

---

Theses and Dissertations

---

Fall 2010

# HVAC system modeling and optimization: a data-mining approach

Fan Tang

*University of Iowa*

Copyright 2010 Fan Tang

This thesis is available at Iowa Research Online: <http://ir.uiowa.edu/etd/895>

---

## Recommended Citation

Tang, Fan. "HVAC system modeling and optimization: a data-mining approach." MS (Master of Science) thesis, University of Iowa, 2010.

<http://ir.uiowa.edu/etd/895>.

---

Follow this and additional works at: <http://ir.uiowa.edu/etd>



Part of the [Industrial Engineering Commons](#)

HVAC SYSTEM MODELING AND OPTIMIZATION: A DATA-MINING  
APPROACH

by  
Fan Tang

A thesis submitted in partial fulfillment  
of the requirements for the Master of Science  
degree in Industrial Engineering  
in the Graduate College of  
The University of Iowa

December 2010

Thesis Supervisor: Professor Andrew Kusiak

Copyright by  
FAN TANG  
2010  
All Rights Reserved

Graduate College  
The University of Iowa  
Iowa City, Iowa

CERTIFICATE OF APPROVAL

---

MASTER'S THESIS

---

This is to certify that the Master's thesis of

Fan Tang

has been approved by the Examining Committee  
for the thesis requirement for the Master of Science degree  
in Industrial Engineering at the December 2010 graduation.

Thesis Committee: \_\_\_\_\_  
Andrew Kusiak, Thesis Supervisor

\_\_\_\_\_  
Yong Chen

\_\_\_\_\_  
Gideon Zamba

To My Parents and Family

Strength alone knows conflict, weakness is below even defeat, and is born  
vanquished.

Swetchine

## ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Professor Andrew Kusiak, for his devotion to this research. He has been the most instrumental person for my academic and research achievements. He provided the motivation, encouragement, guidance and advice which have prepared me for the challenges of future life. I was fortunately exposed to industrial applications while working in the Intelligent Systems Laboratory. This invaluable experience has allowed me to maintain a balance between theory and practice leading to realistic solutions.

I would like to thank Professor Yong Chen and Professor Gedion Zamba for serving on my Thesis Committee and providing valuable suggestions and feedback on my research.

I am also grateful for the financial support from Iowa Energy Center. Discussions with energy experts: Curt Klaassen (Iowa Energy Center), Xiaohui Zhou (Iowa Energy Center), William Haman (Iowa Energy Center), Jiong Zhou (Facilities Management, The University of Iowa) and George Paterson (Facilities Management, The University of Iowa) have provided me invaluable information for this research.

I thank all the members of the Intelligent Systems Laboratory who have worked with me and provided advice, reviews and suggestions. Special thanks to my colleagues: Mingyang Li, who worked with me to solve challenging problems in HVAC system; Zijun Zhang, who provided valuable advices for me; Wenyan Li, who shared the research experience with me; Evan Roz, who discussed with me about data analysis and system modeling; Anoop Verma, who discussed the problems I encountered; and Guanglin Xu, who cooperated with me to deal with existing problems.

And finally, and most importantly, I would like to express my sincere gratitude to my parents and my girlfriend, who solidly supported me in pursuing my career.

## ABSTRACT

Heating, ventilating and air-conditioning (HVAC) system is a complex non-linear system with multi-variables simultaneously contributing to the system process. It poses challenges for both system modeling and performance optimization. Traditional modeling methods based on statistical or mathematical functions limit the characteristics of system operation and management.

Data-driven models have shown powerful strength in non-linear system modeling and complex pattern recognition. Sufficient successful applications of data mining have proved its capability in extracting models that accurately describe the relation of inner system. The heuristic techniques such as neural networks, support vector machine, and boosting tree have largely expanded to the modeling process of HVAC system.

Evolutionary computation has rapidly merged to the center stage of solving the multi-objective optimization problem. Inspired from the biology behavior, it has shown the tremendous power in finding the optimal solution of complex problem. Different applications of evolutionary computation can be found in business, marketing, medical and manufacturing domains. The focus of this thesis is to apply the evolutionary computation approach in optimizing the performance of HVAC system. Energy saving can be achieved by implementing the optimal control setpoints with IAQ maintained at an acceptable level. A trade-off between energy saving and indoor air quality maintenance is also investigated by assigning different weights to the corresponding objective function. The major contribution of this research is to provide the optimal settings for the existing system to improve its efficiency and different preference-based operation methods to optimally utilize the resources.

## TABLE OF CONTENTS

LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
CHAPTER 1 INTRODUCTION .....	1
1.1. Review of analytical approaches in modeling and optimization .....	1
1.2. Review of data-driven approaches in modeling .....	2
1.3. Review of evolutionary computation in system optimization .....	3
1.4. Thesis structure .....	4
CHAPTER 2 SHORT-TERM PREDICTION OF HVAC ENERGY WITH A CLUSTERING APPROACH .....	5
2.1. Introduction .....	5
2.2. HVAC system structure description .....	6
2.3. Data description .....	8
2.4. Modeling of AHU system and AQI sensors .....	9
2.4.1. Parameter selection .....	9
2.4.2. Algorithm selection .....	13
2.4.3. Model construction and validation .....	16
2.5. Clustering-based model of AHU energy .....	19
2.5.1. The modeling architecture of the AHU energy .....	19
2.5.2. Clustering algorithm .....	21
2.6. Case study .....	21
2.6.1. Clustering .....	21
2.6.2. Modeling based on clustering .....	28
2.7. Clustering-based short-term prediction of AHU energy .....	26
2.8. Summary .....	29
CHAPTER 3 MULTI-OBJECTIVE OPTIMIZATION OF HVAC SYSTEM WITH AN EVOLUTIONARY COMPUTATION ALGORITHM .....	30
3.1. Introduction .....	30
3.2. Data description and optimization methodology .....	31
3.2.1 Data description .....	31
3.2.2 Optimization approach .....	33
3.3. HVAC system modeling .....	34
3.3.1. Parameter selection .....	34
3.3.2. Construction and validation of the predictive model .....	37
3.4. Optimization Algorithm .....	39
3.4.1. Model formulation .....	39
3.4.2. Optimization .....	42
3.5. Summary .....	48
CHAPTER 4 MODELING AND OPTIMIZATION OF ENERGY RELATED COMPONENTS IN HVAC SYSTEM .....	49
4.1. Introduction .....	49
4.2. Data description .....	49
4.3. Parameter and algorithm selection .....	50

4.4. Modeling building and validating.....	54
4.5. Optimization model formulation and solving.....	57
4.6. Optimization results and discussion .....	58
4.7. Summary.....	65
<b>CHAPTER 5 MULTI-OBJECTIVE OPTIMIZATION OF HVAC SYSTEM ENERGY MANAGEMENT .....</b>	<b>67</b>
5.1. Introduction.....	67
5.2. Data description and parameter selection.....	67
5.3. Algorithm selection .....	68
5.4. Modeling building and validation.....	71
5.5. Model formulation and solving .....	74
5.5.1 Single-objective optimization model formulation and solving .....	74
5.5.2 Formulation and solving of the quad-objective optimization model .....	75
5.5.3 Weight assignment and solution selection .....	79
5.6. Optimization results and discussion .....	82
5.7. Summary.....	85
<b>CHAPTER 6 CONCLUSION.....</b>	<b>87</b>
<b>REFERENCES .....</b>	<b>89</b>

## LIST OF TABLES

Table 2.1. Data description. ....	9
Table 2.2. Parameter description. ....	11
Table 2.3. Parameter selection result of the corresponding target value by boosting tree algorithm. ....	12
Table 2.4. Training and testing results for models extracted from different data-mining algorithms. ....	15
Table 2.5. Characterization of the four MLP Ensemble models. ....	16
Table 2.6. Correlation coefficients between observed and predicted values of the four models. ....	19
Table 2.7. Clusters based on CHWC-VLV. ....	24
Table 2.8. Clusters based on OA-Temp. ....	24
Table 2.9. Prediction results for four scenarios. ....	26
Table 2.10. Short-term prediction results for Scenario 1 and Scenario 4. ....	32
Table 3.1. Data description. ....	34
Table 3.2. Results of different wrapper algorithms. ....	36
Table 3.3. Parameter description. ....	37
Table 3.4. Prediction results of energy consumption model. ....	38
Table 3.5. Prediction results of room humidity model. ....	39
Table 3.6. Prediction results of room temperature model. ....	39
Table 3.7. Results of different decisions made by their preferences. ....	44
Table 4.1. Data description. ....	50
Table 4.2. Predictor importance produced by the boosting tree algorithm. ....	51
Table 4.3. The results of parameter selection by the wrapper approach. ....	52
Table 4.4. Parameter description. ....	53
Table 4.5. Algorithms selection for building the total energy model. ....	53
Table 4.6. Training and test results of the four models. ....	54

Table 4.7. Energy savings with varied adjustment length of setpoints.....	65
Table 5.1. Data description .....	68
Table 5.2. Parameter description .....	68
Table 5.3. Training and testing accuracy results for models extracted with different data-mining algorithms .....	70
Table 5.4. Correlation coefficient of the observed and predicted values of the four models.....	73
Table 5.5. Instance used in single-objective optimization.....	74
Table 5.6. Description of the eight weight assignment scenarios.....	80
Table 5.7. Solutions for the eight scenarios at some time stamp.....	82

## LIST OF FIGURES

Figure 1.1 Structure of the thesis. ....	4
Figure 2.1. Schematic of zones in the building served by AHU-A and AHU-B. ....	7
Figure 2.2. Schematic of AHU-B of HVAC system. ....	7
Figure 2.3. Outside air temperature of the data collected in the whole period. ....	9
Figure 2.4. Test results obtained from the energy consumption model. ....	17
Figure 2.5. Test results obtained from the indoor temperature model. ....	17
Figure 2.6. Test results obtained from the indoor humidity model. ....	18
Figure 2.7. Test results obtained from the CO2 concentration model. ....	18
Figure 2.8. Cluster-based scheme for model building ....	20
Figure 2.9. Scatter plots showing relationships between energy consumption and selected parameters. ....	22
Figure 2.10. Graphical interpretation of clustering based on CHWC-VLV. ....	23
Figure 2.11. Graphical interpretation of clustering based on OA-Temp. ....	23
Figure 2.12. Mean absolute error for four scenarios. ....	25
Figure 2.13. Mean absolute percentage error for four scenarios. ....	26
Figure 2.14. The architecture of clustering-based short-term prediction of AHU energy. ....	27
Figure 2.15. Short-term prediction of AHU energy for Scenario 1. ....	28
Figure 2.16. Short-term prediction of AHU energy for Scenario 4. ....	28
Figure 3.1. The daily schedule of supply air static pressure and internal load. ....	32
Figure 3.2. Optimization framework. ....	33
Figure 3.3. Change of CHWC-VLV position on 04/06/2010. ....	35
Figure 3.4. Feasible solutions solved by SPEA at some time stamp. ....	43
Figure 3.5. Test results of optimized energy consumption. ....	45
Figure 3.6. Test results of thermal constraints. ....	45
Figure 3.7. Test results of optimized room temperature. ....	46
Figure 3.8. Test results of optimized room humidity. ....	46

Figure 3.9. Recommended supply air temperature set point. ....	47
Figure 3.10. Recommended supply air static pressure set point. ....	47
Figure 4.1. Test results for the chiller energy model. ....	55
Figure 4.2. Test results for the fan energy model. ....	55
Figure 4.3. Test results for the pump energy model. ....	56
Figure 4.4. Test results for the reheat energy model. ....	56
Figure 4.5. The total energy before and after optimization. ....	59
Figure 4.6. The air temperature setpoint before and after optimization. ....	60
Figure 4.7. The supply air static pressure setpoint before and after optimization. ....	60
Figure 4.8. The fan energy before and after optimization. ....	61
Figure 4.9. The pump energy before and after optimization. ....	61
Figure 4.10. The reheat device energy before and after optimization. ....	62
Figure 4.11. The chiller energy before and after optimization. ....	62
Figure 4.12. The cooling output before and after optimization. ....	63
Figure 4.13. Comparison of the IAQ metrics before and after optimization. ....	64
Figure 5.1. Test results from the total energy model ....	71
Figure 5.2. Test results from the facility temperature model. ....	72
Figure 5.3. Test results from the facility relative humidity model ....	72
Figure 5.4. Test results from the facility CO <sub>2</sub> concentration model ....	73
Figure 5.5. Two dimensional solution spaces. ....	78
Figure 5.6. The solution process. ....	78
Figure 5.7. The original and recommended setpoints of the supply air temperature. ....	82
Figure 5.8. The original and recommended setpoints of the supply air static pressure. ....	83
Figure 5.9. The original and the optimized total energy. ....	83
Figure 5.10. The original and the optimized facility temperature ....	84
Figure 5.11. The original and the optimized facility relative humidity ....	84
Figure 5.12. The original and the optimized facility CO <sub>2</sub> concentration. ....	85

## CHAPTER 1

### INTRODUCTION

HVAC system is designed to provide a comfortable and desired environment for the occupants, in addition to meeting any special process requirements, such as indoor air quality. The maintenance of a healthy indoor condition of HVAC system is significant since people spend more than half of their time indoors. The issue of growing energy use has merged to the stage which draws sufficient attentions of not only commercial managers, but also researchers. According to the published statistics, HVAC system frequently consumes over 60% of the energy use in buildings [1, 2]. Therefore, the operation effectiveness and efficiency of HVAC system has become a focus.

The operation of HVAC system is a multi-angle problem. Simply minimizing the energy consumption without considering the indoor air quality control is not acceptable. The optimal control strategies should reduce the system cost and energy use while maintaining the thermal comfort at an allowable level. With both economic cost and occupancy comfort involved, a comprehensive way of system modeling and performance optimization is addressed in this paper.

#### 1.1. Review of analytical approaches in modeling and optimization

The analytical approaches for modeling HVAC system depend on the physics-based models or simulation software. Wang et al [3] presented a simple hybrid model based on the heat transfer mechanism and the energy balance principle to predict the performance of chilled water cooling coils in a static state. Yu et al [4] developed and evaluated the simulation model for dynamic performance of both dry and wet cooling coils based on energy equations and mass balance equations. Yao et al [5] proposed a dynamic model describing the cooling coils' heat and mass exchange using classical control theory. Such analytical models are reliable once certain basic assumptions and simplifications are achieved. However, detailed physics-based models are always computationally expensive due to their complexity and non-linearity, which results in the

troubles in practical application [6]. To overcome the barrier, simulation-based models have been widely investigated. Sufficient surveys and summaries have been addressed regarding the different simulation programs [7-10]. Wang [11] developed dynamic models including multiple components to simulate the realistic performances of the chilling system. The simulation exercises were tested and evaluated by the on-line control of EMCS local strategy and supervisory strategy in different seasons. Crawley et al [12] broadly introduced twenty simulation programs applied nowadays and a basic comparison was made to show the corresponding features and capabilities. Although the merits in simulation-based modeling exist, one restriction is that many components models are steady state which is not suitable for handling high frequency disturbances [13].

To solve optimization problems formulated by analytical models, many nonlinear local optimization techniques can be used. Sun et al [14] developed a comprehensive simulation-based sequential quadratic programming (CSB-SQP) algorithm to optimally control the HVAC system. Rink et al [15] applied the state increment dynamic programming to solve the optimization problem of multi-zone HVAC system which was demonstrated to be efficient in saving energy. Kota et al [16] presented the DDP (differential dynamic programming) technique of optimal control in HVAC systems and they compared its performances with sequential quadratic programming method.

