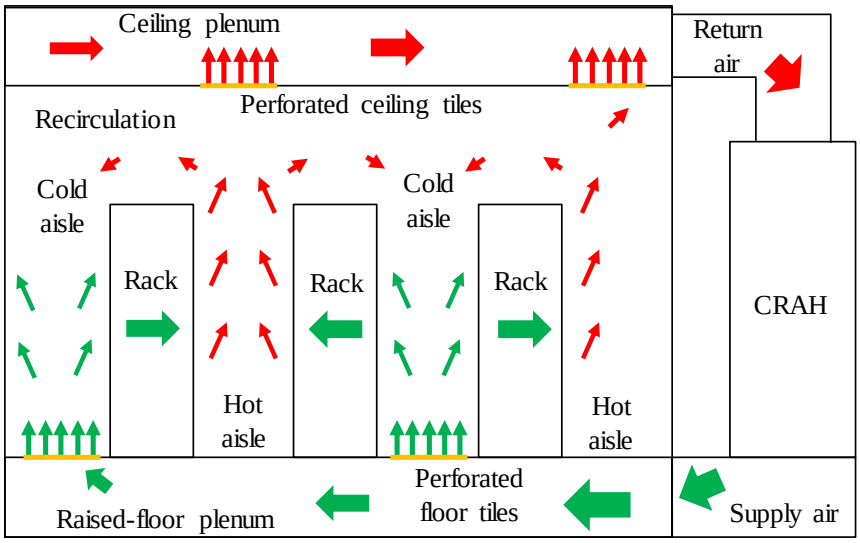


Figure 3-1 Airflow pattern in a typical raised-floor data center

CFD has been widely used to improve the data center thermal management. For example, CFD can be adopted to evaluate the temperature and airflow distribution for scenarios with different designs, such as different layouts of the cold aisle, hot aisle and IT racks. However, one limitation of CFD is the high computational demand, which makes its application infeasible when multiple parameters are studied and hundreds or even thousands of simulations are needed. Thus, the FFD model, which was reported to be 50 times faster than CFD [11], is evaluated in this paper to solve this problem.

3.2.2 Fast Fluid Dynamics

This section introduces conventional FFD for indoor environment modeling and the improved FFD model for data center thermal management. Note that here the "conventional FFD"

represents existing FFD [11, 80-82, 36] as opposed to the proposed FFD described later in this paper. The improvements include changing the solving methods for the advection equation and implementing boundary conditions and special treatments for data center environment. The implementation of the FFD model in OpenCL is also discussed.

3.2.2.1 Governing Equations and Solution Methods

Like CFD, the FFD model solves the same set of governing equations for flows. The Navier-Stokes momentum equation can be generalized and written as:

$$\frac{\partial \boldsymbol{U}_i}{\partial t} = -\boldsymbol{U}_j \frac{\partial \boldsymbol{U}_i}{\partial \boldsymbol{x}_j} + \nu \frac{\partial^2 \boldsymbol{U}_i}{\partial \boldsymbol{x}_j \partial \boldsymbol{x}_j} - \frac{1}{\rho}\frac{\partial P}{\partial \boldsymbol{x}_i} + \boldsymbol{F}_i, \tag{3-1}$$

where $\boldsymbol{U}$ is the velocity vector, t is time, $\boldsymbol{x}$ is the spatial coordinates, ν is the kinematic viscosity, ρ is the density, P is pressure, and $\boldsymbol{F}_i$ is the source term.

The energy equation can be written as:

$$\frac{\partial T}{\partial t} = -\boldsymbol{U}_j \frac{\partial T}{\partial \boldsymbol{x}_j} + \alpha \frac{\partial^2 T}{\partial \boldsymbol{x}_j \partial \boldsymbol{x}_j} + \boldsymbol{S}_T, \tag{3-2}$$

where T is the temperature, α is the thermal diffusivity, and $\boldsymbol{S}_T$ is the thermal source term.

FFD uses a time-split method. Equation 1, for example, is split into three equations in conventional FFD [11]:

$$\frac{U_i^{(1)} - U_i^{(n)}}{\Delta t} = -\boldsymbol{U}_j \frac{\partial \boldsymbol{U}_i}{\partial \boldsymbol{x}_j}, \tag{3-3}$$

$$\frac{U_i^{(2)} - U_i^{(1)}}{\Delta t} = \nu \frac{\partial^2 \boldsymbol{U}_i}{\partial \boldsymbol{x}_j \partial \boldsymbol{x}_j} + \boldsymbol{F}_i, \tag{3-4}$$

$$\frac{U_i^{(n+1)} - U_i^{(2)}}{\Delta t} = -\frac{1}{\rho}\frac{\partial P}{\partial \boldsymbol{x}_i}. \tag{3-5}$$

The three equations are solved sequentially. As shown in Figure 3-2, conventional FFD first solves *Equation 3-3*, namely the advection equation, by a semi-Lagrangian (SL) scheme [83]. *Equation 3-4*, namely the diffusion equation, is solved with an implicit scheme. Finally, *Equation 3-5*, namely the pressure equation, is solved together with the continuity equation:

$$\frac{\partial \boldsymbol{U}_i}{\partial \boldsymbol{x}_i} = 0, \tag{3-6}$$

based on a projection-correction method [84] to ensure mass conservation.

The SL scheme [83] has been widely adopted to solve the advection equation in conventional FFD [11, 38, 40-43]. One of the major advantages of the SL scheme is the fast computing speed, which is achieved by tracing locations at the last time step and calculating velocities through interpolation without any iterative algorithm. However, a major drawback of the SL scheme is that it does not guarantee in-general quantity conservation [85]. In addition, the determination of the time step size is crucial for the stability of simulation as well as the accuracy of the SL scheme [86]. For example, the Courant–Friedrichs–Lewy (CFL) constraints ($\boldsymbol{U}_i \Delta t/\Delta x \leq 1$) should be satisfied to overcome the stability concerns, which may require a small time-step size. Meanwhile, a smaller time step size may also reduce accuracy because of truncation error growth [86]. As a result, additional efforts are required to address these issues. On the other hand, an implicit scheme, such as the first order upwind, is stable and robust even when the CFL number is larger than one. However, the implicit scheme is usually more computationally expensive compared to the SL scheme. Thus, it will increase the computing demand by 8%-20% according to our tests if the SL scheme is replaced by an implicit scheme to solve the advection equation.

In addition to the stability and robustness, the first order upwind scheme has another advantage compared to the SL scheme, which is the unconditional conservation of mass and energy for each cell in the computational domain. With SL, one will have to enforce mass balance at the global level. Even by doing so, the mass balance at cell level is not perfect. This is particularly important in data center applications, because unlike regular buildings where the purpose is to create an environment with general thermal comfort, data center operators and researchers care more about local energy balance, which may influence the generations of local hot spots.

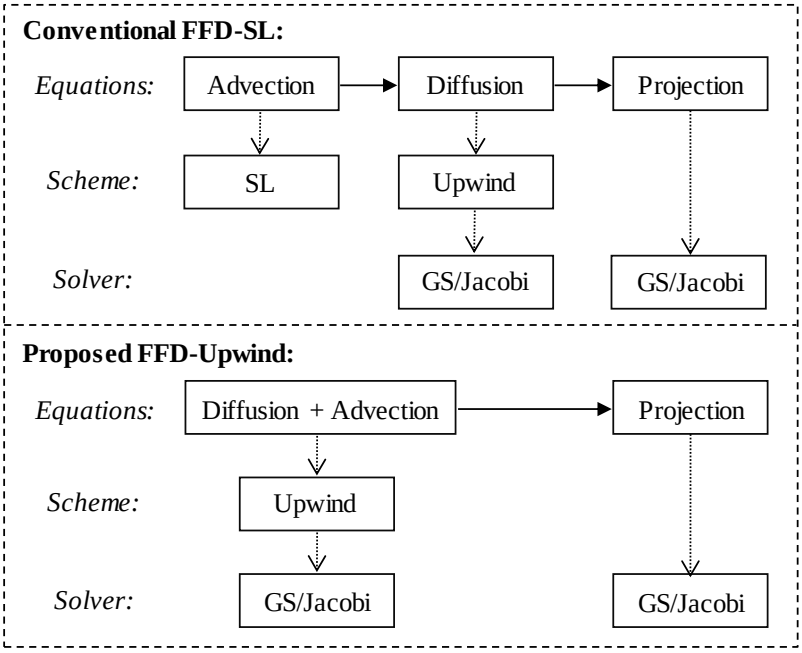


Figure 3-2 Workflow of the convectional FFD-SL and proposed FFD-Upwind

To take advantage of the stability, robustness and unconditional local mass and energy conservation of the first order upwind scheme while not sacrificing computing speed, this paper proposes a new FFD model to simultaneously solve diffusion and advection equations with a first order upwind scheme (*Figure 3-2*). The *Equations 3-3 ~ 3-4* are solved together as:

$$\frac{U_i^{(2)} - U_i^{(n)}}{\Delta t} = \nu \frac{\partial^2 \boldsymbol{U}_i}{\partial \boldsymbol{x}_j \partial \boldsymbol{x}_j} + \boldsymbol{F}_i - \boldsymbol{U}_j \frac{\partial \boldsymbol{U}_i}{\partial \boldsymbol{x}_j}, \quad (3\text{-}7)$$

For clarity, the proposed FFD model is called FFD-Upwind (or the FFD model) and the conventional FFD model is called FFD-SL. As shown in *Figure 3-2*, the FFD-Upwind first assigns coefficients of equation matrix for diffusion and advection sequentially and then solve the equation matrix with a linear Gauss-Seidel (GS) or Jacobi solver. As a result, the FFD-Upwind discards the SL scheme without increasing the computing demand compared to the FFD-SL.

3.2.2.2 Boundary conditions

In this section the general boundary conditions that are commonly used for modeling of indoor environment are briefly introduced. The special boundary conditions for objects in the data center environment are then described including perforated floor tiles and IT racks.

General boundary conditions:

In our study, three general types of boundaries are considered including inlets, outlets, and walls. A Dirichlet boundary condition is applied for the inlets as a fixed velocity. For the outlets, either a Neumann boundary condition with a zero-gradient velocity or a Dirichlet boundary condition with a fixed velocity can be imposed. For the walls, a no-slip wall boundary condition is applied, which assumes the air velocity at the solid wall boundary is zero.

Perforated floor tile:

Since the raised-floor plenum and white space are modeled separately in this paper, the modeling of perforated floor tiles can be done with different considerations depending on which space is studied [13]. The aim of modeling the raised-floor plenum is to predict the air flow rates at perforated tiles. The flow rates are determined by the pressure distribution in the plenum and

the pressure drop when the air flows through the perforated tiles. We adopt a commonly used approach called the porous jump method (also regarded as lumped resistance method). It simplifies the perforated tiles as flow resistances and specifies a step pressure loss across the perforated tiles [31, 13, 87]:

$$\Delta P = \frac{1}{2}\rho f V^2, \tag{3-8}$$

where ΔP is the pressure drop across the tile, V is the velocity approaching the tile, and f is the dimensionless loss coefficient, which can be estimated from manufacturer data or by empirical formulae, for example, the one proposed by Fried and Idelchik [88]:

$$f = \frac{1}{\beta^2}[1 + 0.5(1-\beta)^{0.75} + 1.414(1-\beta)^{0.375}], \tag{3-9}$$

where β is the open-area-ratio of the perforated tile.

When modeling the white space, the goal is to study the airflow and temperature distribution within the space. The perforated tiles are just special inlet boundary conditions. As a result, they are often treated as prescribed uniform-velocity boundaries independent of their open-area-ratio. The magnitude of the velocity is typically determined based on the air flow rate and the total area of the perforated tile. However, in reality, the airflow will be accelerated when it goes through the small openings of the perforated tiles, which results in higher local velocities and a lower-pressure region above the perforated tiles.

Balancing the complexity and accuracy of modeling such perforated tiles should be considered. On one hand, modeling the perforated tiles as fully opened openings with a pressure loss using the porous jump model would omit the jet effect above the tiles, and therefore may lead to inaccurate or even incorrect results [87]. On the other hand, detailed pore-by-pore modeling,

although versatile and generally accurate, is impractical in practice considering the added complexity. Consequently, there are various attempts in this area to achieve a compromise between complexity and accuracy of modeling the perforated tiles. The body force model or modified body force model were used to specify a momentum source above the perforated tiles [89, 87]. Abdelmaksoud, Dang, et al. [89] also developed a quadrants method, which separates the perforated tile to multiple openings with the same total opening area. All these methods were reported to successfully capture the air acceleration through the pores of the perforated tiles. This research adopts the body force method [87] due to its simplicity and good accuracy. An additional force is added into the momentum equations to correct the under-estimated velocity for computational cells just above the perforated floor tiles. Suppose that a perforated tile has an open-area-ratio of β, surface area A (m^2) and air flow rate Q (m^3/s). With the assumption that the momentum source is applied to a computational cell with a height h (m) directly above the tile, we can compute the momentum source as:

$$\boldsymbol{F}_i = \frac{Q^2}{A^2 h}\left(\frac{1}{\beta} - 1\right), \tag{3-10}$$

where the direction of $\boldsymbol{F}_i$ is perpendicular to the surface orientation of the perforated tile. In this paper, we set h as 0.15 m.

IT Rack:

There are generally two approaches to model the IT racks: the open box model and the black box model [90]. The major difference between the two methods is that the airflow inside the rack is modeled in the open box model but is excluded in the black box model. The former method adopts a more detailed model, which may be able to capture the temperature and velocity

stratification across the surfaces of the rack inlet and outlet. The latter method simplifies the rack as a solid box with an inlet and an outlet. The exhaust temperature at the rack outlet is calculated based on the temperature profile of the air flowing into the rack inlet, with an assumed temperature rise to model the effect of heat dissipation inside the rack. Zhang, VanGilder, et al. [91] compared different levels of details for modeling an IT rack including a black box rack, a detailed rack with crude server simulators and a detailed rack with detailed server simulators, and found that the different levels of rack details had little effect on the predicted temperatures for the studied case. Therefore, in the present study, we adopt the black box model for its simplicity and sufficient prediction performance. In the black-box model, individual servers or "slices" are not explicitly resolved. The airflow is assumed to be proportional to total rack power dissipation (P, kW) with 212 m^3/h (125 cfm) of airflow for each kW of power (ψ, m^3/h/kW) [92, 36]. The airflow is spread uniformly over the front and rear of the rack.

$$Q_{in} = P \times \psi \tag{3-11}$$

The air temperature at the rack inlet is assigned with the temperature of its adjacent cells. The vertical temperature gradient at the front (inlet) of the rack is carried through to the rear (exhaust) of the rack with a temperature rise, which is determined by the heat dissipation of IT servers and flow rate of air through the rack:

$$T_{ex} = T_{in} + \frac{q_{server}}{c_p \times \dot{m}} \tag{3-12}$$

3.2.2.3 Software Implementation

The data center can be split into two volumes: the raised-floor plenum and white space [13]. A previous study [93] recommended to model the raised-floor plenum separately and use the

results of flow rates at perforated tiles as boundary conditions in the white space model. For modeling convenience, our current implementation follows the procedure of modeling the raised-floor plenum and the white space separately. This is justified when relatively low open-area-ratio tiles are employed.

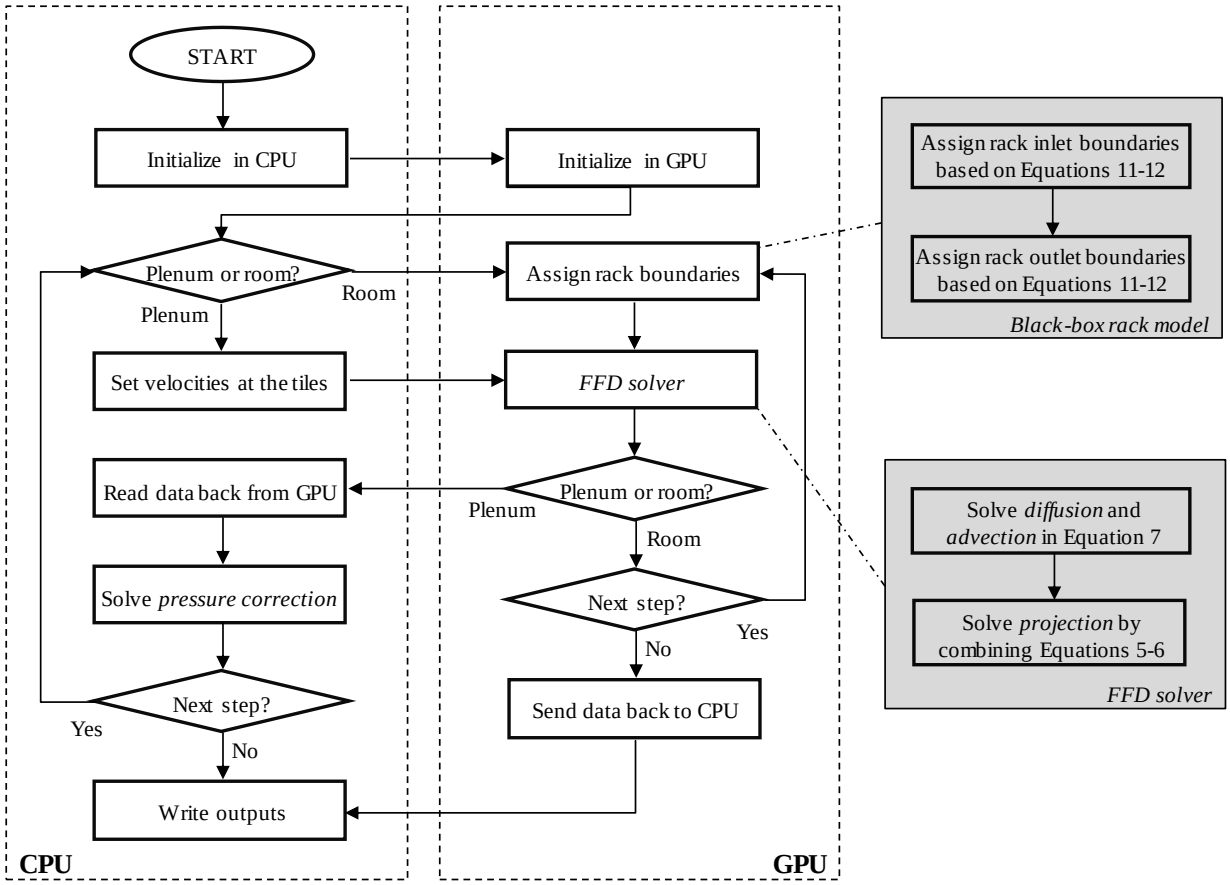


Figure 3-3 Structure of the proposed FFD model

The structure and workflow of the FFD model is shown in *Figure 3-3*. The FFD model is implemented through hybrid programming in C and OpenCL. The host program (i.e. main routine of FFD) runs sequentially on the CPU and the kernels run in parallel on the GPU. The code in OpenCL is used to execute the kernels for assigning boundary conditions and solving the

governing equations. For more details of the structure of the implementation, please refer to Tian, Sevilla, et al. [40]. It is noted that the source codes of this FFD model have been publicly released in [94] , which is the first open source FFD model for data center thermal management.

When modeling the raised-floor plenum, to ensure that the simulated airflow field and air flow rates at tiles simultaneously satisfy *Equations 3-1* and *3-7*, we adopt a pressure correction method proposed by VanGilder, Sheffer, et al. [95]. The FFD first assigns initial values for all air flow rates at tiles and computes the pressure field in the raised floor plenum based on *Equations 3-5* and *3-6*. The FFD model then calculates the air flow rates at tiles according to the obtained pressure field and *Equation 3-7* and checks the mass conservation. Subsequently, the FFD model shifts the pressure in the plenum by a constant value (for example, the pressure at all cells in the plenum will be increased by a constant value if the current pressure field is too small to provide enough outflows) until a converged flow field satisfying momentum and mass conservation is achieved.

When modeling the white space, we adopt a black box model to assign the boundary conditions for the IT racks based on *Equations 3-10* and 3-*11* [36]. All the governing equations are solved based on a Jacobi method on the GPU in parallel. For the modeling of both plenum and white space, the simulation will be terminated when the velocity, temperature and pressure fields become steady state.

3.3 Validation

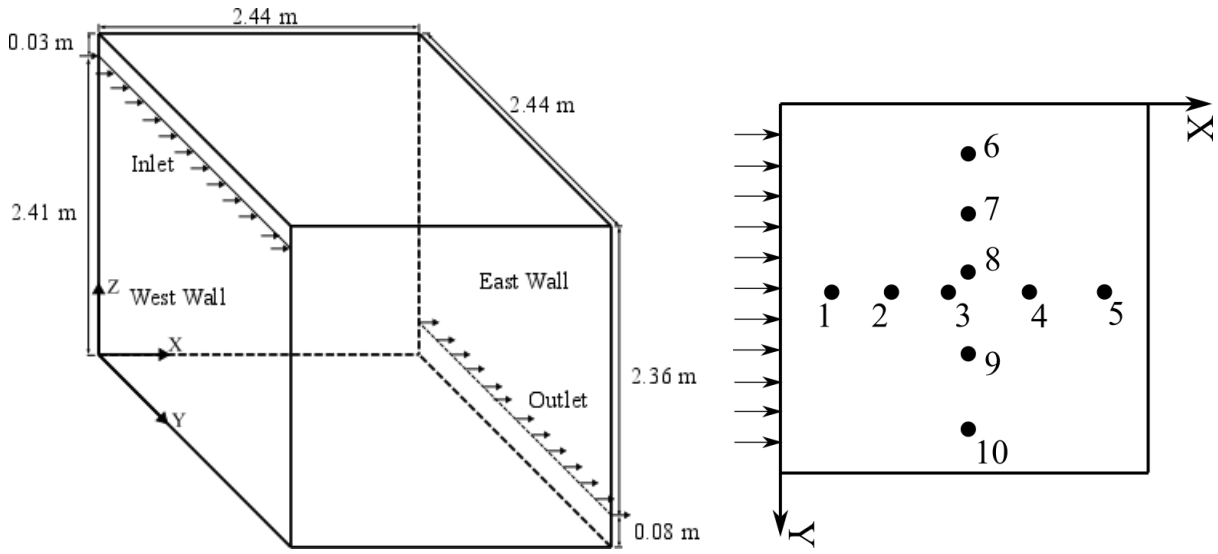
To validate the proposed FFD model, two classic cases for indoor environment modeling and one data center case are selected. The indoor cases are studied to evaluate the proposed FFD model that solves the advection and diffusion equations together, compared to conventional FFD

that solves them separately. The real data center case is then studied to evaluate the capability of the proposed FFD model with new features to simulate data center thermal environment. Experimental data are taken as reference for all the cases.

3.3.1 Validation of the Improved FFD Model

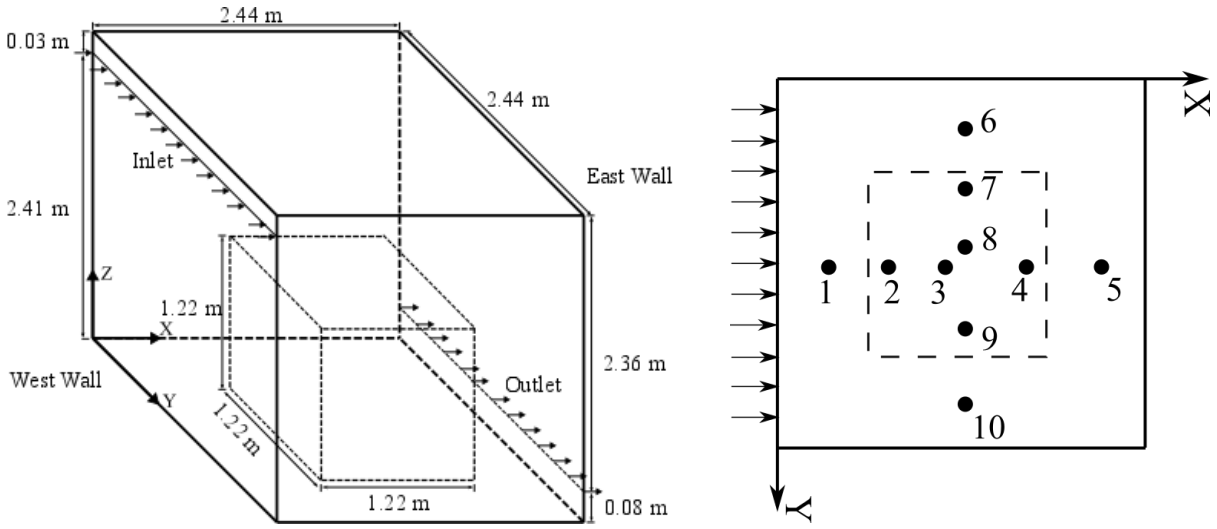
3.3.1.1 Description of the indoor environment cases

The first case is an empty room with forced convection, which is a pure airflow case without heat transfer [79]. The room is 2.44 m (8 ft) long, 2.44 m (8 ft) wide and 2.44 m (8 ft) high with an inlet at the top of the west wall and an outlet at the bottom of the east wall. Other critical dimensions are shown in *Figure 3-4 (a)*. The temperatures of the inlet flow and surfaces in the room are controlled to be the same. Experimental data at ten locations as shown in *Figure 3-4 (b)* are available [79].



(a) (b)

Figure 3-4 (a) Schematic of the forced convection in an empty room and (b) locations of experimental data



(a) (b)

Figure 3-5 (a) Schematic of the mixed convection in a room with a box and (b) locations of experimental data

The second case increases the flow complexity by adding a heated box (1.22 m × 1.22 m × 1.22 m) in the center of the room, in which heat transfer occurs [79]. The size of room and locations of the inlet and outlet are the same with the first case. Other critical dimensions are shown in *Figure 3-5 (a)*. The temperatures of the inlet flow, box surface, ceiling, flow and other walls are 22.2 ºC., 36.7 ºC, 25.8 ºC, 26.9 ºC, and 27.4 ºC, respectively. Experimental data is available for ten locations shown in *Figure 3-5 (b)* [79].

3.3.1.2 Evaluation metrics

Normalized root-mean-square deviation (NRMSD) is adopted to quantify the accuracy of predictions by FFD and CFD with respect to experimental data, which is defined as:

$$NRMSD = \frac{1}{x_{max} - x_{min}} \sqrt{\frac{\sum_{i=1}^{N} (\hat{x}_i - x_i)^2}{N}} \times 100\%, \quad (3\text{-}13)$$

where $\widehat{x_i}$ and x_i are simulated and measured values at point i, respectively. N is the total number of data points. $(x_{max} - x_{\min})$ represents the range of the studied parameter over the computational domain, which is 1.36 m/s when evaluating velocity profiles and 14.5 ºC (36.7 ºC - 22.2 ºC) when evaluating temperature profiles for the two cases.

3.3.1.3 Setup of simulations

Table 3-1 Settings of simulations in FFD-Upwind and FFD-SL

Case	Grid		Simulation Time (s)		Time Step Size (s)	
	FFD-Upwind	FFD-SL	FFD-Upwind	FFD-SL	FFD-Upwind	FFD-SL
Forced Convection in an Empty Room	40×40×40		1350		0.05	
Mixed Convection in a Room with a Box	40×40×40		1350		0.05	

Table 3-1 summarizes the simulation settings employed in both FFD-Upwind and FFD-SL. We perform a similar grid independent study as was done in Tian, Sevilla, et al. [40]. It is found that with the non-uniform structured grid (40×40×40), in which the averaged mesh size is about 6 cm, the simulations achieve grid independent results. Similar conclusions were drawn in previous studies [40, 79]. The mesh is refined at critical locations to capture the gradients and changes in the flow. For example, the inlet as well as near-ceiling areas is refined with a minimum mesh size of about 0.5 cm and the outlet as well as the near-floor area is refined with a minimum mesh size of about 1 cm. The FFD-Upwind and FFD-SL models perform transient-state simulations because of the time-split method used in FFD (with a time step size of 0.05 s and simulation time of 1350 s). Both FFD-Upwind and FFD-SL use a zero-equation turbulence model [96, 97] and are performed on an AMD FirePro™ W8100 GPU.

3.3.1.4 Results

Table 3-2 NRMSD of predictions for cases with FFD-Upwind and FFD-SL

Cases	Simulation Program	Predictions	NRMSD (%)				
			P1	P3	P5	P6	Ave.
Forced Convection in an Empty Room	FFD-SL	Velocity	3.72	7.95	6.59	8.56	6.71
	FFD-Upwind	Velocity	4.93	4.25	6.93	4.62	5.18
Mixed Convection in a Room with a Box	FFD-SL	Velocity	5.22	5.46	4.18	4.38	5.25
		Temperature	5.44	6.68	2.84	7.78	
	FFD-Upwind	Velocity	6.92	4.21	7.19	3.47	4.00
		Temperature	2.33	3.47	2.43	1.99	

Table 3-2 summarizes the NRMSD of predictions for the cases with FFD-Upwind and FFD-SL. The average NRMSDs for the four locations with FFD-Upwind are 5.18% and 4% for the two cases. The FFD-Upwind has slightly better overall prediction accuracy with lower averaged NRMSDs for the two cases compared to FFD-SL. However, it can still be seen that FFD-SL performs better than FFD-Upwind for velocity predictions at the locations P1 and P5. This may be due to the complex flow structure in those areas, which is hard to capture.

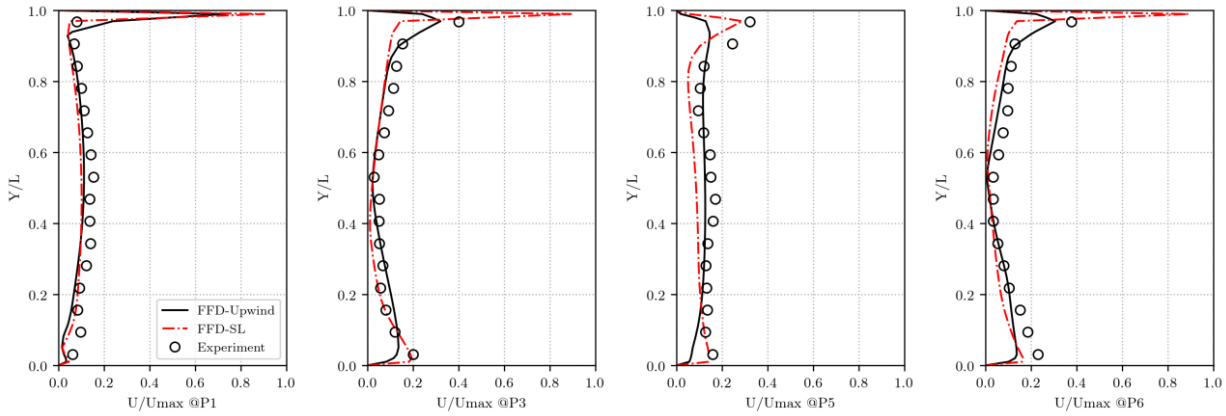


Figure 3-6 Comparison of velocity profiles for forced convection in an empty room

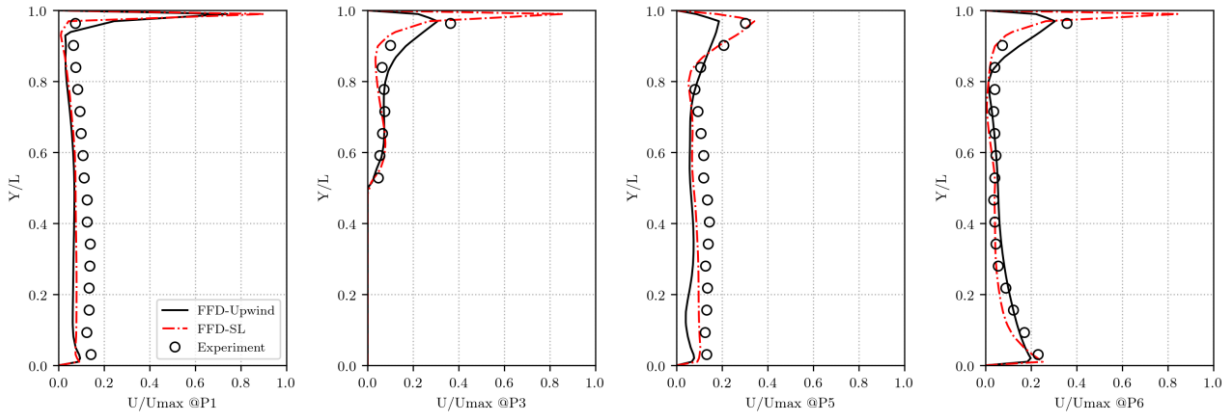


Figure 3-7 Comparison of velocity profiles for mixed convection in an empty room with a box

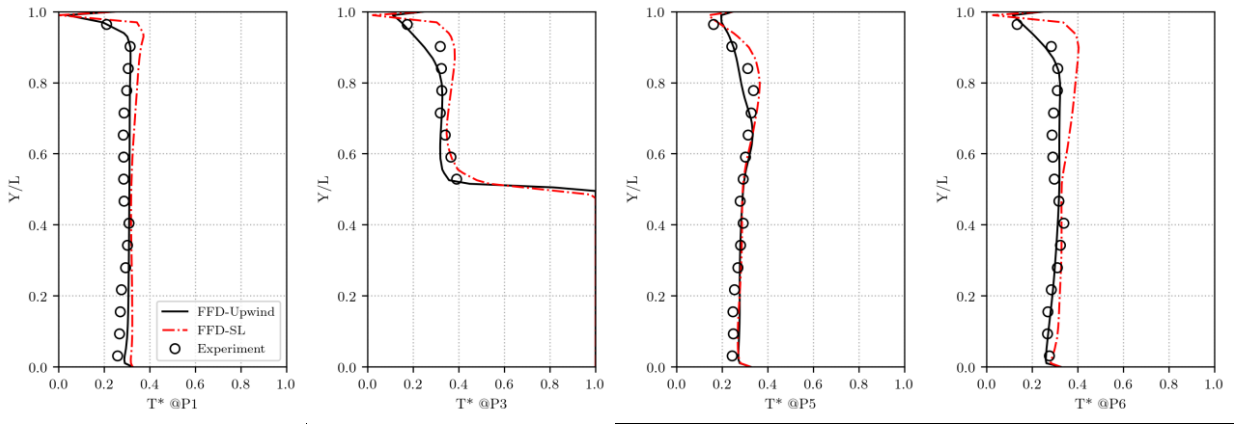


Figure 3-8 Comparison of temperature profiles for mixed convection in an empty room with a box

Figure 3-6 ~ 3-8 compare the velocity and temperature profiles predicted by FFD-Upwind and FFD-SL for the two cases. Generally, all agree with the experimental data [79] for most locations and FFD-Upwind performs slightly better than FFD-SL does. However, some obvious discrepancies can be observed. For example, velocity prediction at P5 by FFD-Upwind and temperature prediction at P6 by FFD-SL for the second case. Both FFD-Upwind and FFD-SL do

not predict the velocity at the areas close to the ceiling and floor precisely. The possible reason is that FFD could not properly estimate the turbulence viscosity due to lack of wall functions even though an approximate wall function [97] is integrated with the zero-equation turbulence model in the FFD models. It is also found that FFD-SL performs better than FFD-Upwind at those areas especially for the location P5 even though FFD-Upwind has a slightly better overall accuracy. This is possible since the prediction accuracy may be influenced by aspects other than the solving methods of FFD, such as configurations of the mesh, settings of physical parameters. It is worthy to note that relatively larger discrepancies in some critical areas were also found in previous studies [80, 98] and even state-of-the-art CFD models with advanced turbulence models could not precisely predict all the flow details [79]. As a simplified alternative to CFD models that targets fast speed, it is acceptable that the proposed FFD model could generally capture the flow dynamics.

For the speed, it takes 244.1 s and 244.9 s for the two cases with FFD-Upwind and takes 260.3 s and 258.0 s for the two cases with FFD-SL. This is because of the new methods to solve equations in the FFD-Upwind as shown in *Figure 3-2*. In FFD-SL, the advection equation is solved with an SL scheme and the diffusion equation is solved with an upwind scheme. In FFD-Upwind, the advection and diffusion equations are solved together with an upwind scheme. As a result, the computing demand in FFD-Upwind is reduced by eliminating the SL scheme. Hence, FFD-Upwind is faster than FFD-SL. To conclude, FFD-Upwind is about 5.7% faster than FFD-SL while achieving a slightly better overall accuracy compared to FFD-SL for the two studied cases.

3.3.2 Validation of New Features for Data Centers

The methodology section introduced the new features in the FFD model for data centers. This section is to validate the new FFD model using a real data center for industrial practice. It is

noted that various research has been done to validate the component models for data centers, such as the black box model [90, 91] and body force method [87]. The focus of this paper is to evaluate if the new FFD model can provide airflow prediction for real-world practice with the current implementation, such as zero-equation turbulence model, body force model and black box rack model. Therefore, we select a middle-size data center in real world instead of a simple data center with well-controlled boundary conditions in the laboratory.

3.3.2.1 Description of the data center case

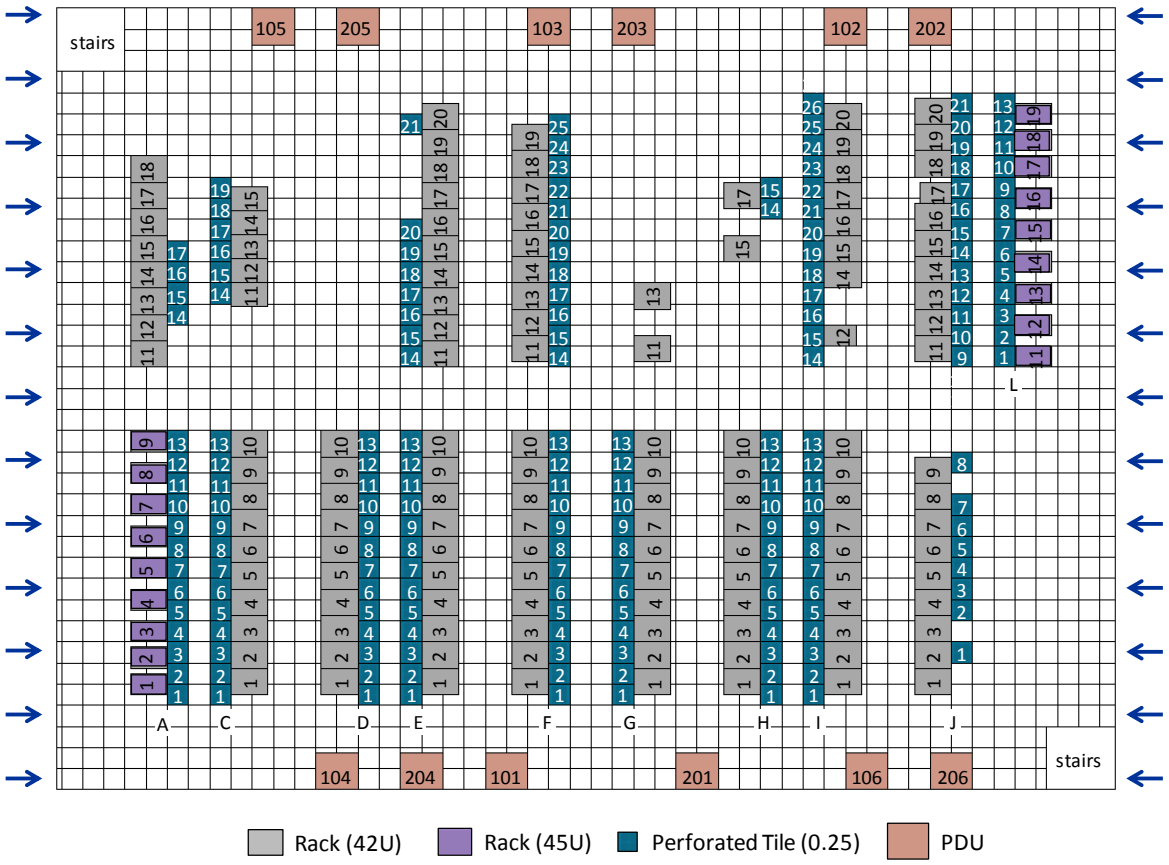


Figure 3-9 Layout of the reference data center

In this section, the new FFD model with the new data center feature is validated using a real medium-size data center with raised-floor and dropped-ceiling architecture located in Massachusetts, U.S.A. As shown in *Figure 3-9*, the reference data center is approximately 30.5 m (100 ft) long, 22.6 m (74 ft) wide and 3.4 m (11 ft) high from the raised floor to the ceiling, with a total white space area of approximately 690 m^2 (7,400 ft^2). Total power consumption by 151 racks and 12 PDUs is approximately 344 kW. Racks G11 and G13 (i.e., the 11^{th} and 12^{th} cabinets in Row G) are empty. There are 18 45U networking racks in Rows 1 and 10, and all the remaining racks in the data center have a capacity of 42U.