### 1.2. Review of data-driven approaches in modeling

Be different from analytical approaches, a data-driven approach is derived from empirical behavior and heuristic searching process of the system. The modeling approach that has drawn the most attention in the last few years seems to be the neural networks [17, 18]. It has a tremendous power in deriving and extracting the accurate patterns from complicated, noisy, and imprecise data. A lot of applications have been achieved by using neural networks to construct the non-linear energy consumption model in HVAC system. A typical application of data driven-methods is predicting steam load in buildings [19]. Another example includes the use of a neural

network to predict heating energy consumption [20]. Kalogirou [21-23] applied neural networks to modeling solar water heating systems, and HVAC system. Kreider and Wang [24] used neural networks to predict the rate of energy use in commercial buildings. One shortcoming of data-driven models is that insufficient data will result in the decrease of model accuracy since the training data may only cover a small range of data patterns.

### 1.3. Review of evolutionary computation in system optimization

The operation of HVAC system is a critical activity in terms of optimizing the control settings to reduce the energy consumption, improving the system efficiency, and preserving the thermal comfort for the occupants. The performance of the existing HVAC system can be largely improved by adjusting the control set points to maximize the overall system capacity and efficiency. Ke and Mumma [25] studied the impact on energy consumption of tuning the supply air temperature set point in a VAV system and found that an optimal supply air temperature setting existed for minimizing the energy cost. Wang et al. [26] proposed a systematic approach for an on-line control strategy of air-conditioning systems. A cost function was formed to weight the energy consumption of the entire system containing fan, pump, and chiller, the indoor thermal comfort, indoor air quality, and the total ventilation rate. The genetic algorithm was applied to search the optimal control settings resulting in reduction of energy based on the incremental dynamic models with self-tuning of the VAV system. Nassif et al. [27, 28] applied the multi-objective evolutionary algorithms to optimize a multi-zone HVAC system, and supervisory control settings were found to reduce the energy consumption as well as maintaining the thermal comfort. Mossolly et al. [29] examined three control strategies on the system component models solved by the genetic algorithm, which is implemented in Matlab. It was proven that huge energy saving could be achieved by varying the system parameters. Magnier and Haghghat [30] used a simulation-based artificial neural network (ANN) to capture the mapping of building behavior, and implemented the ANN model into genetic algorithm for optimization. By doing so, significant improvements regarding energy performance and thermal

comfort were achieved and a large number of potential designs for operating the HVAC system were revealed.

#### 1.4. Thesis structure

Figure 1.1 illustrates the structure of the thesis. Chapter 1 introduces a clustering-based HVAC system modeling and short-term prediction. In Chapter 2, an evolutionary computation algorithm is applied to solve a multi-objective optimization problem in HVAC system. Energy related components such as heat, fan, pump and reheat are optimized by a single objective optimization algorithm, respectively in Chapter 3. Finally, Chapter 5 presents an optimal control strategy for HVAC system energy management.

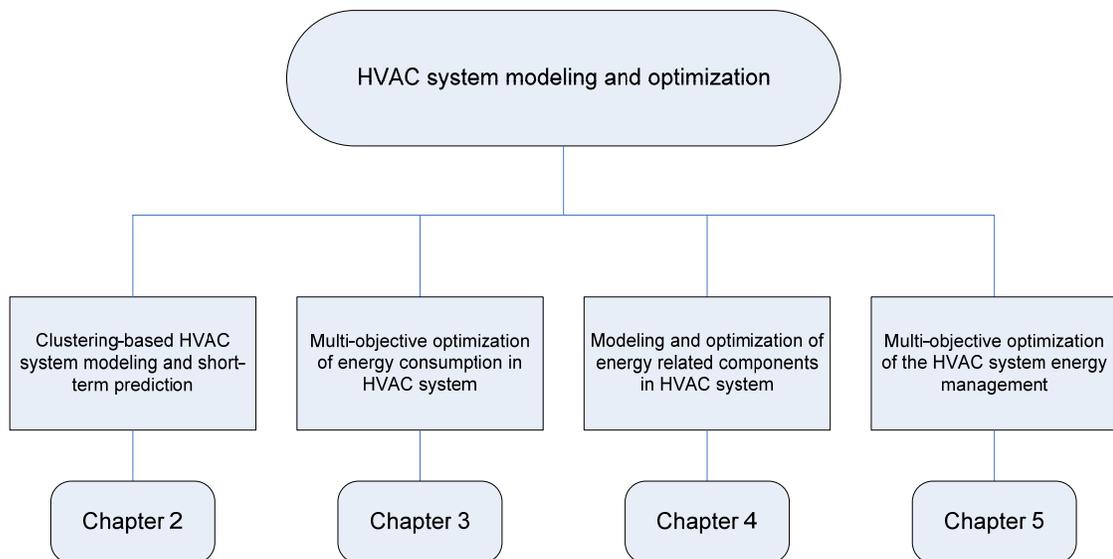


Figure 1.1 Structure of the thesis.

## CHAPTER 2

### SHORT-TERM PREDICTION OF HVAC ENERGY WITH A CLUSTERING APPROACH

#### 2.1. Introduction

The energy used for heating, ventilating, and air-conditioning has become a concern, as it constitutes over 50% of the energy consumed by office buildings in the US [1, 2]. Such energy can be optimized if the underlying model is known. In this research, the energy needed to maintain thermal comfort in an office-type building is studied. This energy is supplied by a heating, ventilating, and air-conditioning (HVAC) system. For the best optimization results, a short-term prediction model is developed. This model is referred to in this section as the HVAC energy model. The HVAC energy model is complex, non-linear, and depends on a number of parameters, e.g., weather, fan speed, and chilled water valve position. Consequently, it is not easy to capture the relationship between the input and output parameters.

As mentioned in the Chapter 1, neural network has shown tremendous power in HVAC system modeling. In addition to the single neural network model, a different type of model combining the data partitioning techniques with neural network algorithm shows a good potential in improving the existing data-driven models. Sfetsos [31] introduced a hybrid clustering model for short-term load forecasting. Sub-models were constructed based on the clusters using neural network. Prediction error was reduced by approximately 7.5% compared with single neural network model. Hiroyuki and Atsushi [32, 33] applied the deterministic annealing clustering model combined with neural networks to predict short-term load of power system. Kusiak and Li [34] applied the clustering method for short-term prediction of wind power. The clustering-based neural network model was developed and it produced accurate prediction even using a small number of inputs.

A four-phase method for the short-term prediction of HVAC energy is presented. In Phase 1, the most important parameters are selected, and a parameter sensitivity analysis is performed. The input data is grouped into clusters in Phase 2. In Phase 3, a multi-layer perceptron (MLP) is constructed in each cluster. In Phase 4, the effectiveness of the proposed clustering approach is tested. The performance of the cluster-based HVAC energy model is discussed and a conclusion is drawn based on the results.

### 2.2. HVAC system structure description

The investigated HVAC system is installed at Energy Resource Station (ERS) in Ankeny, Iowa. It consists of two independent air-handling units (AHU-A and AHU-B) providing the loads for 20 interior zones in the whole building. Each air-handling unit serves 4 test rooms, which are used to collect the original data in this experiment, located in all the directions of the building. For each zone, a variable air volume (VAV) box is connected to the air-handling unit to meet the load of the room thermal comfort. The outside weather conditions are also recorded by the sensors implemented around the building. The experiment in ERS is designated to investigate the impacts of different parameters on the total energy consumption for commercial buildings. Figure 2.1 shows the floor plan of the building served by AHU-A and AHU-B. Figure 2.2 shows the schematic of AHU-B of the HVAC system.

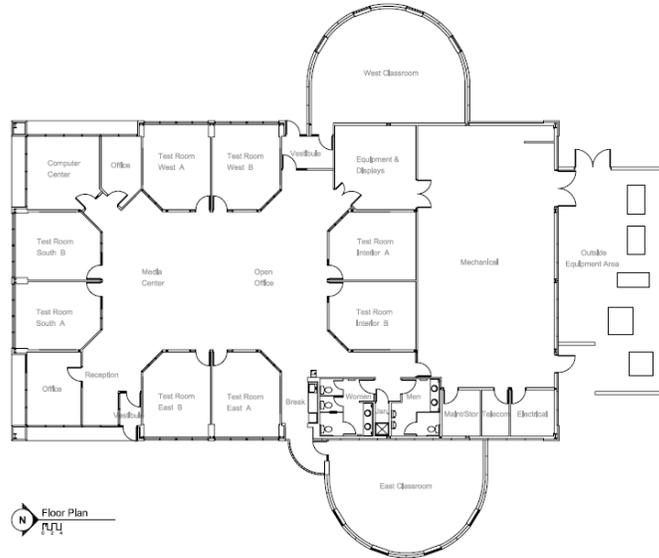


Figure 2.1. Schematic of zones in the building served by AHU-A and AHU-B.

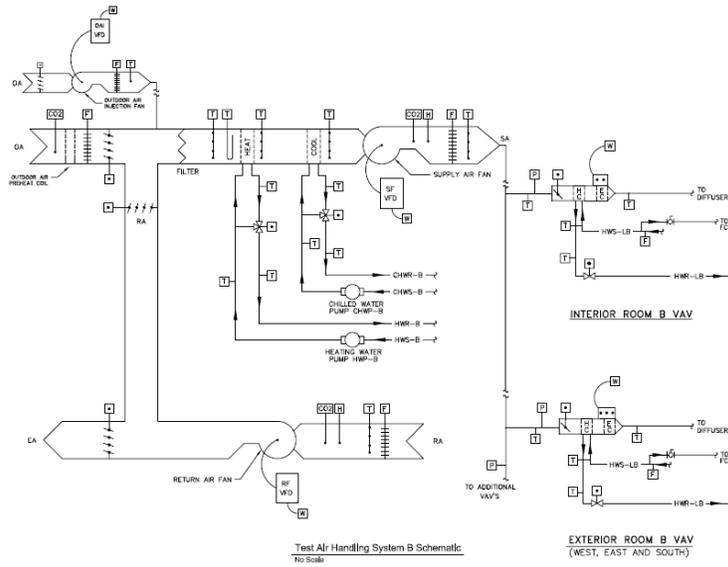


Figure 2.2. Schematic of AHU-B of HVAC system.

### 2.3. Data description

The data collected in this research was obtained from an experiment performed at the ERS of the Iowa Energy Center. Two setpoints, namely the AHU supply air temperature setpoint and static pressure setpoint were adjusted in both AHU-A and AHU-B systems. The supply air temperature (SAT) setpoint varied from 50 °F (10°C) to 65 °F (18.33°C) with 1 °F (0.55°C) increments. The supply air static pressure (SASP) setpoint varied from 1.2 in WG (0.3 kPa) to 1.8 in WG (0.45 kPa) with 0.2 in WG (0.05 kPa) increments. Data on more than 500 parameters was collected at 1 min sampling intervals. Sensors measured air temperature, air humidity, and CO<sub>2</sub> concentration in each thermal zone of HVAC system. Weather patterns such as outside air temperature, humidity, solar normal flux, were also recorded.

The original data was recorded over three different time periods covering summer, winter, and a transient season. In the summer season, data was collected from two experiments performed from August 1 to August 16, 2009 and from September 22 to October 6, 2009. For the winter and transient season, the data was collected from February 3 to February 15, 2010 and from April 1 to April 14, 2010, respectively. A three-day validation experiment was conducted from April 15 to April 17, 2010 with the AHU SAT setpoint set at 55°F and the SASP setpoint set at 1.4 WG. The data collected from the three seasons was combined to obtain the joint data set of 2688 instances from AHU-A and AHU-B. Figure 2.3 illustrates the outside air temperature change during all three time periods. To reduce the error produced by time delay and system error, the original 1 min data was aggregated to 1 h interval data by averaging the values of all of the parameters. After preprocessing, The joint data set of 2688 instances was randomly sampled to produce a training set of 1882 instances (70% of the data) and a test set of 806 instances (30% of the data). Table 2.1 summarizes the data collected in ERS experiment in detail.

Table 2.1. Data description.

Data set	Data Type	Time Period	No. of Instances
1	Summer season	08/01/2009-08/16/2009 & 09/22/2009-10/06/2009	1488
2	Winter season	02/03/2010-02/15/2010	576
3	Transient season	04/02/2010-04/14/2010	624
4	The whole year	cumulating the three separate seasons	2688
5	Prediction	04/15/2010-04/17/2010	144

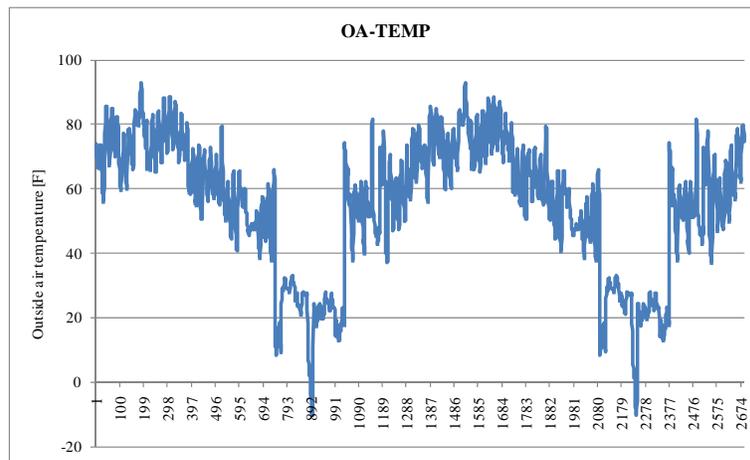


Figure 2.3. Outside air temperature of the data collected in the whole period.

## 2.4. Modeling of AHU system and AQI sensors

### 2.4.1. Parameter selection

Parameter selection is critical in the construction of models. A typical HVAC system may contain hundreds of parameters. Some of the parameters collected are relevant to the output, while others could be irrelevant or redundant. The presence of irrelevant or redundant parameters may mask the primary patterns discovered in data mining. Redundant parameters duplicate the information contained in other parameters, making the model more complex than it should be. Eliminating redundant or less important parameters may improve the accuracy, scalability, and comprehensibility of the resulting model [35].

Based on the domain knowledge, the twenty-one originally selected parameters were divided into four groups: performance parameters, controllable parameters, uncontrollable parameters, and target parameters. Two setpoints, the AHU supply air temperature and the supply air duct static pressure, were selected as the performance parameters to be adjusted for energy optimization. The controllable parameters (e.g., fan speed, chilled water coil valve position (CHWC-VLV), and mixed air temperature (MA-Temp)) were highly correlated to the AHU energy and AQI (air quality index) values, and thus had a large effect on the output results. The weather data represent uncontrollable parameters. Four outputs, the AHU energy consumption, indoor air temperature, indoor air humidity, and indoor CO<sub>2</sub> concentration, were the target parameters to be predicted by the model. Table 2.2 lists all of the parameters and their definitions.

Table 2.2. Parameter description.

Parameter Type	Parameter Name	Description	Unit
<b>Performance Parameter</b>	SAT setpoint	AHU supply air temperature set point	Deg F
	SASP setpoint	Supply air duct static pressure set point	kPa
<b>Controllable Parameter</b>	CHWC-VLV	Chilled water coil valve position	% Open
	SA-Humd	Supply air humidity	% RH
	MA-Temp	Mixed air temperature	Deg F
	CHWC-EWT	Chilled water coil entering water temperature	Deg F
	SA-CFM	Supply air fan speed	CFM
	RA-CFM	Return air fan speed	CFM
<b>Uncontrollable Parameter</b>	OA-Temp	Outside air humidity	Deg F
	OA-Humd	Outside air temperature	% RH
	OA- CO <sub>2</sub>	Outside air CO <sub>2</sub> concentration	PPM
	IR-Radia	Infrared Radiation	B/HFt <sup>2</sup>
	SOL-Horz	Solar normal flux	B/HFt <sup>2</sup>
	SOL-Beam	Solar beam	B/HFt <sup>2</sup>
	BAR-Pres	Barometric Pressure (normalized to sea level)	mBar
	WIND-Vel	Outside wind velocity	MPH
WIND-Dir	Outside wind direction	Deg <sub>n</sub> =0	
<b>Target Parameter</b>	AHU-Energy	Energy consumption of AHU system	kJ
	Indoor-Temp	Indoor temperature	Deg F
	Indoor-Humd	Indoor humidity	% RH
	Indoor- CO <sub>2</sub>	Indoor CO <sub>2</sub> concentration	PPM

To analyze the sensitivity of the parameters and select the most important parameters for the prediction of the target values, the selection of controllable and uncontrollable parameters was performed using the boosting tree algorithm [36]. Because of the different characteristics of the outputs, the input parameters used to build predictive models were ranked. The boosting tree algorithm was applied four times with the corresponding output as the target value. The results are shown in Table 2.3.