Two central Air Handling Units (AHUs) supply cooling airflow through the short sides of the plenum, which is supported using 7/8" stanchions. The total supply air flow rate is 152,000 m^3/hr (89200 cfm). The airflow is then supplied to the IT equipment through 183 perforated floor tiles with 25% open-area-ratio (each 2 ft by 2 ft). The hot exhaust air returns to the CRAH through a dropped-ceiling plenum with 42 perforated ceiling tiles with 83% open-area-ratio.

An on-site measurement was performed for the studied data center. In addition to the data center geometry and types of IT racks, three primary parameters are measured including 1) rack-by-rack powers, 2) perforated tile airflow rates and 3) rack inlet temperatures. Rack-by-rack powers are estimated by two steps. First, all the racks fed by twelve large power distribution units (PDUs) are divided into twelve groups and the total power of each group is equal to the corresponding PDU power. Then, rack-by-rack powers are estimated by scaling the total power in each group of racks by the fraction of occupied U spaces in each rack. A flow hood is used to measure the perforated tile airflow rates. The flow hood is a TSI/Alnor EBT731 and the accuracy of the flow hood is ±3% of reading and ±12 m^3/h at flows >85 m^3/h. When being placed on a

perforated tile, the flow hood adds additional flow resistance which creates a measurement error relative to the true value. To correct for this error, the "2MUP(p=2)" method proposed by VanGilder, Pardey, et al. [99] is adopted. The rack inlet temperatures are measured at 4 points distributed vertically in front of each rack. The sensors are located at heights of 0.53 m, 0.91 m, 1.3 m and 1.68 m, respectively. The horizontal distance between the thermocouples and the rack door is approximately 0.025 m. Four K-type thermocouples are connected to a UEi-DT304 Quad Input IP67 Digital Logging Thermometer with the accuracy of ±[0.1% + 0.5 ^{o}C]. Temperature measurements are averaged over a 1-minute sampling period.

3.3.2.2 Evaluation metrics

To evaluate accuracy of plenum airflow modeling we propose the percentage relative difference (PRD) between simulated tile airflow $\dot{Q}_{sim}$ and experimentally measured tile airflow $\dot{Q}_{exp}$, which is defined as:

$$PRD_{\dot{Q}} = \left|\frac{\dot{Q}_{sim} - \dot{Q}_{exp}}{\dot{Q}_{exp}}\right| \times 100\%. \quad (3\text{-}14)$$

To evaluate accuracy of temperature prediction in the white space, PRD_T is used to represent the PRD between simulated and experimentally measured rack inlet temperatures (T_{sim} and T_{exp}):

$$PRD_T = \left|\frac{T_{sim} - T_{exp}}{\Delta T}\right| \times 100\%, \quad (3\text{-}15)$$

where T_{sim} and T_{exp} are calculated as the average of the temperature at four points along the heights. The reference temperature difference ΔT is assumed to be 14 ^{o}C (25 ^{o}F), which corresponds to a typical magnitude of temperature rise across the racks. In this case, a 5% of PRD_T corresponds to a temperature difference of 0.7 ^{o}C.

3.3.2.3 Setup of simulations

Table 3-3 summarizes the simulation settings of the FFD model and CFD. Both use structured grids for the simulation of plenum and white space. The CFD does not require a time step size or a set simulation time since it performs steady-state simulations, while the FFD model performs transient-state simulations because of the time-split method used in FFD (with a time step size of 0.2 s – 1.0 s and simulation time of 100s – 400s). The CFD uses a $k-\varepsilon$ turbulence model with wall function treatment while the FFD model uses a zero-equation turbulence model [96] with a simplified approximate-wall-function proposed by Dhoot, VanGilder, et al. [97]. The CFD simulation is performed on four cores of an Intel (R) Xeon (R) CPU E3-1220 v6 processor running on a Windows workstation with 32 GB of RAM. The FFD simulation is performed on an AMD FirePro™ W8100 GPU. Except for the refined cells at some critical locations to handle complex geometries, a uniform and structured grid with the mesh size of 0.15 m (6 inches) is adopted for this case in both the FFD and CFD models. The grid independent studies show that the results do not change significantly when the mesh size is below 0.15 m (6 inches) in both models. Similar conclusions were drawn in VanGilder and Zhang [100], which studied ten data center layouts with six grid levels each.

Table 3-3 Settings of simulations in FFD and CFD

Case	**Grid**		**Simulation Time (s)**		**Time Step Size (s)**	
	FFD	CFD	FFD	CFD	FFD	CFD
Plenum	224×161×6	224×161×6	100	N/A	1.0	N/A
White space	200×148×22	200×148×22	400		0.2	

FFD and CFD models use the same settings for the boundary conditions. The walls are assumed to be adiabatic since the heat dissipation from the IT equipment is much larger than the heat flux through walls. The leakage is neglected in the CFD/FFD models as the total raised floor leakage airflow for this data center is negligible according to Pardey, VanGilder, et al. [92]. In the plenum model, the inlet bays are modeled with Dirichlet boundary conditions and the velocities are assigned with measured data. The outlet, i.e. floor perforated tiles, are modeled as two-dimensional flow resistances. In the whitespace model, the perforated tiles become inlets, which are modeled with Dirichlet boundary conditions with air flow rates and temperatures equal to the measurements. An additional momentum force is added to the cells above the floor perforated tiles to correct the under-estimated velocity due to the usage of fully-opened openings to model the perforated inlets in the CFD and FFD models. The outlets of the whitespace, i.e. ceiling perforated tiles, are simplified with fixed flow. The IT racks are modeled as a "black box" [90]. The server flow rates are assumed to be 125 cfm/kW [92]. The server powers are assigned based on the on-site audit results of the rack-by-rack powers.

It is noteworthy that the dropped-ceiling plenum is excluded from the FFD and CFD modeling for modeling convenience and simplicity. We perform CFD simulations to determine the effect of removing the ceiling plenum and find that the temperatures at the rack inlets are not affected. The perforated ceiling tiles are treated as fixed-flow outlets. In the plenum models, the measured supply air flow rates at all side inlets are used as model inputs. In white space models, the measured flow rates and temperatures at perforated floor tiles are used as model inputs.

3.3.2.4 Results

The results of the plenum simulation are depicted in *Figure 3-10.* The predictions of air flow rates through the 183 perforated tiles in the reference data center by FFD and CFD are categorized by PRD from experimentally measured data. Both FFD and CFD has 95.1% of their predictions with a PRD less than 5% from the experimental measurements. Only 7 (3.8%) predictions from FFD and 8 (4.4%) predictions from CFD are within the PRD of 10-20% from experimental measurements. There are only two predictions (1.1%) from FFD and one prediction (0.55%) from CFD with deviation of more than 20% from the experimental measurements.

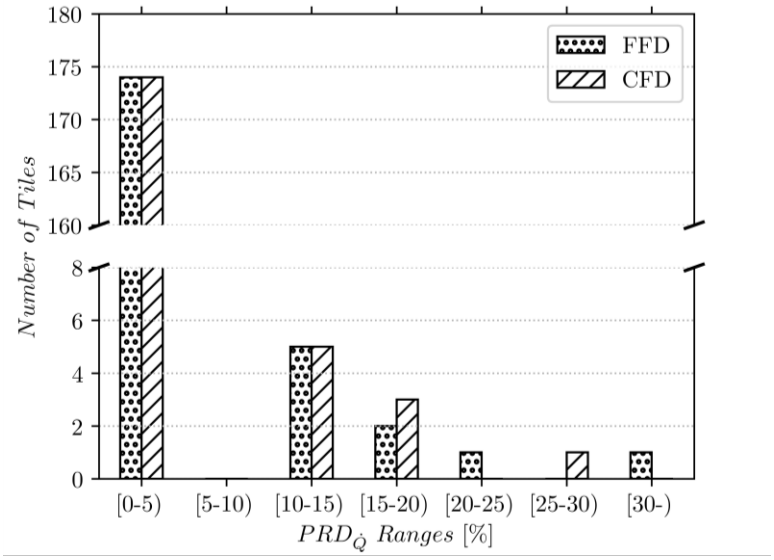


Figure 3-10 Predictions of perforated-tile air flow rates categorized by percentage relative difference (PRD) from measurements

The results of the white space simulations are shown in *Figure 3-11*. The predictions of rack-inlet temperatures at 149 racks (excluding 2 empty racks) in the reference data center by FFD and CFD are categorized by PRD from experimental measurements. 131 (88.0%) predictions from

FFD and 128 (85.9%) predictions from CFD deviate by less than 10% from the experimental measurements. 14 (9.4%) predictions from FFD and 16 (10.7%) predictions from CFD are within the 10%-to-20% relative difference range from experimental measurements. 4 (2.7%) predictions from FFD and 5 (3.4%) predictions from CFD deviate by more than 20% from the experimental measurements.

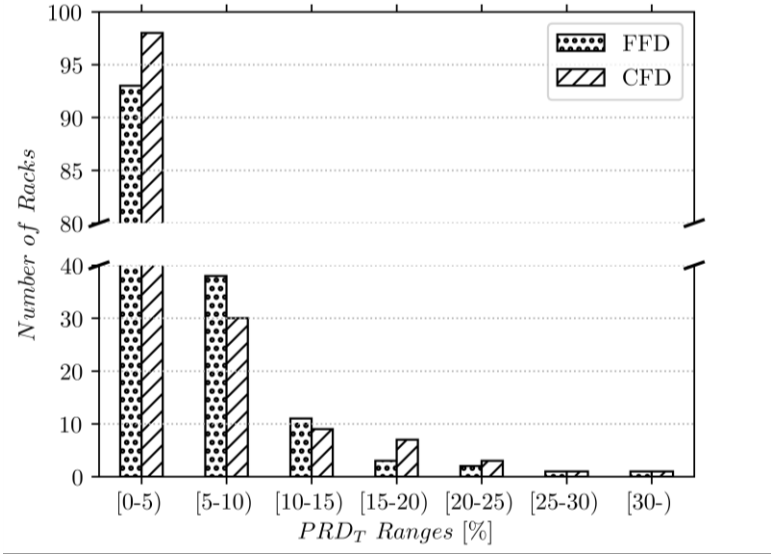


Figure 3-11 Predictions of rack-inlet temperatures categorized by percentage relative difference (PRD) from measurements

Generally, both FFD and CFD capture the rack-to-rack inlet temperature variation well. However, some discrepancies between simulated and measured temperatures can be observed for both FFD and CFD. The reasons may be as follows. First, the dynamic changing of the cooling system might affect the measured air flow rates and temperatures at perforated tiles, given the measurement usually takes several hours. Second, we did not have all required information about this real data center for the models, such as the leakage path through openings in or under the rack

or floor cracks. Moreover, some assumptions in the FFD and CFD models might not be able to represent the real situation. For instance, individual-rack power consumption is crudely estimated based on rack IT population and power measurements from PDUs, which served many racks. The rack airflow is, in turn, crudely estimated to be 212 m^3/h (125 cfm) of airflow for each kW of power dissipation. Furthermore, additional uncertainties may also lie in the measurements since it is very hard to perfectly measure a real data center in operation, which was discussed in [99]. Even so, both FFD and CFD successfully predict the inlet temperature for more than 85% of the racks with a less-than-10% error. To conclude, the proposed FFD model achieves a similar level of sufficient accuracy compared to CFD for the studied case.

For the simulation speed, CFD takes 10.3 minutes and 65.6 minutes while FFD takes 0.17 minutes and 1.07 minutes to simulate the raised-floor plenum and white space, respectively. FFD is approximately 61 times faster compared to CFD. It is noted that the present solution time is obtained based on the current settings of the models for the reference data center and computer configurations. CFD runs on four cores of CPU while FFD runs on a GPU in parallel.

3.4 Case Studies

In this section, the usage of FFD model for data center design and operation is demonstrated by applying it to improve data center thermal management through model-based design and operation. These include 1) an optimal design of data center plenum and perforated tiles; 2) an optimal design of the data center cooling system regarding fan sizing and designed supply air temperature through parametric studies; 3) an optimal operation of the cooling system by optimizing setpoints of supply air flow rate and temperature based on the weather conditions.

3.4.1 Optimal Design of the Plenum and Perforated Tiles

The design of a data center normally involves the determination of several key parameters, such as layout of the IT racks, width of cold aisles and hot aisles, plenum depth and perforated tile open-area-ratio.

In the early design stage, the decision should be made based on the cooling performance, initial cost and operational cost. The initial cost could be roughly estimated using existing database and engineering experience, but it is usually hard to get the quantitative evaluation of the cooling performance and operational cost. CFD has been widely used to do that, but it is limited to evaluating a few scenarios instead of a systematic parametric study of multiple parameters due to its high computational demand.

In this case, we demonstrate the usage of the FFD model for the optimal design of data centers, in which the effect of the design of plenum and perforated tile on the air distribution uniformity is investigated.

3.4.1.1 Evaluation metrics

A real data center can have hundreds of perforated tiles or more. Here we assume the design purpose is to uniformly distribute air flow to perforated tiles. To evaluate the uniformity of the airflow distribution among many perforated tiles, we propose a percentage relative difference from the mean air flow rate of all perforated tiles:

$$PRD_{i,\dot{m}} = \frac{\dot{m}_i - \bar{\dot{m}}}{\bar{\dot{m}}} 100\%, \tag{3-16}$$

where $\dot{m}^i$ is the mass flow rates at perforated tiles i. The n is the total number of perforated tiles. $PRD_{\dot{m}}^i$ represents the percentage relative difference from the mean air flow rate of all perforated

tiles at the perforated tile *i*. Ideally, if the cold air is distributed uniformly to each perforated tile, the $PRD_{\dot{m}}^{i}$ at all perforated tiles should be 0%. If the distribution is non-uniform, we can quantify the non-uniformity by evaluating the values and distribution of $PRD_{\dot{m}}^{i}$ of all perforated tiles.

3.4.1.2 Setup of the case

Table 3-4 Parameter settings in the case for optimal design of the plenum and perforated tiles

Parameters	Unit	Values
Plenum depth	mm	305, 457, 610, 762, 914
Tile open-area-ratio	%	15, 25, 35, 45, 56

As shown in *Table 3-4*, the studied parameters include the plenum depth and perforated tile open-area-ratio. The plenum depth ranges from 305 mm to 914 mm with an interval of 52 mm. The perforated tile open-area-ratio ranges from 15% to 56%. Among them 25% and 56% are commonly used in practice and the others are selected with an interval of 10%. As a result, there are a total of 25 cases for the parametric study. All other parameters are determined based on the operational data. The supply air temperature is 22 ^{o}C, which is calculated by averaging measured supply air temperatures. The supply air flow rate is 1.5×10^{5} m^{3}/h, which corresponds to an air ratio of 2.07 calculated based in *Equation 3-17*. The air ratio is defined as the ratio of total supply air flowrate ($\dot{Q}_{sup}$) over total IT flowrate ($\dot{Q}_{in}$, refer to *Equation 3-11*):

$$Air\ Ratio = \frac{\dot{Q}_{sup}}{\dot{Q}_{in}} \tag{3-17}$$

3.4.1.3 Results

The results from parametric study are shown in *Figure 3-12*. The distribution of the $PRD_{\dot{m}}^{i}$ for all perforated tiles is depicted including maximum, minimum values and standard deviation.

Generally, a greater plenum depth helps improve the airflow uniformity. The influence is more pronounced when the plenum depth is smaller than 610 mm. In terms of the open-area-ratio of perforated tiles, a smaller open-area-ratio helps improve the airflow uniformity. Once again, the influence is more significant when the plenum depth is smaller than 610 mm. The reason is that a deeper plenum and/or more restrictive perforated tiles can lead to a more uniform pressure distribution in the plenum. When the depth of plenum is larger than 610 mm, the pressure distribution in the plenum is already fairly uniform, so further improvement for both parameters may not help as much compared to the cases with a smaller plenum depth. Assuming the design objectives are that maximum and minimum values of $PRD_{\dot{m}}^{i}$ are smaller than 3% and the standard deviation of $PRD_{\dot{m}}^{i}$ is smaller than 1%, the candidate combinations of tile open-area-ratio and plenum depth towards optimal design include 1) 15% + 457 mm and 2) 25% + 610 mm.

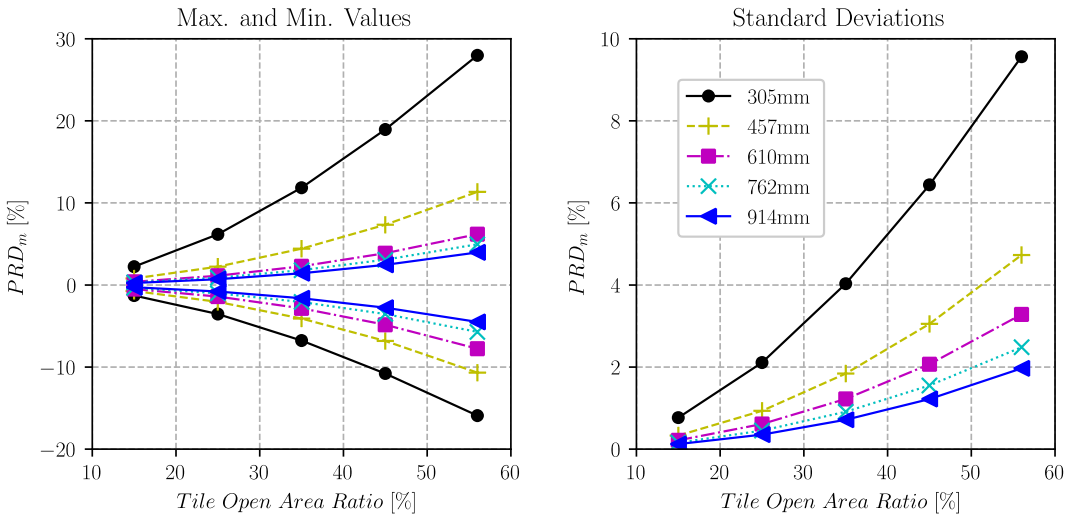


Figure 3-12 Effect of open-area-ratio of perforated tiles and plenum depth on airflow uniformity among perforated tiles

It is worth to point out that the decision in real world could be more complex. Although the airflow uniformity can be improved, a greater plenum depth may increase the capital cost of construction and a smaller open-area-ratio of perforated tiles may increase the operational cost due to a larger pressure drop through the tiles, which increases fan energy consumption. The fan energy can be estimated based on the flow rate and fan head pressure so that the operational cost for fan can be calculated. Therefore, the optimal design should balance the costs and performance. Such a parametric study in the early design stage using FFD simulations could provide quantitative suggestions towards an optimal design.

3.4.2 Optimal Design of the Data Center Cooling System

The previous sections demonstrated optimal design of the data center plenum and perforated tiles using the FFD model. This section demonstrates another optimal design use case of the FFD model, which is the optimal design of the cooling system in the same data center. This is done through a parametric study using FFD simulations, in which we focus on two parameters: fan sizing and designed supply air temperature.

3.4.2.1 Evaluation metrics

There are various metrics to evaluate the thermal environment of a data center, such as rack cooling index (RCI) [101], return temperature index (RTI) [102], supply heat index (SHI) and return heat index (RHI) [103] and capture index (CI) [104]. Among them we adopt the RCI [101], which quantifies the conformance with the data center thermal standards, e.g. ASHRAE thermal guidelines [15], based on the calculation of equipment intake temperatures. ASHRAE guideline categorizes data centers into multiple classes and recommends different thermal standards for different classes. The data center we studied is Class A, which has tightly controlled thermal

environment and mission critical operations. The RCI metric consists of two numbers: RCI_{HI} and RCI_{LO}. The RCI_{HI} can be written as:

$$RCI_{HI} = \left[1 - \frac{\sum_{x=1}^{n} f(x)}{n(T_{max-all} - T_{max-rec})}\right] 100\%, \tag{3-18}$$

$$f(x) = \begin{cases} T_x - T_{max-rec}; & T_x > T_{max-rec} \\ 0; & T_x \leq T_{max-rec} \end{cases}, \tag{3-19}$$

where, T_x is the mean rack-inlet temperature at Rack x; *n* is the total number of racks; $T_{max-rec}$ is the maximum recommended rack-inlet temperature (27 ^{o}C by [15]); $T_{max-all}$ is the maximum allowable rack-inlet temperature (35 ^{o}C by [15]). $RCI_{HI} = 100\%$ means no equipment intake temperature exceeds the maximum recommended value. The data center thermal environment can be regarded as "Good" when RCI_{HI} is larger than 96% and "Acceptable" when RCI_{HI} is in the 91%-to-95% range.

The RCI_{LO} can be written as:

$$RCI_{LO} = \left[1 - \frac{\sum_{x=1}^{n} g(x)}{n(T_{min-rec} - T_{min-all})}\right] 100\%, \tag{3-20}$$

$$g(x) = \begin{cases} T_{min-rec} - T_x; & T_x < T_{min-rec} \\ 0; & T_x \geq T_{min-rec} \end{cases}, \tag{3-21}$$

where, $T_{min-rec}$ is the minimum recommended rack-inlet temperature, which is 18 ^{o}C per ASHRAE [15]. $T_{min-all}$ is the minimum allowable rack-inlet temperature, which is 15 ^{o}C for Class A1 data centers per ASHRAE guideline. $RCI_{LO} = 100\%$ means no equipment intake temperature falls below the minimum recommended value (i.e. 18 ^{o}C per ASHRAE [15]). The data center thermal environment can be regarded as "Good" when RCI_{LO} is larger than 96% and "Acceptable" when RCI_{LO} is in the 91%-to-95% range.

Another metric used in our study is the maximum rack-inlet temperature, which can be written as:

$$T_{in}^{max} = \max_{x \in [1,n]} T_{in}^{x}, \tag{3-22}$$

where, T_{in}^{max} is the maximum rack-inlet temperature; T_{in}^{x} is the intake temperature at Rack x. The limit of T_{in}^{max} can be the maximum allowable rack-inlet temperature, which is 32 ºC for Class A1 data centers per ASHRAE guideline. Other limits of T_{in}^{max} can also be adopted, e.g. 27 ºC according to the requirement of the data center.

3.4.2.2 Setup of the case

Table 3-5 Parameter settings in the case for optimal design of the cooling system

Parameters	Unit	Values
Supply air temperature	ºC	15, 16, 17, 18, 19, 20, 21, 22, 23, 24
Supply air flow rate	m^3/h	3.65×10^4, 5.5×10^4, 7.3×10^4, 9.1×10^4, 1.1×10^5, 1.28×10^5, 1.46×10^5, 1.51×10^5
	Air ratio	0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.07

The settings of parameters for this case are shown in *Table 3-5*. The supply air temperature ranges from 15 ºC to 24 ºC with a 1 ºC interval. We define the air ratio ranging from 0.5 to 2.07. Therefore, there are 80 cases in total for the parametric study. It is noted that the air ratio of 2.07 is determined based on the current settings in the reference data center. Accordingly, the supply air flow rates are calculated for the prescribed air ratios. All other parameters are set according to current configurations in the reference data center.

3.4.2.3 Results

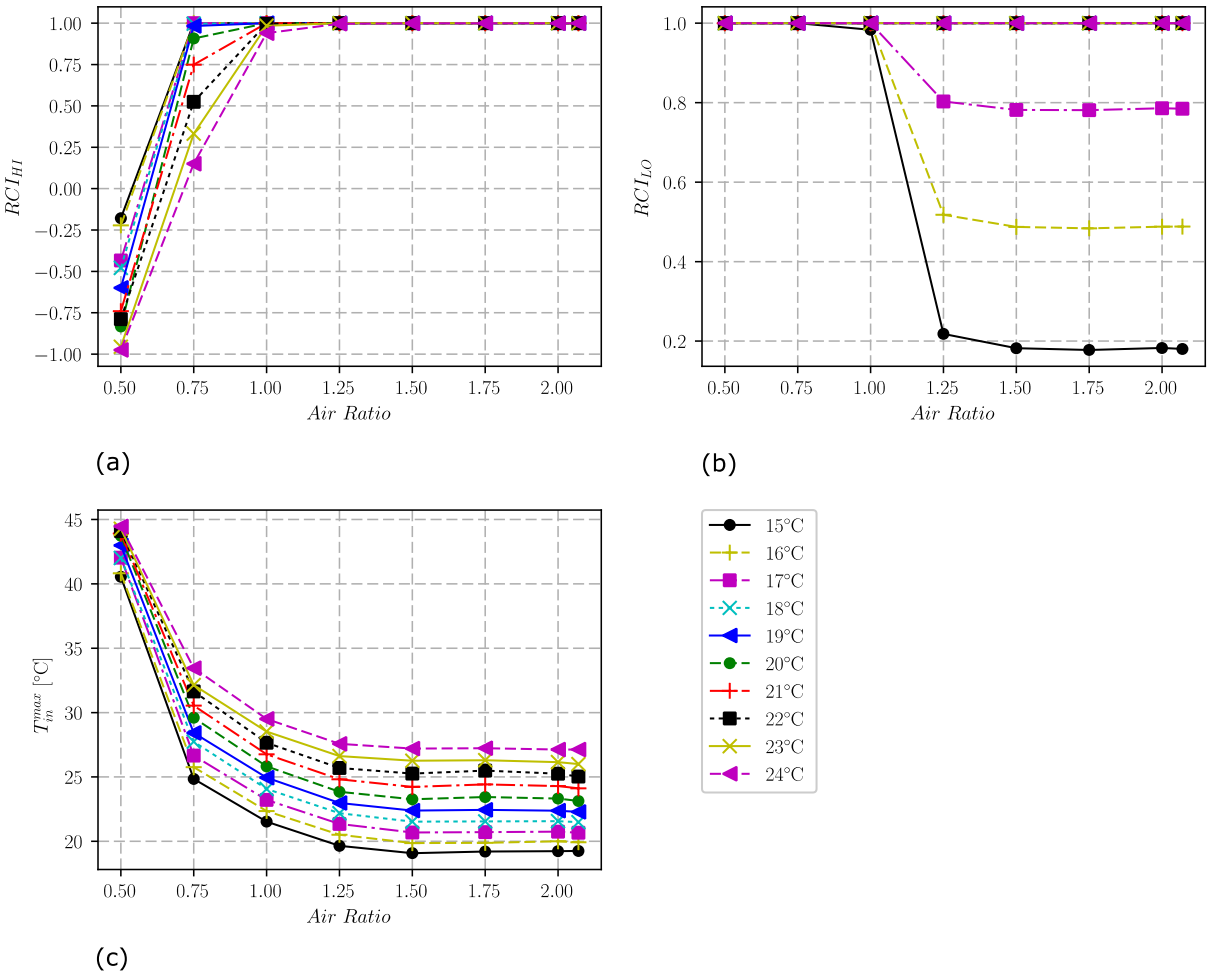


Figure 3-13 Effect of air ratio and supply air temperature on RCI_{LO}, RCI_{HI} and T_{in}^{max}

The effect of the two studied parameters on the cooling performance is investigated regarding three metrics including RCI_{LO}, RCI_{HI} and T_{in}^{max}. As shown in *Figure 3-13 (a)*, when the air ratio is smaller than 1.25, RCI_{HI} increases as the air ratio increases. When the air ratio is greater than 1.25, RCI_{HI} stays constant at 1.0. *Figure 3-13(b)* shows the effect on the RCI_{LO}. For cases with supply air temperature from 15 to 17 °C, RCI_{LO} starts from a constant value of 1.0 when the air ratio is smaller than 0.75 and then falls drastically when the air ratio is larger than 1.0. For other

cases, RCI_{LO} stays constant as the air ratio increases. The obvious turning points for both RCI_{HI} and RCI_{LO} start when the air ratio is around 1.0. As shown in *Figure 3-13(c)*, the T_{in}^{max} decreases as the air ratio increases. The curves under different supply air temperature follow a similar pattern, which become relatively constant when the air ratio is larger than 1.25. The maximum rack-inlet temperatures when reaching constant values are higher than the supply air temperatures by approximately 3 ºC.

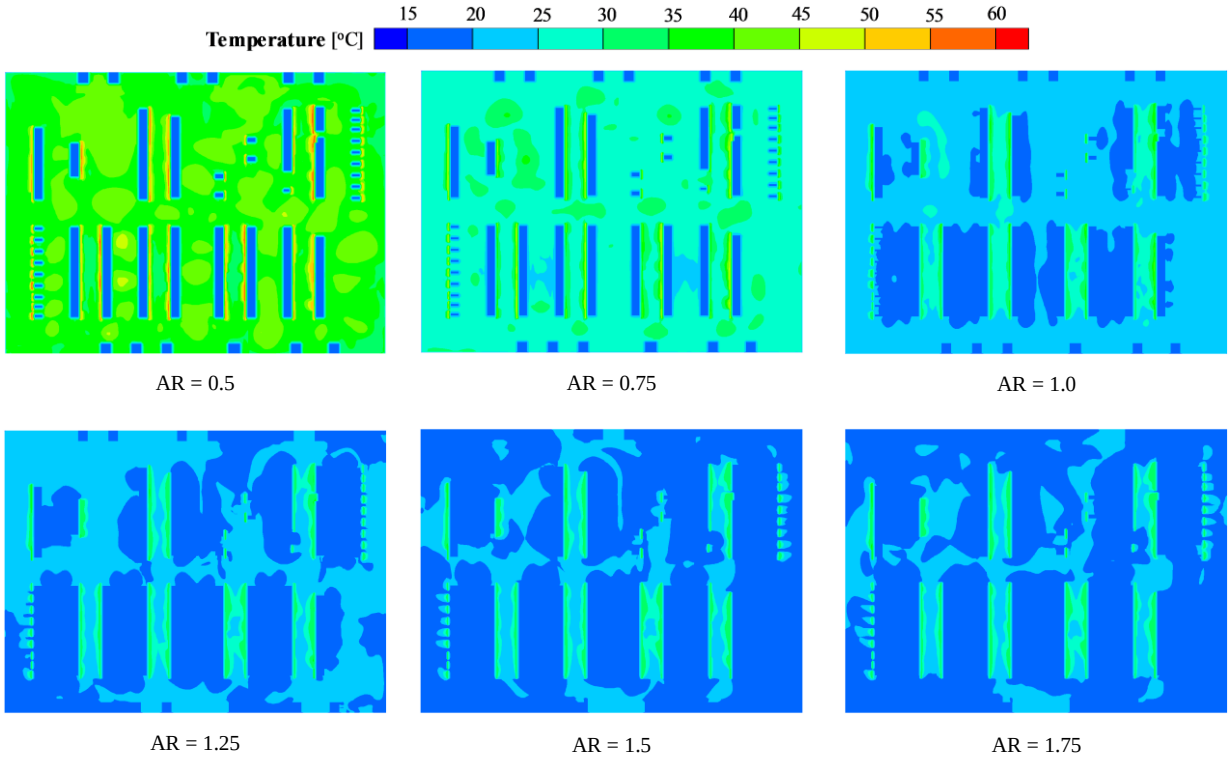


Figure 3-14 Temperature contours at 2/3-RACK height for different air ratios with supply air temperature of 16 ^{o}C

The maximum rack-inlet temperatures are generated at some critical locations, such as Rack-A17 and Rack-A18 (see their locations in *Figure 3-9*), where the cold supply air coming into the racks is richly mixed with hot room air due to lack of perforated tiles close to the racks. The

temperature contours at 2/3-RACK height for different air ratios with supply air temperature of 16 ºC are plotted in Figure 3-14. When the air ratio is less than 1.0, the cooling system supplies less cold air than required by the IT equipment (i.e. total IT server air flow rate), causing recirculation (i.e. mixing cold supply air with hot room air) in front of racks. As a result, the rack-inlet temperatures are higher than the supply air temperature and local hot spots occur if the supply air temperature is not adequately cold. Interestingly, the RCI_{HI} values are negative when the air ratio is 0.5. According to Equations 18 and 19, the RCI_{HI} will be equal to 0 when the inlet temperatures at all racks are the maximum allowable equipment intake temperature (32 ºC).

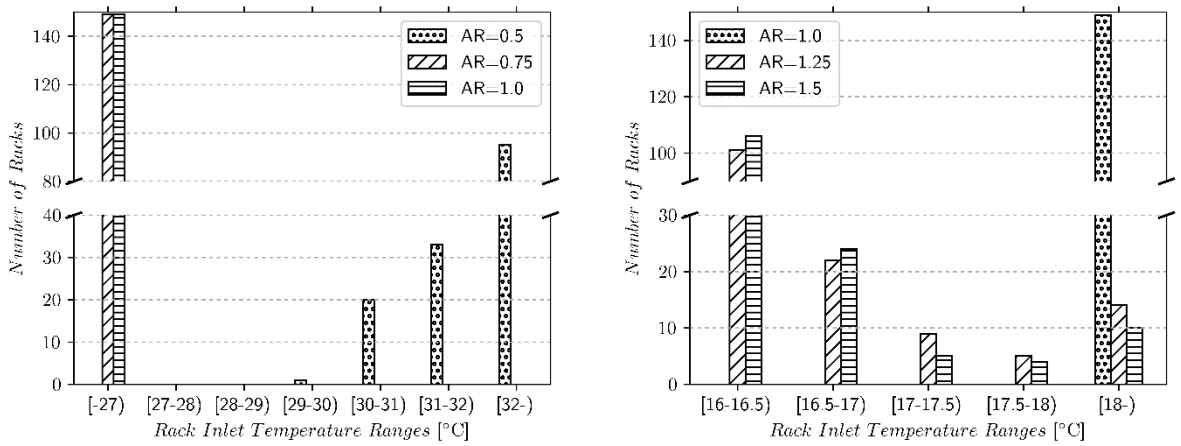


Figure 3-15 Number of racks with rack inlet temperature of different ranges for different air ratios with supply air temperature of 16 ºC

The numbers of racks with inlet temperatures in different ranges are shown in *Figure 3-15*. We find that 95 (63.8%) racks have inlet temperatures higher than 32 ºC and 53 (35.6%) racks have inlet temperatures in the 30-32 ºC range when the air ratio is 0.5, which may lead to the negative value of RCI_{HI}. To conclude, the less cold air the cooling system supplies or the higher the supply air temperature is, the more local hot spots exist and the smaller the RCI_{HI} will be.

When the air ratio continues to increase after 1.0, the rack-inlet temperatures approach closer to the supply air temperature. Therefore, the RCI_{HI} stays constant as 1.0, and RCI_{LO} falls drastically after the air ratio increases past 1.0. This effect is more significant when the supply air temperature is lower. When the supply air temperature is higher than 18 °C, the RCI_{LO} stays constant at 1.0 as the air ratio increases. When the supply temperature is lower than 18 °C, the values of RCI_{LO} stay at 0.2, 0.5, and 0.8 for supply air temperatures of 15 °C, 16 °C, and 17 °C, respectively. Take the case with supply air temperature of 16 °C as an example. Ideally, the rack inlet temperatures should be close to the supply air temperature when the supply air flow rate is large enough. The RCI_{LO} should be equal to 0.33 according to Equations 17 and 18 if the rack inlet temperatures approach close to the supply air temperature. However, there are still several racks with inlet air largely mixed with hot room air. For example, for the case with air ratio of 1.5, there are 10 racks with inlet temperatures that are larger than 18 °C (shown in *Figure 3-15*) including Racks A-11~12, A-16~18, E17~19, H-15 and J-01, which do not have perforated tiles close to them (see *Figure 3-14* and *Figure 3-9*). There are 33 racks with inlet temperatures that are in 16.5-to-18.0 °C range, such as Racks C-11, E-11 and J-2~3, which are located at the edge of each row or have relatively less cold air supply (see *Figure 3-14* and *Figure 3-9*). As a result, the RCI_{LO} is determined to be 0.5 instead of 0.33.

From the results of the parametric study, the optimal air ratio that balances thermal environment and fan energy ranges from 1.25 to 1.5, which corresponds to the fan sizing from 9.1×10^4 to 1.1×10^5 m^3/h. The corresponding optimal designed supply air temperature should be between 20 and 21 °C to create a similar thermal environment as baseline. Please note that there are multiple choices for the two studied parameters to create a similar thermal environment

regarding the three metrics, but the real optimal design should be determined through considering other aspects, such as the energy efficiency of the cooling system.

3.5 Discussions

3.5.1 Accuracy and Speed of the FFD Model

In the validation with two indoor environment cases, the FFD model does not predict the velocity at the areas close to the ceiling and floor precisely. In addition to lack of wall functions in the FFD model, the prediction performance at near-boundary areas may also be influenced by configurations of the mesh and settings of physical parameters. When the turbulent viscosity is estimated improperly, it may lead to improper prediction of the near-wall airflow and further influence the overall prediction performance. Hence, future research is needed to improve the prediction performance at near-boundary areas with the FFD model.

For the data center case, the FFD results are not perfectly consistent with the experimentally measured data. This may be due to various reasons. One possible reason is that there is some physical information (e.g. leakage through tile gap, cables or other obstacles in the plenum) at some local areas that the simulation models fail to capture. Another possible explanation is that the measurements are carried out over a period of hours, during which the airflow and temperature may vary because of control dynamics of the cooling system. Since the scope of this paper is to evaluate the possibilities to use the FFD model for data center thermal management, the validation results of FFD are acceptable since the proposed FFD achieves a similar level of accuracy compared to CFD. Note that the FFD model was originally applied to predict indoor environment with the purpose of fast speed. Even though it showed great success

in different applications [81, 39, 41, 82], the FFD model may not be suitable for applications that require strictly high accuracy without further improvements.

For the computational speed, the FFD model running on GPU is approximately 61 times faster than CFD running on a CPU with four cores. It is noted that the computing time obtained in Section 3.2.4 is highly associated with the computer configurations and convergence parameters, which are case-by-case. Even so, for the case we studied, the 61 times of the speed significantly reduces the computing time of the case study in Section 4. The total computing time of the parametric studies is about 7.6 hours, which is estimated to be 464.8 hours if CFD is used. The improvement from about 20 days to several hours makes model-based design and operation more feasible and practical.