Table 2.3. Parameter selection result of the corresponding target value by boosting tree algorithm.

Energy Consumption		Indoor Temperature	
Variable	Importance	Variable	Importance
CHWC-VLV	1.00	SA-CFM	1.00
OA-Temp	0.95	RA-CFM	0.98
MA-Temp	0.93	MA-Temp	0.82
CHWC-EWT	0.92	SA-Humd	0.75
RA-CFM	0.78	OA-Temp	0.71
SA-Humd	0.77	IR-Radia	0.68
IR-Radia	0.76	OA- CO <sub>2</sub>	0.68
OA- CO <sub>2</sub>	0.73	SOL-Horz	0.60
SA-CFM	0.69	SOL-Beam	0.58
SOL-Beam	0.69	OA-Humd	0.50
SOL-Horz	0.56	CHWC-VLV	0.49
OA-Humd	0.40	CHWC-EWT	0.47
WIND-Dir	0.19	WIND-Dir	0.43
BAR-Pres	0.13	BAR-Pres	0.36
WIND-Vel	0.12	WIND-Vel	0.23
Indoor Humidity		Indoor CO <sub>2</sub> Concentration	
Variable	Importance	Variable	Importance
SA-Humd	1.00	BAR-Pres	1.00
OA-Temp	0.88	IR-Radia	0.83
IR-Radia	0.81	SA-CFM	0.83
CHWC-VLV	0.80	CHWC-EWT	0.79
CHWC-EWT	0.78	MA-Temp	0.78
OA- CO <sub>2</sub>	0.75	WIND-Vel	0.77
MA-Temp	0.71	CHWC-VLV	0.76
SOL-Beam	0.65	RA-CFM	0.72
RA-CFM	0.44	OA-Temp	0.71
WIND-Vel	0.43	OA-Humd	0.69
BAR-Pres	0.43	SA-Humd	0.65
SA-CFM	0.39	WIND-Dir	0.56
SOL-Horz	0.38	SOL-Horz	0.55
WIND-Dir	0.36	SOL-Beam	0.53
OA-Humd	0.26	OA- CO <sub>2</sub>	0.51

Based on the data shown in Table 2.3, the eight parameters with the largest importance metric values were selected. Therefore, two performance parameters, along with the eight controllable and uncontrollable parameters, were ultimately used to build the energy consumption and AQI models.

#### 2.4.2. Algorithm selection

After parameter selection, the predictive models of the AHU system and AQI are expressed in equation (2.1) to (2.4).

$$y_{Energy}(t) = f(x_{SAT\_SPT}, x_{SASP\_SPT}, x_{CHWC\_VIV}, x_{OA\_Temp}, x_{MA\_Temp}, x_{CHWC\_EWT}, x_{RA\_CFM}, x_{SA\_Humd}, x_{IR\_Radia}, x_{OA\_CO_2}) \quad (2.1)$$

$$y_{Humd}(t) = f(x_{SAT\_SPT}, x_{SASP\_SPT}, x_{SA\_Humd}, x_{OA\_Temp}, x_{IR\_Radia}, x_{CHWC\_VIV}, x_{CHWC\_EWT}, x_{OA\_CO_2}, x_{MA\_Temp}, x_{SOL\_Beam}) \quad (2.2)$$

$$y_{Temp}(t) = f(x_{SAT\_SPT}, x_{SASP\_SPT}, x_{SA\_CFM}, x_{RA\_CFM}, x_{MA\_Temp}, x_{SA\_Humd}, x_{OA\_Temp}, x_{IR\_Radia}, x_{OA\_CO_2}, x_{SOL\_Horz}) \quad (2.3)$$

$$y_{CO_2}(t) = f(x_{SAT\_SPT}, x_{SASP\_SPT}, x_{BAR\_Pres}, x_{IR\_Radia}, x_{SA\_CFM}, x_{CHWC\_EWT}, x_{MA\_Temp}, x_{WIND\_Vel}, x_{CHWC\_VIV}, x_{RA\_CFM}) \quad (2.4)$$

Where  $y_{Energy}(t)$ ,  $y_{Humd}(t)$ ,  $y_{Temp}(t)$ ,  $y_{CO_2}(t)$  denote the total energy consumption of AHU system, average indoor temperature, average indoor humidity, and the average indoor  $CO_2$  concentration during 1 hour time period, respectively.

Five data-mining algorithms were used to extract the mapping between inputs and the corresponding outputs: Boosting Tree [36], Random Forest [37], Support Vector Machine (SVM) [38], Multi-layer Perceptron (MLP) [39] and MLP Ensemble [40].

Boosting tree is a machine learning meta-algorithm for supervised learning. Boosting is an iterative procedure used to adaptively modify the distribution of training examples so that the base predictors focus on learning instances misclassified by the previous biased examples.

Random forest is a class of ensemble methods consisting of multiple decision trees, where each tree is generated based on the values of an independent set of random variables. Unlike the adaptive approach used in the boosting tree algorithm, the random variables are generated from a fixed probability distribution.

SVM is a supervised learning algorithm that uses kernel functions. It is used in binary classification and regression. Using specific kernel functions, the original vector space is

transformed into a higher-dimensional space where a separated hyperplane is constructed with the maximum margin.

MLP is also a non-parametric algorithm modeled after cognitive learning for the prediction of patterns that are not part of the training data set. MLP derives relations from complex, noisy, and imprecise data, which are often impossible to model with analytical or parametric techniques. Recognizing that each single MLP may make different and perhaps complementary errors, MLP ensembles are used to pool the results from different MLPs to find a composite system that outperforms any individual classifier.

The five different data-mining algorithms were tested using the joint data set for the construction of predictive models. To evaluate the performance of the different algorithms, the following four metrics (see equations (2.5)-(2.8)) have been used to measure the prediction accuracy of the model: the mean absolute error (MAE), the standard deviation of absolute error (Std\_AE), the mean absolute percentage error (MAPE) and the standard deviation of absolute percentage error (Std\_APE) [41]:

$$MAE = \frac{\sum_{i=1}^N |\hat{y}_i - y_i|}{N} \quad (2.5)$$

$$MAPE = \frac{\sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right|}{N} \quad (2.6)$$

$$Std\_AE = \sqrt{\frac{\sum_{i=1}^N \left( |\hat{y}_i - y_i| - \frac{\sum_{i=1}^N |\hat{y}_i - y_i|}{N} \right)^2}{N-1}} \quad (2.7)$$

$$Std\_APE = \sqrt{\frac{\sum_{i=1}^N (APE(i) - MAPE)^2}{N-1}} \quad (2.8)$$

where  $\tilde{y}$  is the predicted value obtained from the predictive model,  $y$  is the observed target value measured, and  $N$  is the number of data points used for training or testing.

The prediction accuracies of the predictive models in terms of energy consumption, indoor temperature, indoor humidity, and indoor CO<sub>2</sub> concentration are presented in Table 2.4. Both the training and test errors are compared. The MLP ensemble outperformed the other four algorithms. Therefore, it was selected to construct the predictive models.

Table 2.4. Training and testing results for models extracted from different data-mining algorithms.

Algorithm	Data set	Energy model				Temperature model			
		MAE	Std of AE	MAPE	Std of APE	MAE	Std of AE	MAPE	Std of APE
MLP	Training	451.91	381.09	3.80%	4.29%	0.34	0.31	0.48%	0.43%
	Testing	450.27	387.19	3.79%	4.29%	0.42	0.40	0.59%	0.56%
MLP Ensemble	Training	402.75	339.94	3.44%	4.04%	0.32	0.30	0.44%	0.42%
	Testing	412.96	364.02	3.62%	4.80%	0.37	0.37	0.52%	0.52%
Boosted Tree	Training	1053.28	950.93	8.71%	10.11%	0.55	0.51	0.77%	0.73%
	Testing	1091.68	978.15	8.55%	9.19%	0.57	0.55	0.80%	0.78%
Random Forest	Training	980.25	1025.77	7.15%	7.67%	0.49	0.50	0.69%	0.72%
	Testing	1031.50	1090.47	7.63%	8.06%	0.52	0.56	0.74%	0.78%
SVM	Training	840.57	634.81	7.25%	6.83%	0.77	0.58	1.08%	0.83%
	Testing	835.56	650.25	7.01%	7.08%	0.72	0.57	1.02%	0.81%
Algorithm	Data set	Humidity model				CO <sub>2</sub> model			
		MAE	Std of AE	MAPE	Std of APE	MAE	Std of AE	MAPE	Std of APE
MLP	Training	0.69	0.63	2.19%	2.07%	7.88	7.31	1.95%	1.76%
	Testing	0.72	0.65	2.29%	2.26%	9.35	8.91	2.32%	2.19%
MLP Ensemble	Training	0.69	0.63	2.17%	2.03%	6.23	6.32	1.55%	1.54%
	Testing	0.73	0.63	2.29%	2.19%	7.54	7.74	1.86%	1.84%
Boosted Tree	Training	2.78	2.25	8.69%	7.93%	11.55	11.08	2.84%	2.52%
	Testing	2.86	2.31	9.00%	8.55%	11.96	12.17	2.92%	2.74%
Random Forest	Training	1.69	1.56	5.16%	5.28%	15.05	12.88	3.74%	2.98%
	Testing	1.80	1.59	5.78%	6.28%	17.17	14.99	4.23%	3.37%
SVM	Training	1.40	1.02	4.96%	5.51%	16.27	14.62	3.97%	3.25%
	Testing	1.43	1.09	4.78%	5.28%	15.98	12.64	3.92%	2.96%

### 2.4.3. Model construction and validation

The construction of neural network models is a self-adaptive process for minimizing the prediction errors. The total squared error is considered as the cost function. During the training process, the weights of the hidden units are modified to minimize the cost function. For each individual ensemble model, one hundred networks were trained with the number of hidden neurons in the network varying from 10 to 35. The best five networks were then selected.

Detailed characterizations of the four MLP ensemble models with five MLPs are shown in Table 2.5.

Table 2.5. Characterization of the four MLP Ensemble models.

Energy Model			Temperature Model		
Hidden units	Hidden activation function	Output activation function	Hidden units	Hidden activation function	Output activation function
30	Exponential	Identity	27	Hyperbolic function	Identity
23	Hyperbolic function	Logistic	28	Exponential	Logistic
32	Hyperbolic function	Exponential	30	Exponential	Exponential
27	Logistic	Exponential	35	Hyperbolic function	Logistic
18	Exponential	Logistic	34	Exponential	Exponential
Humidity Model			CO <sub>2</sub> Concentration Model		
Hidden units	Hidden activation function	Output activation function	Hidden units	Hidden activation function	Output activation function
17	Hyperbolic function	Logistic	31	Hyperbolic function	Exponential
27	Exponential	Exponential	22	Logistic	Hyperbolic function
21	Exponential	Hyperbolic function	33	Logistic	Logistic
21	Hyperbolic function	Identity	22	Hyperbolic function	Exponential
27	Hyperbolic function	Logistic	21	Hyperbolic function	Identity

The observed and predicted values obtained from the four models of Table 2.5 are shown in Figure 2.4 through Figure 2.7.

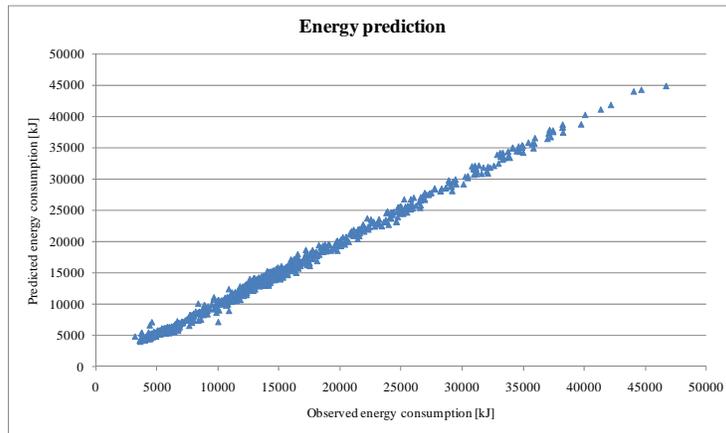


Figure 2.4. Test results obtained from the energy consumption model.

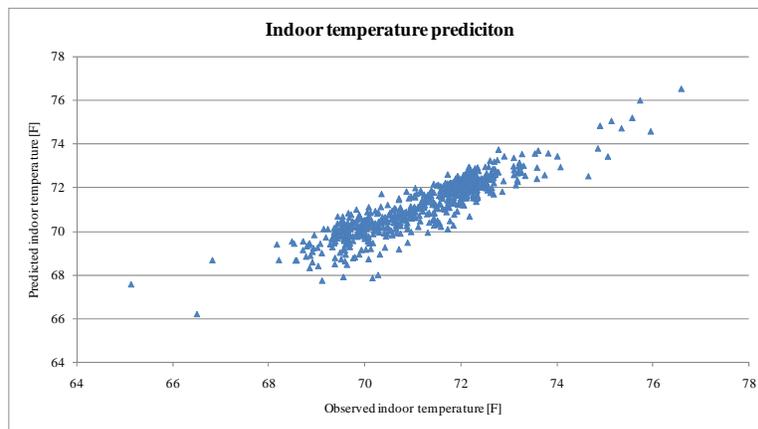


Figure 2.5. Test results obtained from the indoor temperature model.

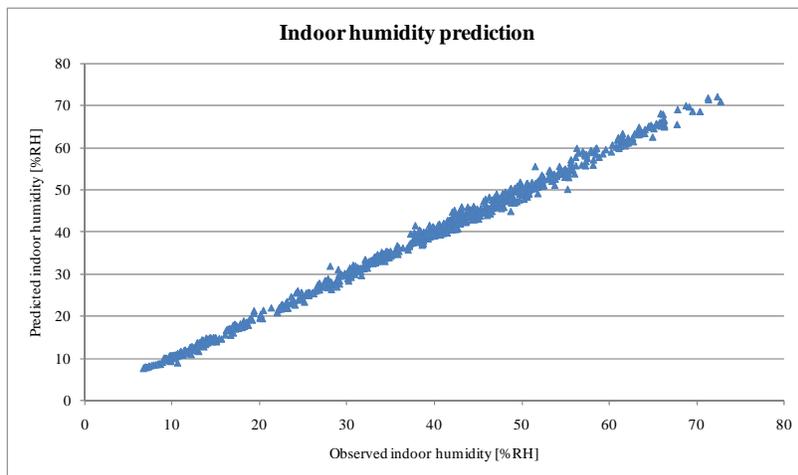


Figure 2.6. Test results obtained from the indoor humidity model.

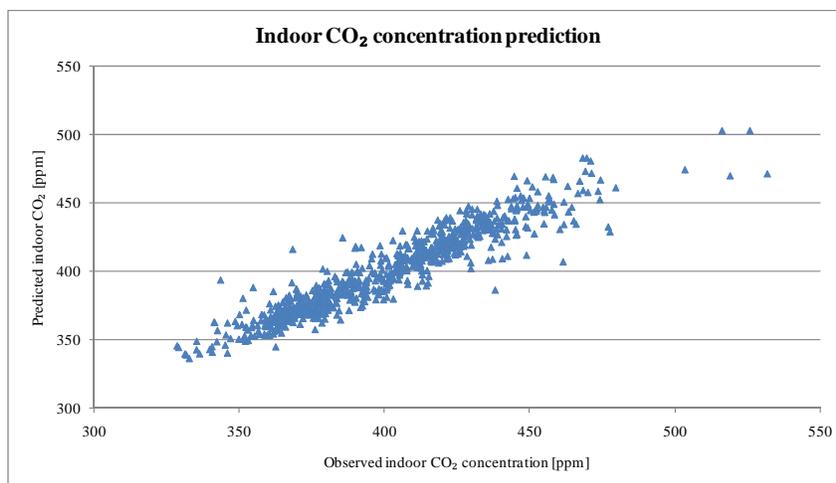


Figure 2.7. Test results obtained from the CO<sub>2</sub> concentration model.

As shown in Figure 2.4 through Figure 2.7, the predicted values of energy consumption and indoor humidity were highly correlated to their observed values. The predicted values of the indoor air temperature and CO<sub>2</sub> concentration were scattered because their thermal characteristics were easily affected by events in the room. Table 6 shows the correlation between

the observed and predicted values for the four prediction models. Table 2.6 shows the correlation between the observed and predicted values for the four prediction models.

Table 2.6. Correlation coefficients between observed and predicted values of the four models.

Variable	Means	Std. Dev.	Correlation Coefficient
Observed Total Energy	14968.42	8584.58	99.79%
Predicted Total Energy	14967.97	8560.55	
Observed Indoor Temp	71.36	1.27	91.29%
Predicted Indoor Temp	71.32	1.20	
Observed Indoor Humd	37.32	15.58	99.81%
Predicted Indoor Humd	37.35	15.54	
Observed Indoor CO <sub>2</sub>	400.56	32.31	94.26%
Predicted Indoor CO <sub>2</sub>	400.03	30.10	

## 2.5. Clustering-based model of AHU energy

### 2.5.1. The modeling architecture of the AHU energy

Figure 2.8 illustrates a clustering scheme, where P is a subset of N parameters representing the controllable and uncontrollable parameters of an AHU. The parameters in set P are denoted as  $P_1, P_2, \dots, P_N$  and are selected as potential candidates to build the model using domain knowledge. Machine learning algorithms (boosting tree, wrapper, and neural network) were applied to select M out of N parameters. The cluster-based scheme for model building involves four phases, as shown in Figure 2.8.

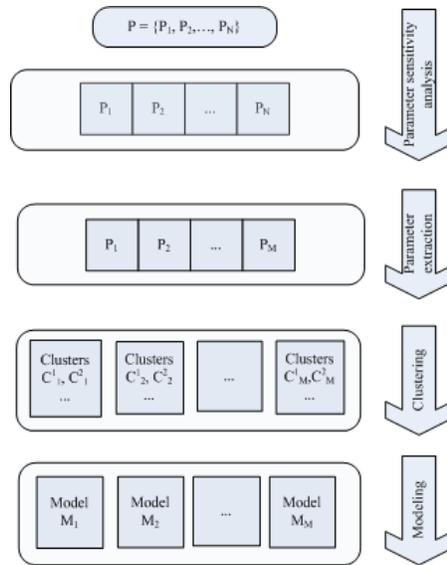


Figure 2.8. Cluster-based scheme for model building

Some machine learning techniques such as boosting tree, wrapper, and neural networks can be conducted for parameter selection. All the parameters will be ranked by a value of importance to the corresponding target. Not all the parameters have significant impact on the target value. Some may be little relevant or even not related. Considering the exhausted computation time of building separate clusters for each parameter, the top parameters with the highest importance are selected to be clustered. For each parameter selected for clustering, several clusters were constructed. The clusters resulting for the most important parameters are as follows:

$$\begin{aligned}
 C^1 &= \{C^1_1, C^1_2, \dots, C^1_a\} \\
 C^2 &= \{C^2_1, C^2_2, \dots, C^2_b\} \\
 &\dots \\
 C^M &= \{C^M_1, C^M_2, \dots, C^M_c\}
 \end{aligned} \tag{2.9}$$

where  $1, 2, \dots, M$  are the most significant parameters in the input space and  $a, b, \dots, c$  denote the numbers of clusters produced for each parameter.

Building a model for each cluster is computationally less expensive than constructing a model from the original data.

### 2.5.2. Clustering algorithm

Clustering the data set results in the clusters  $C_1, C_2, \dots, C_M$  with the corresponding centroids  $y_1, y_2, \dots, y_M$ . The clusters are created by grouping data points so that the distance between a centroid and the data in the cluster is minimized. The distance function is expressed as a squared error function (2.10):

$$D = \sum_{j=1}^k \sum_{i=1}^n \left\| x_i^{(j)} - y_j \right\|^2 \quad (2.10)$$

where  $\left\| x_i^{(j)} - y_j \right\|^2$  is the distance between a data point,  $x_i^{(j)}$ , and the cluster centre,  $y_j$ .

The K-means algorithm is widely used in data mining. A version of the K-means algorithm with a bounded number of clusters (see Step 1) is presented next:

Step 1: Set the range of the initial clusters as [a,b].

Step 2: Classify data into clusters with the closest distance between the center (centroid) and the data for each specified number of clusters.

Step 3: When all objects have been assigned, recalculate the positions of the K centroids.

Step 4: Stop if the convergence criterion is satisfied. Otherwise, return to Step 2.

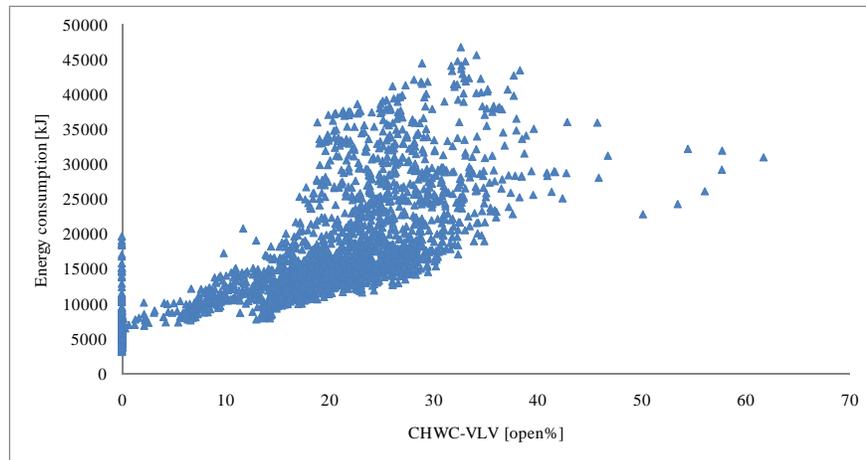
Step 5: Calculate the cost function of clusters with different initial clusters.

Step 6: Choose the optimal clusters when the difference of cost functions between the current and the following clusters reach the threshold.

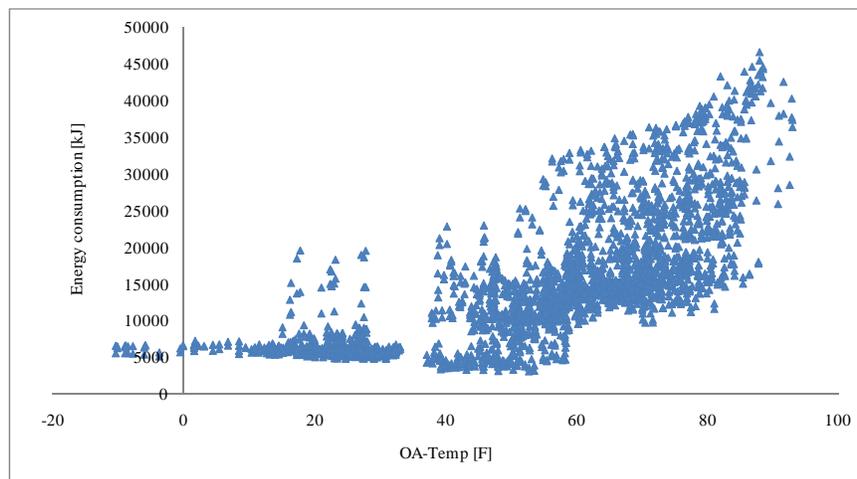
## 2.6. Case study

### 2.6.1. Clustering

To validate the clustering algorithm of Section 2.5.1, the energy consumption model from Section 2.3 was used. As shown in Table 2.3, two parameters, CHWC-VLV and OA-Temp, were ranked as the most important by the boosting tree algorithm. Figure 2.9 (a) and (b) map the relationships between the AHU energy consumption and the two parameters.



(a) CHWC-VLV and energy consumption



(b) OA-Temp and energy consumption

Figure 2.9. Scatter plots showing relationships between energy consumption and selected parameters.

The initial range of clusters was varied from 2 to 25, and the maximum iteration number was 100. Applying the method presented in Section 4.2, the optimal clusters for CHWC-VLV

and OA-Temp were 8 and 6, respectively. The clusters corresponding to the two inputs are shown in in Figure 2.10 and Figure 2.11.

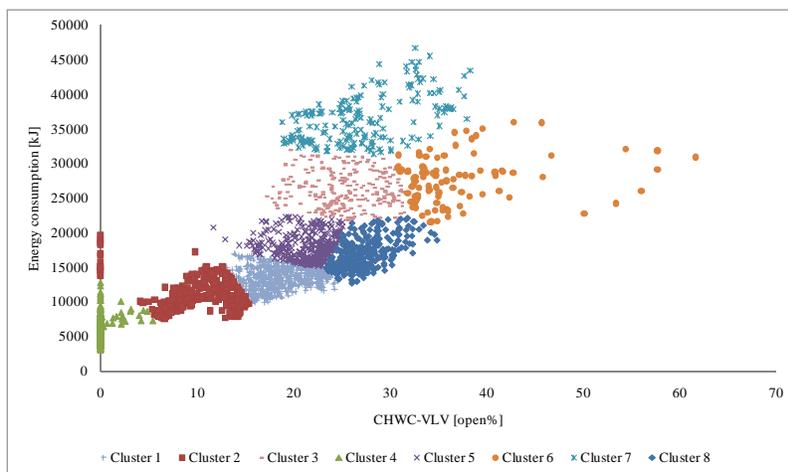


Figure 2.10. Graphical interpretation of clustering based on CHWC-VLV.

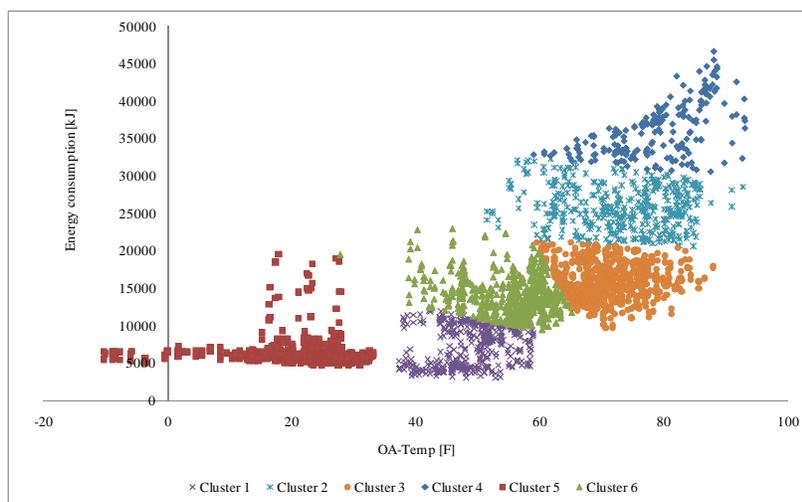


Figure 2.11. Graphical interpretation of clustering based on OA-Temp.

The characteristics of all of the clusters are provided in Table 2.7 and Table 2.8.

Table 2.7. Clusters based on CHWC-VLV.

Clusters	CHWC-VLV (mean)	AHU-Energy (mean)	Number of cases	Percentage (%)
1	18.63	13332.55	520	19.35
2	9.91	11056.69	273	10.16
3	25.28	26117.52	259	9.64
4	0.06	5958.75	762	28.35
5	21.17	18028.55	292	10.86
6	37.10	27603.71	97	3.61
7	27.13	36434.46	182	6.77
8	27.33	16941.22	303	11.27

Table 2.8. Clusters based on OA-Temp.

Clusters	OA-Temp (mean)	AHU-Energy (mean)	Number of cases	Percentage(%)
1	73.07	25764.30	381	14.17
2	70.98	16055.34	593	22.06
3	78.82	36177.72	196	7.29
4	21.36	6459.12	575	21.39
5	55.17	14110.56	580	21.58
6	48.58	7263.33	363	13.50

### 2.6.2. Modeling based on clusters

To evaluate the effectiveness of using clusters, four scenarios were defined. The MLP ensemble was used to build the models. These four scenarios are as follows:

Scenario 1: Model based directly on the joint data set.

Scenario 2: Model based on the three season data sets.

Scenario 3: Model based on clustering CHWC-VLV.

Scenario 4: Model based on clustering OA-Temp.

Scenario 1 involved the use of only one model based on a training data set (70%) and a test data set (30%). In Scenario 2, three models were built based on the three seasons. Scenario 3

and Scenario 4 involved clustering, with eight and six models derived for the two scenarios, respectively.

Figure 2.12 and Figure 2.13 illustrate the mean absolute error and mean absolute percentage error of the training and test results for each model. It can be seen that the models based on clustering outperformed the single model of Scenario 1 and the season-based model of Scenario 2. This confirmed that clustering improves prediction accuracy. Scenario 4, with the input space clustered on the OA-Temp, offered better prediction results than Scenario 3, where CHWC-VLV was clustered. The outside air temperature impacted the energy demand of the AHU system. Compared with Scenario 1, in Scenario 4, the mean absolute percentage error was reduced by 20.35% for training and 11.05% for testing.

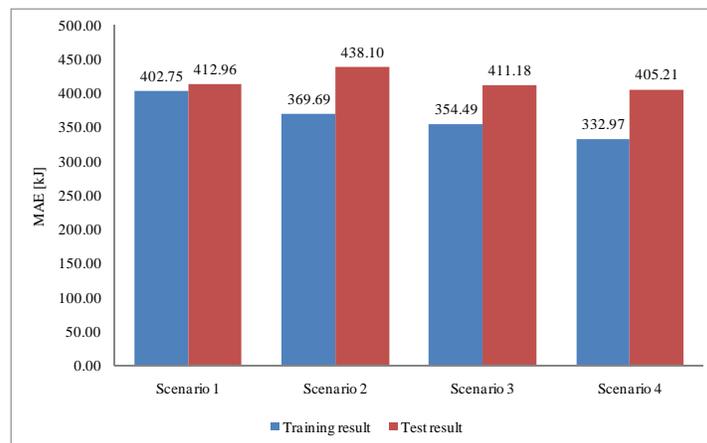


Figure 2.12. Mean absolute error for four scenarios.

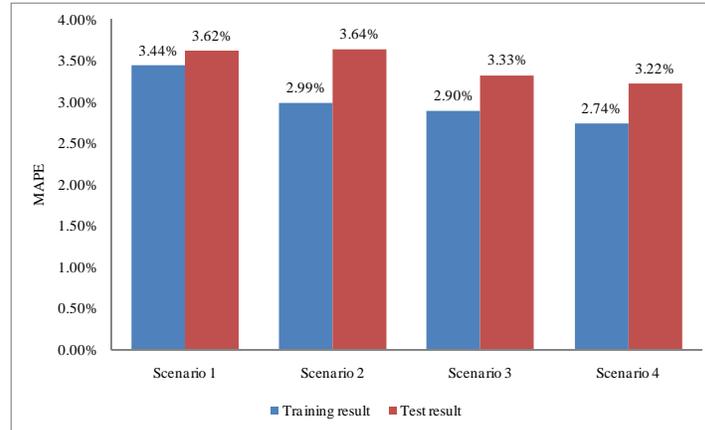


Figure 2.13. Mean absolute percentage error for four scenarios.