3.5.2 Potential Applications Using the FFD Model

The FFD model is promising for applications with high computing demand due to its significant speedup compared to CFD. One application is optimal design of data centers, especially when multiple parameters should be considered in the early design stage, which requires lots of simulations. Another application is optimal operation of data centers. A data-driven model may be adopted to predict critical information of thermal environment since the physics-based model is not fast enough for real-time optimal control. If synthetic data is needed to train the data-driven model, FFD can accelerate this process by providing a dataset much faster than CFD does.

3.5.3 Future Work

The future work may include the following directions. First, more complex component models can be adopted in the FFD model to predict the complex data center thermal environment. This may impose negative influence on the speed and its necessity depends on the requirements of

applications. Second, a few improvement can be made to further accelerate the speed, such as optimizing the OpenCL codes and employing different discretization and solving methods. In addition, the prediction accuracy of the FFD model can also be improved by adopting more complex turbulence models and wall functions.

Another area of further work could be simultaneously modeling the plenum and white space since these two are modeled separately in our current FFD model. There are generally two approaches to achieve that. The first is a co-simulation-based approach where the white space and plenum are simulated separately in parallel. At each time step of the two simulations, data is exchanged between the two simulations. For example, the airflow rates at perforated tiles calculated in plenum model are sent to the white space model, and the pressure distributions above the floor tiles calculated in the whitespace model are sent to the plenum model. The two simulations will be processed until both converge. The other option is an integrated approach, in which the white space and plenum are modeled as one space, and the perforated tiles are modeled as resistances in the Navier Stokes equation. The treatment of the perforated tiles will be different in this approach. Instead of inlets or outlets, the perforated tiles will be treated as momentum sources within the modeling space.

3.6 Conclusion

The new FFD model is first introduced including governing equations, new methods to solve these equations, treatments of special boundary conditions in data centers and the implementation. The new model is then validated with two classical cases for indoor environment modeling and the results show that it achieves better accuracy and faster speed compared to conventional FFD. It is also observed that both FFD models achieve acceptable accuracy, except

for a few localized disparities with experimental data, which might be due to simplified handling of turbulent viscosity near the boundaries. It is also validated with a real medium-size raised-floor data center located in Massachusetts, U.S.A. The results show that the proposed FFD model running on GPU can achieve a similar level of accuracy while being much faster compared to a CFD model running on a CPU with four cores for the studied case. It is worthy to note that some discrepancies between simulation and measurement can still be observed in the data center case. This is acceptable given that the scope of this paper is to develop an open source, adequate and fast alternative to CFD. Subsequently, the FFD model is demonstrated to optimize the design of data center plenum and perforated floor tiles as well as the design of the cooling system through parametric studies. Quantified results are obtained regarding the effect of perforated tile open-area-ratio and plenum depth on the uniformity of airflow among perforated tiles and the effect of supply air temperature and flow rate on the cooling performance, which can be used to improve the design.

With a much faster speed than traditional CFD, the FFD model is promising for carrying out practical model-based design and operation to improve data center thermal management. Even so, future work is still needed to continue to improve the speed for applications such as on-line optimal control. In addition, the prediction accuracy of the FFD model can also be improved by adopting more advanced turbulence models.

Chapter 4 An Online BES-ROM-CFD Co-Simulation Method for Fast Simulation of HVAC Systems with Non-Uniform Thermal Environment

This chapter presents a new method for fast and accurate co-simulation of building energy simulation with indoor environment modeling.

4.1 Introduction

Ventilation systems involving stratified airflow and temperature distributions are designed to create a comfort thermal environment for occupants more effectively and efficiently. The typical applications are stratum ventilation in large spaces, ventilation in spaces with intense heat generation, displacement ventilation and natural ventilation [46, 45, 47]. Coupled simulation of Building Energy Simulation (BES) and computational fluid dynamics (CFD) has been used for studying such systems because of CFD's great capability to model complex airflow and temperature distributions [46]. As shown in *Figure 4-1*, the BES-CFD co-simulation and conventional BES-ROM-CFD co-simulation are widely used in existing research. Here the ROM refers to the reduced order model. The online BES-ROM-CFD co-simulation is a new methodology proposed in this paper.

The BES-CFD co-simulation directly couples BES with CFD during the online stage. The major limitation of this approach is the high computational demand due to the nature of CFD as pointed out in a recent review paper [66]. As a result, previous research only performed BES-CFD co-simulation for a short period, such as typical days, instead of a whole year. For example, Srebric, Chen, et al. [59] developed a coupled airflow-and-energy simulation program to study a

house (6 m × 4 m × 2.5 m) and an atrium (7 m × 4.3 m × 4.5 m), and found that it took about three hours to simulate the two cases for a typical day.

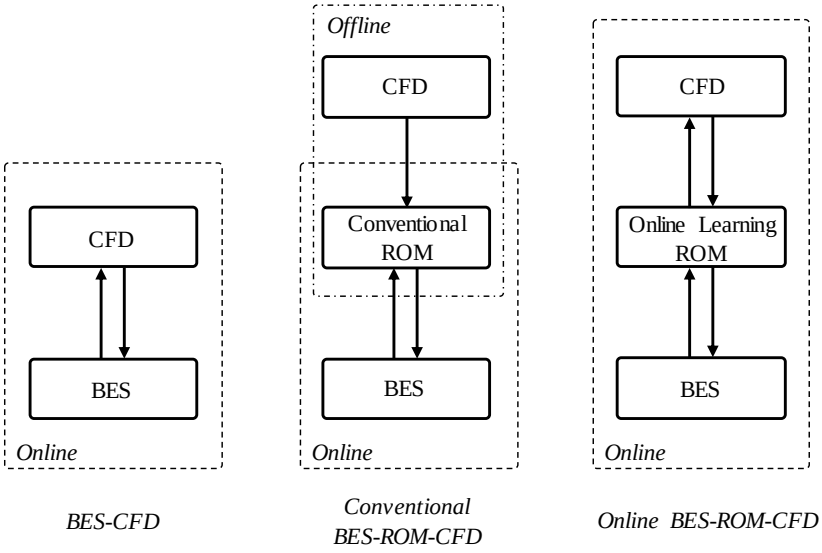


Figure 4-1 Different coupling strategies for co-simulation of BES and CFD

The conventional BES-ROM-CFD co-simulation (also called bin coupling or virtual dynamic coupling [58]) is proposed to address the limitations of BES-CFD. It pre-trains a ROM by CFD simulations on offline stage, then its BES calls the ROM on the online stage. This approach significantly reduces the simulation time during the online stage compared to the BES-CFD co-simulation Zhang, Lam, et al. [105]. The ROM can be curve-fitted functions [58], linear time-invariant (LTI) model [69], proper orthogonal decomposition (POD) method [70], and artificial neural network [71]. However, the ROM model prediction may not be accurate when the model inputs are outside of the training domain. Increasing the training domain may address this issue, but it also significantly increase the time needed for generating the training data using CFD simulations. In addition, it is also difficult to control the ROM model prediction accuracy. For example, Wei, Zhang, et al. [72] found that the representativeness of the data samples from CFD

simulations is critical for the accuracy of the POD interpolation method and multiple iterations may be needed to identify sufficient data samples. As a result, the training of ROMs often needs expert knowledge and requires several iterations to achieve a desired prediction accuracy.

An online learning model is promising to address the limitations in conventional BES-ROM-CFD approaches. In conventional BES-ROM-CFD co-simulation, once the training is completed, the ROM model will not be updated when being used during the online stage. Conversely, an online learning model can be continuously updated during the simulation. This study therefore proposes a new online BES-ROM-CFD co-simulation method that integrates an online learning ROM. The online learning ROM returns prediction results from either regression or executing CFD simulations depending on if the ROM can make accurate predictions through regression.

In comparison with the conventional BES-ROM-CFD co-simulation, several advantages can be expected with the proposed online BES-ROM-CFD co-simulation. First, in the online BES-ROM-CFD co-simulation, the ROM model can provide accurate predictions using the CFD when the inputs of the ROM are outside of the training domain and train itself using this newly generated data. Second, the online learning ROM can adopt error control algorithm, which automatically estimates the error between the ROM and CFD prediction and determines the training needs. Therefore, we can achieve a desired prediction accuracy automatically without expert knowledge from users and multiple iterations of training and testing. Moreover, the training process in the online BES-ROM-CFD co-simulation can be very efficient, since it only focuses on the required sets or segments of ranges of model inputs that occur during the simulation, rather than comprehensively sweeping the entire training domain in the conventional BES-ROM-CFD co-

simulation. In summary, the proposed online BES-ROM-CFD co-simulation is promising to achieve a fast and accurate co-simulation of BES and CFD in an efficient and automatic manner.

To realize the idea of the online BES-ROM-CFD co-simulation, one challenge is the error control of the online learning ROM. In this paper, an In Situ Adaptive Tabulation (ISAT) algorithm is adopted to achieve that. ISAT is an algorithm that approximates nonlinear functions based on local linear approximations from a dynamically adaptive lookup database [53]. An error control algorithm is embedded in ISAT to ensure the prediction error within a user-defined tolerance. Some attempts have been made to use ISAT for fast indoor airflow simulation [106] and optimization of data center workload distribution [107]. This paper further integrates ISAT into BES-CFD co-simulation by implementing a Modelica-ISAT-FFD model, in which a Modelica modeling language [108] is used for BES, ISAT is for online learning ROM, and Fast Fluid Dynamics (FFD) [11] is for CFD. Our work differs from the previous work [106, 107] as they only focuses on indoor airflow prediction and our work extends their work to co-simulation between BES and CFD. In addition, their work adopts a traditional training process, in which ISAT is first trained on the training stage and then used on the evaluation stage. Our work proposes an online learning approach, in which ISAT is trained on the online stage without any pre-training.

The rest of the paper is organized as follows. First, the online BES-ROM-CFD co-simulation method and the implementation of a Modelica-ISAT-FFD model to demonstrate the new method are introduced. Then, the Modelica-ISAT-FFD model is verified and evaluated by comparing it with an existing Modelica-FFD model, which was developed and validated by Zuo, Wetter, et al. [47]. Finally, an annual simulation is performed to demonstrate the performance of the Modelica-ISAT-FFD model for long-term co-simulation.

4.2 Methodology

4.2.1 Modelica-ISAT-FFD Model

To demonstrate proposed online BES-ROM-CFD co-simulation approach, we implement a Modelica-ISAT-FFD model, in which Modelica is selected for BES and FFD is selected for CFD. It is noteworthy that the online BES-ROM-CFD model is not limited to be applied with Modelica and FFD, it can also be utilized to realize the online learning approach for other BES-ROM-CFD models. For example, EnergyPlus-ISAT-Fluent or TRNSYS-ISAT-OpenFOAM can also be applied.

Modelica is an equation-based, object-oriented language that can be used for modeling multi-domain complex systems [108]. In this study, BES is performed through modeling based on the Modelica *Buildings* library, which is an open-source library that contains various modular component and system models for building energy and control systems [12]. In addition to the conventional BES, the Modelica supports rapid prototyping, modeling of heating, ventilation and air-conditioning (HVAC) systems, development and verification of control algorithms, and emulation of the faults at the building-system level [12, 10, 109].

FFD is an alternative to traditional CFD, which solves the same set of governing equations [39]. The primary difference between traditional CFD and FFD is the technique used to solve the governing equations. While traditional CFD commonly uses variants of the Semi-Implicit Method for Pressure Linked Equations (SIMPLE) [37], FFD uses a fractional time method (also called time split method), which breaks the momentum and energy equations into two or three parts and solves them sequentially [11].

In the Modelica-ISAT-FFD model, the ISAT dynamically establishes a lookup table, based on which a new query point (inputs of the model) can be quickly evaluated by local linear approximations [53]. Specifically, during the co-simulation, the ISAT stores FFD results in a database and retrieves predictions from the database if the inputs of the model are within the ranges where a specified error tolerance can be met. If not, the ISAT will execute the FFD simulation to make predictions and update records in the database. The key feature in the ISAT to automatically control the prediction accuracy is the Ellipsoid of Accuracy (EOA) of records (data points) in the database [53]. When a new query point is within the EOA of the nearest record in the database, the prediction can be obtained from a local linear approximation with confident accuracy. Otherwise, the prediction will be made from executing a FFD simulation. The EOAs of records in the database are continuously updated with new data fed to the ISAT model. For more details, please refer to [53, 106, 107].

4.2.2 Coupling Strategies

The workflow and data exchanges of the Modelica-ISAT-FFD are shown in *Figure 4-2*. The coupling process follows a cross quasi-dynamic coupling scheme, which is regarded to be a balance of accuracy and computational speed [66, 68]. The coupling of Modelica with ISAT and FFD follows a master-slave method [47], in which BES is the master, ISAT is the slave and FFD is executed by ISAT when needed. The simulation period of the co-simulation and next data synchronization point are determined in the Modelica. The Modelica performs transient simulations with a time step size of Δt which can be constant or variable depending on the Modelica. The Modelica exchanges data with the ISAT with a time step size of Δt_{syn} (data synchronization time), which can also be fixed or variable.

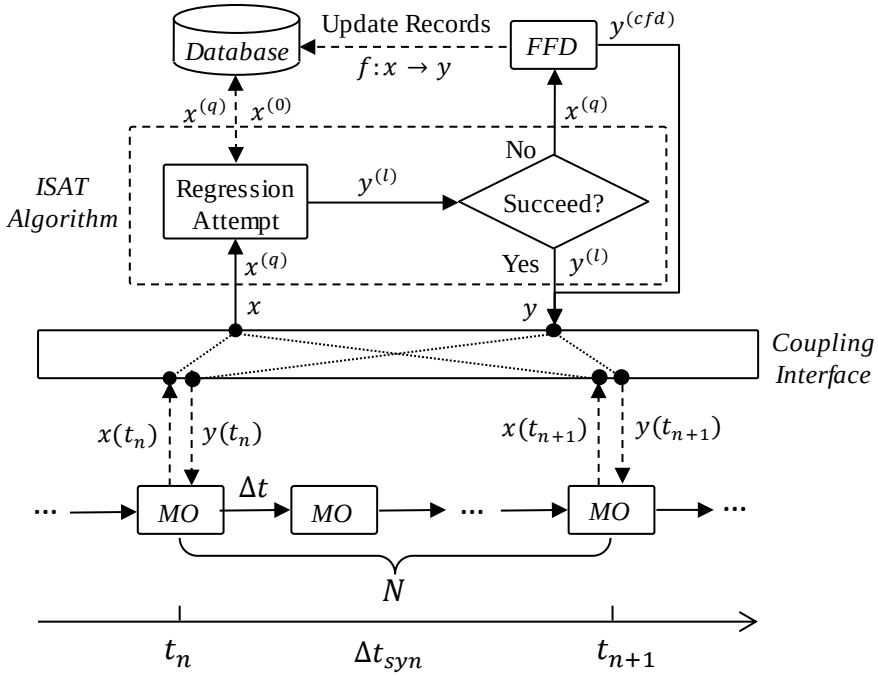


Figure 4-2 Workflow and data exchanges of the Modelica-ISAT-CFD model

The detailed procedures of the coupling process are explained as follows:

Step 0: Initialize Modelica, ISAT and FFD programs.

Step 1: Run Modelica, hold ISAT and FFD

Step 2: If the synchronization point t_n is not reached, continue Modelica simulation and compute states of exchanged variables $\hat{x}$:

$$\hat{x}(t_{n-1} + i\Delta t) = f_{MO}(\hat{x}(t_{n-1} + (i-1)\Delta t), y(t_{n-1})), \tag{4-1}$$

where, $i \in \mathbb{N}$, $i < N$ and f_{MO} represents the nonlinear functions solved in Modelica.

Step 3: If the synchronization point t_n is reached, hold Modelica simulation, compute time-averaged values of exchanged variables, and send $x(t_n)$ from Modelica to the ISAT as a new query point $x^{(q)}$.

$$x(t_n) = \frac{1}{\Delta t_{syn}} \int_{t_{n-1}}^{t_n} \hat{x}(t)dt, \tag{4-2}$$

$$x(t_n) \to x^{(q)}, \tag{4-3}$$

Step 4: Regression attempt: the ISAT searches the nearest record to $x^{(q)}$in the database (assuming it is $x^{(0)}$). It then attempts to estimate the output of the nonlinear function with regard to the query point $x^{(q)}$ using a linear approximation based on the record $x^{(0)}$:

$$y^{(l)} = f\left(x^{(q)}\right) = f\left(x^{(0)}\right) + A(x^{(0)}) \cdot (x^{(q)} - x^{(0)}), \tag{4-4}$$

where, $y^{(l)}$ represents the output of the nonlinear function from a local linear approximation. $f\left(x^{(0)}\right)$ is the mapping $f: x \to y$ with regard to $x^{(0)}$ and $A(x^{(0)})$ is the mapping gradient matrix with regard to $x^{(0)}$, which are stored in the database.

Step 5: If the new query point $x^{(q)}$ is within the EOA of the record $x^{(0)}$ in the database, the regression attempt succeeds, then directly returns the output (*Retrieve*) and proceeds to Step 7.

$$y^{(l)} \to y(t_n), \tag{4-5}$$

Otherwise, go to Step 6.

Step 6: Execute a FFD simulation and compare FFD results $y^{(cfd)}$ with the local linear approximations $y^{(l)}$:

$$\varepsilon = \left\| B\left(y^{(l)} - y^{(ffd)}\right) \right\|_2 \tag{4-6}$$

where, ε is the computed error between $y^{(ffd)}$ and $y^{(r)}$. B is a scaling matrix to make the calculated errors for all outputs comparable, even if they are not within the same order of magnitude.

Update records in the database: if the error ε is less than the user-defined total error tolerance, ε_{total}, a *grow* action will be executed to enlarge the current EOA of the record to include the query point; otherwise, an *add* action will be performed to store the query point and its related data as a new record in database.

Then, return the output:

$$y^{(ffd)} \rightarrow y(t_n), \tag{4-7}$$

Step 7: If $t_n = t_{end}$, then stop; else, go to Step 1.

4.2.3 Software Implementation

As shown in *Figure 4-3*, the Modelica-ISAT-FFD model is developed based on an existing Modelica-FFD model [47] by integrating the ISAT into the coupled model. Rather than through a middleware interface (e.g. Building Controls Virtual Test Bed [110]) or standard interface (e.g. Functional Mock-up Interface [111]), the Modelica is coupled with ISAT through a customized interface. Specifically, the ISAT and FFD codes are compiled as a dynamic link library (.dll) in Windows and a shared library (.so) in Linux, which is loaded by the Modelica program through external "C" functions during the co-simulation. The ISAT algorithm was originally developed in Fortran [53] and is linked to FFD with a wrapper programmed in "C" [106, 107].

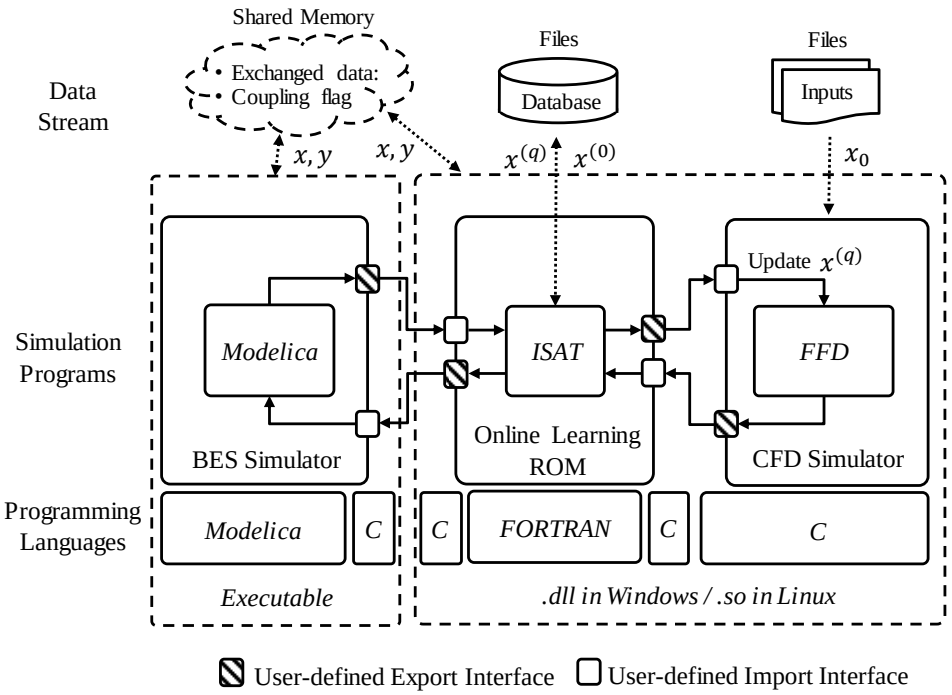


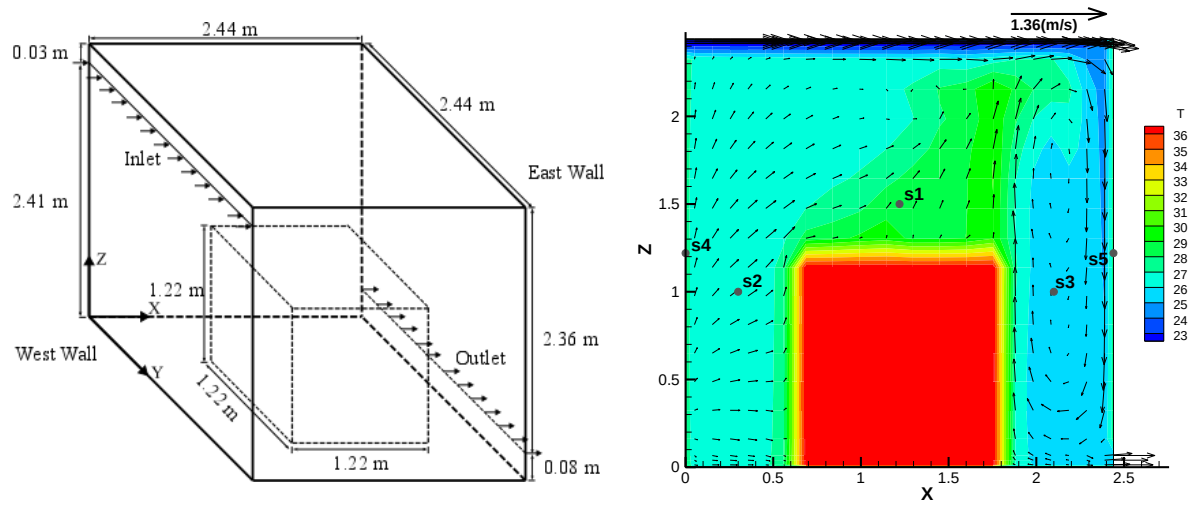
Figure 4-3 Implementation of the Modelica-ISAT-FFD model

The Modelica communicates with the ISAT through shared memory to avoid additional I/O costs in reading/writing files to hard disk for exchanging data. In addition to simulated results, the Modelica and ISAT exchange values of a coupling "flag" to make sure that both programs wait for each other if one program reaches the data synchronization point when the other has not. The records of ISAT are stored in a file for reuse. The FFD reads input information through a file and certain boundary conditions are overwritten and updated during the initialization stage with the data fed by the ISAT. The framework of this implementation can be extended for developing other BES-ISAT-CFD models by replacing the Modelica and FFD models. The customized coupling interfaces may need to be modified to align with the requirements of the specific BES and CFD programs. The source codes of the Modelica-ISAT-FFD model has been publicly released under

an open-source license, which can be found at [112]. The Modelica-ISAT-FFD model will be officially released in Modelica *Buildings* library [12] in the future.

4.3 Numerical Experiments

4.3.1 Case Descriptions



(a) (b)

Figure 4-4 (a) Schematic of the mixed convection case [79] and (b) velocity vectors and temperature contour on a cross-section at Y=1.22m computed by FFD [47]

An existing Modelica-FFD model is taken as a reference model in this study, which was proposed and validated with a mixed convection case in Zuo, Wetter, et al. [47]. In this paper, the same case is studied and the new Modelica-ISAT-FFD model is verified and evaluated by comparing with the existing Modelica-FFD model. As shown in *Figure 4-4* (a), this case simulates airflow in a space with a heated box. The space (2.44 m × 2.44 m × 2.44 m) has an inlet of height 0.03 m along the top of the west wall and an outlet of height 0.08 m along the bottom of the east wall. The heated box (1.22 m × 1.22 m × 1.22 m) is located in the center of the space on the floor.

The heated box is meant to represent a heat source, such as occupants. The experiment for this case was designed to study airflow in an aircraft cabin with significant internal heat load [79]. The original settings of the boundary conditions in the experiment are described as follows [79]. The velocity and temperature at the inlet are 1.36 m/s and 22.2 ^{o}C, respectively. The temperatures of the ceiling and floor are 25.8 ^{o}C and 26.9 ^{o}C, respectively. The temperatures of the other walls are 27.4 ^{o}C. The temperature of the surface of the heated box is 36.7 ^{o}C.

This case has a typical stratified airflow and temperature distribution. As shown in *Figure 4-4* (b), mixed convection occurs in this space due to forced convection caused by the airflow from the inlet, as well as natural convection caused by temperature differences among the inlet air, walls, and heated box. The cold air is injected from the upper left corner and forms a circulation pattern between the east wall and heated box, while a thermal plume rises above the heated box due to buoyant effects. An obvious stratified airflow and temperature distribution is formed in the indoor space. Five virtual sensors are placed in the space and temperatures are extracted in these locations to evaluate the capabilities of the coupled model to capture the non-uniform airflow and temperature distribution.

Based on this mixed convection case, we propose five case studies to evaluate the new Modelica-ISAT-FFD model as described in *Table 4-1*. Cases 1-3 comprehensively evaluate the new Modelica-ISAT-FFD model by comparing with the existing Modelica-FFD model, and determine appropriate settings of parameters including the coupling frequency and user-defined error tolerance. After that, an annual simulation is performed to demonstrate the performance of the new model for long term co-simulation. A space heating case is also conducted to show the capability of the new model for control studies.

Table 4-1 Descriptions of case studies

Cases	Models	Inputs of ISAT	Outputs of ISAT	Error Tolerance [ºC]	Data Sychroniza-tion Time [s]	Simulation Period
Case 1: Detailed evaluation with constant boundary conditions	Modelica-FFD	-	-	-	4	800 s
	Modelica-ISAT-FFD	T_{Wall} $T_{Ceiling}$ T_{Floor} T_{Inlet} $\dot{m}_{Inlet}$	T_{Room} $T_{Occupied}$ $T_{s1} \sim T_{s5}$ $\dot{Q}_{E\ Wall}$ $\dot{Q}_{Floor}$	0.2	800	
Case 2: Different coupling frequencies	Modelica-FFD	-	-	-	4	1 day
	Modelica-ISAT-FFD	T_{Wall}	T_{Room} $T_{Occupied}$	0.2	900; 1800; 3600; 7200; 14400	
Case 3: Different error tolerances	Modelica-FFD	-	-	-	3600	31 days
	Modelica-ISAT-FFD	$T_{E/N\ Wall}$ $T_{Ceiling}$	T_{Room} $T_{Occupied}$	0.2; 0.4; 0.6, 0.8		
Case 4: Long-term co-simulation	Modelica-ISAT-FFD	$T_{E/N\ Wall}$ $T_{Ceiling}$	T_{Room} $T_{Occupied}$	0.6	3600	365 days
Case 5: Space heating with feedback loop control	Modelica-ISAT-FFD	$T_{E/N\ Wall}$ $T_{Ceiling}$	T_{Room} $T_{Occupied}$	0.6	300	31 days

Case 1 performs a detailed evaluation of the Modelica-ISAT-FFD model with constant boundary conditions. The settings of the boundary conditions in this case are consistent with that in previous work that validated the existing Modelica-FFD model [39]. All key boundary

conditions are taken as inputs of ISAT. Various types of information are returned from ISAT to evaluate the capability of the Modelica-ISAT-FFD model to precisely capture non-uniform indoor environment and heat fluxes through envelopes as the Modelica-FFD model does. A detailed comparison of the results from the Modelica-ISAT-FFD and Modelica-FFD models are conducted in this case. Since the boundary conditions are constant, we only simulate this case for one data synchronization period.

Case 2 evaluates the Modelica-ISAT-FFD model with different coupling frequencies. Before further exploring the model, it is essential to identify a proper data synchronization time, which may vary for different applications. For example, Zhai, Chen, et al. [57] used 1 hour to for building energy analysis and Zuo, Wetter, et al. [47] used 4 seconds to study control dynamics. This can be determined from a parametric study with varying data synchronization time, which is determined by continuously halving a higher value from 14,400. The lower limit is determined to be 900 s because it becomes insufficient for the indoor airflow to reach steady state if 900 s is further halved to 450 s. The Modelica-FFD model with a data synchronization time of 4 s is taken as the ground truth since it has sufficiently high coupling frequency. Note that a data synchronization time of 4 seconds with the Modelica-FFD model is meaningful since FFD performs transient simulations. The simulation period for this case is one day. The values of the ISAT inputs (i.e. wall temperatures) are assumed to be a typical daily profile that follows a hypothetical sine function.

Case 3 evaluates the Modelica-ISAT-FFD model with different settings of the error tolerance. The prediction accuracy and computational costs are assessed through comparisons with the Modelica-FFD model. This case simulates a typical application scenario, in which Modelica

sends boundary conditions (e.g. wall temperatures) to FFD and FFD returns the prediction of indoor environment (e.g. temperature at occupied zone) to Modelica. The relatively shorter simulation period (i.e. 31 days of January) is used for this case because of the excessive computing time required for the existing Modelica-FFD model to perform an annual simulation, which is estimated to be about 50 days.

Using the data synchronization time and error tolerance determined in Cases 2 and 3, Case 4 performs an annual simulation with the Modelica-ISAT-FFD model to assess its performance on long term co-simulation. In this case, we did not include the Modelica-FFD model due to its high computational costs. All other settings are the same as that in Case 2. Detailed analysis on the online learning process is provided for this case.

Case 5 investigates the performance of the Modelica-ISAT-FFD model for control studies. This case compares the control performance of using different control variables including averaged room temperature, which is widely used with standalone BES, and occupied zone temperature, which can only be done by coupling external models to predict the non-uniform environment in BES. A shorter data synchronization time period, 300 s, is adopted to capture the minutes-level short-term thermal dynamics for control studies.

To make a fair comparison between the Modelica-ISAT-FFD model and Modelica-FFD model, we made the following assumptions. In the previous Modelica-FFD model, FFD continuously performs transient simulations until the end of the co-simulation to capture the seconds-level short-term airflow and thermal dynamics. However, it may not be needed for scenarios with long data synchronization time. For example, the widely-used BES-CFD coupling scheme adopts a data synchronization time of 1 hour and CFD returns steady state results to BES

during the co-simulation [54, 57, 59]. Therefore, here in Cases 3 and 4 that adopt a data synchronization time of 1 hour, FFD simulations are treated as "steady state" CFD simulations, which converge by simulating a sufficiently long period. In the Modelica-FFD model, the FFD simulation stops when it converges and returns steady state results for airflow prediction to Modelica as it is in the Modelica-ISAT-FFD model.

4.3.2 Model Implementation

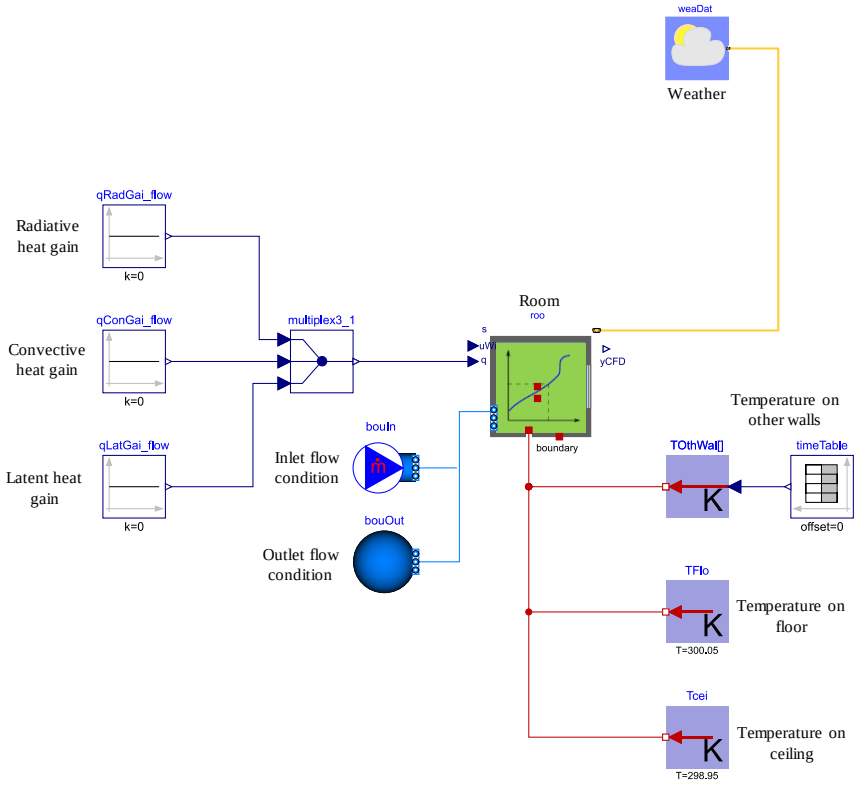


Figure 4-5 Diagram of Implementations for the Modelica-ISAT-FFD model in Modelica

The Modelica models for the Modelica-ISAT-FFD model of Cases 1 and 2 are implemented as shown in *Figure 4-5*. To enable the model exchange, the Modelica-FFD model and Modelica-ISAT-FFD model share a similar interface on the Modelica side. The major

difference lies within the room model. In the Modelica-FFD model, the room model exchanges data with FFD, for example, Modelica sends the boundary conditions to FFD and FFD sends its simulation results back to Modelica [12, 47]. In the Modelica-ISAT-FFD model, the room model exchanges data with ISAT and then ISAT determines if the prediction should be returned through retrieving from the database or executing an FFD simulation.

In Case 1, the room model receives boundary conditions for the temperatures of the floor, ceiling, remaining walls as well as the temperature and mass flow rate at the inlet. The radiative, convective and latent heat gains are set to zero. The outputs of the room model include averaged room temperature, averaged temperature at occupied zone, temperatures at five sensor locations, and heat fluxes through walls. The data synchronization time in the Modelica-FFD is set to be 4 s, which is the same as in the previous research [47]. For the Modelica-ISAT-FFD model, the data synchronization time is set to be identical with the simulation period of the co-simulation (i.e. 800 s) since the boundary conditions are assumed to be constant in this case. From our tests, 800 s is enough for the airflow to become steady state for this case. FFD simulations use a non-uniform 20 × 20 × 20 mesh and time step size of 0.1 s. The detailed settings of the FFD model as well as the validation of the FFD model with experimental data can be found in [47].

The Case 2 utilizes a similar Modelica model as Case 1. It differs from Case 1 in two aspects. First, the inputs and outputs of ISAT are different as described in *Figure 4-5*. Second, Case 2 uses a dynamic boundary condition for the walls rather than a constant one in Case 1. The dynamic boundary condition is assigned to represent a typical daily profile:

$$T_{wall}(t) = 5\sin\left(\frac{2\pi}{3600\times 24}t - \frac{\pi}{2}\right) + 25, \tag{4-8}$$

where, $T_{wall}(t)$ is the wall temperature at the time t. The wall temperature starts from 20 ºC at 12:00 am of a day, increases to 30 ºC at 12:00 pm, and then goes back to 20 ºC at the end of the day.

The Cases 3 and 4 consider a more realistic scenario with exterior constructions and varying weather profile. They assume the studied room with mixed convection is located in the northeast corner on the top floor of a building, such that there are three exterior constructions: the north wall, east wall and ceiling. The north and east walls are constructed with 120 mm of brick while the ceiling has 200 mm of concrete and 100 mm of insulation. The interior surface temperatures of the north and east walls are assumed to have the same dynamic temperature values, since they are constructed with the same materials and their exteriors are exposed to the same outdoor environment. The remaining walls are interior partition walls adjacent to conditioned indoor rooms, with the interior surfaces of these walls treated as a fixed temperature value.

Figure 4-6 shows the Modelica model for Cases 3 and 4. The interior surfaces of the interior partition walls, including the south and west walls and the floor, are assumed to be 25 ºC. The room model receives inputs for the interior surface temperatures of the exterior walls, which are calculated by solving a one-dimensional heat transfer problem between indoor and outdoor on the Modelica side. The outdoor weather, including temperature, wind speed and solar irradiation, is included to determine the convective and radiative heat transfer between the building exterior and outdoor environment. The thermal properties of the building materials are also included to determine the heat conduction through the walls. Lastly, convection from the room air is considered when determining the temperature of these walls. The supply air flow rate and temperature are set to typical values for a ventilation system of 0.024 kg/s and 22.2 ºC. The

additional radiative, convective and latent heat gains are assumed to be zero. This case is simulated for the month of January using weather data of Chicago, IL.

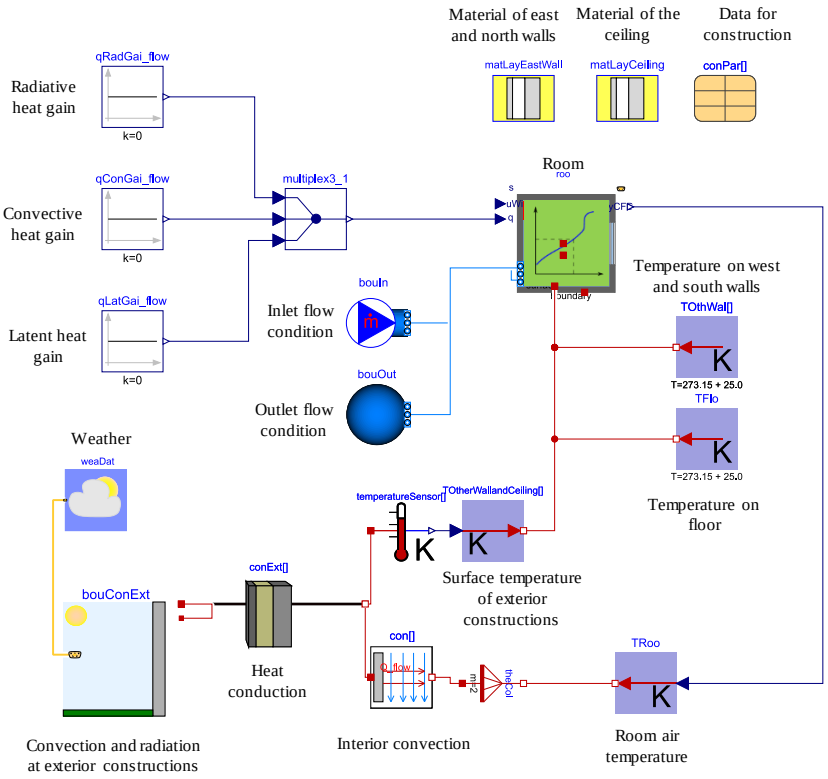


Figure 4-6 Diagram of Modelica models for Cases 2 and 3

For the ISAT model, two inputs are set including the interior surface temperatures of E/N walls and ceiling, and the two outputs are the average temperature in the room and in the occupant zone. The data synchronization time for both Modelica-ISAT-FFD and Modelica-FFD models is 3600 s. The error tolerance is defined to be 0.6 ^{o}C for Cases 2 and 3 in the ISAT model.