Table 2.9 summaries the prediction results for Scenario 1 through Scenario 4.

Table 2.9. Prediction results for four scenarios.

Model set	Data set	MAE	Std of AE	MAPE	Std of APE
Scenario 1	Training	402.75	339.94	3.44%	4.04%
	Testing	412.96	364.02	3.62%	4.80%
Scenario 2	Training	369.69	321.37	2.99%	2.96%
	Testing	438.10	391.58	3.64%	4.19%
Scenario 3	Training	354.49	310.39	2.90%	2.78%
	Testing	411.18	356.70	3.33%	3.21%
Scenario 4	Training	332.97	289.24	2.74%	2.80%
	Testing	405.21	353.96	3.22%	3.05%

### 2.7. Clustering-based short-term prediction of AHU energy

The prediction data set in Table 2.1 is used for the short term prediction of the AHU energy. Because Scenario 4 performed the best, it was selected to construct a prediction model. Figure 2.14 shows the architecture of clustering-based short-term prediction. The raw data set was first clustered using the clustering algorithm constructed in Scenario 4. Each point was

assigned to the corresponding cluster by minimizing the distance to the nearest cluster center. A model representing each cluster was used to validate the prediction results. Table 2.10 shows the prediction results for Scenario 1 and Scenario 4. Table 2.10 shows the prediction results of the two models.

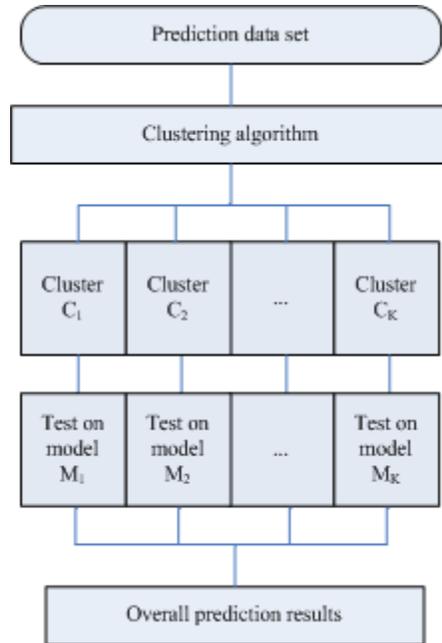


Figure 2.14. The architecture of clustering-based short-term prediction of AHU energy.

Table 2.10. Short-term prediction results for Scenario 1 and Scenario 4.

Model set	MAE	Std of AE	MAPE	Std of APE
Scenario 1	818.71	606.84	9.83%	8.97%
Scenario 4	792.58	731.12	8.63%	8.34%

The results in Table 2.10 show that the clustering-based model outperformed the single model (original data) in the prediction of the AHU energy. Compared to Scenario 1, Scenario 4

reduced the mean absolute percentage error by 12.21%. Figure 2.15 and Figure 2.16 show the prediction results for Scenario 1 and Scenario 4.

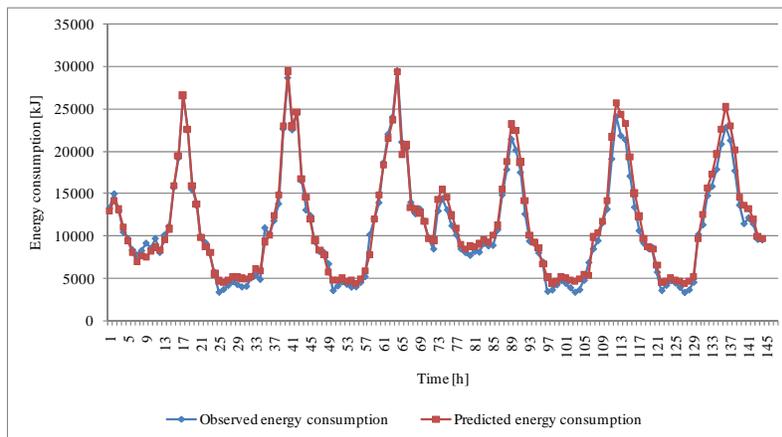


Figure 2.15. Short-term prediction of AHU energy for Scenario 1.

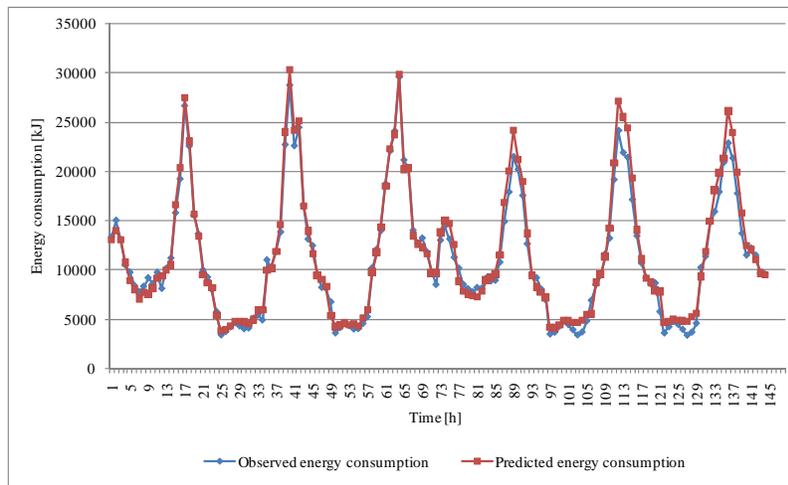


Figure 2.16. Short-term prediction of AHU energy for Scenario 4.

### 2.8. Summary

Data-mining algorithms were used to model the energy consumption of an AHU, as well as the indoor temperature, indoor humidity, and indoor CO<sub>2</sub> concentration. Of the many parameters available in this research, the most relevant parameters with respect to the target output were selected. The AHU energy model and the AQI models derived with the MLP Ensemble method outperformed the models derived using the other data mining algorithms considered in this research. To reduce the prediction errors and computation cost, a clustering scheme was applied. The MLP Ensemble algorithm applied to the clustered data provided the most accurate results. Data clustering reduced the mean absolute percentage error by 11.05% compared with the single model derived from the original data. The short-term prediction model derived from the clustered data offered improved prediction accuracy and reduced computation time.

## CHAPTER 3

### MULTI-OBJECTIVE OPTIMIZATION OF HVAC SYSTEM WITH AN EVOLUTIONARY COMPUTATION ALGORITHM

#### 3.1. Introduction

With regard to the strategic management of HVAC system, one of the major focuses is on effective and efficient energy management. The operation of HVAC system is a critical activity in terms of optimizing the control settings to reduce the energy consumption, improving the system efficiency, and preserving the thermal comfort for the occupants. The performance of the existing HVAC system can be largely improved by adjusting the control set points to maximize the overall system capacity and efficiency.

Zheng et al. [42] formulated the thermal process in a variable air volume (VAV) box with constraints on zone humidity. This provided daily operating strategies achieving optimal outdoor air-flow rates and energy savings. Fong et al. [43] discussed energy reduction by using an evolutionary programming approach to suggest optimal settings in response to the dynamic cooling loads and changing weather conditions. Nassif et al. [27, 28] applied evolutionary algorithms to one-objective and two-objective optimization of an HVAC system, and the supervisory control strategies resulted in energy savings. Kusiak and Li [44] applied an evolutionary strategy algorithm to solve a bi-objective optimization model to minimize the cooling output while maintaining the corresponding thermal properties.

The objective of this work is to investigate the effects of different control settings on energy consumption in an existing multi-zone office building while thermal comfort is sustained. The models built with data-mining algorithms were implemented inside an optimization model. Considering the specific requirements and effects of the thermal comfort, a weight-based constraints function was formed to satisfy the preferences. A strength Pareto evolutionary algorithm (SPEA) was employed to search for the optimal control settings of the HVAC system.

A preference function was formed for operators to find the local optimal solutions. According to the preference function, the operations of control strategies could largely affect the system performance by finding a trade-off between energy saving and the maintenance of indoor thermal comfort. Energy usage of the overall system regarding different preference settings were also compared and discussed.

### 3.2. Data description and optimization methodology

#### 3.2.1 Data description

The data set 3 in Table 2.1 is used. During this particular period of the year, cooling coils and heating coils in the AHU system work simultaneously to provide the essential loads on account of the dramatic variability of outside weather temperature. The AHU supply air temperature set point and static pressure set point were adjusted for both air handling units, AHU-A and AHU-B as described in Section 2.3. To simulate the impact of people and the lighting in the thermal zones, the internal load was divided into four stages reflecting the different thermal states at different time. Figure 3.1 shows the daily schedule of the internal load and supply air static pressure. The purpose of this experiment was to find the optimal set points minimizing the energy consumption while maintaining an acceptable level of thermal comfort.

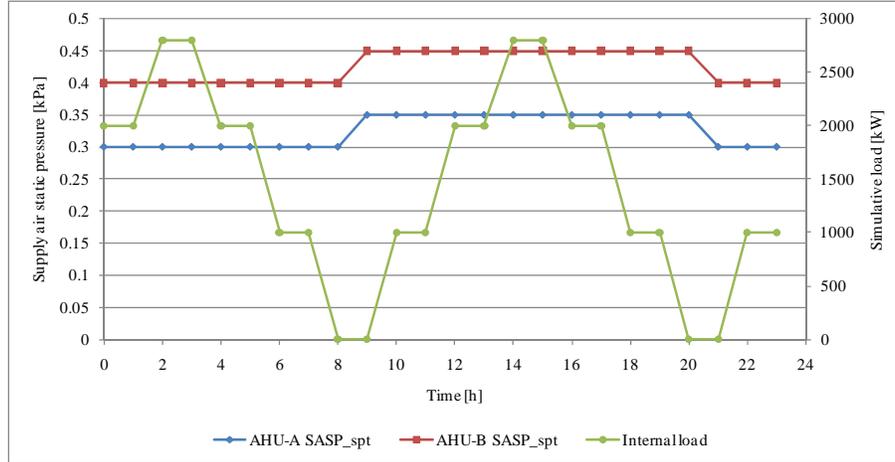


Figure 3.1. The daily schedule of supply air static pressure and internal load.

The total energy consumption in the HVAC system of the building was composed of four major parts: the energy consumption of AHUs  $E_{\text{Heat}}$ , the energy consumption of water pump  $E_{\text{Pump}}$ , the energy consumption of supply fan and return fan  $E_{\text{Fan}}$ , and the reheating process in VAV box  $Q_{\text{Reheat}}$ . Energy consumption for domestic hot water, appliances, and lighting were not counted because they were not expected to significantly change throughout the optimization. The complete energy consumption objective could thus be described as follows:

$$E_{\text{Total}} = E_{\text{Heat}} + E_{\text{Fan}} + E_{\text{Pump}} + Q_{\text{Reheat}} \quad (3.1)$$

The energy consumption of AHUs, water pump, and fans could be calibrated by the meters originally installed in the system. In the VAV box, the reheat load, which supplies conditioned air for a specific thermal zone to meet the comfort temperature of the zone envelope, accounts for the total energy consumption. By tuning the valve position and the dampers in the VAV box, the hot water flows through the coils adjusted to the actual requirements of the zone comfort. The reheat load is computed from equation (3.2) [45].

$$Q_{\text{Reheat}} = cm(T_{\text{VAV\_EAT}} - T_{\text{VAV\_DAT}}) \quad (3.2)$$

The metrics used to access thermal comfort in this experiment is evaluated by the indoor temperature and humidity. Based on the requirement of management, the room temperature

should be maintained between 70 Deg °F (21.11 Deg °C) and 72 Deg °F (22.22 Deg °C), and the room humidity should be ranged from 30% to 60%. The practical system sometimes runs out of control and creates discomfort. Since discomfort should never occur inside the zones, the objective of  $y_{\text{Room temp}}$  and  $y_{\text{Room humd}}$  were taken as constraints. The penalty function in the optimization algorithm successfully handled the constraints in this study.

### 3.2.2 Optimization approach

The optimization framework of this study is summarized in Figure 3.2. It is divided into three sequential steps. First the data reflecting all the characteristics of the building was collected from ERS in time series. The original data should be preprocessed and transformed into available format for analysis. Then parameter selection and algorithm selection were conducted to choose the suitable parameters for building the energy model. Different data mining algorithms were investigated and the one with highest accuracy was selected as the optimal algorithm. Finally, SPEA [46] implemented with the optimal algorithm was run to evaluate potential solutions and the best was selected by weighing the preference of management.

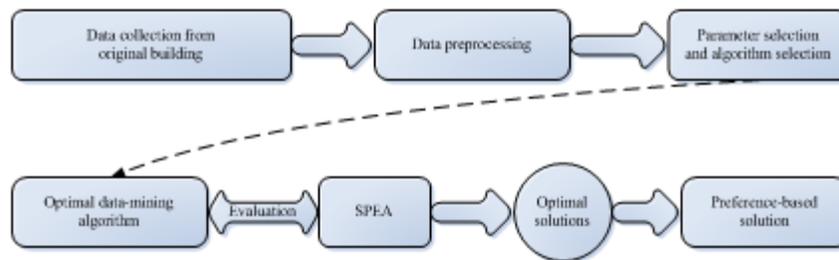


Figure 3.2. Optimization framework.

### 3.3. HVAC system modeling

The performance of HVAC system becomes much more complex and diverse in the transient season than any other single season (either cooling or heating season). Considering both AHU-A and AHU-B systems, there are totally 672 (fourteen-day data, 1 h interval data from April 1 to April 14, 2010) observations recorded in this period. After the preprocessing of the whole data, the dataset was randomly sampled and roughly divided into a training set (85% of data) and a testing set (15% of data). The data description is presented in Table 3.1.

Table 3.1. Data description.

Data Set	Description	No. of Instances
1	Model training; a random sample of 85% of the preprocessing data	572
2	Model test; the remaining 15% of the preprocessing data	100

#### 3.3.1 Parameter selection

Based on domain knowledge and the requirement of experiment, 15 parameters were selected as candidates for parameter selection including both internal loads and outside weather conditions. Besides the two control parameters and the internal load, 3 important parameters inside the AHU system namely, chilled water coil valve position (CHWC-VAV), supply air humidity (SA-Humd), and mixed air temperature (MA-Temp) and 9 parameters with respect to the outside weather patterns were chosen for further parameter selection. Due to the PI control implemented in the AHU system, the CHWC-VLV can roughly represent the control mode of the system by tuning the valve position. When the outside temperature is above the supply air temperature set point, the cooling mode will be turned on, otherwise, the heating mode will be operated. Figure 3.3 explains the control process of AHU system.

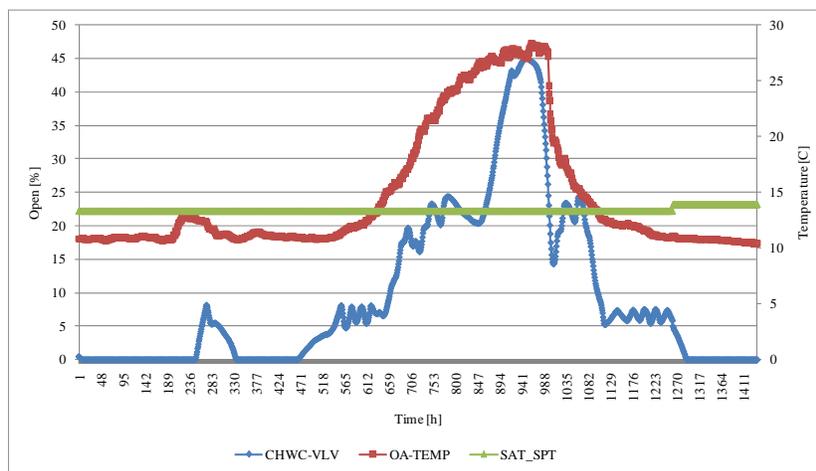


Figure 3.3. Change of CHWC-VLV position on 04/06/2010.