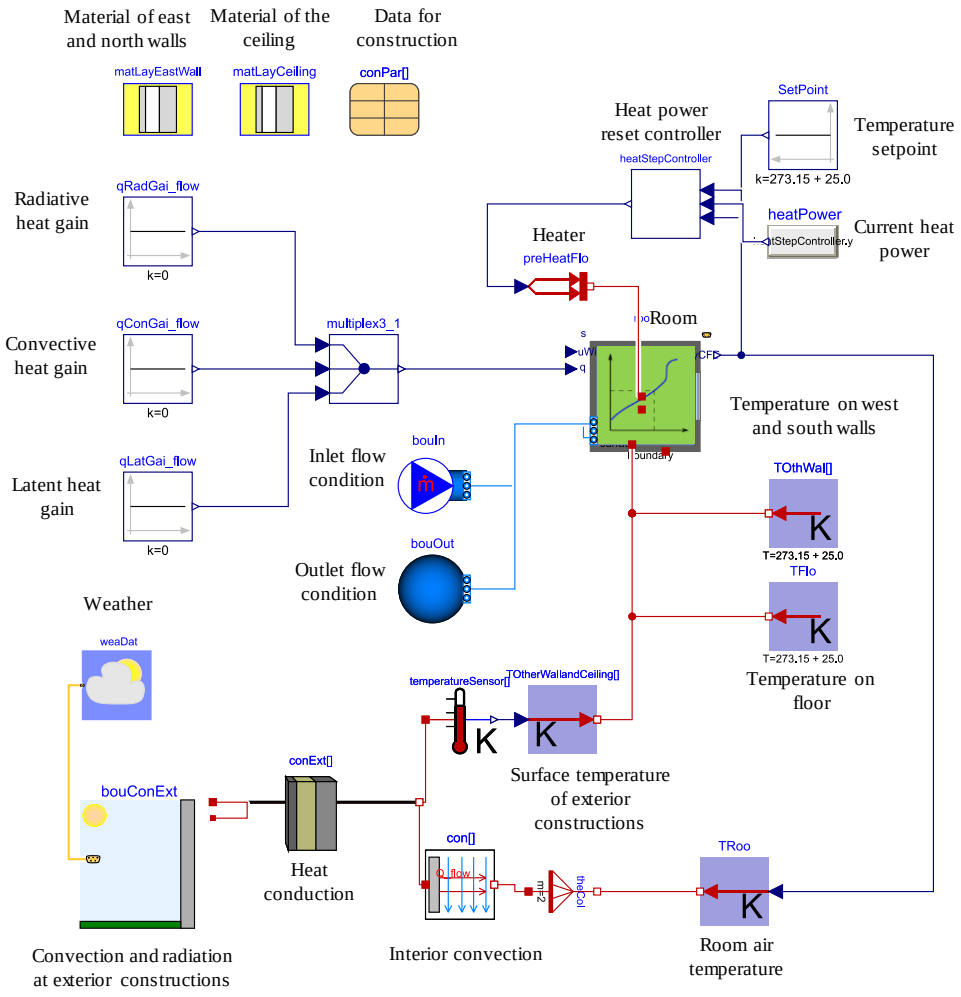


Figure 4-7 Diagram of Modelica models for space heating of the mixed convection case

In Case 5, a space heating case is studied to illustrate the capability of the Modelica-ISAT-FFD model for feedback loop control. The implementation of the case in Modelica is shown in *Figure 4-7*, which adds an electrical heater and a heating power controller. The control objective of the heating system is to maintain the average room temperature at 25 ^{o}C with a dead band of ±1 ^{o}C. The heating power controller resets the heating power by steps.

4.3.3 Results

Case 1: Detailed Evaluation with Constant Boundary Conditions

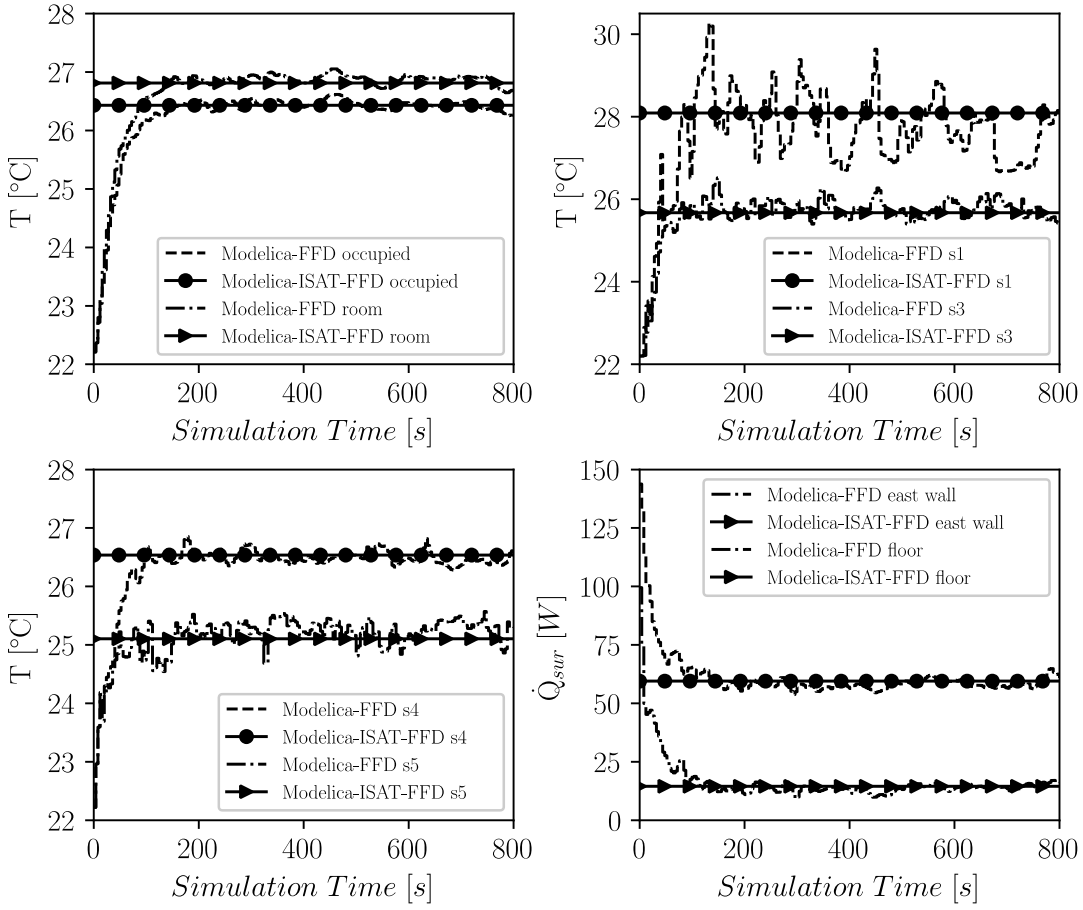


Figure 4-8 Comparison of the predicted indoor airflow conditions in the mixed convection case by the Modelica-ISAT-FFD and the Modelica-FFD

Figure 4-8 shows the simulation results using the Modelica-ISAT-FFD model and Modelica-FFD model. Note that the Modelica-ISAT-FFD model predicts steady-state results of the airflow, while the Modelica-FFD performs transient airflow simulations. The temperatures predicted in both models include room temperature, temperature at occupied zone, and temperatures at different sensors. It can be found that the Modelica-ISAT-FFD steady-state

predictions are generally consistent with the time-averaged FFD results as they reach steady-state. The room temperature is slightly higher than the temperature at occupied zone because of the vertical stratification of temperature caused by the buoyancy effect. Both models capture the stratified air temperature distribution at different sensor locations in the room. For example, both models predict a warmer temperature at Sensor 1, which is caused by the hot air rising from the heated box.

The last plot in *Figure 4-8* shows the heat fluxes through the east wall and floor. The Modelica-ISAT-FFD steady-state results are again consistent with the time-averaged Modelica-FFD results as they approach steady state. The heat flux through the east wall differs significantly from that through the floor. This is mainly because the temperature difference between the construction surface and adjacent indoor air are different for the east wall and floor due to the non-uniform temperature distribution as shown in *Figure 4-4* (b). To conclude, the Modelica-FFD model can be advantageous for capturing short-term dynamics by performing transient simulations, while the Modelica-ISAT-FFD model can be used for predicting steady-state results generally consistent with those from Modelica-FFD simulations.

Case 2: Different Coupling Frequencies

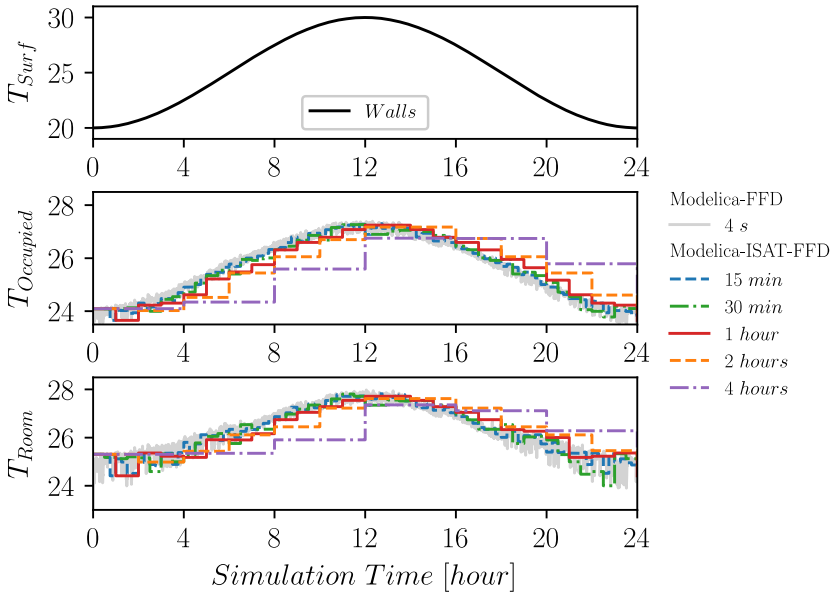


Figure 4-9 Comparison of the Modelica-ISAT-FFD simulation with different coupling frequencies and the Modelica-FFD simulation for the mixed convection case

Figure 4-9 shows the simulation results for the mixed convection case in a typical day. Although the Modelica-FFD results oscillate significantly due to the dynamic characteristics of the flow, their time-averaged values generally match with the steady-state temperature predictions from the Modelica-ISAT-FFD model, especially for the coupling frequencies with data synchronization time of equal or less than 1 hour. The Modelica-ISAT-FFD predictions delay by one data synchronization time compared to the Modelica-FFD results. This is because of the quasi-dynamic coupling scheme used in the model. For example, from t_{n-1} to t_n, the ISAT predictions are made based on exchanged data at the data synchronization point t_{n-1}, and remain constant until the next data synchronization point t_n. As a result, when the coupling frequency is low, e.g.

the data synchronization time is 4 hours, the accuracy of the Modelica-ISAT-FFD predictions decline significantly due to the considerable delay.

The differences between the Modelica-ISAT-FFD predictions and the Modelica-FFD results are quantified by Normalized Mean Bias Error (NMBE), e_{NMBE}:

$$e_{NMBE} = \frac{\sum_{i=1}^{n}\left|\frac{1}{\Delta t}\int_{t_0+(i-1)\Delta t}^{t_0+i\Delta t} y^{(isat)}(t)dt - \frac{1}{\Delta t}\int_{t_0+(i-1)\Delta t}^{t_0+i\Delta t} y^{(ffd)}(t)dt\right|}{n\left|y_{max}^{(ffd)} - y_{min}^{(ffd)}\right|}, \quad (4\text{-}9)$$

where $y^{(isat)}$ and $y^{(cfd)}$ are the Modelica-ISAT-FFD and the Modelica-FFD results, respectively. To reduce the effect of the short-term oscillation in Modelica-FFD results, we divide the simulation period into small segments and calculate the e_{NMBE} based on time-averaged values for each segment. The size of each segment, Δt, is 5 minutes in this study, which results in the total number of segments, n, being 288. The mean bias error is normalized by dividing the range of the values in Modelica-FFD results, which is 3.9 ºC for $T_{Occupied}$ and 4.1 ºC for T_{Room} for this case. This is to avoid underestimating the percentage error by using absolute values, which are about 24-28 ºC.

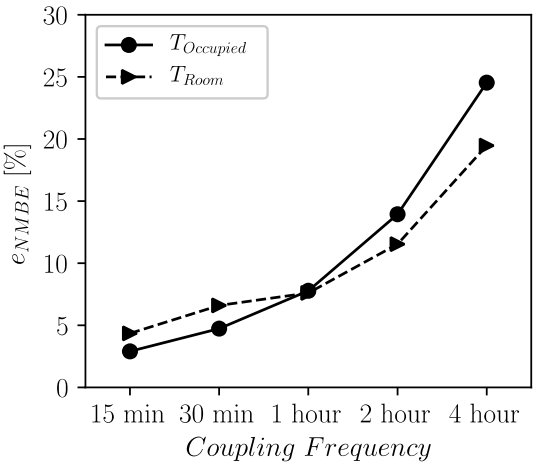


Figure 4-10 NMBEs of predictions by the Modelica-ISAT-FFD simulation with different coupling frequencies

The NMBEs of predictions by the Modelica-ISAT-FFD simulation with different coupling frequencies are shown in *Figure 4-10*. Both results for $T_{Occupied}$ and T_{Room} follow a similar pattern, in which the NMBEs increase when the data synchronization time increases. This is more noticeable in particular when the data synchronization time is larger than 1 hour. We also tested the NMBEs for other outputs, e.g. heat flux through walls, and found a similar pattern. The prediction accuracy of $T_{Occupied}$ is slightly more sensitive to the coupling frequency. This may be due to different air volumes and thermal mass capacities for these two outputs. To balance the accuracy and computational costs, the data synchronization time of 1 hour, which achieves NMBEs of 7.7-7.8% for both outputs, is determined for the following Cases 2 and 3.

Case 3: Different Error Tolerances

The EnergyPlus weather data for the month of January in Chicago, IL [12] is shown in *Figure 4-11*. The top plot shows the ambient dry bulb temperature and black sky temperature throughout the month, where the black sky temperature is included to determine the radiative exchange between the sky and building exterior. The second and third plots show the wind direction in radians and wind speed in m/s, respectively. The wind profile is critical to computing the convective heat transfer between the building exterior and outdoor environment. The final plot shows the solar irradiation on the ceiling and exterior walls of the building. The plot has spikes during the day when the sun is present, and there is consistently more incident irradiation on the ceiling compared to the exterior walls during these spikes.

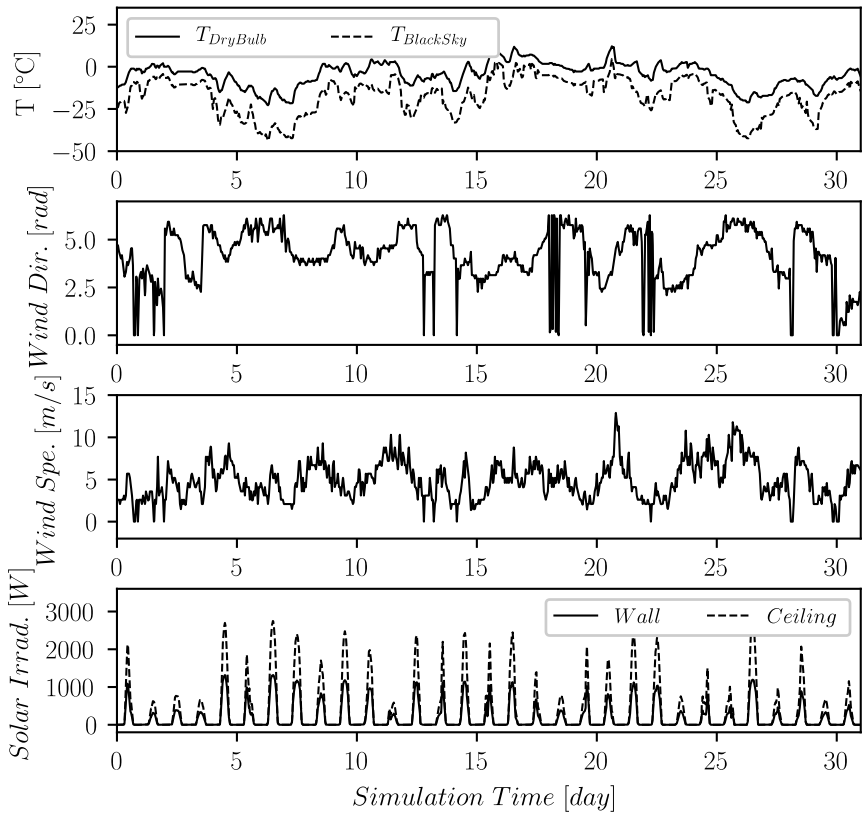


Figure 4-11 Weather profiles in January of the studied case in Chicago.

The five plots in the left column in *Figure 4-12* show the interior surface temperatures of the E/N walls and ceiling, which are defined as the inputs of the ISAT in the Modelica-ISAT-FFD model. These temperatures are calculated by solving a one-dimensional heat transfer problem in the Modelica model. All results follow similar patterns, such as the surface temperatures of the ceiling being much higher and more stable than that of the E/N walls because of the materials and insulation layer added in the ceiling. The results for the Modelica-ISAT-FFD model with different settings of error tolerances show little differences, which can be explained because even the predicted room temperatures can differ as much as 0.8 ºC. Such a deviation has very little impact

on the calculation of the interior surface temperatures, since the temperature differences between the room air and the interior surface are much larger than 0.8 ºC.

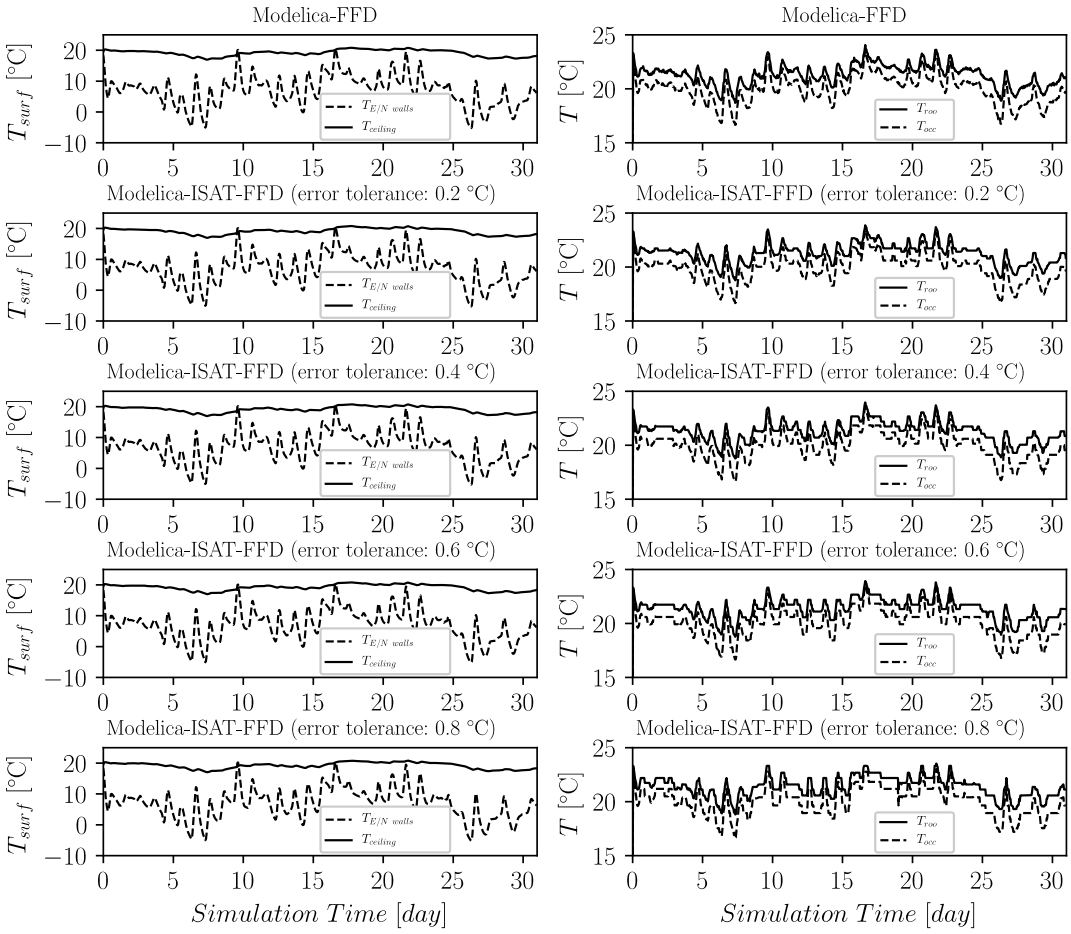


Figure 4-12 Surface temperatures at exterior constructions and temperature predictions by the Modelica-FFD model and the Modelica-ISAT-FFD model with different error tolerances

The five plots in the right column in *Figure 4-12* show the average room temperature and the average occupied zone temperature predicted by the Modelica-ISAT-FFD and Modelica-FFD. The results from both models are very similar and the predicted average room temperatures are higher than the average occupant zone temperatures. This is because both models predict the

temperature stratification in the room, so it is able to capture the warmer air rising towards the ceiling, resulting in the occupied zone temperature to be lower than the average room temperature. Interestingly, some discrepancies occur in the results of the Modelica-ISAT-FFD model with different settings of error tolerances. For example, for the predicted results on Days 2-3 and 24-25, it can be clearly seen that the prediction resolution decays as the error tolerance becomes loose. This becomes more obvious when then the error tolerance is 0.8 ºC, in which some unexpected oscillations occur.

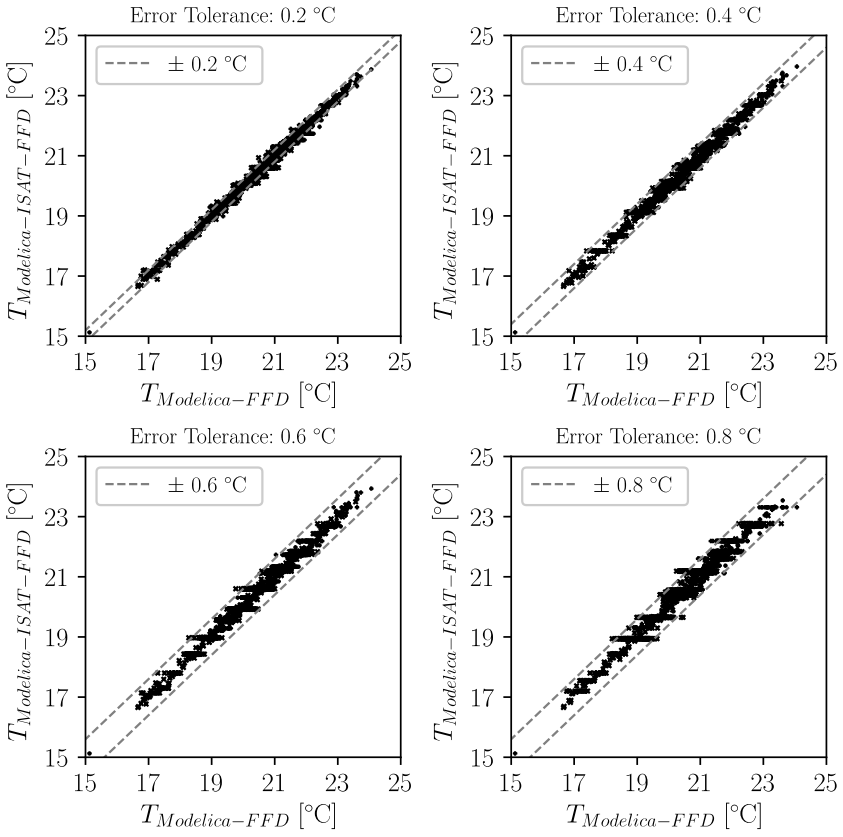


Figure 4-13 Comparison of temperatures predicted by the Modelica-ISAT-FFD model with different error tolerances and the Modelica-FFD model

The temperatures predicted by the Modelica-ISAT-FFD model and the Modelica-FFD model are compared in *Figure 4-13*. The results from both models generally match with each other. Taking the predictions from the Modelica-FFD model as the ground truth, the prediction errors of the Modelica-ISAT-FFD model are mostly within the pre-defined error tolerances. On one side, reducing the error tolerance in Modelica-ISAT-FFD can make the prediction more accurate. The average error reduces from 0.21 ^{o}C for 0.8 ^{o}C to 0.11 ^{o}C for 0.2 ^{o}C, as shown in *Table 4-2*. On the other side, reducing the error tolerance will also make the error control more challenging with more predictions outside of the user defined error tolerance (e.g. the violation increases from 0.5% for 0.8 ^{o}C to 15.7% for 0.2 ^{o}C, with an average value of 4.5% for all cases).

Table 4-2 Number of actions and estimated computing time

Models	**Error Tolerance [^{o}C]**	**Number of Actions**					**Solution Time [hour]**	**Average Error [^{o}C]**
		FFD Calls				**Regressions**		
		FFD	**Add**	**Grow**	**Total**	**Retrieve**		
Modelica-FFD	-	744	0	0	744	0	103.3	-
Modelica-ISAT-FFD	0.2	0	78	160	238	506	33.1	0.11
	0.4	0	28	124	152	592	21.1	0.14
	0.6	0	20	99	119	625	16.5	0.15
	0.8	0	15	84	99	645	13.8	0.21

The simulation was performed using one core of Intel® Xeon® CPU E5-1603 with 2.80GHz. The number of actions and estimated computing time for this case with different models are shown in *Table 4-2*. In the Modelica-FFD model, all the predictions are obtained from FFD

simulations, which result in a computing time of 103.3 hours. With the Modelica-ISAT-FFD model, most predictions are obtained from retrieving from the database, which do not require FFD simulations. As a result, the Modelica-ISAT-FFD models with error tolerances of 0.2 ^{o}C, 0.4 ^{o}C, 0.6 ^{o}C and 0.8 ^{o}C save 68%, 79.7, 84%, and 86.6% of the computing time, respectively, compared to the Modelica-FFD model. To balance the accuracy and computational costs, the error tolerance of 0.6 ^{o}C is used to perform the annual simulation in Case 3.

Case 4: Long-Term Co-Simulation

In this case, an annual simulation is performed to evaluate the long-term performance of the Modelica-ISAT-FFD model. We did not run this case with the Modelica-FFD model due to the huge computing costs, but its computing time is estimated as a benchmark.

The weather and predicted temperatures of the Modelica-ISAT-FFD model are shown in *Figure 4-14*. The first plot shows the outdoor dry bulb temperature profile during the year, which ranges from -20 ^{o}C to 30 ^{o}C. The following second and third plots show the surface temperatures of exterior constructions. As was also found in Case 1, the surface temperatures of the ceiling are more stable than that of the E/N walls because of the construction materials. The fourth and fifth plots show the predicted average room temperature and average occupied zone temperature, respectively. They generally follow similar patterns with the surface temperatures of exterior constructions. Similar to Case 1, the average room temperatures are generally higher than the average occupied zone temperatures. Interestingly, it is found that the differences between these two are much less during the summer. This may be because the higher surface temperatures of exterior constructions in summer result in a more uniform temperature distribution in the room.

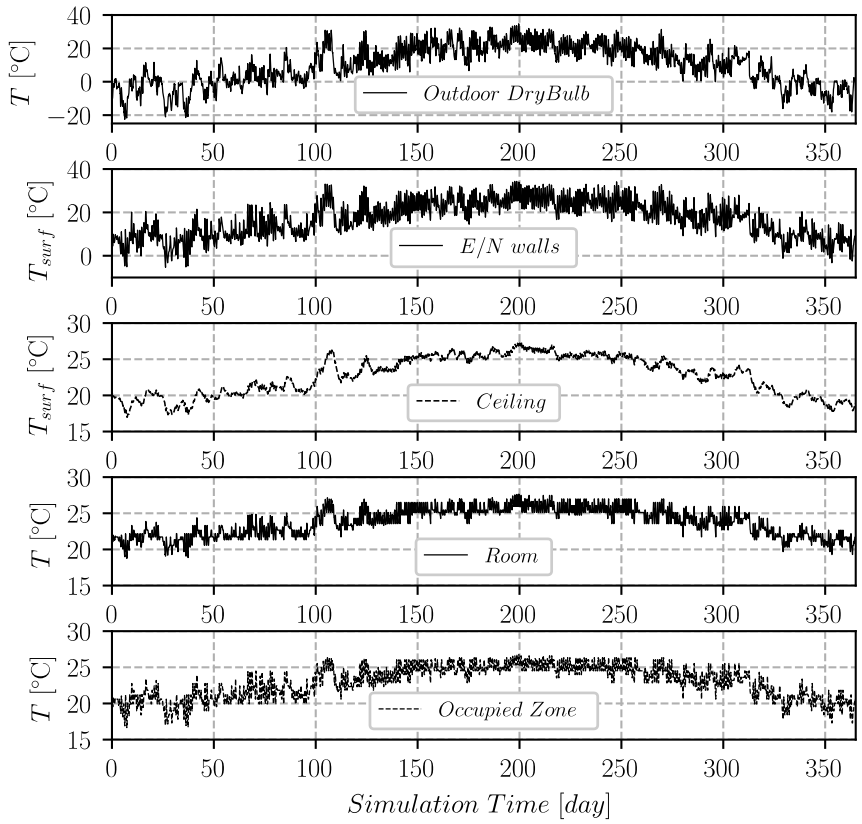


Figure 4-14 Weather and predicted temperatures with the Modelica-ISAT-FFD model

The ISAT training performance during the simulation is shown in *Figure 4-15*. The first plot shows the number of actions including *grow*, *add* and *retrieve* for each day. The *grow* and *add* actions occur frequently at the beginning of the simulation from Day 1 to Day 15 due to lack of data in the database, but then a few *grow* and *add* actions occur in the following days. Starting from Day 100, *grow* and *add* actions occur frequently again. This is because the weather varies drastically at that time, so that ISAT inputs are beyond the ranges of existing records in the database, preventing the predictions from being directly retrieved from the database. The occurrence of several other small spikes during the simulation shares the same reason.

The second and third plots show the number of FFD runs (*grow* and *add*) versus the number of regressions (*retrieve*) including the numbers of accumulated actions as well as percentages over the total actions. It can be found that about 25-50% of total actions are FFD runs at the first 20 days, which means the Modelica-ISAT-FFD model is only about 50-75% faster than the Modelica-FFD model if the simulation period is less than 20 days. The percentage of FFD runs decreases to about 10% after Day 50, then slightly increases from Day 100, and finally becomes about 5%. This indicates that a longer simulation period tends to better exploit the speed advantages of the Modelica-ISAT-FFD model.

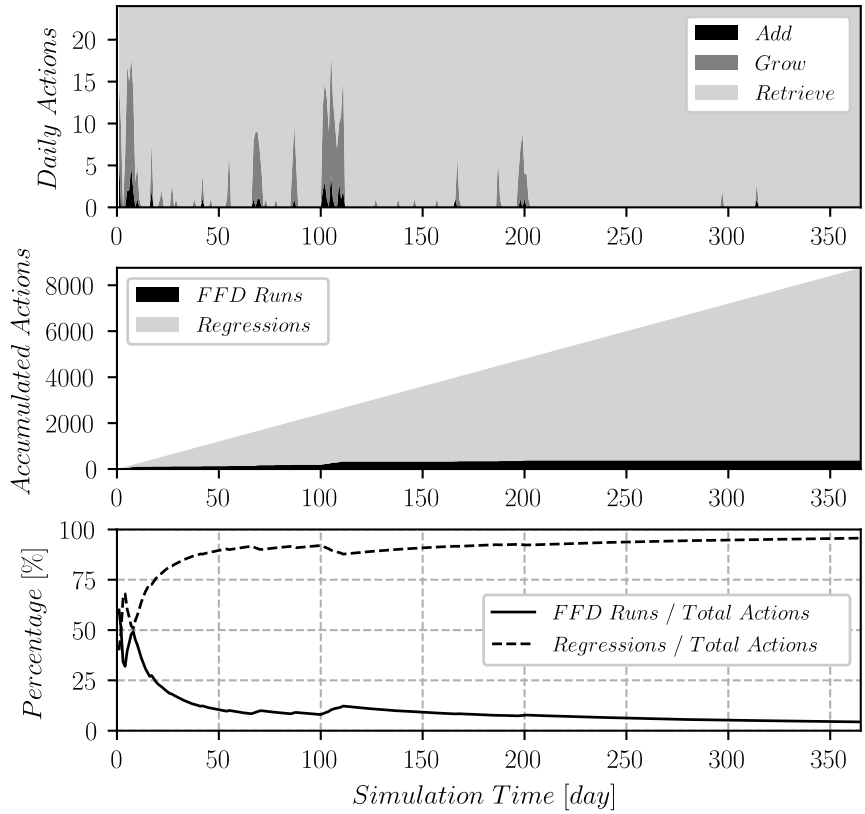


Figure 4-15 Training performance in the annual simulation with the Modelica-ISAT-FFD model

Table 4-3 Number of actions and estimated computing time

Models	Error Tolerance [°C]	Number of Actions					Solution Time [day]
		FFD Calls				Regressions	
		FFD	Add	Grow	Total	Retrieve	
Modelica-FFD	-	8760	0	0	8760	0	50.7[a]
Modelica-ISAT-FFD	0.6	0	46	344	390	8380	2.2

a: The computing time of the Modelica-FFD model shown in this table was determined from an estimation, in which FFD was called every hour for the whole year resulting in an 8,760-times of FFD runs in total.

The total number of actions and estimated computing time are summarized in *Table 4-3*. The actions and computing time with the Modelica-FFD model are estimated as a reference case. The Modelica-ISAT-FFD model saves 95.7% of computing time compared to the Modelica-FFD model. This significant acceleration makes the long-term BES-CFD co-simulation feasible while still controlling the prediction accuracy of the airflow.

Case 5: Space Heating with Feedback Loop Control

Two scenarios are studied in this case. The first one uses the Modelica-ISAT-FFD model with the control variable of average temperature at occupied zone. The second one uses the Modelica-ISAT-FFD model with the control variable of average room temperature. The first plot in *Figure 4-16* shows the average temperature at occupant zone. It can be found that it is controlled at 25 ± 0.75 °C in the first scenario, but at a lower value in the second scenario. This is because when the average room air temperature is controlled at around 25 °C, the average temperature at occupied zone is lower than that due to temperature vertical stratification in the room. Consequently, the heating powers as well as the daily heating load in the first scenario are higher

than that in the second scenario, which result in a total heating load of 956.4 kWh in the first scenario and a total heating load of 704.8 kWh in total in the second scenario.

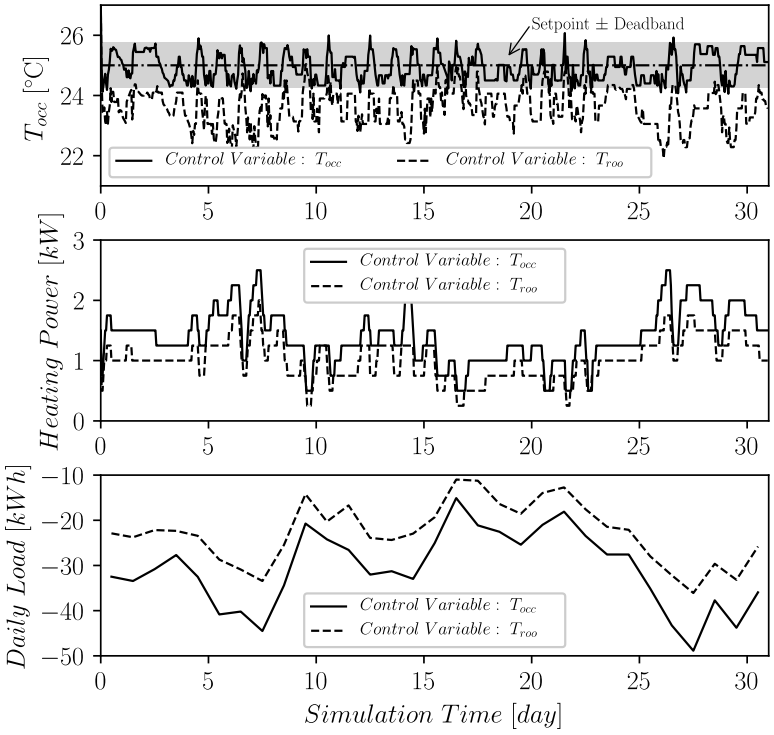


Figure 4-16 Occupant zone temperatures and heating powers in different cases

The number of actions and estimated computing time are summarized in *Table 4-4*. The actions and computing time with the Modelica-FFD model are estimated as benchmark, which are the same for the two scenarios. The Modelica-ISAT-FFD model saves 80.2%-81.2% of computing time for the two scenarios respectively, compared to the Modelica-FFD model.

Table 4-4 Number of actions and estimated computing time

Models	Control Variable	Number of Actions				Computing Time [day]
		Directly Calling FFD	Add	Grow	Retrieve	
Modelica-FFD[a]	T_{occ}	8,928	0	0	0	20.7
Modelica-FFD[a]	T_{roo}	8,928	0	0	0	20.7
Modelica-ISAT-FFD	T_{occ}	0	96	1601	7231	3.9
Modelica-ISAT-FFD	T_{roo}	0	107	1651	7170	4.1

a: The computing time of the Modelica-FFD model shown in this table was determined from an estimation, in which FFD was called every 300 s for January resulting in an 8,928 times of FFD runs in total.

4.4 Discussions

To evaluate the performance, we conduct a comprehensive comparison between the new Modelica-ISAT-FFD model and the existing Modelica-FFD model, in which the accuracy of the Modelica-ISAT-FFD model is evaluated by comparing it with the Modelica-FFD model rather than using experimental data. This is justified since the existing Modelica-FFD model has already been validated with experimental data in Zuo, Wetter, et al. [47]. In addition, these comparisons can clearly show the advantages of the proposed new BES-ISAT-CFD coupling scheme against the conventional BES-CFD coupling scheme.

The first advantage of the Modelica-ISAT-FFD model is the capability of error control, which is usually a challenge with conventional ROMs. The results of the case studies show that the prediction errors with the Modelica-ISAT-FFD model are generally well controlled within the pre-defined error tolerances. Therefore, with the Modelica-ISAT-FFD model, one can set different preferences for airflow prediction accuracy for different applications. Given that a stricter setting

of the error tolerance usually leads to a higher computing cost by calling more FFD simulations, this feature provides the flexibility to meet the requirement of airflow prediction accuracy while minimizing the computing costs.

The second advantage of the Modelica-ISAT-FFD model is the fast speed. Compared to the existing Modelica-FFD model, the Modelica-ISAT-FFD model saves the computing time up to 95.7% for an annual simulation. The savings of computing time are case-by-case, but generally a higher error tolerance helps improve the speed. A longer simulation period with the Modelica-ISAT-FFD model may also help offer more computational savings than the Modelica-FFD model from obtaining more retrieve actions.

Another advantage of the Modelica-ISAT-FFD model is the online learning. With this feature, training with pre-computed CFD results is not required before the co-simulation, which is usually needed for conventional ROMs. The accuracy can also be guaranteed even when the query point is outside of the training domain, which is challenging for conventional ROMs. Moreover, the online training can be more efficient and effective compared to the conventional training method. The online training is performed per request during the co-simulation, while conventional training should cover the whole training domain by training the ROMs with a comprehensive dataset generated by CFD before the co-simulation.