Wrapper algorithms [47, 48] that use induction learning as evaluation functions were applied to select the most important parameters. A wrapper algorithm searches the space of all the possible attributes and evaluates each subset by building a model on this subset. The subset with highest estimation accuracy is chosen as the optimal solutions after a number of iterative processes. Due to the high computational cost of iterations, wrapper algorithms are always more valuable to a small amount of data. Three different classifiers, namely pace regression, support vector machine (SVM) regression, and linear regression were used with the uniform genetic searching method.

Since a single wrapper algorithm might dominate the features in some aspects, different combinations of the classifiers and searching method in wrapper will provide more robust results than a single method. Each wrapper algorithm was performed with 10-fold cross-validation to eliminate the errors in the heuristic searching process. The total number of each candidate selected within 10-fold cross-validation represents the importance to the target output. Table 3.2 shows the selection results of different wrapper algorithms.

Table 3.2. Results of different wrapper algorithms.

Pace Regression+Genetic		SMO Regression+Genetic		Linear Regression+Genetic	
10(100 %)	CHWC-VLV	10(100 %)	CHWC-VLV	10(100 %)	CHWC-VLV
10(100 %)	SA-Humd	10(100 %)	SA-Humd	10(100 %)	SA-Humd
6( 60 %)	MA-Temp	4( 40 %)	MA-Temp	4( 40 %)	MA-Temp
4( 40 %)	BAR-Pres	3( 30 %)	BAR-Pres	2( 20 %)	BAR-Pres
3( 30 %)	IR-RADIA	6( 60 %)	IR-RADIA	1( 10 %)	IR-RADIA
10(100 %)	OA-Humd	10(100 %)	OA-Humd	10(100 %)	OA-Humd
10(100 %)	OA-Temp	10(100 %)	OA-Temp	10(100 %)	OA-Temp
0( 0 %)	OA-CO2	2( 20 %)	OA-CO2	1( 10 %)	OA-CO2
7( 70 %)	SOL-Beam	10(100 %)	SOL-Beam	8( 80 %)	SOL-Beam
10(100 %)	SOL-Horz	10(100 %)	SOL-Horz	10(100 %)	SOL-Horz
2( 20 %)	WIND-Dir	1( 10 %)	WIND-Dir	2( 20 %)	WIND-Dir
5( 50 %)	WIND-Vel	5( 50 %)	WIND-Vel	3( 30 %)	WIND-Vel

From the results, 5 parameters out of 12 namely CHWC-VLV, SA-Humd, OA-Humd, OA-Temp, and SOL-Horz were chosen 10 times during the 10-fold cross-validation for the different combinations of wrapper algorithms. Therefore 5 uncontrollable variables along with 3 controllable variables were finally chosen for building the energy consumption model. The previous time stamp of internal load was also involved because of the effect of memory and system delay. The parameter description is shown in Table 3.3.

Table 3.3. Parameter description.

Parameter Type	Parameter Name	Description	Unit
Controllable Variables	SAT set point	AHU supply air temperature set point	Deg C
	SASP set point	Supply air duct static pressure set point	kPa
Uncontrollable Variables	Internal load (t)	System internal load	/
	Internal load(t-1)	The previous state of system internal load	/
	CHWC-VLV	Chilled water coil valve position	%Open
	SA-Humd	Supply air humidity	% RH
	SOL-Horz	Solar normal flux	B/HFt2
	OA-Humd	Outside air humidity	% RH
	OA-Temp	Outside air temperature	Deg C

### 3.3.2. Construction and validation of the predictive model

After parameter selection and dimensionality reduction, the predictive model of the HVAC system is expressed in equation (3.3)—(3.5).

$$y_{Energy}(t) = f(x_{SAT\_Spt}, x_{SASP\_Spt}, x_{Load(t)}, x_{Load(t-1)}, x_{CHWC\_VLV}, x_{SA\_Humd}, x_{SOL\_Horz}, x_{OA\_Humd}, x_{OA\_TEMP}) \quad (3.3)$$

$$y_{Temp}(t) = f(x_{SAT\_Spt}, x_{SASP\_Spt}, x_{Load(t)}, x_{Load(t-1)}, x_{CHWC\_VLV}, x_{SA\_Humd}, x_{SOL\_Horz}, x_{OA\_Humd}, x_{OA\_TEMP}) \quad (3.4)$$

$$y_{Humd}(t) = f(x_{SAT\_Spt}, x_{SASP\_Spt}, x_{Load(t)}, x_{Load(t-1)}, x_{CHWC\_VLV}, x_{SA\_Humd}, x_{SOL\_Horz}, x_{OA\_Humd}, x_{OA\_TEMP}) \quad (3.5)$$

where  $y(t)$  is the output to be optimized;  $x$  represents all the inputs of this predictive model.

Five data-mining algorithms were used to extract the mapping between inputs and outputs: standard Chi-square Automatic Interaction Detector (CHAID) [49], Boosting Tree [36], Random Forest [37], Multivariate Adaptive Regression Splines (MARSplines) [50], and Neural Networks (NN) [39].

The standard CHAID is one of the oldest tree classification methods originally proposed by Kass, which will build non-binary trees allowing multiple splits of nodes and can be used for detection of interaction between variables in regression and classification analysis.

MARSplines is a non-parametric regression procedure that makes no assumption about the underlying functional relationship between the dependent and independent variables.

Motivated by the recursive partitioning approach, MARSplines builds the relation from a set of

coefficients and functions that are entirely derived from the regression data, making it particularly suitable for problems with higher input dimensions.

The five different data-mining algorithms have been tested for the construction of predictive models. The prediction performances of three predictive models regarding energy consumption, room humidity, and room temperature are presented from Table 3.4 to Table 3.6. Both training and validation errors are compared by the five different data mining algorithms. The NN outperform other four algorithms and therefore it is selected as the optimal solution to construct the predictive models.

Table 3.4. Prediction results of energy consumption model.

Algorithm	Data set	MAE	MAPE	Std_AE	Std_MAPE
Stand CHAID	Training	3344.51	0.26	2997.46	0.24
	Validation	3331.03	0.27	2391.07	0.17
Boosted Tree	Training	2537.49	0.20	2119.52	0.17
	Validation	2323.96	0.20	1809.40	0.17
Random Forest	Training	2145.67	0.17	1755.75	0.14
	Validation	2189.19	0.18	1752.73	0.14
MARSplines	Training	2745.52	0.25	2026.13	0.25
	Validation	2694.88	0.27	1921.01	0.29
Neural Networks	Training	743.71	0.07	690.33	0.07
	Validation	1094.35	0.10	870.47	0.10

Table. 3.5. Prediction results of room humidity model.

Algorithm	Data set	MAE	MAPE	Std_AE	Std_MAPE
Stand CHAID	Training	3.320	0.104	2.933	0.099
	Validation	3.361	0.105	2.784	0.092
Boosted Tree	Training	1.396	0.042	1.191	0.038
	Validation	1.488	0.046	1.383	0.046
Random Forest	Training	1.436	0.044	1.218	0.042
	Validation	1.521	0.048	1.383	0.047
MARSplines	Training	0.870	0.026	0.709	0.020
	Validation	1.093	0.037	1.212	0.071
Neural Networks	Training	0.352	0.011	0.303	0.011
	Validation	0.606	0.019	0.760	0.028

Table. 3.6. Prediction results of room temperature model.

Algorithm	Data set	MAE	MAPE	Std_AE	Std_MAPE
Stand CHAID	Training	0.4981	0.0070	0.4671	0.0065
	Validation	0.5664	0.0079	0.4901	0.0069
Boosted Tree	Training	0.4180	0.0058	0.3305	0.0046
	Validation	0.3857	0.0054	0.2956	0.0041
Random Forest	Training	0.3551	0.0049	0.3663	0.0050
	Validation	0.3791	0.0053	0.3760	0.0052
MARSplines	Training	0.5382	0.0075	0.3923	0.0054
	Validation	0.5003	0.0070	0.3791	0.0053
Neural Networks	Training	0.2288	0.0032	0.2176	0.0030
	Validation	0.3305	0.0046	0.3647	0.0050

### 3.4. Optimization Algorithm

#### 3.4.1. Model formulation

The total energy consumption was minimized by implementing the optimal control settings achieved by optimization algorithm. Considering the thermal comfort, the constraints of control parameters were addressed and a constraint function with penalty was formed to optimize the objectives. In order to save the computation time, the NN algorithm was implemented into

the SPEA prior to solving the optimization model. The optimization model is formed through the identification of the model parameters, the objective functions, and the constraints.

### **Model parameters**

The model parameters of the HVAC system have been determined by the wrapper algorithm described above. Table 3 lists each parameter used for the optimization process. The two controllable parameters---the AHU supply air temperature and the supply air duct static pressure set points, are to be varied to obtain the optimal solutions. As uncontrollable input parameters are essentially independent of the controllable ones, the values of uncontrollable variables, such as supply air humidity, outside air temperature and other outside weather patterns, can be fixed in seeking the optimal control settings at each time stamp.

### **Objective functions**

The objective of the optimization model is to minimize the energy use as well as maintain the thermal comfort at an acceptable level. The total energy consumption, including the fan, pump, and reheat power, is computed from equation (3.1). The input-output relationship was expressed by the HVAC system model presented above in equations (3.3), (3.4) and (3.5). The energy objective function is to be minimized while the other two - temperature and humidity objective functions - are treated as constraints to satisfy the indoor air quality of the AHU system.

### **Constraints**

The constraints in the model are identified by assigning the lower and upper bounds of control parameters and set up an acceptable range of room humidity and temperature objectives. The value of the supply air temperature set point, the supply air duct static pressure set point, indoor room temperature, and room humidity are restricted within the limits:

Supply air temperature must vary between 51 Deg °F (10.56 °C) to Deg 64 °F (17.78°C).

Supply air static pressure must vary between 1.2 WG (0.3 kPa) to 1.8 WG (0.45 kPa).

Room temperature must be maintained between 70 Deg °F (21.11 Deg °C) and 72 Deg °F (22.22 Deg °C).

Room humidity must be controlled between 30%--60% according to the comfort.

Consequently, the optimization model can be expressed as minimizing the objective function (3.3) with control parameters varying within their bounds, meanwhile maintaining the room temperature and humidity. The model is presented as follows:

$$\begin{aligned}
 & \min(y_{Energy}(t)) \\
 & \quad x_{SAT\_Spt}, x_{SASP\_Spt} \\
 & \text{subject to:} \\
 & y_{Energy}(t) = f(x_{SAT\_Spt}, x_{SASP\_Spt}, x_{Load(t)}, x_{Load(t-1)}, x_{CHWC\_VLV}, x_{SA\_Humd}, x_{SOL\_Horz}, x_{OA\_Humd}, x_{OA\_TEMP}) \\
 & y_{Temp}(t) = f(x_{SAT\_Spt}, x_{SASP\_Spt}, x_{Load(t)}, x_{Load(t-1)}, x_{CHWC\_VLV}, x_{SA\_Humd}, x_{SOL\_Horz}, x_{OA\_Humd}, x_{OA\_TEMP}) \\
 & y_{Humd}(t) = f(x_{SAT\_Spt}, x_{SASP\_Spt}, x_{Load(t)}, x_{Load(t-1)}, x_{CHWC\_VLV}, x_{SA\_Humd}, x_{SOL\_Horz}, x_{OA\_Humd}, x_{OA\_TEMP}) \\
 & 10.56 \leq x_{SAT\_Spt} \leq 17.78 \\
 & 0.3 \leq x_{SASP\_Spt} \leq 0.45 \\
 & 20.56 \leq y_{Temp}(t) \leq 22.78 \\
 & 30\% \leq y_{Humd}(t) \leq 60\%
 \end{aligned} \tag{3.6}$$

For such a nonlinear multi-objective optimization problem, it is challenging for large data sets to get the results, since the procedure of computation is complicated and time consuming. Due to the variability of room humidity and temperature, a constraint function can be established by assigning the two objective functions into one objective function, which is also to be minimized. The transformed constraint function can be described in equation (3.7).

$$y_{Constraint_s}(t) = \max\{0, 20.56 - y_{Temp}(t)\} + \max\{0, y_{Temp}(t) - 22.78\} + \max\{0, 30 - y_{Humidity}(t)\} + \max\{0, y_{Humidity}(t) - 60\} \tag{3.7}$$

As the constraints are satisfied, each of the four terms in equation (3.7) will remain 0 and the sum equals 0. However, since room humidity and temperature are not at the same scale, the constraint function may not accurately reflect the influences of them. It is possible room humidity may dominate the result because of its high value. The values of room humidity and temperature are then normalized to eliminate the deviation. The revised function is shown in equation (3.8).

$$y_{Constraint_s}(t) = \frac{\max\{0, 20.56 - y_{Temp}(t)\} + \max\{0, y_{Temp}(t) - 22.78\}}{(y_{Temp\_upperbound} - y_{Temp\_lowerbound})} + \frac{\max\{0, 30 - y_{Humidity}(t)\} + \max\{0, y_{Humidity}(t) - 60\}}{(y_{Humidity\_upperbound} - y_{Humidity\_lowerbound})} \tag{3.8}$$

In the practical operation of HVAC system, it is possible that some constraint is of more importance or at least over than some other constraint. In different time period of the year, managers may emphasize on different objectives. Room temperature should be paid more attention to room humidity in winter since the air humidity is really low and is not a big issue. Taking the effects of different constraints, weights should be assigned to each constraint to meet the specific preference of HVAC system. The final constraint function is presented below:

$$y_{Constraint}(t) = w_1 * \{ \max\{0, 20.56 - y_{Temp}(t)\} + \max\{0, y_{Temp}(t) - 22.78\} \} / (y_{Temp\_upperbound} - y_{Temp\_lowerbound}) + w_2 * \{ \max\{0, 30 - y_{Humidity}(t)\} + \max\{0, y_{Humidity}(t) - 60\} \} / (y_{Humidity\_upperbound} - y_{Humidity\_lowerbound}) \quad (3.9)$$

$$w_1 + w_2 = 1 \quad (3.10)$$

Based on discussion above, the optimization problem was transformed into a bi-objective model. Finally the optimization can be modified to equation (3.11):

$$\begin{aligned} & \min(y_{Energy}(t)) \\ & \quad x_{SAT\_Spt}, x_{SASP\_Spt} \\ & \quad \text{subject to:} \\ y_{Energy}(t) = & f(x_{SAT\_Spt}, x_{SASP\_Spt}, x_{Load}(t), x_{Load}(t-1), x_{CHWC\_VLV}, x_{SA\_Humd}, x_{SOL\_Horz}, x_{OA\_Humd}, x_{OA\_TEMP}) \\ y_{Constraint}(t) = & w_1 * \{ \max\{0, 20.56 - y_{Temp}(t)\} + \max\{0, y_{Temp}(t) - 22.78\} \} / (y_{Temp\_upperbound} - y_{Temp\_lowerbound}) + \\ & w_2 * \{ \max\{0, 30 - y_{Humidity}(t)\} + \max\{0, y_{Humidity}(t) - 60\} \} / (y_{Humidity\_upperbound} - y_{Humidity\_lowerbound}) \\ & 10.56 \leq x_{SAT\_Spt} \leq 17.78 \\ & 0.3 \leq x_{SASP\_Spt} \leq 0.45 \end{aligned} \quad (3.11)$$

### 3.4.2. Optimization

In this research, a genetic algorithm search method named Strength Pareto Evolutionary Algorithm (SPEA) was used to solve the model (3.11). Considering that energy consumption and thermal constraints are two objectives aimed to be both minimized, non-dominated results were found at each time stamp. Figure 3.4 shows the feasible elitist optimization results of the model at some time stamp.