Nevertheless, the Modelica-ISAT-FFD also has limitations. First, it can only predict steady-state results while the Modelica-FFD model can capture the short-term dynamics in a scale of a few seconds. In this study, ISAT is used for predicting steady-state results, but it involves a transient process in the Modelica simulation. This is based on the assumption that the conditions do not change drastically between two data synchronization points, which is similar as [64]. Also,

even though the Modelica-ISAT-FFD has shown great improvement on the simulation speed compared to the existing Modelica-FFD model, the simulation speed can be further improved to handle full scale cases in the real world. One approach is to run the time consuming FFD on GPU [40, 38].

4.5 Conclusions

To realize fast and accurate BES-CFD co-simulation, this study proposed an online BES-ROM-CFD approach by integrating an online learning ROM with the BES-CFD coupled model. A Modelica-ISAT-FFD model was implemented to demonstrate the new approach and it was verified with a mixed convection case by comparing its results with that from an existing Modelica-FFD model, which was validated in previous research. The results showed that the steady state results of indoor environment and heat flux through walls predicted in the Modelica-ISAT-FFD generally match with that predicted by the Modelica-FFD model, although it cannot capture short-term dynamics as the Modelica-FFD does. Three cases were studied to further evaluate the performance of the new model regarding the prediction accuracy of the ROM and computing time. The results showed that the Modelica-ISAT-FFD model saves up to 95.7% of computing time for an annual simulation while generally well controlling the prediction errors within the user-defined error tolerances compared to the Modelica-FFD model.

To conclude, the proposed online learning approach significantly accelerates the BES-CFD co-simulation with well controlled prediction accuracy of non-uniform airflow and thermal environment. To allow wide use of the proposed methodology, the Modelica-ISAT-FFD model has been publicly released under a free open-source license.

Chapter 5 Towards Holistic Optimization of Data Center Cooling Systems and Airflow Management: Adaptive Online BES-ROM-CFD Co-Simulation Powered by Distributed Computing

This chapter further improves the online BES-ROM-CFD method and based on that proposes a holistic optimization platform. A real middle size data center is adopted to demonstrate the capability of the proposed optimization platform.

5.1 Introduction

Data centers, which house a large amount of mission-critical IT equipment, consume about 1.5% of the global electricity in 2010 [113]. With the fast growing markets on cloud computing, data centers' share on global electricity was estimated to be as high as 3-13% in 2030 [3]. The cooling systems, which consume about 24%-60% of total data center energy consumption [2], are crucial for reliable operation of IT equipment in data centers. However, many data centers are significantly overcooled to ensure the reliable operation of IT equipment, which leads to a low cooling efficiency [5]. To improve the effectiveness and efficiency of data center cooling systems, simulation-based optimization [6] can be adopted, in which the simulation models are critical to achieve the optimal solution.

Among all data center modeling techniques, computational fluid dynamics (CFD) has been one of the most widely-used methods to evaluate and improve data center airflow management and cooling effectiveness [13]. CFD can be used to understand the detailed airflow and temperature fields in data centers. For example, Healey, VanGilder, et al. [36] found that with careful

calibration, CFD is able to make reliable predictions of perforated-tile air flow rates and rack-inlet temperatures for a middle-size data center. With this capability, CFD has been widely adopted to evaluate different design considerations. For example, Shrivastava [114] investigated the impact of different configurations of floor and ceiling supply diffusers and floor and ceiling returns through a what-if scenario evaluation. However, these research focused on the designs and few considered the off-design operations even though it has a significant energy impact as pointed out by Demetriou, Khalifa, et al. [115].

To support optimal off-deign operations, various high-fidelity building energy simulation (BES) models were developed. For example, eQuest [8], EnergyPlus [7, 17], TRNSYS [18, 115], and Modelica [10] were adopted to develop detailed models of cooling systems and optimize the off-design operations for data centers. If there is no special treatment, these BES programs commonly adopt a multizone room model with a well-mixed assumption, which makes it challenging to handle non-uniform thermal environment [47, 66]. However, the data center thermal requirements pertain to the air entering the IT equipment [15], which usually involves non-uniform thermal environment, therefore, the BES models may fail to guarantee the thermal guidelines can be met. For example, Fu, Zuo, et al. [10] optimized the cooling systems using a Modelica model with a well-mixed room model. They assumed the plenum air temperature is equal to the IT rack inlet temperature without considerations of any recirculation and leakage hot air mixed with the supply cold air. This assumption may not be valid for data centers with open-cold-aisle configurations. Even though a strict constraint may be added to force the optimal solution from cooling system optimizations to meet the thermal requirements, it may lead to the local optima problem.

To address these issues, a holistic optimization can be performed by simultaneously considering the cooling system and airflow management. This can be achieved by using a BES-CFD coupled model [66]. However, dynamically coupling BES with CFD is too computationally expensive to perform an annual simulation within a reasonable computing time. For example, Zhang, Lam, et al. [54] studied a naturally ventilated building with a EnergyPlus-Fluent co-simulation model, and found that it took about 11 hours to simulate the case for a day [54]. It becomes even more challenging when a simulation-based optimization is performed. For example, Tian, Han, et al. [65] optimized the thermostat location in a small office (5.16 m × 3.65 m × 2.43 m) based on a Modelica-FFD coupled model. It took more than five hours to finish the optimization even though only 1 hour is simulated for the case and an extremely coarse grid (8 cells × 5 cells × 5 cells) is used in the FFD simulations. Consequently, a huge computational cost is expected if one optimizes a data center with annual simulations, which has much more complex grids and longer simulation time than the case studied in [65]. Consequently, to our best knowledge, no research adopted a BES-CFD co-simulation to perform a holistic optimization of data centers.

To realize a holistic optimization of data centers, various efforts have been made to couple BES with different variants of simplified airflow models. First, a constant value can be used to represent the non-uniform thermal environment. For example, [116] developed an analytic model to study the cooling system and airflow management in a holistic view, in which the air temperature rise from the perforated tile to the rack inlet temperatures was assumed to be a constant value, which can be determined from CFD simulations. Sun, Hong, et al. [117] made similar assumptions in an EnergyPlus data center model. However, this temperature rise depends on some parameters in cooling systems, IT racks and others, which may vary during the operation, such as

the supply air flow rate, IT loads, and room airflow pattern and average temperature. To take different scenarios into considerations, regression models or other reduced order models (ROMs) can be developed from a parametric study of CFD simulations. For example, Billet, Healey, et al. [118] utilized a simplified power consumption model to minimize the energy consumption by adjusting the total supply air flow rate, in which regression models was developed to predict the maximum rack inlet temperature. Demetriou and Ezzat Khalifa [70] optimized the IT load placement with a dynamic cooling system model, in which a ROM using proper orthogonal decomposition (POD) is used for predicting the rack inlet temperature distributions. However, these methods are only reliable for operating conditions within the training domain. The perdition accuracy may significantly decay if the operating conditions are outside of the training domain. The Chapter 4 introduced an online BES-ROM-CFD method that can address this issue. But it is still not fast enough for practical use in real applications.

To further improve the online BES-ROM-CFD method, this dissertation proposes an adaptive online BES-ROM-CFD method, in which adaptive coupling frequencies are used to reduce the number of ROM-CFD calls during the co-simulation. The co-simulation is further powered by the distributed computing technology, which allows that BES-ROM runs on a CPU and CFD runs on a GPU in parallel during the co-simulation. By reducing the number of ROM-CFD calls and maximizing the capacity of computing hardware of a computer, the new method significantly accelerates the online BES-ROM-CFD co-simulation. Based on that, a holistic optimization platform is proposed. To demonstrate the proposed optimization platform, an adaptive Modelica-ISAT-FFD model connected with GenOpt is implemented.

The rest of this chapter is organized as follows. First, the adaptive online BES-ROM-CFD method powered by distributed computing technology as well as the optimization platform based on that are introduced. Then, the implementation of the optimization platform based on an adaptive Modelica-ISAT-FFD model is described. After that, the new coupled model is verified with a real data center case. Finally, an optimization study is conducted with the same data center to demonstrate the performance of the proposed optimization platform.

5.2 Methodology

5.2.1 Improved Modelica-ISAT-FFD Model

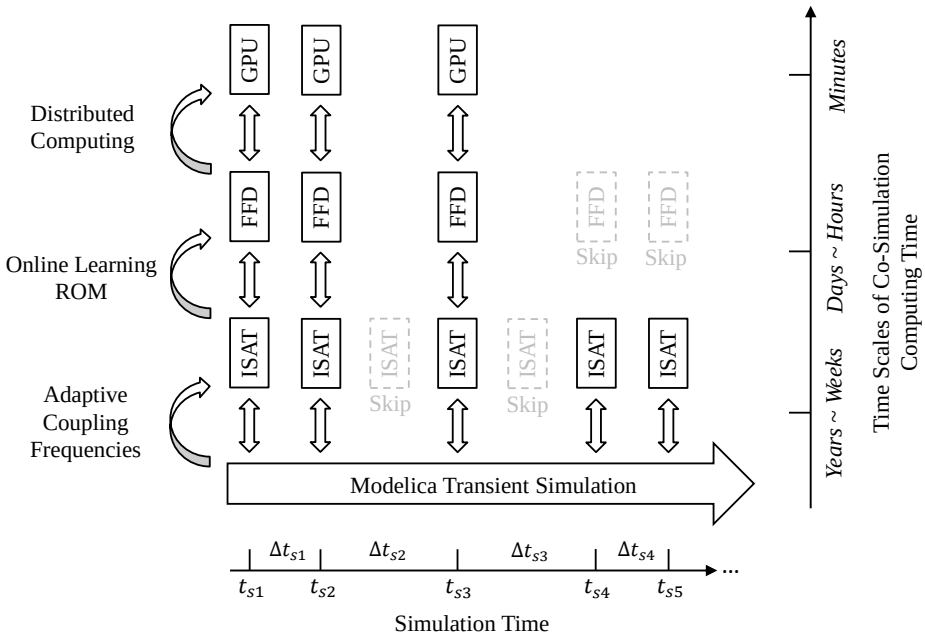


Figure 5-1 Coupling stratergy of the proposed adpative Modelica-ISAT-FFD model powered by distributed computing

To realize the practical use of BES-CFD co-simulaiton based optimization, the key is to significantly reduce the computational costs of the BES-CFD co-simulaiton to an acceptable level.

Since the bottleneck of the BES-CFD co-simulaiton speed lies in the computationally expensive CFD simulations, there are two ways to accelearate the co-simulaiton. One the one hand, it can be done by reducing the number of CFD simulations. On the other hand, accelerating the CFD simulaiton itself can also help. To achieve both, this paper proposes an adaptive BES-ISAT-CFD coupled model powered by disbributed computing.

Figure 5-1 shows the stratergy of the proposed adpative Modelica-ISAT-FFD model powered by distributed computing. First, an adaptive coupling frequency stratergy is proposed to reduce the number of calls of external solvers (i.e. FFD or ROMs) for steady state FFD results from the Modelica. The idea is that if the data sent from Modelica to FFD or ROMs does not change considerably in a certain period, the calls of FFD or ROMs will be skiped and Modelica adopts steady state FFD results returned at the beginning of that period. As a result, this stratergy avoids redundent FFD calls, which, otherwise, lead to high additional computatinal costs. Even if ROMs are coupled with BES, this stratergy reduces the computing time of exhanging data between two programs and loading ROMs.

Second, an online learning ROM, namely ISAT, is adopted to further reduce the number of FFD calls while ensuring the accuracy. One the one hand, ISAT can significantly reduce the computational costs by performing linear approximations instead of calling FFD, which is similar as a conventional ROM. On the other hand, with the online learning feature, ISAT calls FFD for an accuracy prediction when a linear approximation can not provide sufficiently accuract results.

By using the above two techniques, the time scale of perfoming a BES-CFD co-simulation can be reduced to days ~ hours. However, this may still not be fast enough since dozens of or even hundreds of runs of simulations are required to evalute the objective function during the

optimization. To further significantly speed up the FFD simulaitons, a distributed computing technique is integrated in the coupled model. The Modelica simulation, which is much faster than CFD, runs on a CPU and CFD, which is the speed bottleneck, runs on a GPU in parallel.

5.2.2 Co-Simulation Based Optimization Platform

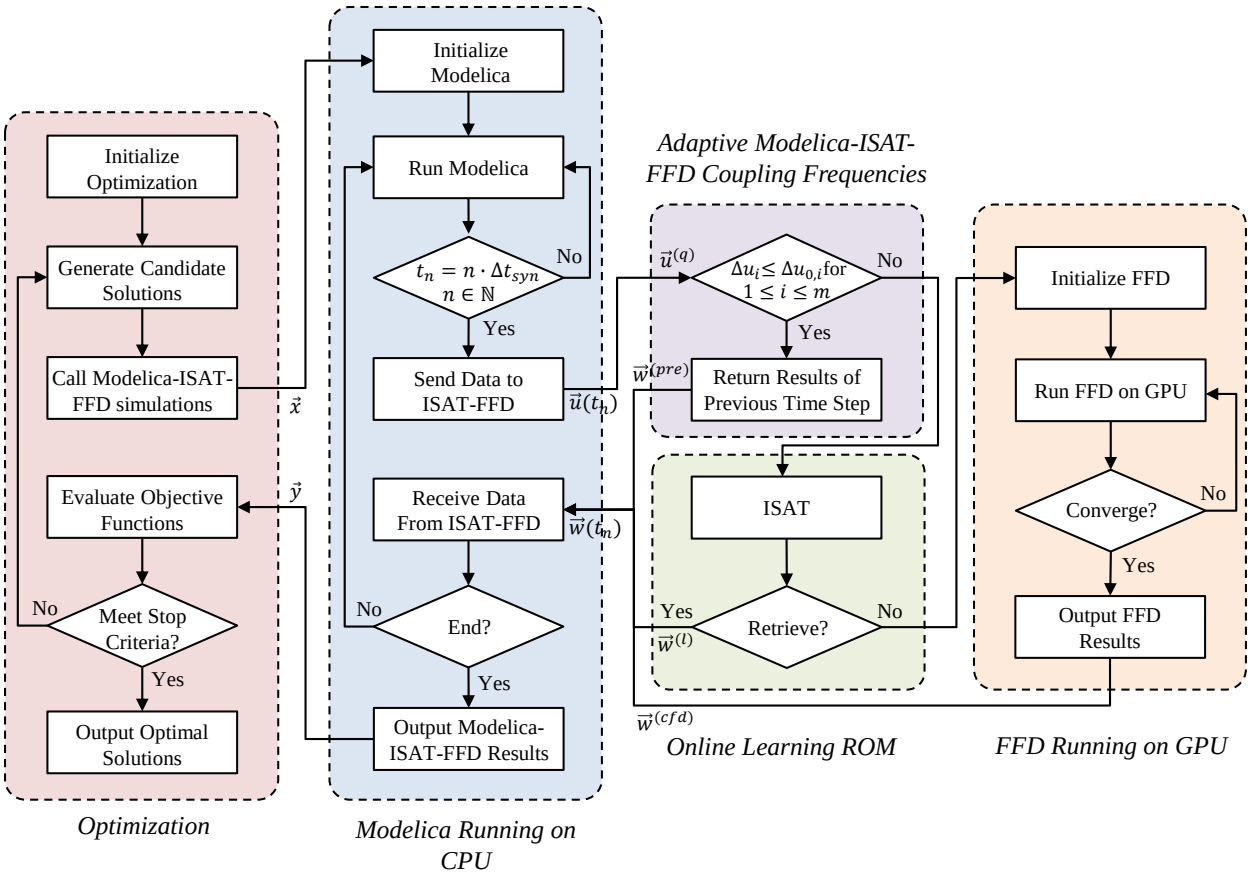


Figure 5-2 The optimization platform based on an adaptive Modelica-ISAT-FFD powered by distributed computing

An optimization platform is then proposed by linking the adaptive Modelcia-ISAT-FFD coupled model with an optimization program. By integrating all these techniques, we finally make such a BES-CFD co-simulation based optimization become computationally feasible. It is

noteworthy that this paper only focuses on data centers, but the proposed methodology can be used for other applications, such as optimal design and operation of displacement ventilations.

The workflow of the optimization platform is shown in *Figure 5-2*. Assume we have the following multi-objective optimization problem:

$$\min_{\vec{x}} \sum_{i=0}^{N} \int_{t_1}^{t_2} \omega_i \cdot y_i(\vec{x}) \mid y_i \in \vec{y}, \tag{5-1}$$

where $\vec{x}$ and $\vec{y}$ are vectors of optimization variables and objectives, respectively. w_i is the weight for the objective y_i.

During the optimization, the optimization engine generates candidate solutions $\vec{x}$ and call Modelcia-ISAT-FFD simulations to evaluate the objective functions $\vec{y}$ with the candidate solutions. The key feature of the optimization platform is the Modelcia-ISAT-FFD model with adaptive coupling frequencies and distributed computing. The workflow is as follows.

Initialize and run Modelica. When the data synchronization time t_n is reached:

$$t_n = n \cdot \Delta t_{syn} \mid n \in \mathrm{N}, \tag{5-2}$$

BES calculates $\vec{u}(t_n)$:

$$\vec{u}(t_n) = \vec{g}(\vec{x}, \vec{w}(t_{n-1})), \tag{5-3}$$

where $\vec{u}$ is the vector of data sent from Modelica to FFD and $\vec{w}$ is the vector of data sent from FFD to Modelica. $\vec{g}$ represents the functions solved in Modelica to calculate $\vec{u}$.

BES sends $\vec{u}(t_n)$ to ISAT-FFD:

$$\vec{u}(t_n) \rightarrow \vec{u}^{(q)} \tag{5-4}$$

To return the data $\vec{w}(t_n)$ from FFD to Modelica, there are three scenarios.

Scenario 1: when the boundary conditions at t_n updated from Modelica do not have notable changes compared to that at t_{n-1}:

$$\Delta u_i = |u_i(t_n) - u_i(t_{n-1})| \leq \Delta u_{0,i} \text{ for } 1 \leq i \leq m \tag{5-5}$$

where $\Delta u_{0,i}$ is the user-defined tolerance for the change of ith boundary condition from t_{n-1} to t_n. m is the number of elements in $\vec{u}$.

Then, return $\vec{w}(t_n)$ with the results at previous time step:

$$\vec{w}^{(pre)} = \vec{w}(t_{n-1}) \rightarrow \vec{w}(t_n) \tag{5-6}$$

Scenario 2: If the condition of Scenario 1 is not true, call ISAT. If the ISAT can retrieve the outputs from the database, then return $\vec{w}(t_n)$ from performing a linear approximation based on the closest record $\vec{u}^{(0)}$ to $\vec{u}^{(q)}$ in the database:

$$\vec{w}^{(l)} = f\left(\vec{u}^{(0)}\right) + A\left(\vec{u}^{(0)}\right) \cdot \left(\vec{u}^{(q)} - \vec{u}^{(0)}\right) \rightarrow \vec{w}(t_n), \tag{5-7}$$

where $y^{(l)}$ represents output of the nonlinear function from a local linear approximation. $f\left(x^{(0)}\right)$ is the mapping $f: x \rightarrow y$ with regard to $x^{(0)}$ and $A(x^{(0)})$ is the mapping gradient matrix with regard to $x^{(0)}$, which are stored in the database.

Scenario 3: If the ISAT fails to retrieve the outputs from the database, then return $\vec{w}(t_n)$ from executing a FFD simulation:

$$\vec{w}^{(cfd)} = \vec{h}(\vec{x}, \vec{u}(t_n)) \rightarrow \vec{w}(t_n), \tag{5-8}$$

where $\vec{h}$ represents the functions solved in FFD to calculate $\vec{w}$.

Until now, the data synchronization between Modelica and FFD at t_n is completed. Then, continue to run Modelica and repeat above process for the next data synchronization t_{n+1} until the end of the Modelica -ISAT-FFD simulation.

After finishing the Modelica -ISAT-FFD simulation, the results are sent to the optimization engine to evaluate the objective functions $\vec{y}$. The optimization engine repeats above process from generating candidate solutions to evaluating objective functions by calling the BES-ISAT-CFD simulations until finding the optimal solution.

5.2.3 Software Implementation

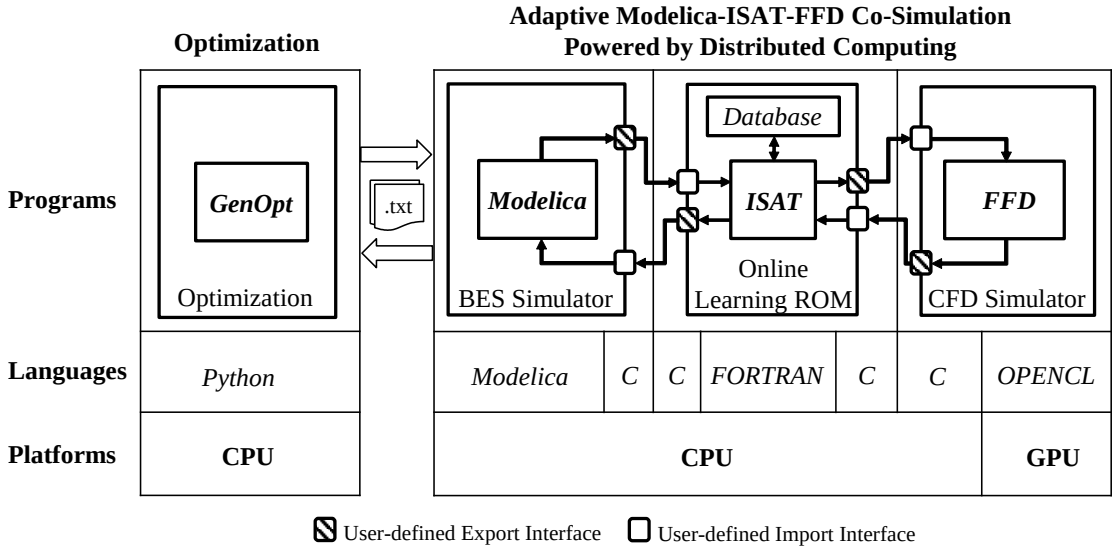


Figure 5-3 Implementation of the co-simulation based optimization platform

The adaptive BES-ISAT-CFD model is demonstrated with a Modelica-ISAT-FFD model. The implementation of the Modelica-ISAT-FFD was described in Chapter 4. The adaptive coupling frequencies are realized from revising the interface between Modelica with ISAT. The new feature for distributed computing is realized by implementing the FFD solver with a hybrid C and OpenCL programming languages. The FFD model is linked with the ISAT model through a wrapper in C. The optimization platform is implemented through a standard approach in GenOpt [119]. In each generation of the optimization, the GenOpt generates candidate solutions for the

optimization variables and the coupled simulations of Modelica with ISAT and FFD are performed to evaluate the objective function for these candidate solutions.

5.3 Verification

5.3.1 Case Description

The studied data center is a real middle-size data center with raised-floor and dropped ceiling configuration located in Massachusetts. The total white space floor area of the reference data center is approximately 690 m^2 (7,400 ft^2). As shown in *Figure 3-9*, the reference data center is approximately 30.5 m (100 ft) long, 22.6 m (74 ft) wide and 3.4 m (11 ft) high from the raised floor to the ceiling. Total power consumption by the 151 racks and 12 PDUs is approximately 344 kW. Racks G11 and G13 (i.e., the 11^{th} and 12^{th} cabinets in Row G) are empty. There are 18 45U networking racks in Rows 1 and 10, and all the remaining racks in the data center have a capacity of 42U.

The central Air Handling Units (AHUs) supply cooling airflow through the short sides of the plenum, which is supported using 7/8" stanchions. The total supply air flow rate is 152,000 m^3/hr (89200 cfm). The airflow is then supplied to the IT equipment through 183 25%-open-area tiles (each 2 ft by 2 ft). The heated airflow returns to the CRAH through a dropped ceiling plenum with 42 approximately-83%-open-area tiles.

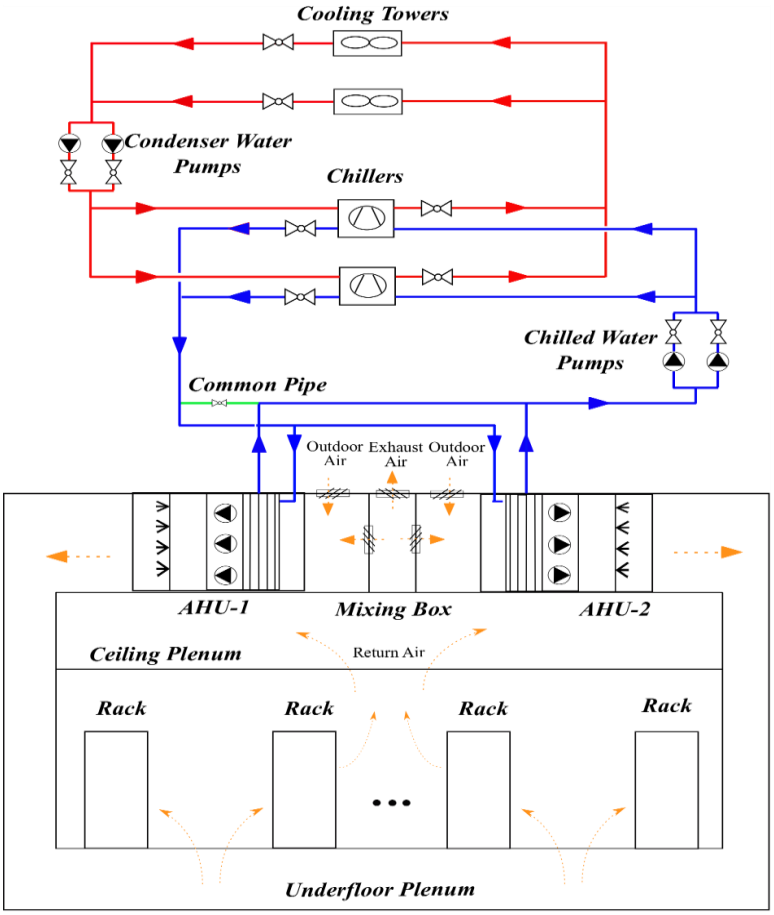


Figure 5-4 Schematics of the cooling system in the data center

The schematic drawing of the HVAC system of the subject data center is shown in *Figure 5-4*. There are three chillers, chilled water pumps, condenser water pump and cooling towers in the schematic drawing, but only two of them run in Lead/Lag configuration and the other one is for backup. The chilled water is distributed to the AHU-1 and the AHU-2. The cooling system compromises of the following components:

- 2 chillers which operate in a Lead/Lag configuration.
- 2 variable speed primary chilled water pumps which operate in a Lead/Lag configuration.
- 2 constant speed condenser water pumps which operates in a Lead/Lag configuration.

- 2 cooling towers, each tower with a variable frequency driver (VFD) driven fan and basin heater which operate in a Lead/Lag configuration.
- 2 AHUs with speed controls based on the static pressure in the plenum.

The cooling system configures a chilled water system with airside economizers (ASEs), which operates in three cooling modes: (1) free cooling (FC) mode, where only ASEs are activated; (2) partial mechanical cooling (PMC) mode, where the chilled water system and ASEs work simultaneously; (3) fully mechanical cooling (FMC) mode, where only the chilled water system is used. The transition conditions between different cooling modes are shown in **Error! Reference s ource not found.**, which are determined based on the weather, control settings and system conditions.

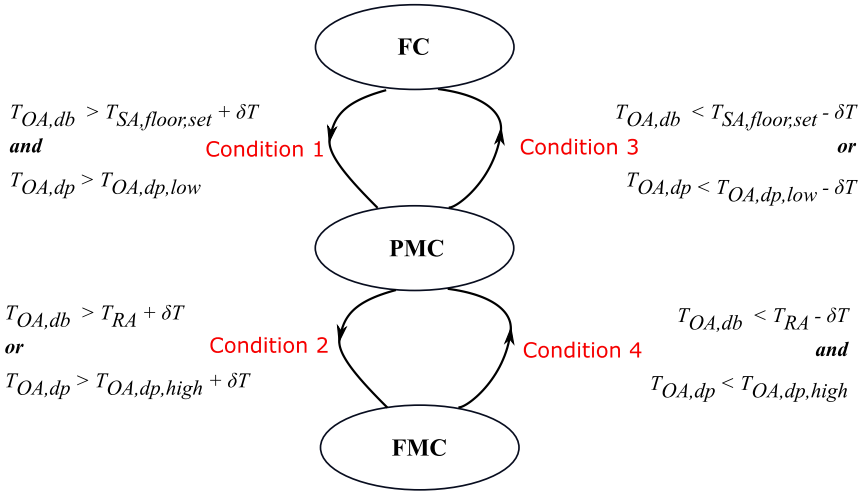


Figure 5-5 State graph of the cooling mode controller

The ASEs are activated and the cooling system stages from the FMC mode to the PMC mode, when the firing condition 4 in *Figure 5-5* is met:

$$T_{OA,db} < T_{RA} - \delta T \text{ and } T_{OA,dp} < T_{OA,dp,high}, \quad (5\text{-}9)$$

and deactivated (from the PMC mode to the FMC mode) when the condition 2 in *Figure 5-5* is met:

$$T_{OA,db} > T_{RA} + \delta T \text{ and } T_{OA,dp} > T_{OA,dp,high} + \delta T, \quad (5\text{-}10)$$

where $T_{OA,dp}$ is the dew point temperature of outdoor air, $T_{OA,db}$ is the dry bulb temperature of outdoor air, T_{RA} is dry bulb temperature of the return air, $T_{OA,dp,high}$ is the high cutoff limit for the dew point temperature, and δT is the temperature dead band. While $T_{OA,dp}$, $T_{OA,db}$, and T_{RA} are read from measured data, $T_{OA,dp,high}$ and δT are set as 14 ℃ and 1.1 ℃ in the current system respectively.

The chillers are activated and the cooling system stages from the FC mode to the PMC mode when the condition 1 in *Figure 5-5* is met:

$$T_{OA,db} > T_{SA,floor,set} + \delta T \text{ and } T_{OA,dp} > T_{OA,dp,low}, \quad (5\text{-}11)$$

and deactivated (from the PMC mode to the FC mode) when the condition 3 in *Figure 5-5* is met:

$$T_{OA,db} < T_{SA,floor,set} - \delta T \text{ and } T_{OA,dp} < T_{OA,dp,low} - \delta T, \quad (5\text{-}12)$$

where $T_{SA,floor,set}$ is the temperature setpoint of the supply air in the underfloor plenum, and $T_{OA,dp,low}$ is the low cutoff limit of the dew point temperature. $T_{SA,floor,set}$ is set as 22.2 ℃ in the baseline and it is dynamically adjusted by the optimal controller to maximize the operating time of the free cooling mode in the optimization case. The $T_{OA,dp,low}$ is set as 12 ℃ in the current study. Note that the humidity of the data center is controlled by using limits of dew point temperature of outdoor air. For more detailed description of the cooling system, please refer to Fu, Zuo, et al. [10].

5.3.2 Modelica Model

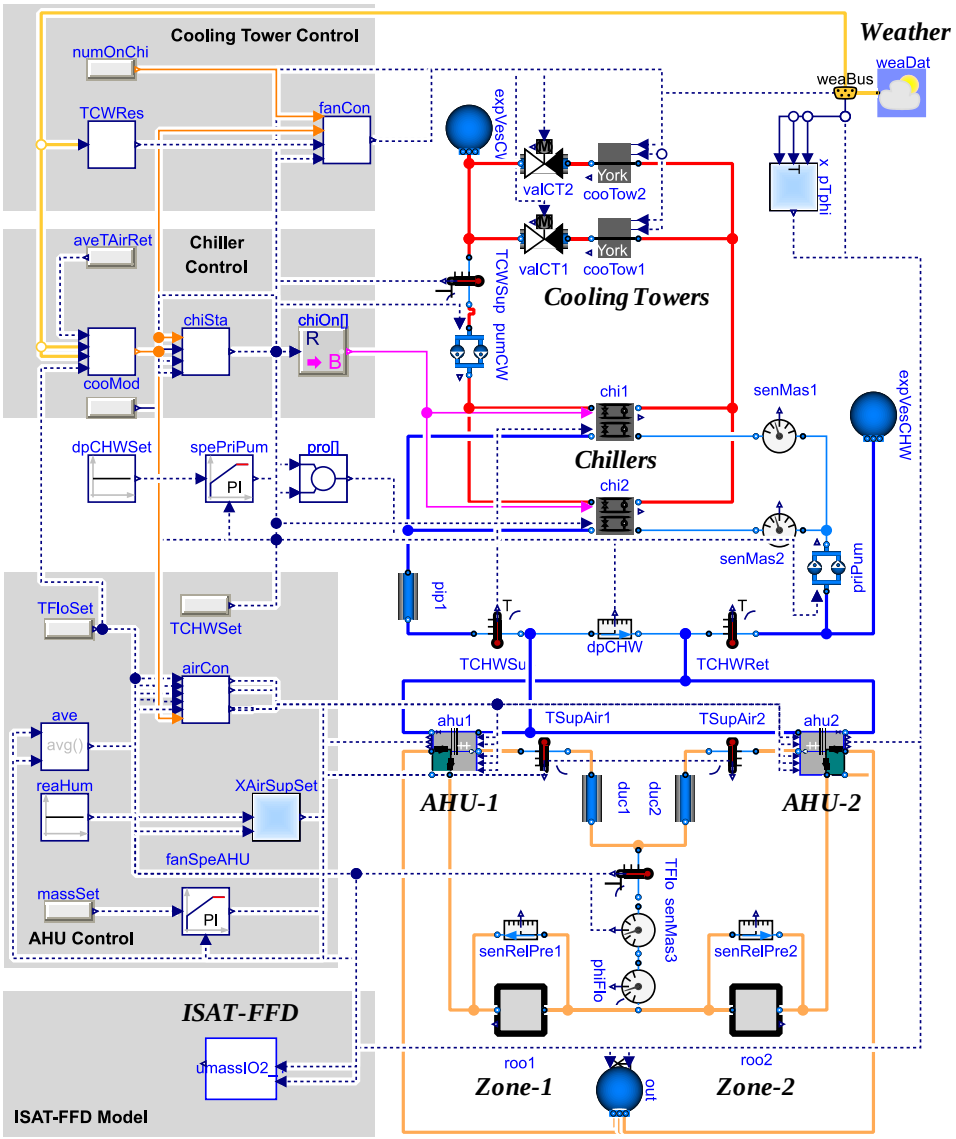


Figure 5-6 Implementation of the cooling system model in Modelica

The cooling system is modeled in Modelica based on Modelica Buildings library [12], as shown in *Figure 5-6*. The major component models include chillers, AHUs, cooling towers, chilled water pumps and condenser water pumps. The control system consists of the cooling mode control

and local controls. In *Figure 5-6*, the *Cooling Tower Control* represents the staging control and the speed control of the cooling tower fans. The *Cooling Mode Control* describes the cooling mode controller. The differential pressure control and the flowrate control of the evaporators are modelled in the *Primary Pump Control*. The supply air temperature and the chilled water supply temperature reset strategy, and the reheater control are described in the *AHU Control*. The settings of the reheater control are different in this study compared to Fu, Zuo, et al. [10]. To save the reheater energy, the reheater is only activated when the supply air temperature is lower than the allowable lower threshold according to [15]. For more detailed description of the cooling system, please refer to Fu, Zuo, et al. [10].

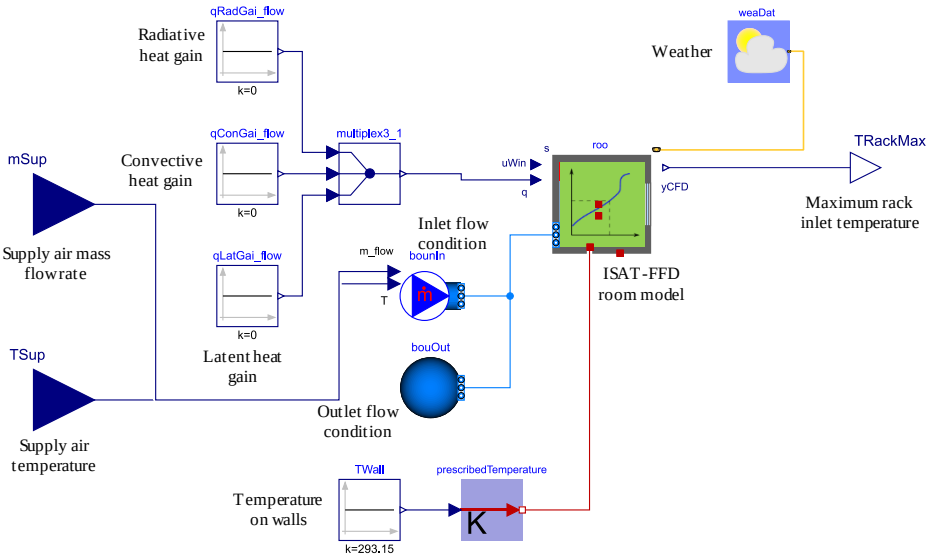


Figure 5-7 Diagram of the Modelica model for the ISAT-FFD model

The key component model to realize the Modelica-ISAT-FFD model is the *ISAT-FFD* model as shown in *Figure 5-6*. The inputs of the *ISAT-FFD* model are connected with the sensors of supply air mass flow rate and temperature. The output of the *ISAT-FFD* model is the maximum

rack inlet temperature, which can be used for control. The detailed implementation of the *ISAT-FFD* model is shown in *Figure 5-7*. The radiative, convective and latent heat gains are set to zero. The boundary conditions of walls are defined as constant values since the cooling loads injected through walls are much less than that caused by heat dissipation from IT servers and therefore are neglected in this study. The room model exchanges data with the ISAT model and the ISAT model determines to return the prediction through retrieving from the database or calling FFD to perform a full simulation.

5.3.3 FFD Model

The plenum and white space of the data center are modeled separately in the FFD model with structured grids for the simulation. The FFD model performs transient-state simulations with a time step size of 0.2 s – 1.0 s and simulation time of 100s – 400s. A zero-equation turbulence model is adopted [96, 97]. The FFD simulation is performed on an AMD FirePro^{TM} W8100 GPU. For more details of the FFD model, please refer to Section 3.3.2.

The FFD model was validated with experimental measurements in Section 3.3.2. It was found that 95.1% of the FFD predictions deviate by less than 5% from the experimental measurements for the plenum and 88.0% predictions of the FFD predictions deviate by less than 10% from the experimental measurements. Therefore, the FFD model generally capture the airflow distribution in the plenum and rack-to-rack inlet temperature variation well. The FFD model achieves a similar level of sufficient accuracy compared to CFD even though some discrepancies between simulated and measured temperatures can be observed. For more details, please refer to [120].

5.3.4 Results

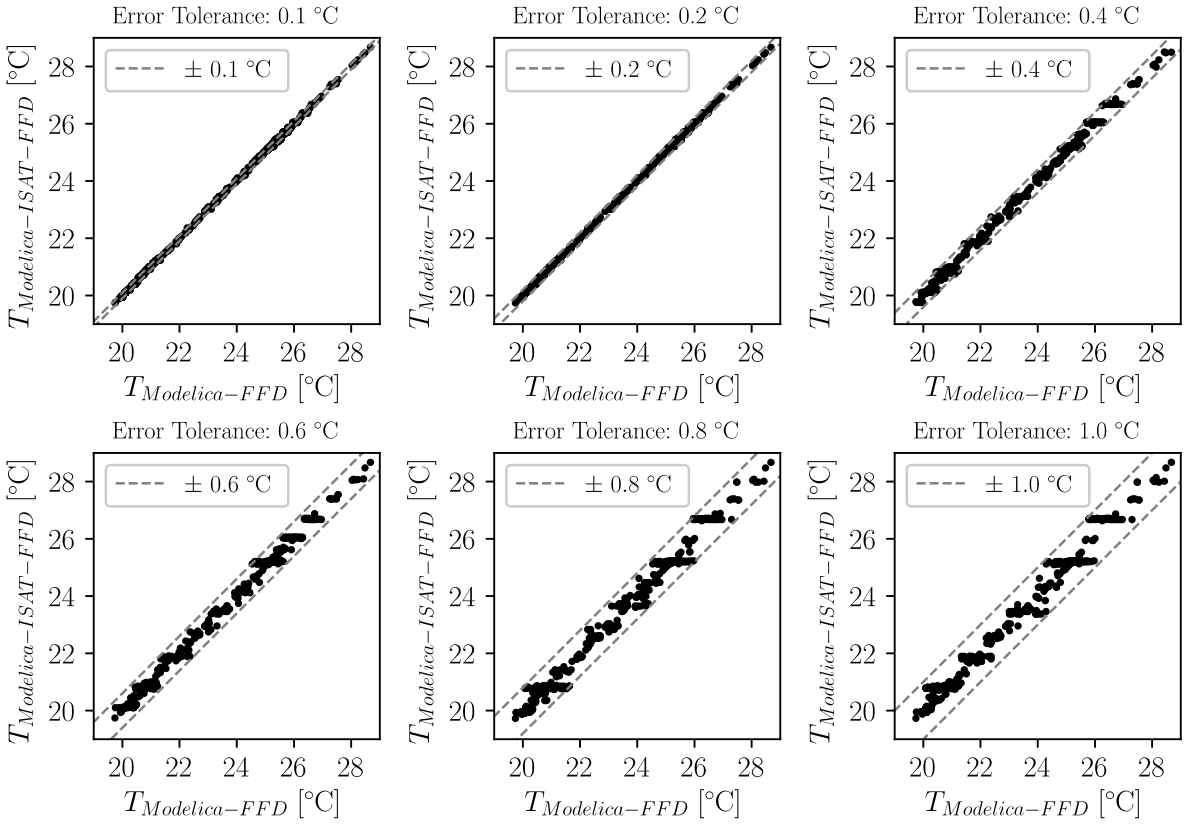


Figure 5-8 Prediction error of the Modelica-ISAT-FFD model with different settings of error tolerances compared to the Modelica-FFD model

Figure 5-8 compares the predictions from the Modelica-ISAT-FFD model with different settings of error tolerances and the Modelica-FFD model in details. It is found that the errors of predictions with the Modelica-ISAT-FFD model against the Modelica-FFD model are within the user-defined error tolerances. Generally, reducing the error tolerance in Modelica-ISAT-FFD can make the prediction more accurate. Even though there are some predictions outside of the user-defined error tolerance, the prediction accuracy with the Modelica-ISAT-FFD is generally well controlled.

Table 5-1 Number of actions and estimated computing time

Models	Error Tolerance [°C]	Number of Actions				Computing Time [hour]
		Directly Calling FFD	Add	Grow	Retrieve	
Modelica-FFD	-	8,760	0	0	0	146
Modelica-ISAT-FFD	0.1	0	50	32	8,678	1.4
	0.2		29	32	8,699	1.0
	0.4		18	24	8,718	0.7
	0.6		13	21	8,726	0.6
	0.8		9	28	8,723	0.6
	1.0		8	22	8,730	0.5

The number of actions and estimated computing time for this case with different models are shown in *Table 5-1*. With the adaptive Modelica-ISAT-FFD model, most predictions are obtained from retrieving from the database, which do not require FFD simulations. Taking advantages of the distributed computing, the Modelica-ISAT-FFD models with error tolerances of from 0.1 °C to 1.0 °C save the computing time by from 99.04% to 99.66%, compared to the Modelica-FFD model. To balance the accuracy and computational costs, the error tolerance of 0.4 °C is used for the following case studies.

5.4 Case Studies

5.4.1 Description of the Holistic Optimization

5.4.1.1 Optimization Problem

In this study, the optimization problem is to minimize the energy consumption of data center cooling systems with air-side economizer (ASE) while meeting the thermal requirements of data centers [15]. The optimization variables is the supply air flowrate:

$$\min_{\dot{m}_{sup}} \sum_{i=0}^{N} \int_{t_1}^{t_2} E_i, \tag{5-13}$$

s.t. $\dot{m}_{sup,min} \leq \dot{m}_{sup} \leq \dot{m}_{sup,max}$

$T_{rack,max} \leq 27℃$

where, the ranges of $\dot{m}_{sup}$ is constrained according to the configurations of the cooling systems. The maximum rack inlet temperature should be equal or lower than 27 ^{o}C according to the thermal guideline by ASHRAE [15]. The control of the $T_{rack,max}$ is achieved through a supply air temperature setpoint reset controller.

5.4.1.2 Supply Air Temperature Setpoint Reset Control

The non-uniform thermal environment is controlled through a supply air temperature setpoint reset control as shown in *Figure 5-9*. The setpoint of the supply air temperature is adjusted according to the $T_{rack,max}$, which is predicted by the ISAT-FFD model based on the system status. The control objective is to maintain the $T_{rack,max}$ at 26 ±1 ^{o}C. When the $T_{rack,max}$ is higher than the target, the setpoint of supply air temperature will be lowered and vice versa. The key

component in this controller is the ISAT-FFD model that predicts the critical information of non-uniform thermal environment.

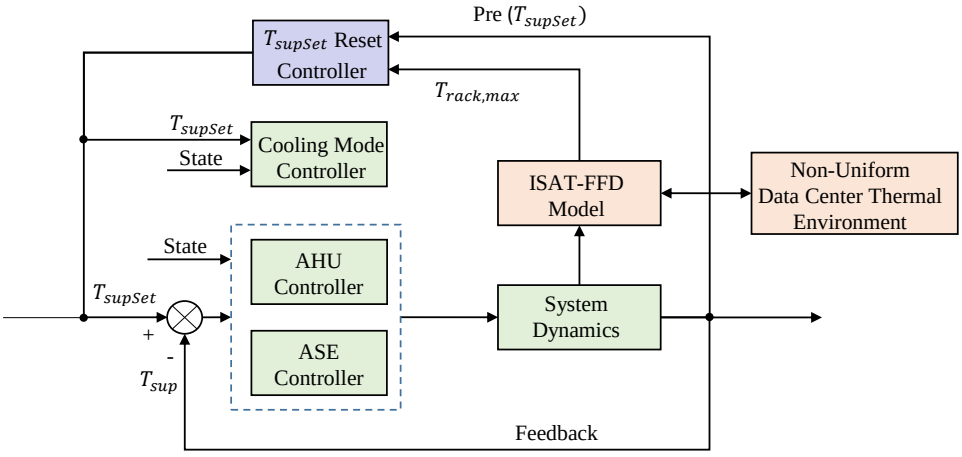


Figure 5-9 Setpoint of the supply air temperature reset control

5.4.2 Evaluation of the Optimization Platform

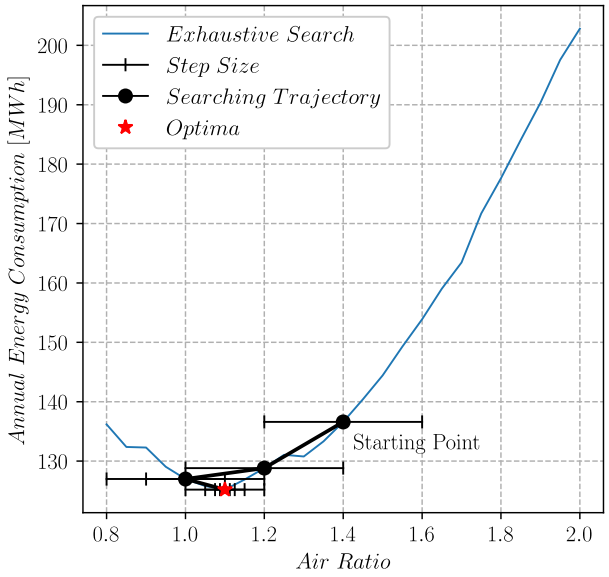


Figure 5-10 Searching trajectory with the GPS optimization method

The optimization is implemented based on GenOpt [119] as introduced Section 6.2.2. The generalized pattern search (GPS) optimization method provided in GenOpt [119] is adopted in this study. At each iteration of the GPS optimization, the pattern either moves to the point which best minimizes its objective function, or reduces the step size by a user-defined factor (2 in this study) if no point can be found to be better than the current base point, until the desired accuracy has been achieved [121]. The range the optimization variable, namely air ratio, is from 0.8 to 2.0. The initial step size to perform the GPS algorithm is defined as 0.2.

The GPS based optimization is evaluated by comparing its results with that from an exhaustive search. The starting point in the GPS optimization is at 1.4 and it moves to 1.2, 1.0 and finally 1.1 during the optimization. It is found that the GPS optimization finds the global optima successfully.

Table 5-2 Number of actions and estimated computing time for the optimization with the different models

Optimization With Different Models	**Coupling Frequencies**	**Distributed Computing**	**Offline Stage**		**Online Stage**			**Estimated Computing Time**
			Add	Grow	Directly Calling FFD	Add	Grow	
Modelica-FFD	Constant	No	0	0	1024920	0	0	585 years
Modelica-FFD	Adaptive	No	0	0	113880	0	0	65 years
Modelica-ISAT-FFD	Adaptive	No	334	708	0	18	74	47.25 days
Modelica-ISAT-FFD	Adaptive	Yes	334	708	0	18	74	0.78 days

Table 5-2 shows the total number of actions and estimated computing time for the optimization with the different models. Using the original Modelica-FFD model with constant coupling frequencies and without distributed computing, FFD is called at each data synchronization time resulting in a 1024920 times of FFD calls in total and corresponding 585 years of computing time. Taking advantage of the adaptive coupling frequencies, FFD is called for 113880 times, which leads to a reduced computing time to 65 years. Taking advantage of the online learning, FFD is called for only 92 times, which leads to a significantly reduced computing time to 47.25 days. Powered by the distributed computing, an additional 98% of reduction on the computing time is achieved. As a result, the total computing time eventually reaches as less as 0.78 days, which becomes acceptable for practical use in real applications.

5.4.3 Optimization Results

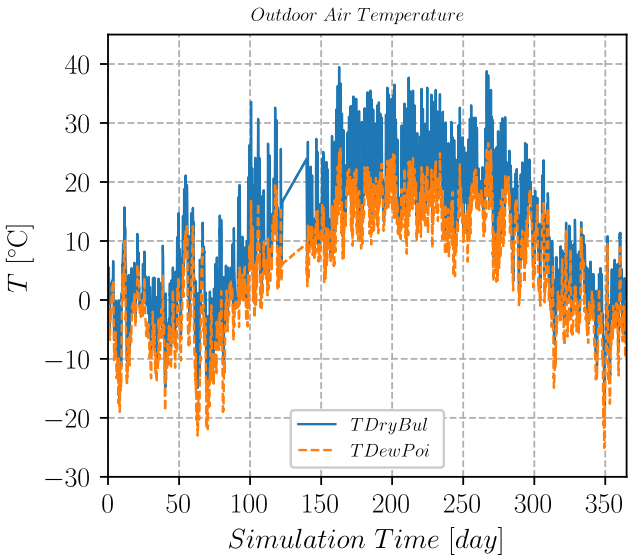


Figure 5-11 Weather profiles for the studied case in Massachusetts

The baseline has constant setpoints for supply air temperature and mass flow rate as description in Section 6.3.1. The optimization is performed with different optimization frequencies including annually, monthly, weekly and daily optimization. The studied case is a real data center in Massachusetts as described in Section 6.3.1. The weather profiles are downloaded from the closest meteorological station from the data center and shown in *Figure 5-11*. The dry bulb temperatures are as low as more than -20 ºC in winter and as high as 40 ºC in summer.

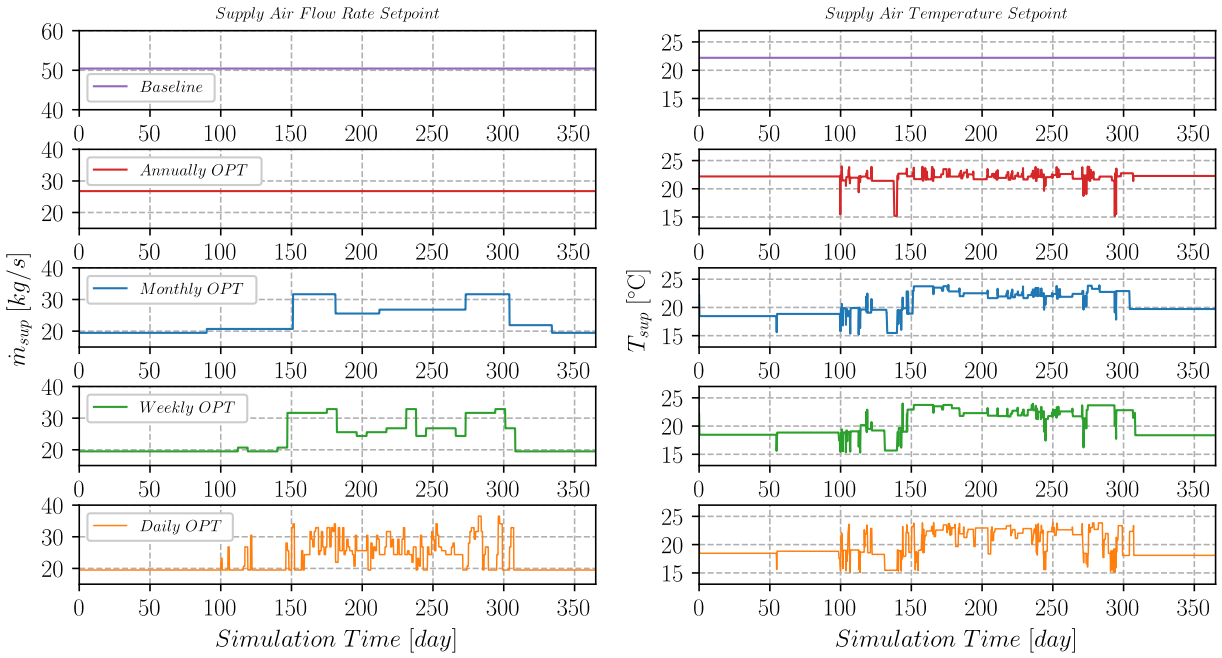


Figure 5-12 Control actions in baseline and optimization cases

The optimization results are shown in *Figures 5-13~5-17* in terms of the control actions, performance of energy efficiency and thermal environment control. Control actions in baseline and optimization cases with different optimization frenqucies are shown in *Figure 5-12*. In the optimization cases except the annually optimization, the optimal setpoints of supply air mass flow rate are determined to be at the lower limit in winter and with higher values in summer. This is

because the fan energy is dominant in winter since the cooling system operates on the free cooling mode and consequently a supply air mass flow rate at the lower limit saves the fan energy to the maximum extent. Correspondingly, the setpoints of supply air temperature are adjusted to a lower level to meet the thermal requirements of data centers. On the other hand, in the summer, higher supply air flow rates and corresponding higher supply air temperatures are determined to get more free cooling and save chiller energy.

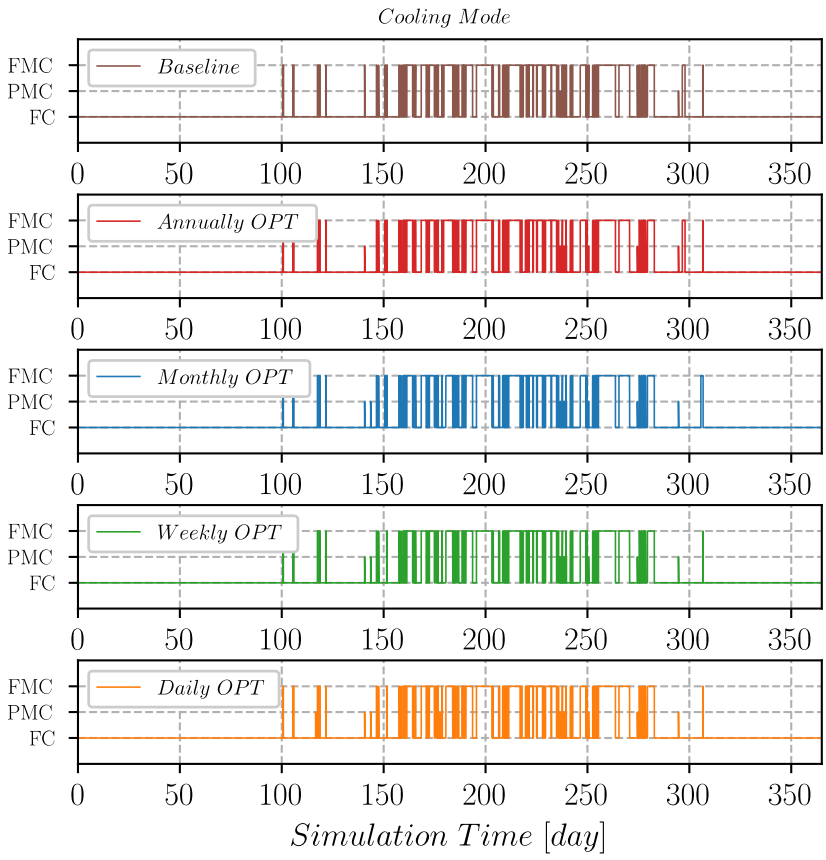


Figure 5-13 Cooling modes in baseline and optimization cases

The Cooling modes in baseline and optimization cases with different optimization frenqucies are shown in *Figure 5-13*. The baseline runs with a high supply air flow rate and temperature so that it performs well regarding the usage of free cooling to save energy. As a result,

the optimization cases are intended mostly to save energy through reducing the fan energy and meanwhile avoiding or reducing the sacrifice of free cooling instead of getting more free cooling. The hours running on the free cooling mode in the optimization cases are generally similar as that in baseline.

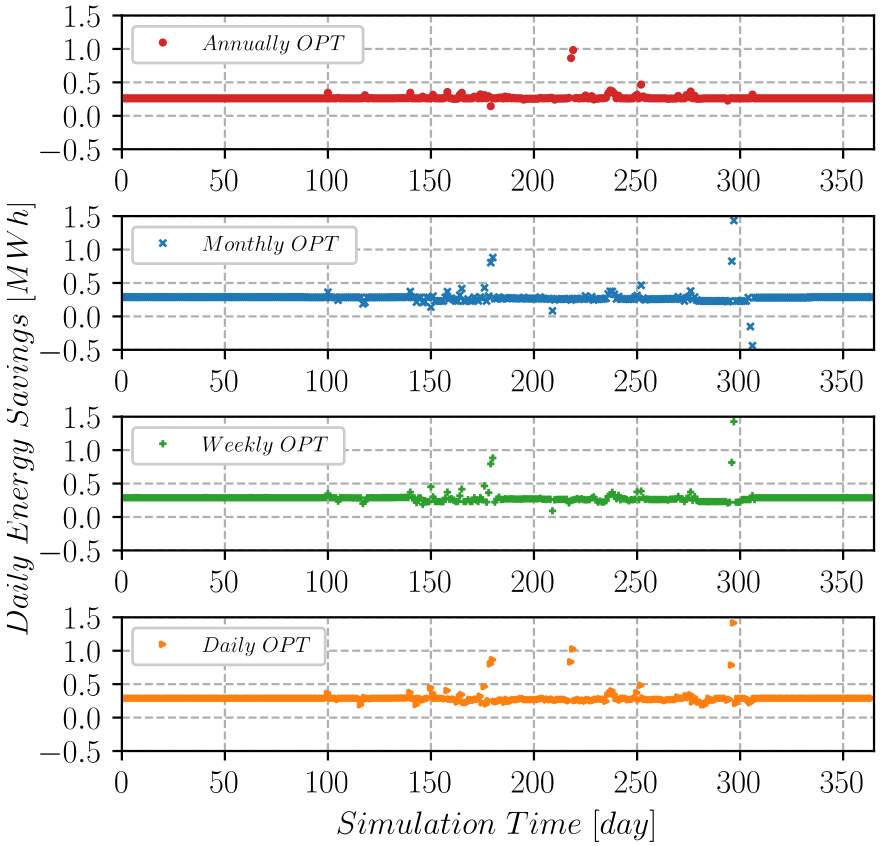


Figure 5-14 Daily energy savings for optimization cases with different optimization frenqucies compared to the baseline

The *Figure 5-14* shows the daily energy savings for optimization cases with different optimization frenqucies compared to the baseline. As analyzed above, most energy is saved from fan energy. This is also shown in *Figure 5-15*. Interestingly, obvious higher or lower energy savings are achieved on some days than that on other days. This is because the changes of cooling

modes the cooling system runs on those days. For example, significant energy can be saved if the cooling system switches from mechanical cooling to free cooling from adjusting the setpoints of the supply air temperature.

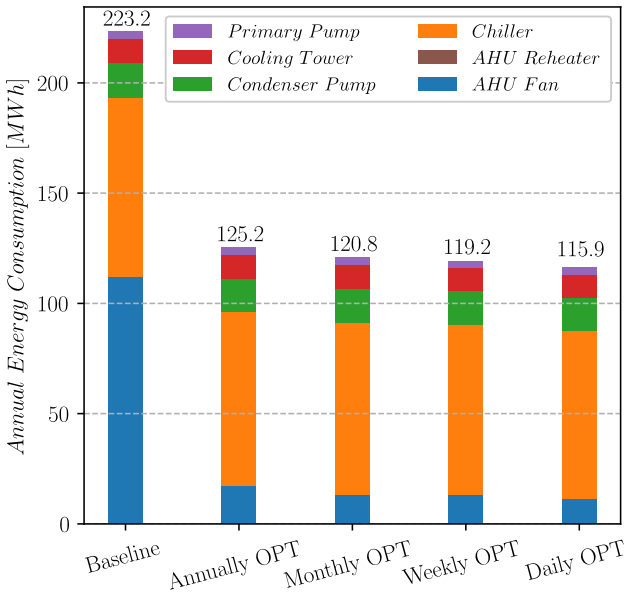


Figure 5-15 Annual energy consumption in baseline and optimization cases

As shown in Figure 5-15, the chiller and fan consume most of the energy in baseline, but only fan energy is reduced from optimizations. The chiller energy depends on the operating time of chiller. In baseline, it operates with a large supply air flow rate and high supply air temperature, which leads to that the baseline has almost maximum free cooling and minimum operating time of chiller. Therefore, the energy saving potential from chiller is limited. As a result, the energy savings from simulation-based optimizations are mostly obtained from the fan energy. Interestingly, it is found that annual optimization results are close to daily optimization results. By conducting an exhaustive search for two typical days, we found that the optimal setpoint for a specific day from annual optimization only has slightly worse performance than that from daily

optimization for both winter and summer conditions. In other words, the optimal setpoint from annual optimization is already fairly good in term of saving fan energy compared to that from daily optimization.

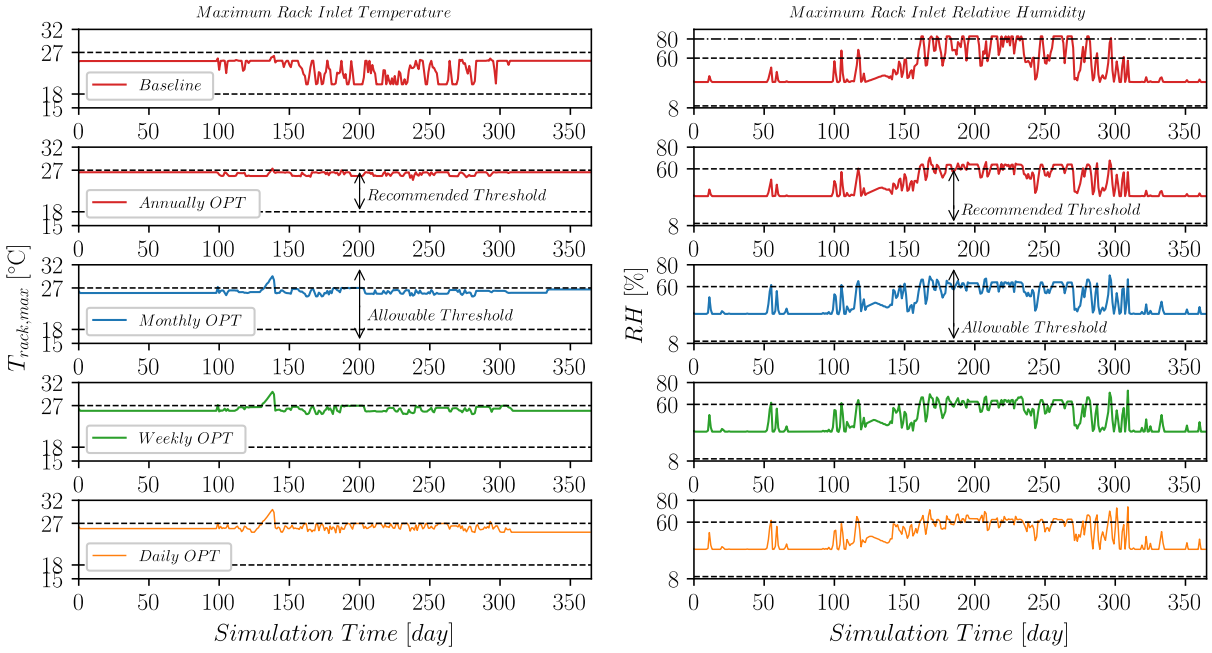


Figure 5-16 Maximum rack inlet temperature and relative humidity in baseline and optimization cases with different optimization frequencies

Except the energy efficiency, another important aspect for a holistic optimization is the airflow management. *Figure 5-16* shows the maximum rack inlet temperature and relative humidity in baseline and optimization cases with different optimization frequencies. As described in Section 2.1.2, the allowable and recommended ranges for rack inlet temperature are 15-32 ºC and 18-27 ºC, respectively, and the allowable and recommended ranges for humidity are 8-80% and -9-15 ºC DP and 60% RH, respectively. Most rack inlet temperatures and relative humidities are controlled within the recommended ranges. Though some are beyond the recommended ranges,

all the rack inlet temperatures and relative humidities are within the allowable ranges. To conclude, the holistic optimization saves energy by as much as 48.1% and meanwhile the thermal environment is well controlled to meet the requirements of data centers.

5.5 Discussions

For the optimization of the two parameters including supply air temperature and flow rate, some considerations are as follows. Ideally, we can optimize both parameters simultaneously. However, since the Modelica-ISAT-FFD co-simulation is still time consuming (though much faster than existing methods), we want to reduce the number of co-simulations as much as possible during the optimization. This is achieved by a dimension reduction in the optimization problem, in which only one optimization variable (supply air flow rate) is considered. By do this, the optimization only requires about 20 times of co-simulation to find the optima, which can be 100 or even more if two variables are optimized.

The idea of the dimension reduction is that if we optimize both parameters, the optimal value of the supply air temperature to minimize the energy consumption while satisfying the optimization constrains should be the highest temperature that can meet the thermal requirements (no generation of hot spots). This is because a higher supply air temperature setpoint is always good to get more chances for free cooling leading to a higher cooling efficiency in the studied case. Given that, we design a supply air temperature setpoint reset control, which can reset the supply air temperature setpoint to the highest value that can meet thermal requirements. This allows us to exclude the supply air temperature setpoint in the optimization and still be able to set it at the optimal value. Above all, the optimal setpoint of the supply air flow rate is obtained from simulation-based optimization on the offline stage (before operation), which remains constant at

each optimization period, e.g. one month or one week. The optimal setpoint of the supply air temperature is dynamically determined based on a local reset control during the operation.

In practical, it is more common to use a constant supply air flow rate and adjust the supply air temperature in the data center cooling systems. For example, our studied data center uses a fixed setpoint of supply air flow rate throughout the year. This is because the adjustment of supply air flowrate will change the airflow pattern in the data center, and this transition process may bring uncontrollable risks to generate hot spots. Note that this dissertation also demonstrates simultaneously optimizing these two parameters in Chapter 6 by using a real-time optimal control strategy.

5.6 Conclusions

This study proposed a holistic solution to simultaneously optimize the data center cooling system and airflow management based on an adaptive Modelica-ISAT-FFD model powered by distributed computing. A real middle size data center was adopted to validate the new model and we found that the rack inlet temperatures predicted by the proposed model are generally within the pre-defined error tolerances compared to the existing Modelica-FFD model. The total computing time of the holistic optimization was significantly reduced to as less as 0.78 days with the proposed model, while it is estimated to take many years with conventional Modelica-FFD models. The results of the case studies showed that the holistic optimization saves the annual energy consumption by as much as 48.1% and meanwhile hot spots involving non-uniform thermal environment are well controlled for reliable operations of data centers.

Chapter 6 Machine Learning Assisted Expert System for Optimal Control of Data Center Cooling Systems with Air Side Economizer

This chapter presents a machine learning assisted expert system (MLES) for optimal control of cooling systems with air side economizer in data centers. A model predictive control (MPC) method is also implemented to evaluate the performance of the proposed MLES approach.

6.1 Introduction

Data centers consume approximately 2% of total electricity in the United States, and about half of that is used for cooling. Many data centers are overcooled for reliable operation since overheating of IT servers can reduce performance, permanently damage the hardware, or in rare cases start fires. For instance, the cooling system normally supplies a fixed amount of cold air that is designed for handling the nominal IT load unless variable frequency drives (VFDs) are installed, which may lead to overcooling and a lower cooling efficiency at the partial load scenario. This study therefore aims to propose an optimal control strategy that dynamically optimize the cooling efficiency of data centers.

Free cooling has been considered as one of the most prominent ways to improve energy efficiency of data centers [122]. Among free cooling technologies, direct ASE has been adopted by about 40% of data centers utilizing free cooling [123], in which the outdoor air is used for cooling the IT servers when the weather satisfies certain conditions. Various efforts have been made to improve cooling efficiency by using ASE. Chen, Zhang, et al. [124] proposed a control strategy of ventilation cooling technology with ASE to maximize the energy savings while

ensuring the thermal environment in a telecommunication base station. [8] utilized simulation based methods to analyze the energy saving potential of using ASE in different climate zones. However, these research employed rule based control strategies, which use constant supply air flow rate. Such a fixed-flow fan may limit the energy saving potential since the fan energy may become dominant when the cooling system operates in free cooling mode. To dynamically optimize the supply air flow rate, one of the major challenges is to ensure the reliable operation by preventing the generation of hot spots, which usually involve non-uniform airflow pattern influenced by supply air flow rate and temperature, room temperature, leakage path and IT loads in data centers.

Model predictive control (MPC) has been widely used to realize the optimal control [125]. Zhou, Wang, et al. [126] implemented a MPC to minimize the cooling power while meeting the thermal requirements by dynamically tuning supply air temperature, flow rate and openings of adaptive vent tiles, in which an analytical model was developed for predicting rack inlet temperatures based on mass and heat balance principles. The coefficients and constant term in the analytical model represent the airflow pattern. For example, a constant C represents the rack inlet temperature increase caused by recirculation, which can be identified from system identification experiments. However, this simplification is not valid when the airflow pattern changes drastically. For example, when the air ratio (ratio of the total IT flow rate to the total supply air flow rate) increases from 1.0 to 2.0, the recirculation air entering the rack will be reduced significantly. Some simplifications are also made for predicting the energy consumption in the MPC. For example, they assumed the chiller power follows a linear relationship with the supply air temperature. This may be reasonable when the chiller is continuously running. But this is not valid if an ASE is

adopted, with which the chiller may be turned on or off under different cooling modes. To summarize, though the MPC can achieve good performance theoretically, formulating a precise and sufficiently fast model for MPC is not easy especially for complex systems, such as data center cooling systems with ASE, which involve non-uniform thermal environment and complex control logics.

As a model-free supervisory control method, an expert system is promising to address this problem. The key of the expert system is the capability to determine the optimal actions that minimize the cooling energy while preventing generation of hot spots for reliable operation. To realize that, this study proposes a novel machine learning assisted expert system (MLES) optimal control method. A model predictive control (MPC) method is also implemented to evaluate the performance of the proposed MLES optimal control. The rest of the chapter is organized as follows. First, the methodology is introduced including a MPC as the benchmark and the proposed MLES. Then two cases studies are performed to demonstrate the performance of the proposed MLES for optimal control of the cooling system with ASE in a real data center.

6.2 Methodology

In this study, the optimization problem is to minimize the energy consumption of data center cooling systems with air-side economizer (ASE) while meeting the thermal requirements of data centers [15]. Two optimization variables including supply air temperature and mass flowrate are considered.

$$\min_{T_{sup},\dot{m}_{sup}} \sum_{i=0}^{N} \int_{t_1}^{t_2} E_i, \tag{6-1}$$

s.t. $\dot{m}_{sup,min} \leq \dot{m}_{sup} \leq \dot{m}_{sup,max}$

$$T_{sup,min} \leq T_{sup} \leq T_{sup,max}$$

$$T_{rack,max} \leq 27℃$$

where, the ranges of $\dot{m}_{sup}$ and T_{sup} are constrained according to the configurations of the cooling systems. The maximum rack inlet temperature should be equal or lower than 27 ^{o}C according to the thermal guideline by ASHRAE [15].

The challenge of this problem is the control of the $T_{rack,max}$. For a data center with open-aisle configuration, the temperature of air entering the rack is supply cold air mixed with hot room air. Therefore, the $T_{rack,max}$ usually needs be predicted by detailed airflow simulation, such as CFD, instead of a well-mixed room model in conventional BES programs. In this study, a machine learning (ML) model is trained to handle this issue. The inputs of the ML model are IT load and supply air flowrate, and the output is the highest supply air temperature that can meet the thermal requirements. The training database is generated from a parametric study of FFD simulations.

6.2.1 Model Predictive Control (MPC) Approach

A conventional model predictive control (MPC) approach is implemented in this study as benchmark. The implementation generally follows a standard MPC approach [127] as shown in *Figure 6-1*. In each time step, an optimization is performed to determine the optimal actions, in which the objective function is evaluated through model-based simulations with the time period that covers the prediction horizon. In this study, the time step size and prediction horizon are defined as 10 minutes and 20 minutes as they achieve the best performance. The optimization workflow generally follows a similar procedure as is in Chapter 5. The difference is that the ISAT

model is replaced with a machine learning model to support realization of optimization in real-time.

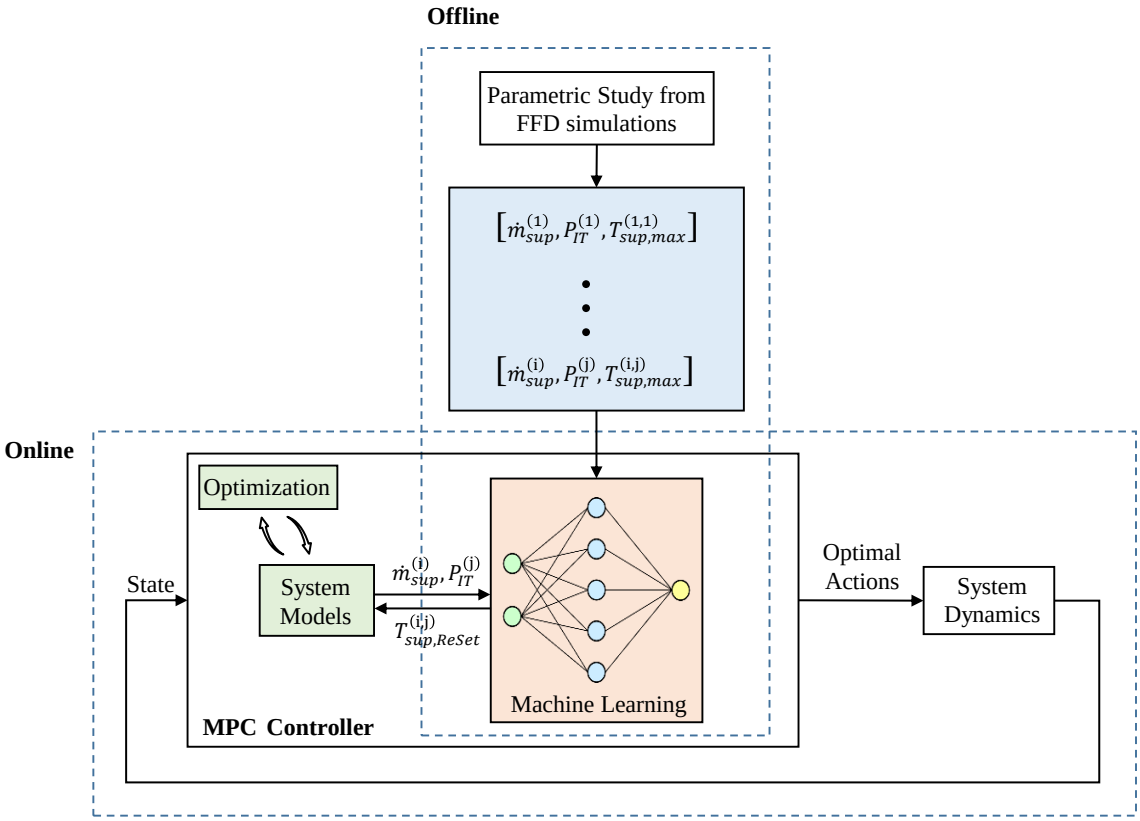


Figure 6-1 Schematic diagram of the model predictive control (MPC) approach

The difference here from a conventional MPC is the involvement of the ML model. The objective function is formulated as follows:

$$\min_{\dot{m}_{sup}} \sum_{i=0}^{N} \int_{t_1}^{t_2} E_i, \tag{6-2}$$

s.t. $\dot{m}_{sup,min} \leq \dot{m}_{sup} \leq \dot{m}_{sup,max}$

$T_{sup} = f(\dot{m}_{sup}, P_{IT})$ such that $T_{rack,max} \leq 27℃$

where, the range of $\dot{m}_{sup}$ is constrained according to the configurations of the cooling systems. The setpoint of T_{sup} is determined according to the ML model that predicts the highest supply temperature for a given air ratio (supply air flow rate) with a certain IT load level (rack power) to meet the data center thermal requirements [15].