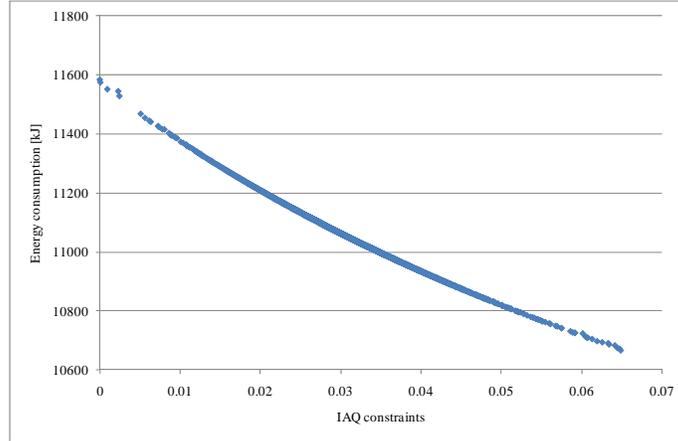


Figure 3.4. Feasible solutions solved by SPEA at some time stamp.

Since a penalty function was implemented in the optimization algorithm in terms of finding out the optimal solutions, the bounds set for supply air temperature and static pressure set points may be slightly violated. Because of the contradiction between energy saving and maintenance of air quality, a decrease in constraints indicating better air quality results in an increase in energy demand. When the operation of HVAC is aim to maximize the energy saving (the bottom right point), the thermal discomfort will arise. If the decision is to minimize the thermal discomfort (the up left point), more energy are consumed to meet the demand. According to the different preferences of management, a preference-based objective function can be built with different weights assigning to individual objective. The following two equations show the process.

$$\text{Objective} = w_1 * \frac{(\text{objective1} - \text{objective1}_{\min})}{(\text{objective1}_{\max} - \text{objective1}_{\min})} + w_2 * \frac{(\text{objective2} - \text{objective2}_{\min})}{(\text{objective2}_{\max} - \text{objective2}_{\min})} \quad (3.12)$$

$$w_1 + w_2 = 1 \quad (3.13)$$

where  $w_1$  and  $w_2$  represent the weight given to objective1 and objective2, respectively; objective1 represents the energy use and objective2 represents the violation of thermal comfort.

After assigning the weights, the optimal result is found by minimizing the final objective. Table 7 lists the results of different decisions made by their preferences. The preference-based optimization algorithm provides more flexibility for the management of HVAC system and accurately meets the requirements of different operation strategy. The adjustment of HVAC system optimization makes it more robust in dealing with the outside weather change than the traditional control strategy.

Table 3.7. Results of different decisions made by their preferences.

Strategy	Weight	SAT_spt	SASP_spt	Energy consumption	Constraints	Room temperature	Room humidity
A	$w_1=0$ $w_2=1$	52.72	1.48	11583.1	0.000	71.998	43.603
B	$w_1=0.5$ $w_2=0.5$	52.08	1.46	11102.7	0.027	72.090	42.713
C	$w_1=1$ $w_2=0$	51.52	1.34	10667.5	0.065	69.784	58.054

Strategy A and C are the two extrema of preference-based optimization which entirely focus on minimizing the energy consumption and thermal discomfort, respectively. A trade-off can be found between the two situations balancing the energy use and thermal comfort. A brief description of weights assignment of the two objectives is shown in equation (3.14).

$$w = \begin{cases} w_1=0 \ w_2=1, & \text{Minimize constraints to strictly maintain thermal comfort} \\ w_1=a \ w_2=1-a, & \text{Find tradeoff to balance energy use and thermal comfort} \\ w_1=1 \ w_2=0, & \text{Minimize energy use without considering the thermal comfort} \end{cases} \quad (3.14)$$

To compare the optimal results with different preference settings to the actual energy consumption, two extrema (Strategy A and C) were picked which represented minimizing thermal discomfort and energy use, respectively. Figure 3.5 shows the actual and optimal energy demand after optimization. The energy saving by applying the optimization algorithm is roughly ranged from 12% (Strategy A) to 30% (Strategy C). In the process of finding the optimal solutions, the thermal comfort was more or less violated to meet the system demand. Figure 3.6

shows the values of thermal constraints of the two extrema. Strategy A emphasizing on minimization of thermal discomfort, had a very low violation of indoor air quality. The constraints mostly were obtained during the optimization process. Strategy C saved more energy based on the sacrifice of thermal comfort.

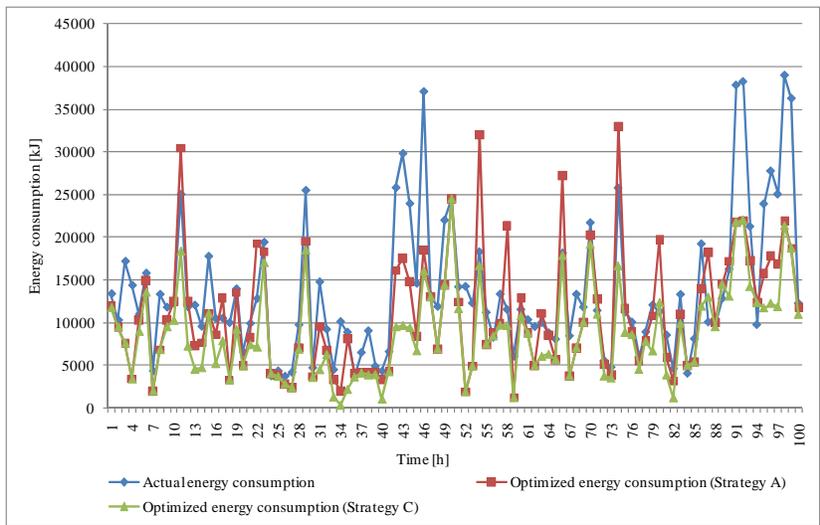


Figure 3.5. Test results of optimized energy consumption.

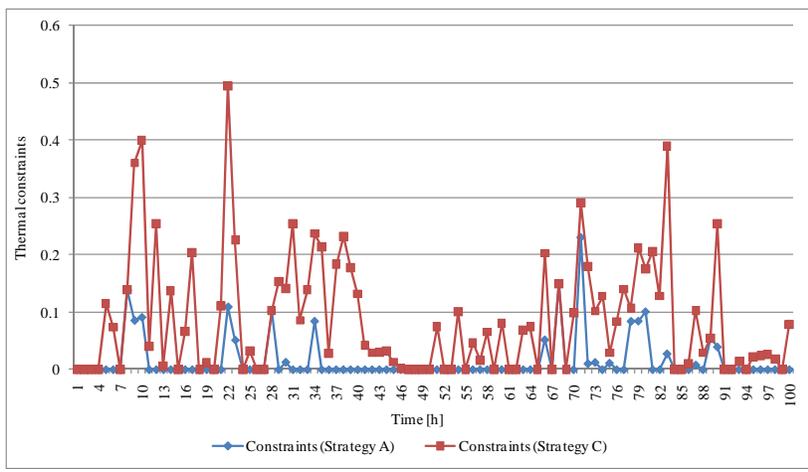


Figure 3.6. Test results of thermal constraints.

The optimized room temperature and humidity are presented in Figure 3.7 and Figure 3.8. According to the Figures, the room temperature is optimized by forcing the uncontrollable points to the restricted range, although strategy C has more points out of bound than strategy A. The indoor air quality is improved after optimization algorithm.

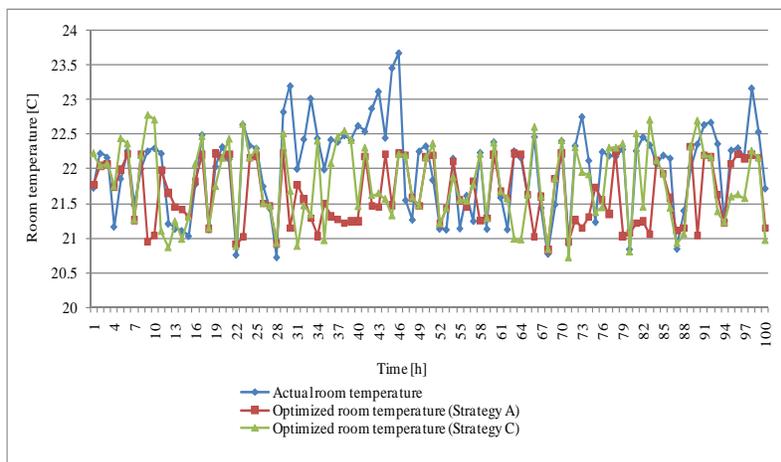


Figure 3.7. Test results of optimized room temperature.

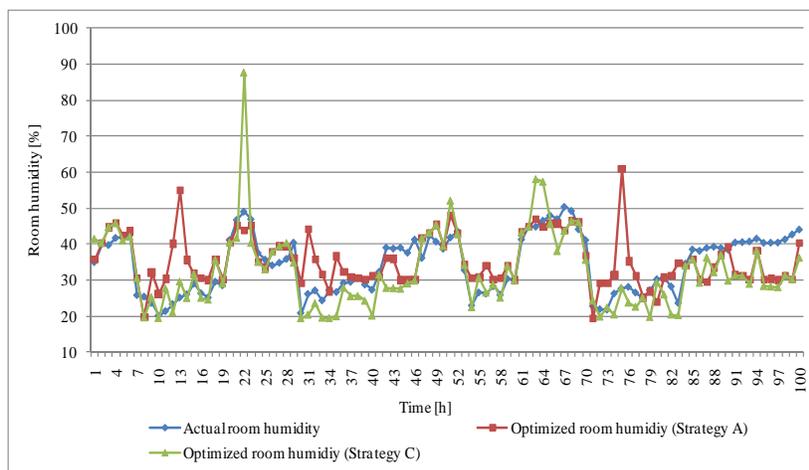


Figure 3.8. Test results of optimized room humidity.

Figure 3.9 and Figure 3.10 show the recommended control settings of supply air temperature and static pressure set points. Based on the observation of test results, the system is out of control when the supply air temperature set point is above 60 Deg °F (15.56 °C). The control strategy C is more fluctuant compared with strategy A. Taking the system delay into consideration, strategy A is more robust for realization in practice.

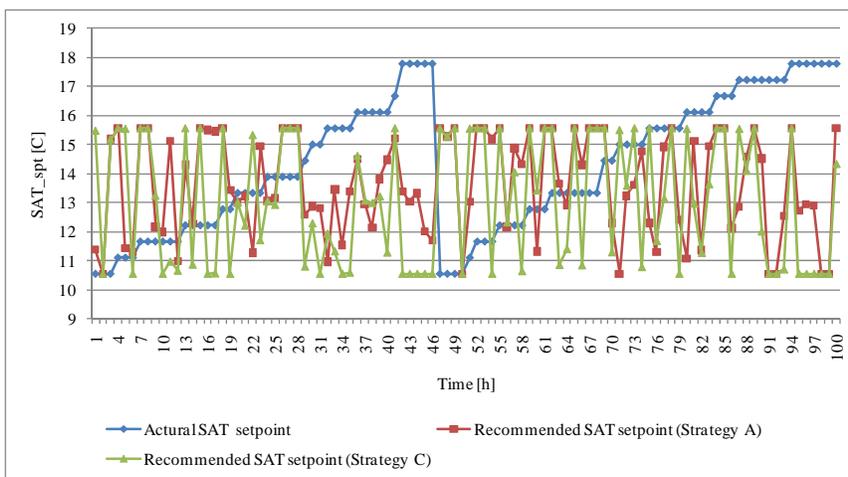


Figure 3.9. Recommended supply air temperature set point.

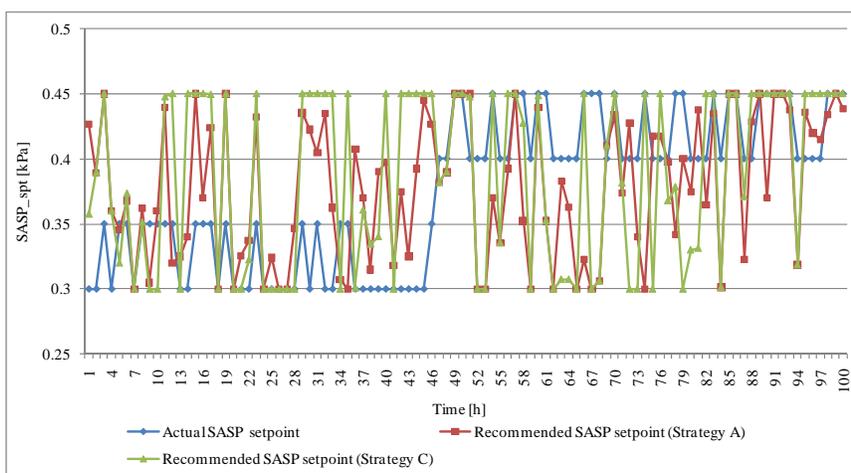


Figure 3.10. Recommended supply air static pressure set point.

What is interesting to note from Figure 3.5 to Figure 3.10 is the relation between the thermal comfort and the energy consumption. The increase of energy consumption can lead to relative reductions in thermal comfort, otherwise, the inverse situation occurs. This case highlights the major advantage of a true multi-objective optimization algorithm, which help the decision makers find out a balance between the two contradictive objectives. A comprehensive understanding of the operation system is achieved as well as the potentiality of each preference.

### 3.5. Summary

This chapter presented an optimization methodology based on a data-driven method, which combined Neural Network and a multi-objective evolutionary algorithm. First, the experimental data was studied and some important parameters were chosen as the inputs. Several data-mining algorithms were then discussed and finally NN outperform other algorithms and was selected for building the predictive models. The NN proved to be able to accurately construct the mapping of the actual data between inputs and outputs. The optimal control settings were achieved by SPEA with the implementation of NN inside, resulting in large energy saving of observed HVAC system.

Regarding the optimization results, the recommended control settings could lead to significant improvements of the energy use and the performance of thermal comfort of the building. The optimal solutions also revealed a large number of potential control strategies which could be used as a tool for decision makers to find a suitable balance between energy and thermal comfort. It would be an interesting research for future to deeply study the accuracy of data-driven models. The implementation of human intelligence in the preference function to simulate the thoughts of decision makers will be also further investigated. Further and more systematic studies are required to evaluate the performance of optimization, especially with multi-constraints problems.

## CHAPTER 4

### MODELING AND OPTIMIZATION OF ENERGY RELATED COMPONENTS IN HVAC SYSTEM

#### 4.1. Introduction

The body of literature on improving the operation and efficiency of HVAC systems is extensive. Numerous models developed based on the first principles, simulation, and optimization, have been published. Ma *et al.* [51] developed pressure-drop models for different water networks in a central air-conditioning system and proposed an optimal sequence control of a pump to reduce the energy consumption. A simple yet accurate model for a cooling coil unit for real-time control and optimization was presented by Wang *et al.* [5]. Fong *et al.* [43] used the simulation program TRNSYS [52] to determine optimum settings of chilled water and supply air temperature to save energy. Lu *et al.* [53] formulated a mix-integer nonlinear constrained model to optimize setpoints for differential pressure of ducts. A methodology for optimizing setpoints of the VAV (variable air volume) box was presented in Huang *et al.* [2].

In this chapter energy consumption is optimized with data-driven models derived with data-mining algorithms. The energy models built by the data-mining algorithms are integrated into an energy optimization model. To minimize the energy consumption, a Particle Swarm Optimization (PSO) algorithm is employed to find the near optimal solutions for control settings in response to different internal loads and other uncontrollable variables. Energy usage of the overall system and each individual component are also compared and discussed in detail.