6.2.2 Machine Learning Assisted Expert System (MLES) Approach

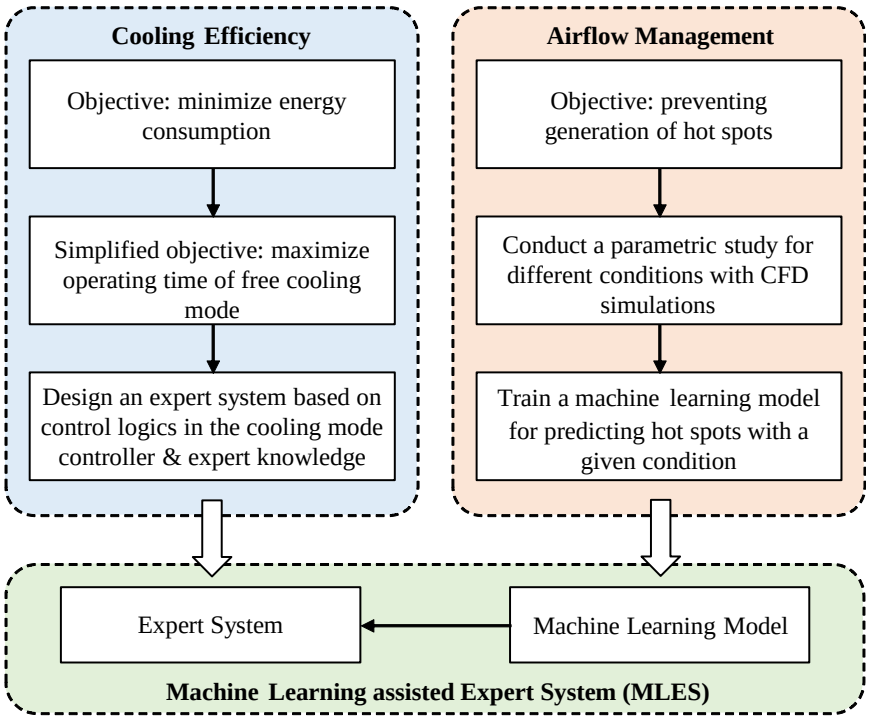


Figure 6-2 Workflow of the proposed MLES approach

The workflow of the proposed MLES is shown in *Figure 6-2*. First, the objective function of the optimization that minimizes the energy consumption of the cooling systems with ASE is simplified to maximize the operating hours of the free cooling mode:

$$\max_{T_{sup},\dot{m}_{sup}} \sum_{i=0}^{N} \int_{t_1}^{t_2} T_{i,FC}, \tag{6-3}$$

s.t. $\dot{m}_{sup,min} \leq \dot{m}_{sup} \leq \dot{m}_{sup,max}$

$T_{sup} = f(\dot{m}_{sup}, P_{IT})$ such that $T_{rack,max} \leq 27℃$

where, $T_{i,FC}$ is the total time that the cooling system runs on the free cooling mode, and the range of $\dot{m}_{sup}$ is constrained according to the configurations of the cooling systems. The setpoint of T_{sup} is determined according to the ML model that predicts the highest supply temperature for a given air ratio (supply air flow rate) with a certain IT load level (rack power) to meet the data center thermal requirements [15].

Then, an expert system is designed based on control logics in the cooling system and expert knowledge. To support the decision making in ES for reliable operation, a machine learning model is trained from CFD simulations to predict and control hot spots. The schematic diagram is shown in *Figure 6-3*. There are two key parts: one is the expert system (ES) and the other is the machine learning model (ML). The ES determines the optimal actions based on a pre-defined decision making process without calculating the cooling energy for candidate solutions and performing an optimization to find the optima. The decision making process in ES can be different for different cases. It is configured as shown *Figure 6-5* for the case demonstrated in this study.

The constraints of the optimization regarding the thermal requirements are satisfied through the ML model. The ML model is trained on the offline stage and called by the ES on online stage. Generally, the highest supply temperature increases with a greater air ratio. The MLES adjusts the setpoints of supply air flowrate and temperature to make the cooling system operate on the free cooling mode as much as possible and meanwhile saving the fan energy as much as possible. The detailed description will be introduced in Section 6.3.2.

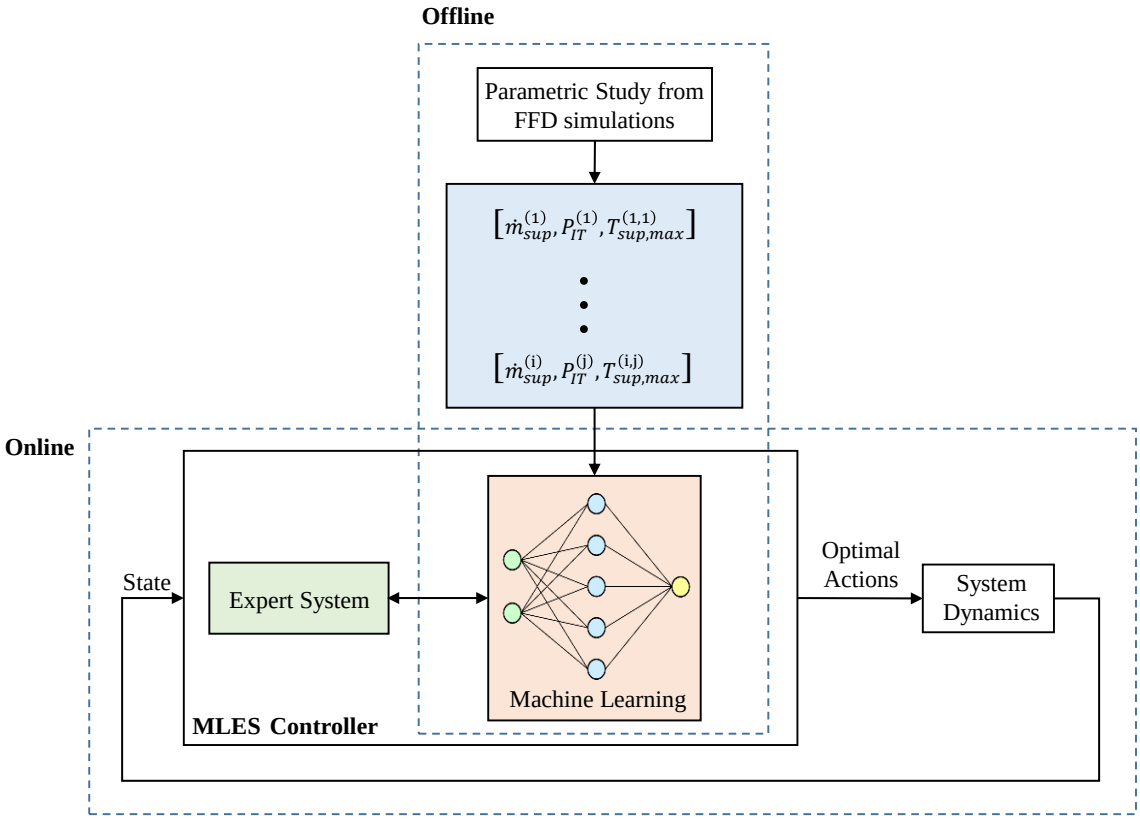


Figure 6-3 Schematic diagram of the proposed MLES appraoch

6.3 Case Studies

6.3.1 Case Description

The Layout of the subject data center room is shown in *Figure 3-9*. The data center is a single-room of 690 m^2, containing 151 IT racks and 12 floor-mounted PDUs that collectively consume 344 kW. The white space is 30.5 m wide and 22.6 m long, with a ceiling height of 3.4 m. The data center has a conventional cold-aisle and hot-aisle configuration. The front face of each rack orients toward a cold aisle. Each cold aisle is supplied with cold air from perforated floor tiles set into a raised floor and fed by an underfloor plenum. The exhaust air from the IT racks is directed into open hot aisles and then enters a ceiling plenum through the ceiling tiles. After that, the

exhaust air will mix with outdoor air in a mixed plenum and then return into the Air Handler Unit (AHU).

The schematic drawing of the HVAC system of the subject data center is shown in *Figure 5-4*. There are three chillers, chilled water pumps, condenser water pump and cooling towers in the schematic drawing, but only two of them run in Lead/Lag configuration and the other one is for backup. The chilled water is distributed to the AHU-1 and the AHU-2. Refer to Section 5.3.1 and Fu, Zuo, et al. [10] for more details.

The cooling system features a chilled water system with airside economizers (ASEs), which provides free cooling in cold weather. The cooling system operates in three cooling modes: (1) free cooling (FC) mode, where only ASEs are activated; (2) partial mechanical cooling (PMC) mode, where the chilled water system and ASEs work simultaneously; (3) fully mechanical cooling (FMC) mode, where only the chilled water system is used. The transition conditions between cooling modes are determined based on the weather data, control settings and system conditions, which are shown in *Figure 5-5*. For example, the FC mode switches to PMC mode when the outdoor dry bulb temperature is higher than the raised-floor plenum temperature setpoint (which determines supply air temperature) plus a dead band and the outdoor dew point temperature is higher than its predefined low cutoff limit plus a dead band. For more detailed description of the cooling system, please refer to Section 5.3.1 and Fu, Zuo, et al. [10].

6.3.2 Model Implementation

The cooling system is modeled in Modelica based on Modelica Buildings library [12], as shown in *Figure 6-4*. The major component models include chillers, AHUs, cooling towers,

chilled water pumps and condenser water pumps. The control system consists of the cooling mode control and local controls.

The new feature in the Modelica model is the MLES controller, which is implemented in Python and dynamically communicates with Modelica model through a Python interface in Modelica Buildings library [12].

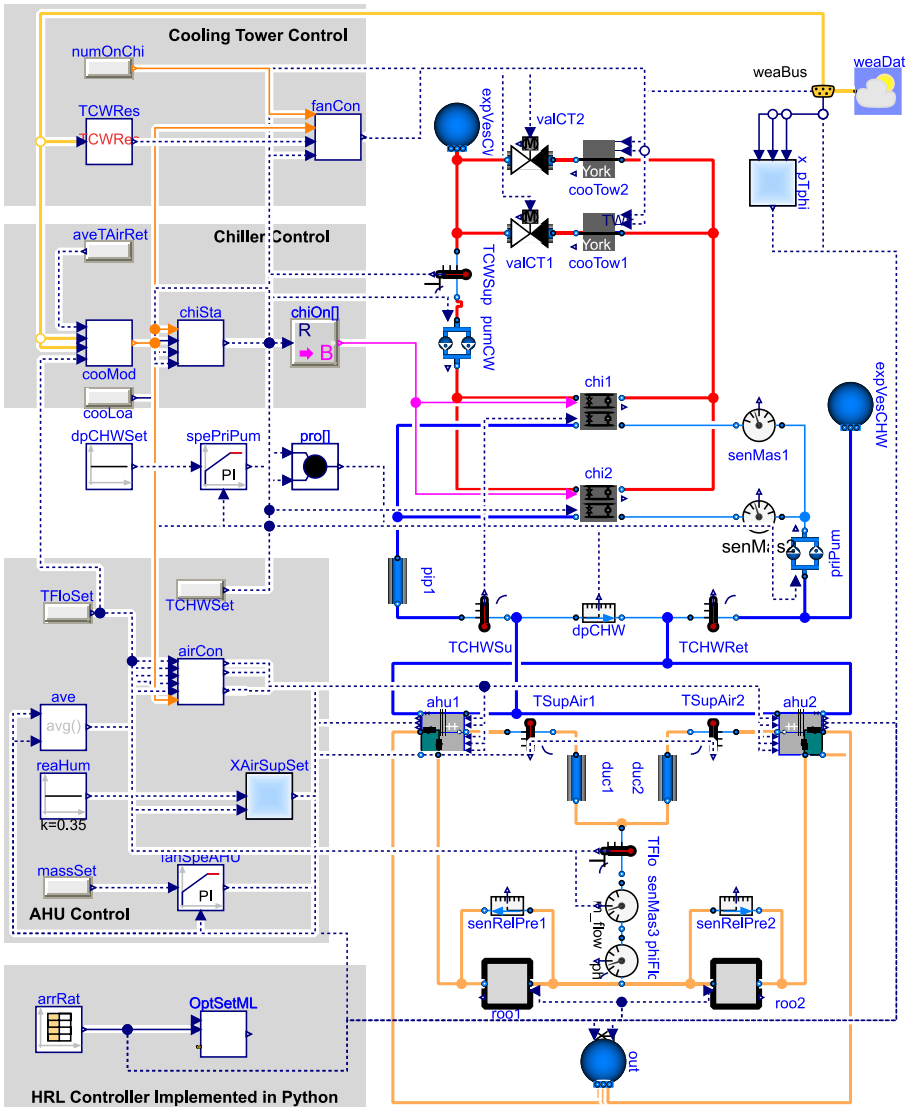


Figure 6-4 Implementation of the cooling system model in Modelica

The inputs of the MLES include:

- Previous status of cooling mode,
- Arrival rate of IT loads,
- Outdoor dry bulb temperature,
- Outdoor dew point temperature,
- Low limit of dew point temperature defined in cooling mode controller,
- High limit of dew point temperature defined in cooling mode controller,
- Deadbands defined in cooling mode controller;

The outputs of the MLES include:

- Optimal setpoint of supply air temperature,
- Optimal setpoint of supply air mass flow rate.

The MLES controller determines the optimal control actions through a machine learning assisted expert system based decision making process as shown in *Figure 6-5*. As described in Section 6.2.2, the objective of the controller is to maximize the operating hours of the free cooling mode.

There are three steps for the decision making process in the MLES. In Step 1, the controller evaluate if the cooling could keep on or switch to the free cooling mode if the supply air mass flow rate is determined to be at high limit. If no, that means the cooling system will have to operate on the mechanical cooling mode no matter how the optimization variables are adjusted. Then the optimal actions will be provided with optimization variables at lower limit to save energy from the fan since little energy can be saved from the chiller. If yes, go to Step 2.

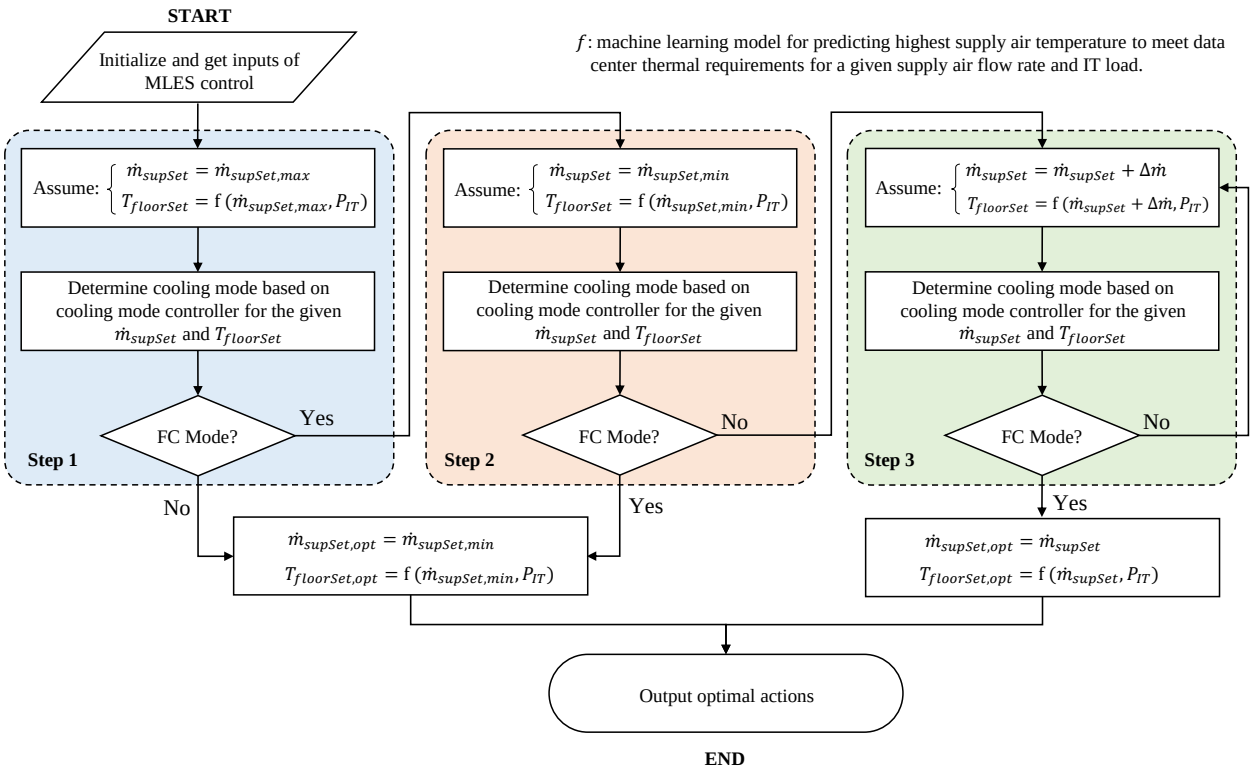


Figure 6-5 The machine learning assisted decision making process in the MLES

In Step 2, the controller evaluates if the cooling could keep on or switch to the free cooling mode if the supply air mass flow rate is determined to be at low limit. If yes, that means the cooling system could run on the free cooling mode with any values of the optimization variables and consequently the fan energy is dominant. Then the optimal actions will be provided with optimization variables at lower limit to save the fan energy as much as possible. If no, that means there should be an optimal setpoint for the supply air mass flow rate between the lower and higher limits, go to Step 3.

In Step 3, a "for" loop is performed to find the optima setpoint, with which the cooling system can run on the free cooling mode and meanwhile the fan energy can be saved as much as possible.

6.3.3 Machine Learning Models for Airflow Management

6.3.3.1 Parametric Study with FFD Simulations

A comprehensive parametric study is performed to generate data for training of machine learning models. As shown in *Table 6-1*, the studied parameters include supply air temperature, supply air flowrate and IT load. The supply air temperature ranges from 15 ^{o}C to 25 ^{o}C with an interval of 2 ^{o}C. The supply air flow rate varies, which correspond to the air ratios from 0.8 to 2.0. The IT load ranges from 2.0 kW/rack to 5.0 kW/rack. This results in 1,014 times of FFD simulations in total. The simulation settings of the FFD model for the studied data center was introduced in Section 3.3.2. The FFD model was also validated in Section 3.3.2. For more details, please refer to [120].

Table 6-1 Settings of the parametric study

Parameters	**Unit**	**Values**
Supply air temperature	^{o}C	15, 17, 19, 21, 23, 25
Supply air flow rate	Air ratio	0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7 ,1.8, 1.9, 2.0
IT load	kW/rack	2.0, 2.25, 2.5, 2.75, 3.0, 3.25, 3.5, 3.75, 4.0, 4.25, 4.5, 4.75, 5.0

Taking advantages of the fast speed of the FFD model, the parametric study can be completed within a day. The results of the parametric study is shown in *Figure 6-6*. Maximum rack inlet temperatures are extracted from the results of FFD simulations for different IT loads, air ratios and supply air temperatures. Generally, the relationship between the maximum rack inlet

temperature and the air ratio follows similar pattern for different IT loads and supply air tempertaures.

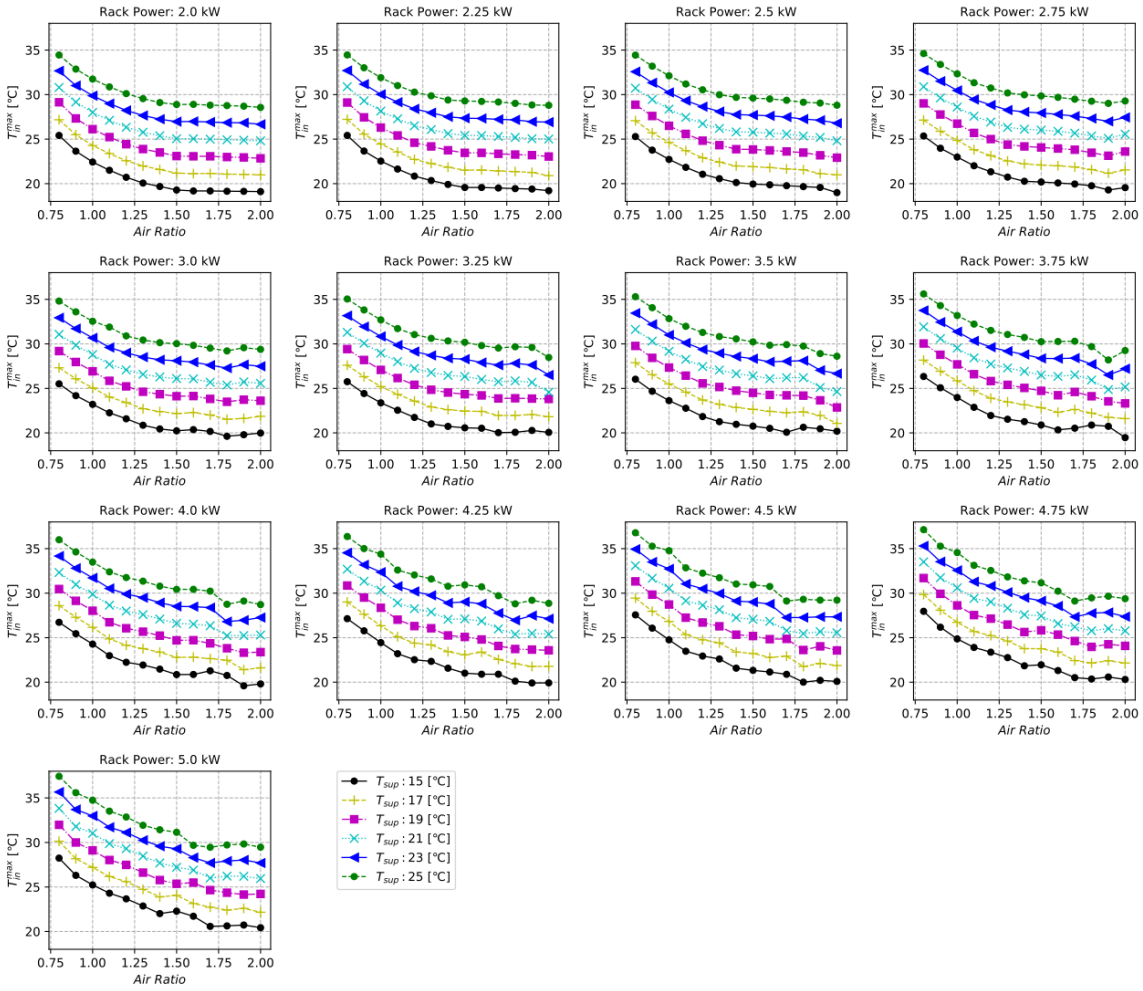


Figure 6-6 Maximum rack inlet temperatures with different IT loads and air ratios predicted by FFD simulations

Based on the results of the parametric study, the highest supply air temperatures under different IT loads and air ratios to meet thermal requirements of data centers are determined as shown in *Figure 6-7*. The machine learning model generated based on these data can be used to

adjust supply air temperature to meet the data center thermal requirements by being integrated into the cooling system simulation and optimization.

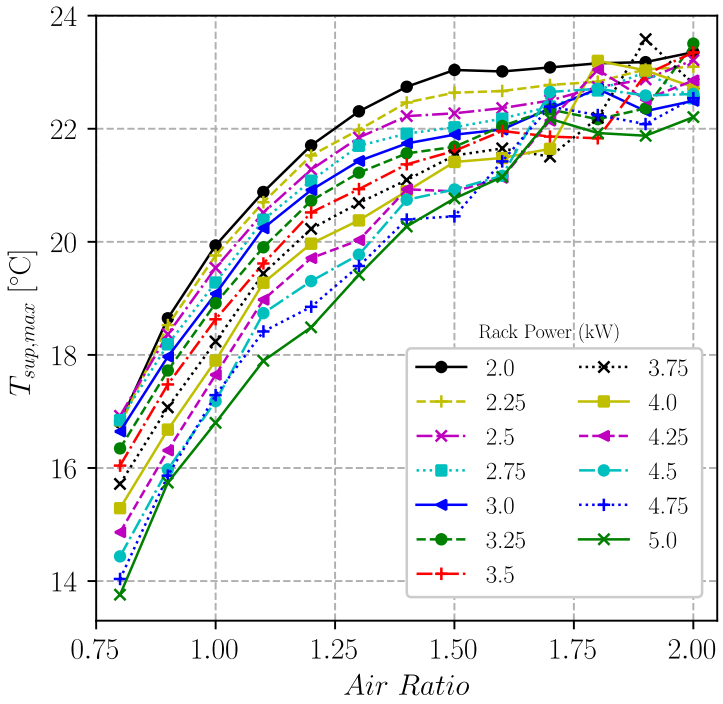


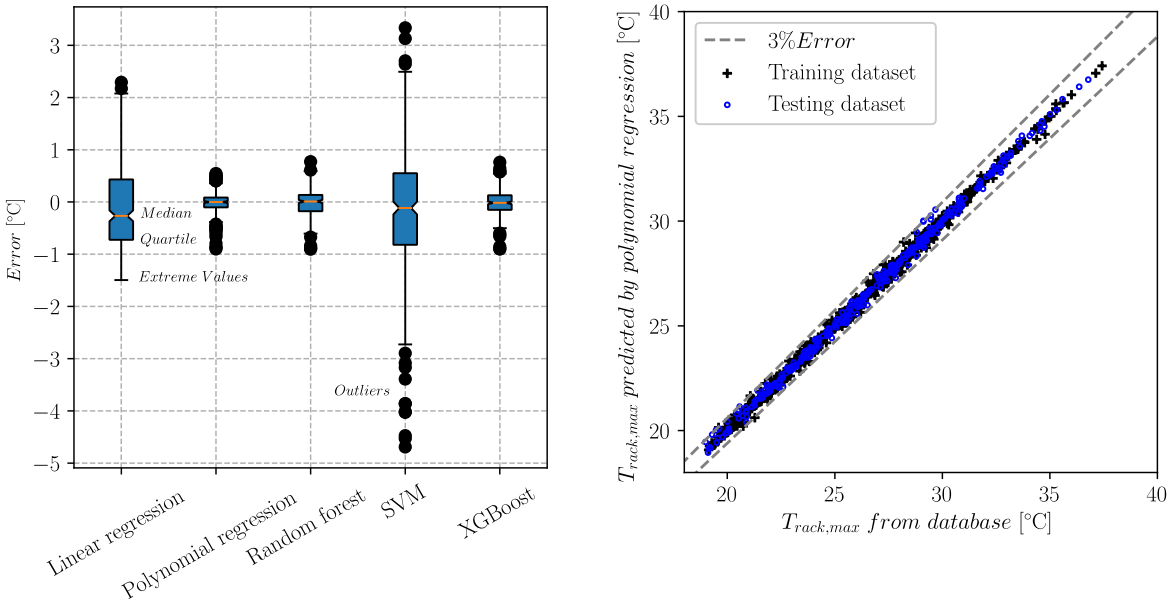
Figure 6-7 Highest supply air temperatures under different IT loads and air ratios to meet thermal requirements of data centers

6.3.3.2 Training of Machine Learning Models

Table 6-2 Inputs and outputs of the machine learning models

Machine Learning Models	Inputs	Outputs
ML-1	1. IT load 2. Air ratio 3. Supply air temperature	1. Maximum rack inlet temperature
ML-2	1. IT load 2. Air ratio	1. Highest supply air temperature that can meet the data center thermal requirements

As described in *Table 6-2*, two machine learning models are described including ML-1 and ML-2. The ML-1 is to evaluate the data center thermal environment regarding the maximum rack inlet temperature with the inputs include IT load, air ratio and supply air temperature. The ML-2 is to provide the highest supply air temperature that can meet the data center thermal requirements for a given IT load and air ratio.



(a) Error distribution with different ML models (b) Polynomial regression model

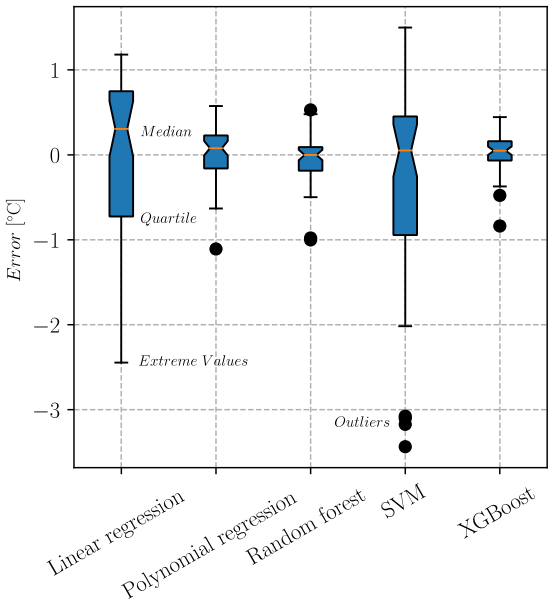
Figure 6-8 Training of the ML-1 that predicts the maximum rack inlet temperatures under different IT loads, supply air temperatures and air ratios

Different machine learning models are evaluated for the two proposed problems as described in *Table 6-2*. The machine learning models [128] include:

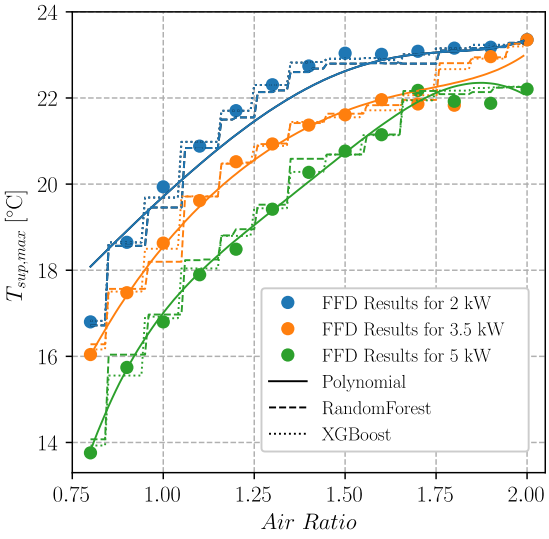
- Linear regression model: a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables).

- Polynomial regression model: a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x.
- Random forest model: an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.
- Support vector machine: supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.
- XGBoost: a software library which provides a gradient boosting framework.

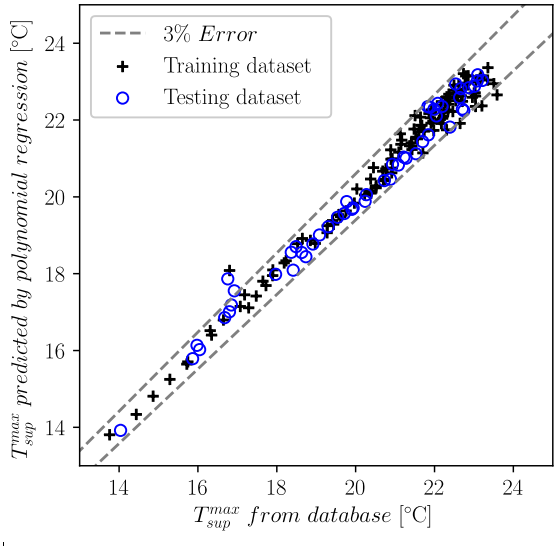
All descriptions for these machine learning models are from Wikipedia. The data generated from the parametric study is divided in training dataset and testing dataset. The training dataset is determined by randomly selecting 70% of the data in the original dataset and remaining data in the original dataset is defined as testing dataset. The training performance of ML-1 and ML-2 models is shown in *Figure 6-8* and *Figure 6-9*, respectively. *Figure 6-8* (a) compares the prediction errors with different machine learning models and it is found that the polynomial regression, random forest and XGBoost models have the best prediction accuracy. The polynomial regression model is selected in this study and the detailed comparison between its predictions and true values is shown in *Figure 6-8* (b). It is found that almost all the predictions are within 3% error compared to the true values in the database.



(a) Error distribution with different ML models



(b) Detailed analysis



(c) Polynomial regression model

Figure 6-9 Training of the ML-2 that predicts highest supply air temperatures under different IT loads and air ratios to meet thermal requirements of data centers

The same evaluations with different machine learning models are performed for the ML-2 case. Similar findings are obtained for the ML-2 case as the ML-1 case. The *Figure 6-9* (b) shows the detailed analysis for the best three machine learning models. The polynomial regression model has the best overall performance if considering new data between existing data points in the dataset though the random forest and XGBoost models have smaller errors with respect to the existing data in the dataset. It is also found that almost all the predictions with the polynomial regression model are within 3% error compared to the true values in the database for the ML-2 case.

6.3.4 Detailed Control Performance Evaluation with Typical Days

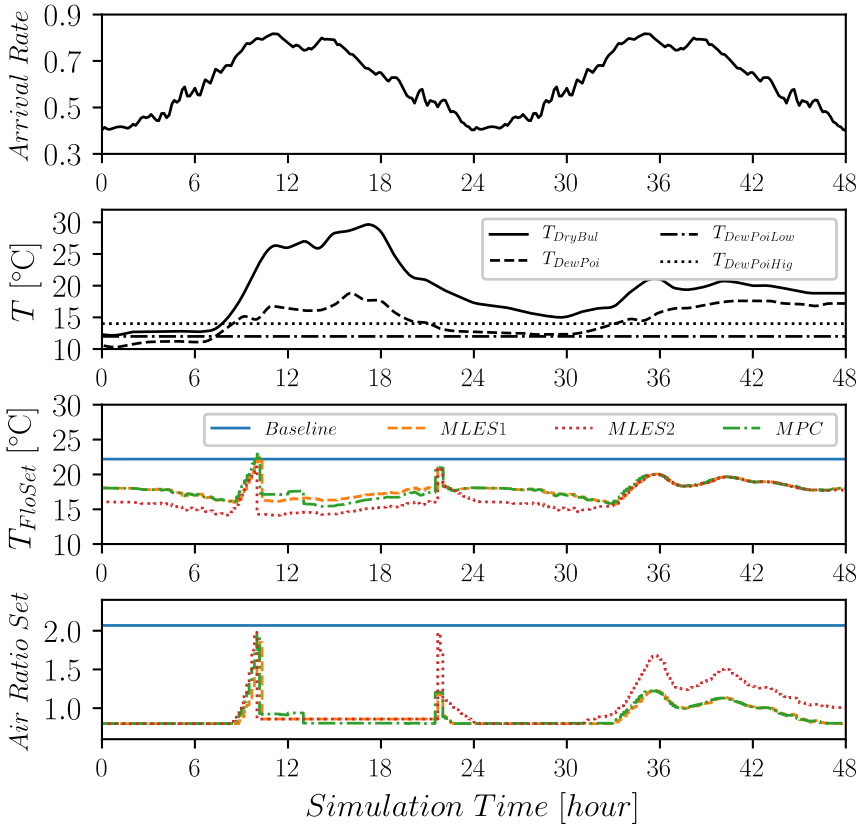


Figure 6-10 Control inputs and actions in different systems

In this section, a detailed evaluation for the proposed MLES real-time optimal control is performed with typical days from July 25-26. Two cases using the MLES control with different settings of target maximum rack inlet temperature are studied. In MLES-1, the setpoint of maximum rack inlet temperature is defined as the recommended high threshold 27 ^{o}C suggested by ASHRAE [15]. In MLES-2, the setpoint of maximum rack inlet temperature is defined with a more conservative setting, which is 2 ^{o}C lower (25 ^{o}C) than the recommended high threshold suggested by ASHRAE [15]. A case with a conventional MPC approach is also conducted as a reference case, in which a 10 minutes of time step and 20 minutes of prediction horizon are used. The MPC is assumed to be an ideal MPC, which has an ideal model and perfect predictions for future weather and IT loads.

The control inputs and actions in different systems are shown in *Figure 6-10*. In baseline, the setpoints for supply air temperature (T_{FloSet} in *Figure 6-10*) and supply air mass flow rate (Air Ratio Set in *Figure 6-10*) are constant. In MLES-1, MLES-2 and MPC, the setpoints are dynamically adjusted to achieve the best energy performance while meeting the thermal requirements. The optimal setpoints determined by MLES-1 and MPC almost match with each other. MLES-2, which has a more conservative setting, results in a lower supply air temperature or higher supply air flow rate to better meet the thermal requirements.

Figure 6-11 shows the control performance of different systems. Both MLES and MPC approaches save fan energy through dynamically adjusting the setpoints of supply air flow rate and temperature without sacrificing free cooling, which realizes in significant energy savings compared to the baseline. The thermal environment is well controlled in all the cases and among them, the MLES-2 performs better than the MLES-1 and MPC as expected.

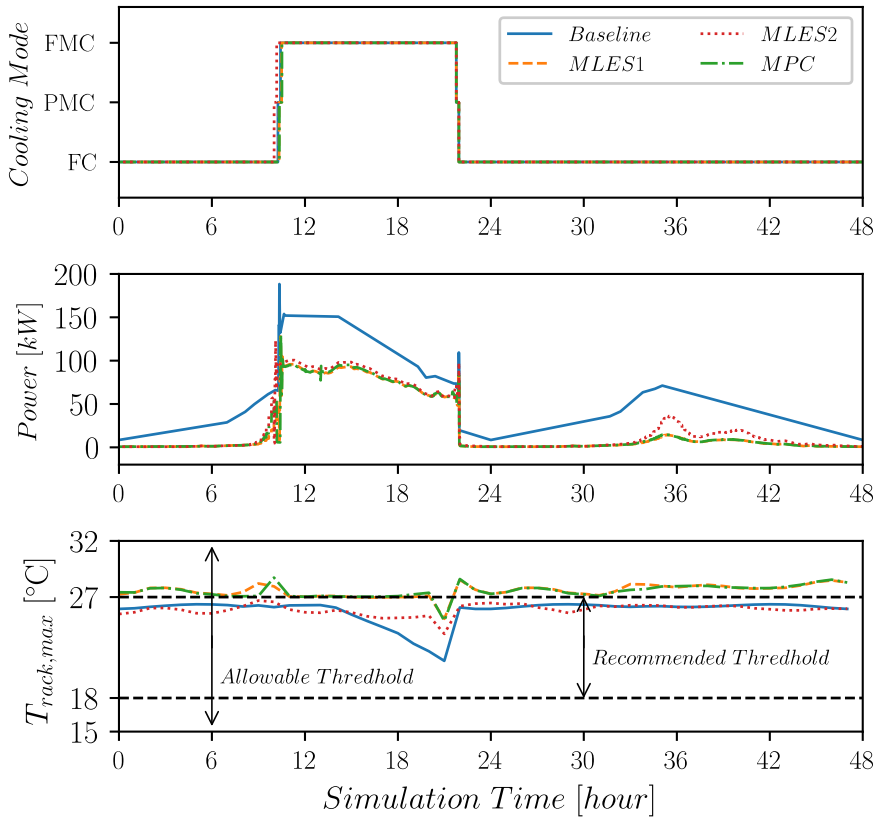


Figure 6-11 Control performance of different systems

Regarding the computing time, it takes about 2.5 minutes to determine the optimal actions for one time step with the MPC, but only less than 1 second with the MLES. Above all, the proposed MLES real-time optimal control could achieve similar performance as a well-designed ideal MPC does, but is much faster than the MPC.

6.3.5 Control Performance Evaluation with Annual Simulation

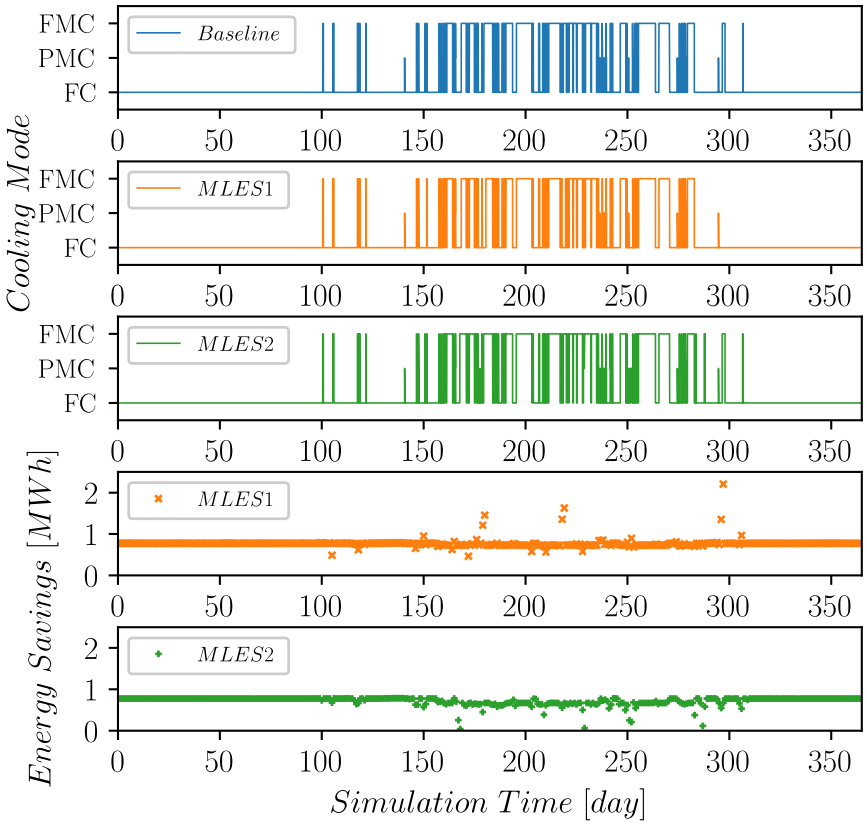


Figure 6-12 Control performance and daily energy savings in different systems

The performance of the MLES real-time optimal control is also evaluated with an annual simulation. The weather profiles were shown in Section 6.4.3. *Figure 6-12* shows the control performance and daily energy savings in different systems. The MLES cases are intended mostly to save energy through reducing the fan energy and meanwhile avoiding or reducing the sacrifice of free cooling. The hours running on the free cooling mode in MLES cases are generally similar as that in baseline. Interestingly, obvious higher energy savings are achieved on some days than that on other days. This is because more free cooling is achieved with MLES control than baseline

on those days. For example, on the days around Day 300, MLES-1 case got more free cooling than the baseline.

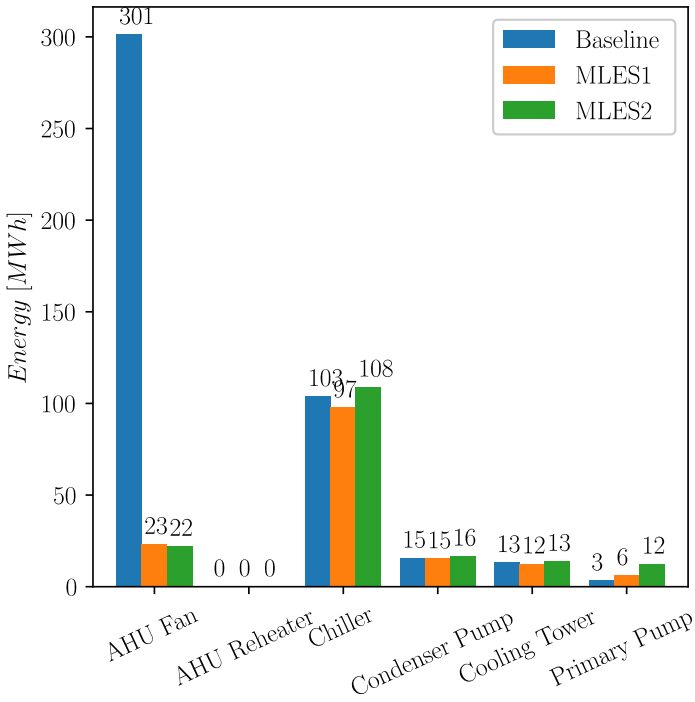


Figure 6-13 Annual energy consumption for different components of the cooling system for different systems

Figure 6-13 summarizes the annual energy consumption for different components of the cooling system for different systems. Most energy savings are achieved from the AHU fans in MLES cases. MLES-1 saves chiller energy by 6 MWh (6%) because it allows the cooling system to run more on the free cooling mode through a real-time optimal setpoints of supply air temperature. MLES-2 consumes more chiller energy than the baseline and MLES-1 because of a conservative settings for a more robust thermal environment control. Consequently, the MLES-1 and MLES-2 saves the annual energy consumption by 282.6 MWh (64.6%) and 263.3 MWh (60.2%).

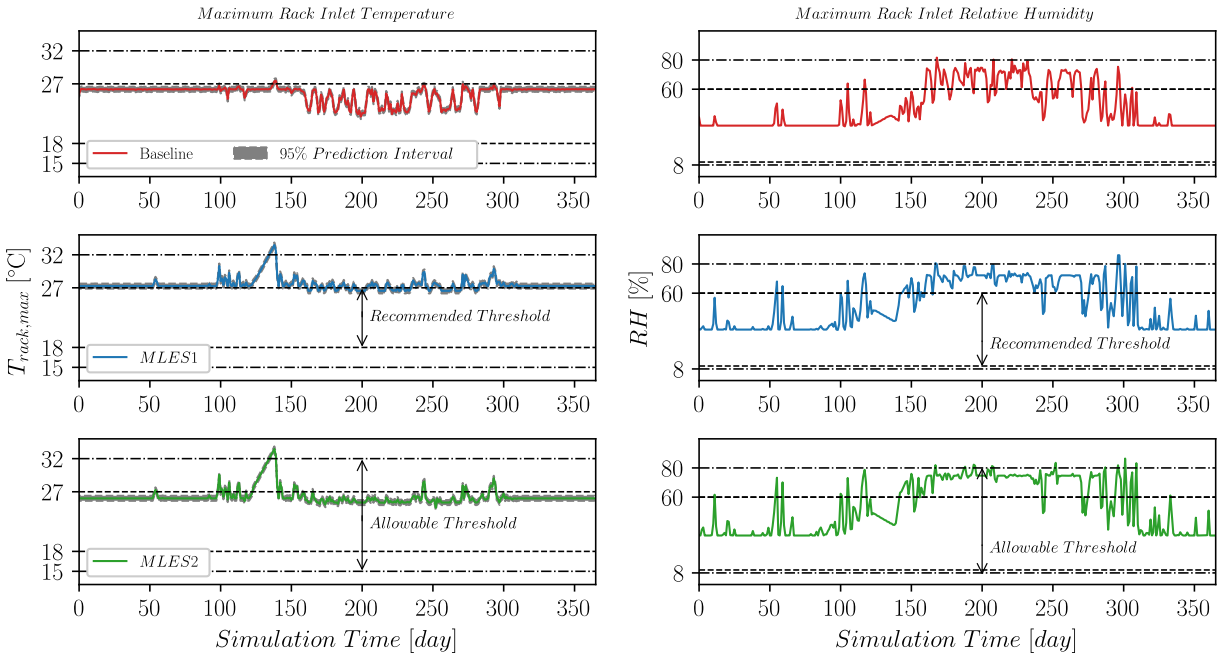


Figure 6-14 Data center thermal environment created by different systems

Except for the energy performance, another important aspect is the impact on the airflow and thermal management. *Figure 6-14* shows the maximum rack inlet temperature and relative humidity in baseline and MLES cases. As described in Section 2.1.2, the allowable and recommended ranges for rack inlet temperature are 15-32 ºC and 18-27 ºC, respectively, and the allowable and recommended ranges for humidity are 8-80% and -9-15 ºC DP and 60% RH, respectively. Most rack inlet temperatures and relative humidities are controlled within the recommended and allowable ranges. But it can be found that the rack inlet temperatures exceed the high limit of the allowable ranges at around Day 140. The reason lies in the weather data. As can be seen in *Figure 5-11*, the segment of the weather data from Day 120 to Day 140 is created from interpolation because of lack of data. For those days, the outdoor dew point temperatures are lower than 10 ºC but the outdoor dry bulb temperatures are about 20 ºC. This results in that the

cooling system operate on the free cooling mode because of the low outdoor dew point temperatures. But the rack inlet temperature can be very high due to the high outdoor dry bulb temperatures. To conclude, the MLES real-time optimal control saves the annual energy consumption by 60.2%-64.6% and meanwhile the thermal environment is generally well controlled to meet the requirements of data centers.

6.4 Discussions

In addition to the good control performance, fast speed in real-time and ease to implement, the proposed MLES optimal control strategy has another advantage over a MPC in term of the uncertainty. MPC needs to perform model-based optimization for a future horizon. Therefore, the future weather predictions, future load predictions, and model accuracy bring a lot of uncertainties to MPC. However, MLES is model-free optimal control method, which makes decisions based on expert knowledge, the previous system status, and current IT loads and weather conditions. So, the uncertainties in MPC do not exist in MLES. Based on algorithm in MLES, one of the potential uncertainties may come from the changes of IT loads and weather conditions between two activations of MLES. For example, after calling MLES and determining the optimal setpoints, the setpoints will keep constant until next time of calling MLES. Therefore, during this period, if the weather or IT loads change drastically, it will bring uncertainties to the systems. This study uses a very small time interval to call MLES to reduce this effect. Another possible uncertainty comes from the accuracy of the ML model. This uncertainty can be reduced by improving parameter tuning for ML models, adopting more advanced ML models, and creating a more accurate and larger training dataset.

6.5 Conclusion

This study proposed a novel robust and easy-to-implement MLES method to achieve a real-time optimal control for data center cooling systems. First, a comprehensive parametric study was conducted with more than 1,000 CFD simulations. Then, ML models are trained with the data generated from the parametric study, in which different machine learning models were evaluated. To realize the real-time optimal control, a MLES based decision making process was proposed. A MPC approach was also implemented to evaluate the performance of the proposed MLES optimal control. A detailed control performance evaluation shows that the proposed MLES real-time optimal control could achieve similar performance as a well-designed ideal MPC approach, but is much faster than the MPC. An annual simulation shows that the proposed MLES optimal control saves the annual energy consumption by 60.2%-64.6% and meanwhile the non-uniform thermal environment is generally well controlled to meet the requirements of data centers.

Chapter 7 Conclusions and Future Work

7.1 Conclusions

This dissertation proposes novel modelling methods, optimization framework and optimal control methods to support improving cooling efficiency while maintaining reliable operation for data centers in the real world. The major scientific contributions and conclusions are as follows:

1) **Developed a new open-source Fast Fluid Dynamics (FFD) model, which allows fast and accurate prediction of data center airflow management**. First, the new FFD model shows slightly better accuracy and faster speed compared to existing FFD models. Second, to simulate a real data center, the new FFD model achieves similar accuracy compared to CFD and is 61 times faster than CFD. Third, FFD shows good capability on handling realistic cases with complex boundary conditions. Fourth, the new FFD model has been publicly released.
2) **Proposed a new online BES-ROM-CFD co-simulation method, which significantly accelerates BES-CFD co-simulation while well controlling prediction accuracy**. A Modelica-ISAT-FFD model is implemented to demonstrate the new co-simulation method. First, for the accuracy, the new model can generally control the prediction error within user-defined tolerances compared to an existing Modelica-FFD model. Second, for the speed, an annual simulation shows that the new model saves up to 95.7% of computing time against the existing Modelica-FFD model.
3) **Proposed an optimization platform based on an adaptive online BES-ROM-CFD co-simulation powered by distributed computing, which makes co-simulation-based**

optimization become computationally practical and sufficiently accurate for real applications. A real middle-size data center is adopted to demonstrate the capability of the proposed optimization platform. First, the computing time of the holistic optimization reaches as less as 0.78 days, which is estimated to take many years based on an existing Modelica-FFD model. Such an improvement makes a co-simulation based optimization become practical for real applications, which is potentially to have great impact for other ventilation systems involving stratified airflow and thermal environment. Second, the results of the case studies show that the holistic optimization saves the annual energy consumption by as much as 48.1% and meanwhile the non-uniform thermal environment is generally well controlled to meet the thermal requirements of data centers.

4) **Proposed a robust and easy-to-implement machine learning assisted expert system method, which realizes real-time optimal control for data center cooling systems**. First, a detailed control performance evaluation shows that the proposed MLES optimal control could achieve similar performance as a well-designed ideal MPC approach does, but is much faster than the MPC. Second, an annual simulation shows that the proposed MLES optimal control saves the energy consumption by up to 64.6% while maintaining reliable operation of data centers. Third, this methodology is potentially to help realize optimal control for other ventilation systems involving stratified airflow and thermal environment.

7.2 Limitations and Future Work

Based on findings and conclusions in this dissertation, some limitations and potential future research are proposed as follows:

1) The proposed Modelica-ISAT-FFD model can only predict steady state results for indoor thermal environment, and cannot capture seconds-level short term airflow and thermal dynamics. This is because the data drive model, ISAT, is trained with steady state FFD results. Future work may investigate how to use data driven models for transient problems.
2) The current work mostly focused on data center applications. Future work may evaluate the proposed models, optimization platform and optima control methods for other types of buildings that involve non-uniform thermal environment, such as residential buildings with natural ventilation, offices with displacement ventilation.
3) The current work validated the Modelica-FFD model for the data center case in a separate manner instead of a dynamically coupling way. This is due to limited resources and time, as well as the consideration on the focus of this dissertation. Since the Modelica-FFD model has already been validated and proved in previous research, this study simply assumed it is correct, and put more efforts on demonstrating the advantages of our proposed new models by taking the existing Modelica-FFD model as benchmark. It can be a direction for future research to validate the Modelica-FFD model or even the Modelica-ISAT-FFD model in a dynamically coupling way.
4) Other future work may include (a) Systematically evaluating and ameliorate the training method to further improve the training efficiency of the Modelica-ISAT-FFD model; (b) Performing BES-CFD co-simulation in community and city scales to support large scale building energy and urban thermal environment simulations.

BIBLIOGRAPHY

[1] A. Shehabi, S. Smith, D. Sartor, R. Brown, M. Herrlin, J. Koomey, E. Masanet, N. Horner, I. Azevedo, W. Lintner, United states data center energy usage report, (2016).

[2] M. Salim, R. Tozer, Data Centers' Energy Auditing and Benchmarking-Progress Update, ASHRAE Transactions, 116 (1) (2010).

[3] A.S. Andrae, T. Edler, On global electricity usage of communication technology: trends to 2030, Challenges, 6 (1) (2015) 117-157.

[4] T. Evans, The Different Technologies for Cooling Data Centers White Paper 59 Version 2, in, Schneider Electric's Data Center Science Center, 2012.

[5] Q. Tang, S.K.S. Gupta, G. Varsamopoulos, Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach, IEEE Transactions on Parallel and Distributed Systems, 19 (11) (2008) 1458-1472.

[6] A.-T. Nguyen, S. Reiter, P. Rigo, A review on simulation-based optimization methods applied to building performance analysis, *Applied Energy*, 113 (2014) 1043-1058.

[7] Y. Pan, R. Yin, Z. Huang, Energy modeling of two office buildings with data center for green building design, *Energy and Buildings*, 40 (7) (2008) 1145-1152.

[8] K.-P. Lee, H.-L. Chen, Analysis of energy saving potential of air-side free cooling for data centers in worldwide climate zones, *Energy and Buildings*, 64 (2013) 103-112.

[9] M. Kummert, W.M. Dempster, K. McLean, Transient thermal analysis of a data centre cooling system under fault conditions, in: 11th International Building Performance Simulation association Conference and Exhibition, Building Simulation 2009, 2009.

[10] Y. Fu, W. Zuo, M. Wetter, J.W. VanGilder, X. Han, D. Plamondon, Equation-based object-oriented modeling and simulation for data center cooling: A case study, *Energy and Buildings*, 186 (2019) 108-125.

[11] W. Zuo, Q. Chen, Real-time or faster-than-real-time simulation of airflow in buildings, *Indoor Air*, 19 1 (2009) 33-44.

[12] M. Wetter, W. Zuo, T.S. Nouidui, X. Pang, Modelica buildings library, *Journal of Building Performance Simulation*, 7 (4) (2014) 253-270.

[13] J. Rambo, Y. Joshi, Modeling of data center airflow and heat transfer: State of the art and future trends, *Distributed and Parallel Databases*, 21 (2-3) (2007) 193-225.

[14] R.R. Schmidt, E.E. Cruz, M. Iyengar, Challenges of data center thermal management, *IBM Journal of Research and Development*, 49 (4.5) (2005) 709-723.

[15] ASHRAE, Data Center Power Equipment Thermal Guidelines and Best Practices, in, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2016.

[16] Y. Zhu, Applying computer-based simulation to energy auditing: A case study, *Energy and buildings*, 38 (5) (2006) 421-428.

[17] S.-W. Ham, J.-W. Jeong, Impact of aisle containment on energy performance of a data center when using an integrated water-side economizer, *Applied Thermal Engineering*, 105 (2016) 372-384.

[18] A. Agrawal, M. Khichar, S. Jain, Transient Simulation of Wet Cooling Strategies for a Data Center in Worldwide Climate Zones, *Energy and Buildings*, 127 (2016) (2016) 352–359.

[19] A. Shehabi, A. Horvath, W. Nazaroff, S. Ganguly, A. Gadgil, K. Traber, H. Price, R. Engineers, Energy Implications of Economizer Use in California Data Centers in: ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, U.S.A., 2008, pp. 319-330.

[20] M. Wetter, Modelica-based modelling and simulation to support research and development in building energy and control systems, *Journal of Building Performance Simulation*, 2 (2) (2009) 143-161

[21] M. Wetter, M. Bonvini, T.S. Nouidui, Equation-based languages–A new paradigm for building energy modeling, simulation and optimization, *Energy and Buildings*, 117 (2016) 290-300.

[22] M.T. Radosevic, J. Hensen, A.T.M. Wijsman, Distributed building performance simulation—a novel approach to overcome legacy code limitations, *HVAC&R Research*, 12 (S1) (2006) 621-640.

[23] M. Trcka, M. Wetter, J. Hensen, Comparison of co-simulation approaches for building and HVAC/R system simulation, in: Proceedings of the International IBPSA Conference, Beijing, China, 2007.

[24] M. Wetter, Generic Optimization Program User Manual Version 3.0. 0, *Lawrence Berkeley National Laboratory*, (2009).

[25] M. Wetter, A View on Future Building System Modeling and Simulation in: J.L.M. Hensen, R. Lamberts (Eds.) Building Performance Simulation for Design and Operation, UK, Routledge, 2011.

[26] P. Fritzson, Principles of object-oriented modeling and simulation with Modelica 3.3: a cyber-physical approach, John Wiley & Sons, 2014.

[27] M. Wetter, A Modelica-based model library for building energy and control systems, *Lawrence Berkeley National Laboratory*, (2010).

[28] B. Eisenhower, K. Gasljevic, I. Mezic, Control-oriented dynamic modeling and calibration of a campus theater using Modelica, *SimBuild 2012*, (2012).

[29] D. Lee, B. Lee, J.W. Shin, Fault Detection and Diagnosis with Modelica Language using Deep Belief Network, in: Proceedings of the 11th International Modelica Conference, Versailles, France, September 21-23, 2015, Linköping University Electronic Press, 2015, pp. 615-623.

[30] S.V. Patankar, Airflow and cooling in a data center, *Journal of Heat transfer*, 132 (7) (2010) 073001.

[31] J.W. VanGilder, R.R. Schmidt, Airflow uniformity through perforated tiles in a raised-floor data center, in: ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference, American Society of Mechanical Engineers, 2005, pp. 493-501.

[32] C.M. Healey, J.W. VanGilder, Z.R. Sheffer, X.S. Zhang, Potential-Flow Modeling for Data Center Applications, in: ASME 2011 Pacific Rim Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Systems, American Society of Mechanical Engineers Digital Collection, 2011, pp. 527-534.

[33] Q. Tang, T. Mukherjee, S.K. Gupta, P. Cayton, Sensor-based fast thermal evaluation model for energy efficient high-performance datacenters, in: 2006 Fourth International Conference on Intelligent Sensing and Information Processing, IEEE, 2006, pp. 203-208.

[34] Z. Song, B.T. Murray, B. Sammakia, Multivariate prediction of airflow and temperature distributions using artificial neural networks, in: ASME 2011 Pacific Rim Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Systems, American Society of Mechanical Engineers Digital Collection, 2011, pp. 595-604.

[35] B. Elhadidi, H.E. Khalifa, Application of Proper Orthogonal Decomposition to Indoor Airflows, *ASHRAE Transactions*, 111 (1) (2005).

[36] C.M. Healey, J.W. VanGilder, Z.M. Pardey, Perforated Tile Airflow Prediction: A Comparison of RANS CFD, Fast Fluid Dynamics, and Potential Flow Modeling, in: ASME 2015 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems collocated with the ASME 2015 13th International

Conference on Nanochannels, Microchannels, and Minichannels, American Society of Mechanical Engineers, 2015, pp. V001T009A024-V001T009A024.

[37] S. Patankar, Numerical heat transfer and fluid flow, CRC press, 1980.

[38] W. Zuo, Q. Chen, Fast and informative flow simulations in a building by using fast fluid dynamics model on graphics processing unit, *Building and Environment*, 45 (3) (2010) 747-757.

[39] W. Zuo, Q. Chen, Validation of fast fluid dynamics for room airflow, in: IBPSA Building Simulation 2007, 2007.

[40] W. Tian, T.A. Sevilla, W. Zuo, A systematic evaluation of accelerating indoor airflow simulations using cross-platform parallel computing, *Journal of Building Performance Simulation*, 10 (3) (2017) 243-255.

[41] M. Jin, W. Zuo, Q. Chen, Simulating natural ventilation in and around buildings by fast fluid dynamics, *Numerical Heat Transfer, Part A: Applications*, 64 (4) (2013) 273-289.

[42] W. Liu, R. You, J. Zhang, Q. Chen, Development of a fast fluid dynamics-based adjoint method for the inverse design of indoor environments, *Journal of Building Performance Simulation*, 10 (3) (2017) 326-343.

[43] A. Katal, M. Mortezazadeh, L.L. Wang, Modeling building resilience against extreme weather by integrated CityFFD and CityBEM simulations, *Applied Energy*, 250 (2019) 1402-1417.

[44] W. Tian, J.W. VanGilder, X. Han, C.M. Healey, M.B. Condor, W. Zuo, A New Fast Fluid Dynamics Model for Data-Center Floor Plenums, *ASHRAE Transactions*, 125 (2019).

[45] W. Tian, J. VanGilder, M. Condor, X. Han, W. Zuo, An Accurate Fast Fluid Dynamics Model for Data Center Applications, in: 2019 18th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), IEEE, 2019, pp. 1275-1281.

[46] Q. Chen, Ventilation performance prediction for buildings: A method overview and recent applications, *Building and Environment*, 44 (4) (2009) 848-858.

[47] W. Zuo, M. Wetter, W. Tian, D. Li, M. Jin, Q. Chen, Coupling indoor airflow, HVAC, control and building envelope heat transfer in the Modelica Buildings library, *Journal of Building Performance Simulation*, 9 (4) (2016) 366-381.

[48] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W.F. Buhl, Y.J. Huang, C.O. Pedersen, R.K. Strand, R.J. Liesen, D.E. Fisher, M.J. Witte, EnergyPlus: creating a new-generation building energy simulation program, *Energy and Buildings*, 33 (4) (2001) 319-331.

[49] S.A. Klein, J.A. Duffie, W.A. Beckman, TRNSYS-A transient simulation program, *Ashrae Trans.*, 82 (1976) 623.

[50] J. Clarke, D. McLean, ESP-A building and plant energy simulation system, *Strathclyde: Energy Simulation Research Unit, University of Strathclyde*, (1988).

[51] T. Kalamees, IDA ICE: the simulation tool for making the whole building energy and HAM analysis, *Annex*, 41 (2004) 12-14.

[52] C. Rode, K. Grau, Whole building hygrothermal simulation model, *TRANSACTIONS-AMERICAN SOCIETY OF HEATING REFRIGERATING AND AIR CONDITIONING ENGINEERS*, 109 (1) (2003) 572-582.

[53] S.B. Pope, Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation, *Combustion Theory and Modelling*, 1 (1) (1997) 41-63.

[54] R. Zhang, K.P. Lam, S.-c. Yao, Y. Zhang, Coupled EnergyPlus and computational fluid dynamics simulation for natural ventilation, *Building and Environment*, 68 (2013) 100-113.

[55] S. Zhang, Y. Cheng, C. Huan, Z. Lin, Heat removal efficiency based multi-node model for both stratum ventilation and displacement ventilation, *Building and Environment*, 143 (2018) 24-35.

[56] A.C. Megri, F. Haghighat, Zonal modeling for simulating indoor environment of buildings: Review, recent developments, and applications, *HVAC&R Research*, 13 (6) (2007) 887-905.

[57] Z. Zhai, Q. Chen, P. Haves, J.H. Klems, On approaches to couple energy simulation and computational fluid dynamics programs, *Building and Environment*, 37 (8-9) (2002) 857-864.

[58] Q. Chen, J. Van der Kooi, Accuracy—A program for combined problems of energy analysis, indoor airflow, and air quality, *ASHRAE Transactions*, 94 (1988) 196-214.

[59] J. Srebric, Q. Chen, L.R. Glicksman, A coupled airflow and energy simulation program for indoor thermal environmental studies, *ASHRAE Transactions*, 106 (1) (2000) 465-476.

[60] W.S. Dols, L. Wang, S.J. Emmerich, B.J. Polidoro, Development and application of an updated whole-building coupled thermal, airflow and contaminant transport simulation program (TRNSYS/CONTAM), *Journal of Building Performance Simulation*, 8 (5) (2015) 326-337.

[61] J. Lau, Q. Chen, Floor-supply displacement ventilation for workshops, *Building and Environment*, 42 (4) (2007) 1718-1730.

[62] A. Pappas, Z. Zhai, Numerical investigation on thermal performance and correlations of double skin façade with buoyancy-driven airflow, *Energy and Buildings*, 40 (4) (2008) 466-475.

[63] G. Villi, W. Pasut, M. De Carli, CFD modelling and thermal performance analysis of a wooden ventilated roof structure, in: Building Simulation, Springer, 2009, pp. 215-228.

[64] Z. Du, P. Xu, X. Jin, Q. Liu, Temperature sensor placement optimization for VAV control using CFD–BES co-simulation strategy, *Building and Environment*, 85 (2015) 104-113.

[65] W. Tian, X. Han, W. Zuo, Q. Wang, Y. Fu, M. Jin, An optimization platform based on coupled indoor environment and HVAC simulation and its application in optimal thermostat placement, *Energy and Buildings*, 199 (2019) 342-351.

[66] W. Tian, X. Han, W. Zuo, M.D. Sohn, Building energy simulation coupled with CFD for indoor environment: A critical review and recent applications, *Energy and Buildings*, (2018).

[67] Z. Sun, S. Wang, A CFD-based test method for control of indoor environment and space ventilation, *Building and Environment*, 45 (6) (2010) 1441-1447.

[68] Z.J. Zhai, Q.Y. Chen, Performance of coupled building energy and CFD simulations, *Energy and Buildings*, 37 (4) (2005) 333-344.

[69] D. Kim, J. Braun, E. Cliff, J. Borggaard, Development, validation and application of a coupled reduced-order CFD model for building control applications, *Building and Environment*, 93 (2015) 97-111.

[70] D.W. Demetriou, H. Ezzat Khalifa, Thermally aware, energy-based load placement in open-aisle, air-cooled data centers, *Journal of Electronic Packaging*, 135 (3) (2013).

[71] R. Qin, D. Yan, X. Zhou, Y. Jiang, Research on a dynamic simulation method of atrium thermal environment based on neural network, *Building and environment*, 50 (2012) 214-220.

[72] Y. Wei, T.T. Zhang, S. Wang, Prompt design of the air-supply opening size for a commercial airplane based on the proper orthogonal decomposition of flows, *Building and Environment*, 96 (2016) 131-141.

[73] J.D. Moore, J.S. Chase, P. Ranganathan, R.K. Sharma, Making Scheduling" Cool": Temperature-Aware Workload Placement in Data Centers, in: USENIX annual technical conference, General Track, 2005, pp. 61-75.

[74] Q. Tang, S.K. Gupta, G. Varsamopoulos, Thermal-aware task scheduling for data centers through minimizing heat recirculation, in: Cluster Computing, 2007 IEEE International Conference on, IEEE, 2007, pp. 129-138.

[75] X.S. Zhang, J.W. VanGilder, C.M. Healey, A real-time data center airflow and energy assessment tool, in: ASME 2009 InterPACK Conference collocated with the ASME 2009 Summer Heat Transfer Conference and the ASME 2009 3rd International Conference on Energy Sustainability, American Society of Mechanical Engineers, 2009, pp. 793-798.

[76] M.M. Toulouse, G. Doljac, V.P. Carey, C. Bash, Exploration of a potential-flow-based compact model of air-flow transport in data centers, in: ASME 2009 International Mechanical Engineering Congress and Exposition, American Society of Mechanical Engineers, 2009, pp. 41-50.

[77] D.W. Demetriou, H.E. Khalifa, Thermally aware, energy-based load placement in open-aisle, air-cooled data centers, *Journal of Electronic Packaging*, 135 (3) (2013) 030906.

[78] J. Moore, J.S. Chase, P. Ranganathan, Weatherman: Automated, online and predictive thermal mapping and management for data centers, in: 2006 IEEE International Conference on Autonomic Computing, IEEE, 2006, pp. 155-164.

[79] M. Wang, Q. Chen, Assessment of various turbulence models for transitional flows in an enclosed environment (RP-1271), *HVAC&R Research*, 15 (6) (2009) 1099-1119.

[80] M. Jin, Q. Chen, Improvement of fast fluid dynamics with a conservative semi-Lagrangian scheme, *International Journal of Numerical Methods for Heat & Fluid Flow*, (2015).

[81] W. Liu, M. Jin, C. Chen, R. You, Q. Chen, Implementation of a fast fluid dynamics model in OpenFOAM for simulating indoor airflow, *Numerical Heat Transfer, Part A: Applications*, 69 (7) (2016) 748-762.

[82] M. Mortezazadeh, L.L. Wang, Solving city and building microclimates by fast fluid dynamics with large timesteps and coarse meshes, *Building and Environment*, (2020) 106955.

[83] R. Courant, E. Isaacson, M. Rees, On the solution of nonlinear hyperbolic differential equations by finite differences, *Communications on Pure and Applied Mathematics*, 5 (3) (1952) 243-255.

[84] A.J. Chorin, A numerical method for solving incompressible viscous flow problems, *Journal of Computational Physics*, 2 (1) (1967) 12-26.

[85] A. Staniforth, J. Côté, Semi-Lagrangian integration schemes for atmospheric models—A review, *Monthly Weather Review*, 119 (9) (1991) 2206-2223.

[86] M. Mortezazadeh, L. Wang, An adaptive time-stepping semi-Lagrangian method for incompressible flows, *Numerical Heat Transfer, Part B: Fundamentals*, 75 (1) (2019) 1-18.

[87] V.K. Arghode, Y. Joshi, Rapid modeling of air flow through perforated tiles in a raised floor data center, in: Fourteenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), IEEE, 2014, pp. 1354-1365.

[88] E. Fried, I. Idelchik, Flow resistance, A Design Guide for Engineering, Hemisphere Publ, in, Co, 1989.

[89] W.A. Abdelmaksoud, T.Q. Dang, H.E. Khalifa, R.R. Schmidt, M. Iyengar, Perforated tile models for improving data center CFD simulation, in: 13th InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, IEEE, 2012, pp. 60-67.

[90] J.Z. Zhai, K.A. Hermansen, S. Al-Saadi, The Development of Simplified Rack Boundary Conditions for Numerical Data Center Models, *ASHRAE Transactions*, 118 (2) (2012).

[91] X. Zhang, J.W. VanGilder, M. Iyengar, R.R. Schmidt, Effect of rack modeling detail on the numerical results of a data center test cell, in: 2008 11th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, IEEE, 2008, pp. 1183-1190.

[92] Z.M. Pardey, J.W. VanGilder, C.M. Healey, D.W. Plamondon, Creating a calibrated CFD model of a midsize data center, in: ASME 2015 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems collocated with the ASME 2015 13th International Conference on Nanochannels, Microchannels, and Minichannels, American Society of Mechanical Engineers, 2015, pp. V001T009A029-V001T009A029.

[93] R. Schmidt, E. Cruz, Cluster of high-powered racks within a raised-floor computer data center: Effect of perforated tile flow distribution on rack inlet air temperatures, *Journal of Electronic Packaging*, 126 (4) (2004) 510-518.

[94] X. Han, W. Tian, J.W. VanGilder, M. Condor, W. Zuo, doetools, in, https://github.com/doetools/isat_ffd/releases, 2018.

[95] J.W. VanGilder, Z.R. Sheffer, X.S. Zhang, C.M. Healey, Potential Flow Model for Predicting Perforated Tile Airflow in Data Centers, *ASHRAE Transactions*, 117 (2) (2011).

[96] Q. Chen, W. Xu, A zero-equation turbulence model for indoor airflow simulation, *Energy and buildings*, 28 (2) (1998) 137-144.

[97] P. Dhoot, J. VanGilder, Z. Pardey, C. Healey, Zero-Equation Turbulence Models for Large Electrical and Electronics Enclosure Applications, in: Proceedings of 2017 ASHRAE Winter Conference, 2017.

[98] M. Jin, W. Zuo, Q. Chen, Improvements of fast fluid dynamics for simulating air flow in buildings, *Numerical Heat Transfer, Part B: Fundamentals*, 62 (6) (2012) 419-438.

[99] J.W. VanGilder, Z.M. Pardey, C. Healey, Measurement of perforated tile airflow in data centers, *ASHRAE Transactions*, 122 (2016) 88.

[100] J.W. VanGilder, X.S. Zhang, Coarse-grid CFD: The effect of grid size on data center modeling, *ASHRAE Transactions*, 114 (2008) 166.

[101] M.K. Herrlin, Rack cooling effectiveness in data centers and telecom central offices: The rack cooling index (RCI), *Transactions-American Society of Heating Refrigerating and Air conditioning Engineers*, 111 (2) (2005) 725.

[102] M.K. Herrlin, Improved data center energy efficiency and thermal performance by advanced airflow analysis, in: Proceedings of Digital Power Forum, 2007, pp. 10-12.

[103] R. Sharma, C. Bash, C. Patel, Dimensionless parameters for evaluation of thermal design and performance of large-scale data centers, in: 8th AIAA/ASME Joint Thermophysics and Heat Transfer Conference, 2002, pp. 3091.

[104] J.W. Vangilder, S.K. Shrivastava, Capture Index: An Airflow-Based Rack Cooling Performance Metric, *ASHRAE Transactions*, 113 (2007) 126.

[105] R. Zhang, K.P. Lam, S.-C. Yao, Y. Zhang, Annual coupled EnergyPlus and computational fluid dynamics simulation of natural ventilation, *Proceedings of SimBuild*, 5 (1) (2012) 314-321.

[106] W. Tian, T.A. Sevilla, D. Li, W. Zuo, M. Wetter, Fast and self-learning indoor airflow simulation based on in situ adaptive tabulation, *Journal of Building Performance Simulation*, 11 (1) (2018) 99-112.

[107] X. Han, W. Tian, W. Zuo, J.W. VanGilder, Optimization of Workload Distribution of Data Centers Based on a Self-Learning In Situ Adaptive Tabulation Model, in: Proceedings of the 16th Conference of International Building Performance Simulation Association (Building Simulation 2019), Rome, Italy, 2019.

[108] S.E. Mattsson, H. Elmqvist, Modelica-An international effort to design the next generation modeling language, *IFAC Proceedings Volumes*, 30 (4) (1997) 151-155.

[109] S. Huang, W. Zuo, M.D. Sohn, Amelioration of the cooling load based chiller sequencing control, *Applied Energy*, 168 (2016) 204-215.

[110] M. Wetter, Co-simulation of building energy and control systems with the Building Controls Virtual Test Bed, *Journal of Building Performance Simulation*, 4 (3) (2011) 185-203.

[111] T. Blochwitz, M. Otter, M. Arnold, C. Bausch, C. Clauss, H. Elmqvist, A. Junghanns, J. Mauss, M. Monteiro, T. Neidhold, The functional mockup interface for tool independent exchange of simulation models, in: Proceedings of the 8th International Modelica Conference, Linköping University Press, 2011, pp. 105-114.

[112] W. Zuo, M. Wetter, J.W. VanGilder, X. Han, Y. Fu, C. Faulkner, J. Hu, W. Tian, M. Condor, Improving Data Center Energy Efficiency through End-to-End Cooling Modeling and Optimization, in, 2020.

[113] S.V. Garimella, T. Persoons, J. Weibel, L.-T. Yeh, Technological drivers in data centers and telecom systems: Multiscale thermal, electrical, and energy management, *Applied Energy*, 107 (2013) 66-80.

[114] S.K. Shrivastava, Cooling analysis of data centers: CFD modeling and real-time calculators, State University of New York at Binghamton, 2008.

[115] D.W. Demetriou, H.E. Khalifa, M. Iyengar, R.R. Schmidt, Development and experimental validation of a thermo-hydraulic model for data centers, *HVAC&R Research*, 17 (4) (2011) 540-555.

[116] M. Iyengar, R. Schmidt, Analytical modeling for thermodynamic characterization of data center cooling systems, *Journal of Electronic Packaging*, 131 (2) (2009).

[117] K. Sun, T. Hong, X. Luo, Development of a New Prototype Energy Model for Data Centers, *ASHRAE Transactions*, 125 (2) (2019) 137-145.

[118] L.M. Billet, C.M. Healey, J.W. VanGilder, Z.M. Pardey, Data Center Cooling Efficiency with Simulation-Based Optimization, in: International Electronic Packaging Technical Conference and Exhibition, American Society of Mechanical Engineers, 2015, pp. V001T009A025.

[119] M. Wetter, Generic Optimization Program User Manual Version 3.1.0, in, 2011.

[120] X. Han, W. Tian, J. VanGilder, W. Zuo, C. Faulkner, An Open Source Fast Fluid Dynamics Model for Data Center Thermal Management, Energy and Buildings, (2020) 110599.

[121] R. Hooke, T.A. Jeeves, ``Direct Search"Solution of Numerical and Statistical Problems, Journal of the ACM (JACM), 8 (2) (1961) 212-229.

[122] H. Zhang, S. Shao, H. Xu, H. Zou, C. Tian, Free cooling of data centers: A review, Renewable and Sustainable Energy Reviews, 35 (2014) 171-182.

[123] J. Kaiser, J. Bean, T. Harvey, M. Patterson, J. Winiecki, Survey results: Data center economizer use, White Paper, (2011).

[124] Y. Chen, Y. Zhang, Q. Meng, Study of ventilation cooling technology for telecommunication base stations: Control strategy and application strategy, Energy and Buildings, 50 (2012) 212-218.

[125] S. Wang, Z. Ma, Supervisory and optimal control of building HVAC systems: A review, HVAC&R Research, 14 (1) (2008) 3-32.

[126] R. Zhou, Z. Wang, C.E. Bash, A. McReynolds, C. Hoover, R. Shih, N. Kumari, R.K. Sharma, A holistic and optimal approach for data center cooling management, in: Proceedings of the 2011 American Control Conference, IEEE, 2011, pp. 1346-1351.

[127] A. Afram, F. Janabi-Sharifi, Theory and applications of HVAC control systems–A review of model predictive control (MPC), Building and Environment, 72 (2014) 343-355.

[128] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, API design for machine learning software: experiences from the scikit-learn project, arXiv preprint arXiv:1309.0238, (2013).