#### 4.2. Data description

The data set 1 in table 2.1 is used in this research. In the ERS facility, sensors measure air temperature, humidity, and air flow rate at different locations of the HVAC system. Energy consumption of devices like pumps and fans is also recorded. Since at each day a selected combination of setpoints was implemented, arbitrarily dividing all data into training and test sets

based solely on time could distort the models extracted with data-mining algorithms. Therefore, sampled data is used in parameter and algorithm selection as well as model building. The data set description is presented in Table 4.1.

Table 4.1. Data description.

Data Set	Description	No. of Instances
1	Parameter selection; algorithm selection; random sampling from preprocessed data	129 observations
2	Model training; a random sample of 85% of the preprocessed data	658 observations
3	Model test; the remaining 15% of the data (excluding the training data)	116 observations

#### 4.3. Parameter and algorithm selection

The total energy consumption from four major sources, namely the chillers, fans, pumps, and VAV box, is computed every hour. Besides the controllable variables mentioned above, some uncontrollable parameters may impact the hourly energy consumption. Based on the domain knowledge, such candidate variables are as follows: cooling coil entering air temperature, chilled water entering water temperature, infrared radiation, solar normal flux, and the outside air temperature. To derive the patterns from high resolution data (1-min data), two commonly used statistical measures of mean and standard deviation are employed. Thus, ten additional parameters are selected for the additional possible inputs of the energy model. Among the ten variables, some may be redundant or even irrelevant to the energy consumption. Therefore, parameter selection is performed to eliminate uncontrollable parameters of lesser importance. In this section, the boosting tree algorithm [36] and wrapper [47, 48] are used in performing the parameter selection on data set 1 shown in Table 4.1. Table 4.2 lists the predictor importance produced by the boosting tree algorithm.

Table 4.2. Predictor importance produced by the boosting tree algorithm.

Parameter	Importance
Chilled water entering temperature (stdev)	100
Chilled water entering temperature (mean)	77
Solar normal flux (mean)	68
Solar normal flux (stdev)	57
Outside air temperature (mean)	56
Cooling coil entering air temperature (mean)	51
Infrared radiation (stdev)	43
Cooling coil entering air temperature (stdev)	42
Outside air temperature (stdev)	40
Infrared radiation (mean)	37

Considering the expensive computational cost, pace regression is used as the evaluator, and a genetic algorithm is used as the search algorithm in this section. The population size is set at 20, and the maximum number of iterations is 20. The crossover probability is 0.6, and the mutation probability is 0.033. The 10-fold cross validation results of the wrapper approach are presented in Table 4.3.

Table 4.3. The results of parameter selection by the wrapper approach.

Number of Folders	Parameter
10	Chilled water entering temperature (mean)
10	Chilled water entering temperature (stdev)
10	Outside air temperature (mean)
10	Solar normal flux (mean)
10	Solar normal flux (stdev)
8	Infrared radiation (mean)
4	Cooling coil entering air temperature (mean)
4	Cooling coil entering air temperature (stdev)
2	Infrared radiation (stdev)
1	Outside air temperature(stdev)

Five parameters are selected ten times during the 10-fold cross validation. They match the five most significant predictors selected by the boosting tree algorithm. Therefore, considering the parameter selection results of two methods, five uncontrollable variables along with three controllable variables are selected in building the total energy model. Note that the internal load at the previous time interval is also considered. Table 4.4 shows the parameter descriptions of inputs selected for building the energy model.

Table 4.4. Parameter description.

Parameter	Description	Unit
$x_1$	Internal load	Discrete
$x_2$	Internal load at previous time interval	Discrete
$x_3$	Supply air temperature setpoint	F
$x_4$	Supply air static pressure setpoint	WG
$v_1$	Chilled water entering temperature (mean)	F
$v_2$	Chilled water entering temperature (stdev)	F
$v_3$	Outside air temperature (mean)	F
$v_4$	Solar normal flux (mean)	B/HFt2
$v_5$	Solar normal flux (stdev)	B/HFt3

After parameter transformation and selection, the total energy model can be expressed as:

$$y = f(x_1, x_2, x_3, x_4, v_1, v_2, v_3, v_4, v_5) \quad (4.1)$$

To construct a mapping among these variables, six data-mining algorithms, namely CHAID, boosting tree, random forest, MARSplines, MLP, MLP Ensemble and SVM, have been applied. As shown in Table 4.1, date set 1 is used for selecting the most appropriate algorithm in building the model. Table 4.5 shows the prediction accuracy of models built by different data-mining techniques.

Table 4.5. Algorithms selection for building the total energy model.

Algorithm	MAE	Std of MAE	MAPE	Std of MAPE
CHAID	2292.1561	2438.1754	11.23%	11.30%
Boosting tree	1943.2698	1658.7942	9.75%	7.60%
Random forest	1864.6265	1983.4052	9.46%	8.36%
MARSplines	1702.8106	1464.0291	8.67%	6.54%
MLP	915.0163	977.6062	4.66%	4.20%
MLP Ensemble	719.1084	689.6606	3.77%	3.50%
SVM	1531.7237	1297.2686	7.62%	5.61%

The MLP ensemble over-performs other data-mining algorithms and produces the most accurate total energy model. Therefore, the MLP ensemble is selected for building energy models.

#### 4.4. Modeling building and validating

In Section 4.3, nine variables have been selected for building the total energy model by the MLP ensemble. Since the total energy consumption involves four sources, namely chillers, fans, pumps, and the VAV reheat device, building four separate models will benefit analysis of optimization results. The total energy model can be simplified as the sum of four sub-models. Data sets 2 and 3 shown in Table 4.1 are used for training and testing the model. Table 4.6 shows the training and test results of the four models.

Table 4.6. Training and test results of the four models.

Chiller Energy Model				
	MAE	Std of AE	MAPE	Std of APE
Training	541.0405	513.6529	5.03%	5.23%
Testing	645.0991	539.3706	5.99%	5.79%
Fan Energy Model				
	MAE	Std of AE	MAPE	Std of APE
Training	286.9700	305.8468	5.49%	6.90%
Testing	383.2912	488.5904	7.21%	12.62%
Pump Energy Model				
	MAE	Std of AE	MAPE	Std of APE
Training	6.9087	20.6677	0.23%	0.80%
Testing	6.2690	5.1443	0.21%	0.17%
Reheat Energy Model				
	MAE	Std of AE		
Training	53.3107	113.6242		
Testing	49.1659	112.1540		

For reheat energy, some of the observed values are zero, which is due to the fact that the reheating valve is open only when the room temperature is below the heating setpoint of 70°F. The absolute percentage error is meaningless for those instances. Therefore, only one statistical measure, the absolute error, is provided for the reheat energy model. Detailed test results of the four models are presented in Figures 4.1 to 4.4.

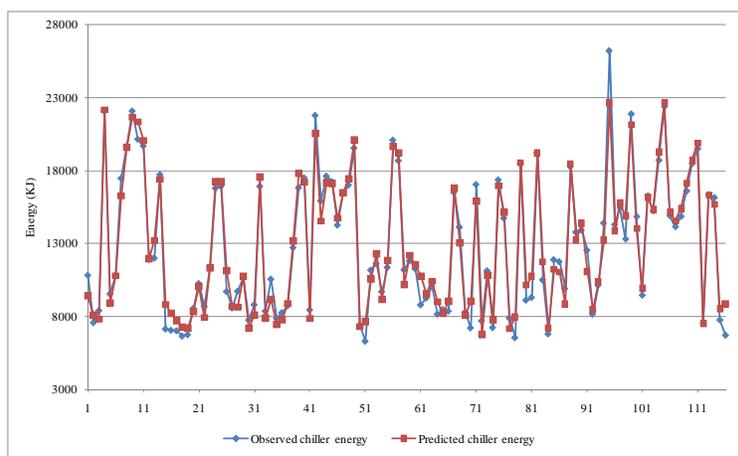


Figure 4.1. Test results for the chiller energy model.

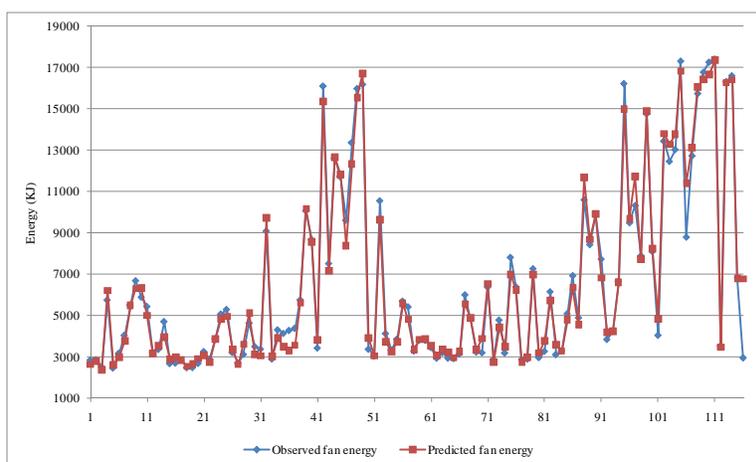


Figure 4.2. Test results for the fan energy model.

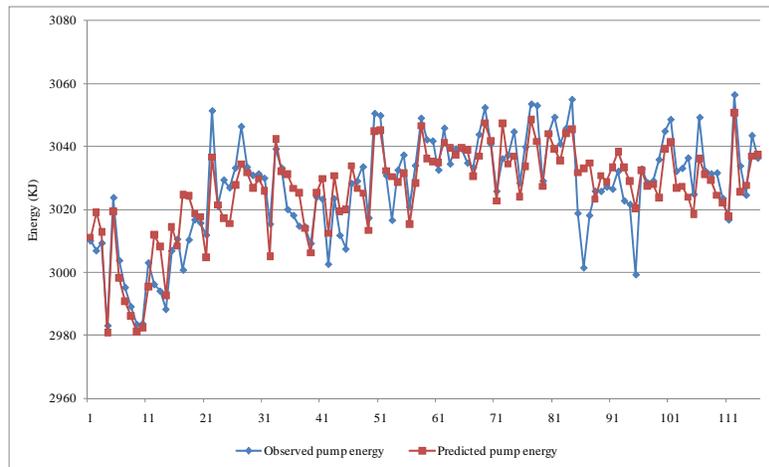


Figure 4.3. Test results for the pump energy model.

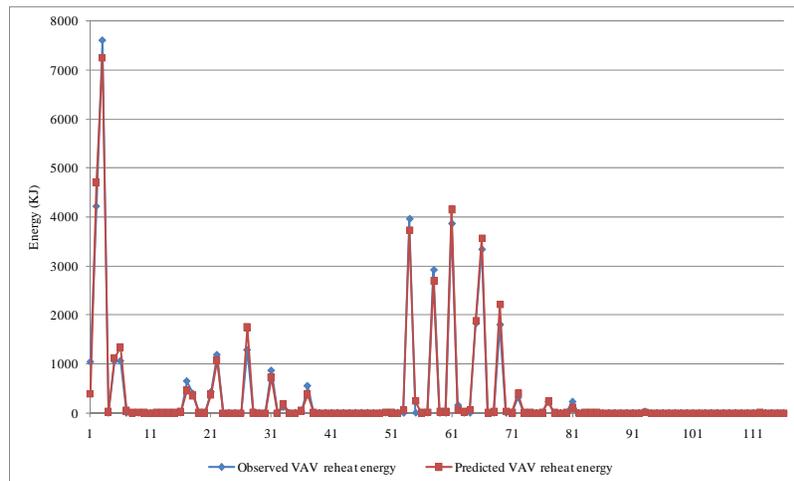


Figure 4.4. Test results for the reheat energy model.

As demonstrated in Figures 4.1 to 4.3, estimates obtained from the four sub-models are very close to the actual values of the specific energy usage. The MLP ensemble algorithm can

successfully identify the nonlinear relationship among the energy consumption, controllable variables, and uncontrollable parameters.

#### 4.5. Optimization model formulation and solving

The sub-models learned by the MLP ensemble are used to construct the overall optimization model. To minimize the total energy consumption, the single objective is expressed as the sum of the chiller energy, fan energy, pump energy, and reheat energy. The decision variables are the supply air temperature setpoint and the static pressure setpoint. The supply air temperature setpoint is constrained from 50°F to 65°F, and the static pressure setpoint is constrained from 1.2 WG to 1.8 WG. The single objective optimization problem is presented in (4.2):

$$\begin{aligned}
 & \min_{x_1, x_2} Obj \\
 s.t. \quad & Obj = \sum_{i=1}^4 f_i(x_1, x_2, x_3, x_4, v_1, v_2, v_3, v_4, v_5) \\
 & 50 \leq x_1 \leq 65 \\
 & 1.2 \leq x_2 \leq 1.8
 \end{aligned} \tag{4.2}$$

where  $f_i(\cdot)$  refers to the energy model, room temperature model, room humidity model and CO<sub>2</sub> model, respectively.

The nonlinearity, complexity and opaqueness of the energy models built in this section pose a challenge for solving by traditional mathematical programming methods. Particle Swarm Optimization (PSO) [54] is a stochastic optimization technique inspired by flocks of birds. As one of the swarm intelligence algorithms, it has a well-balanced mechanism to enhance and adapt global and local exploration abilities. PSO excels in solving single objective optimization [55, 56] with its fast convergence advantage. Therefore, the PSO algorithm is used in this paper to solve model (4.2).

The standard canonical PSO algorithm is presented next:

- Step 1: Initialize  $n$  particles positions  $p_i \in R^m$  and velocities  $q_i \in R^m$
- Step 2: Find current best position  $\hat{p}_i$  of each particle and let  $\hat{g}$  be the global best  

$$\hat{p}_i \leftarrow p_i, \hat{g} \leftarrow \arg \min(f(p_i))$$
- Step 3: For each particle, update the particle velocities and positions  

$$q_i \leftarrow \omega q_i + c_1 \text{rand}() (\hat{p}_i - p_i) + c_2 \text{rand}() (\hat{g} - p_i)$$

$$p_i \leftarrow p_i + q_i$$
- Step 4: Update the local bests and the global best
- Step 5: If the stop criterion is satisfied, global best  $\hat{g}$  is the final optimal solution with fitness  $f(\hat{g})$ ; otherwise, return to Step 3.

Here, the dimension for each particle's position  $p_i$  and velocity  $q_i$  is 2. The first element of  $p_i$  refers to the decision variable of the temperature setpoint and is drawn from uniform distribution  $U(50,65)$ . The second element of  $p_i$  refers to the decision variable of the static pressure setpoint and is drawn from uniform distribution  $U(1.2,1.8)$ . Initial values for elements of the velocity vectors are set as 0. For the velocity updating part,  $\omega$  is the inertia factor, which is used to balance the global and local search. Two randomly generated coefficients are drawn from uniform distribution  $U(0,1)$ .  $c_1, c_2$  are learning factors which control the influence of the social and cognitive components. Here, set  $\omega = 0.95$  and  $c_1, c_2 = 2$ . When the updated particle's position is out of bound, its corresponding speed changes towards the opposite direction, and its position is recomputed. For example, assume that the position vector and the velocity vector of one particle is  $[65, 1.25]^T$  and  $[1.5, -0.1]^T$ , respectively, before updating. The updated position vector is  $[66.5, 1.15]^T$ , and both elements are out of their decision boundaries. The velocity vector is set to the opposite direction, and the updated velocity and position vectors are  $[63.5, 1.35]^T$  and  $[-1.5, 0.1]^T$ , respectively.

#### 4.6. Optimization results and discussion

To demonstrate the energy saving potential, 50 consecutive points (50 h long period) from the training data set have been selected to illustrate the optimization process. For each instance, an optimization model (4.2) is formulated and solved by the PSO algorithm. After tuning the related parameters in PSO, the initial population size is set of 40, while the maximum number of

iterations is set of 20 to find the near optimal solution. Figure 4.5 shows the results for the total energy optimization.

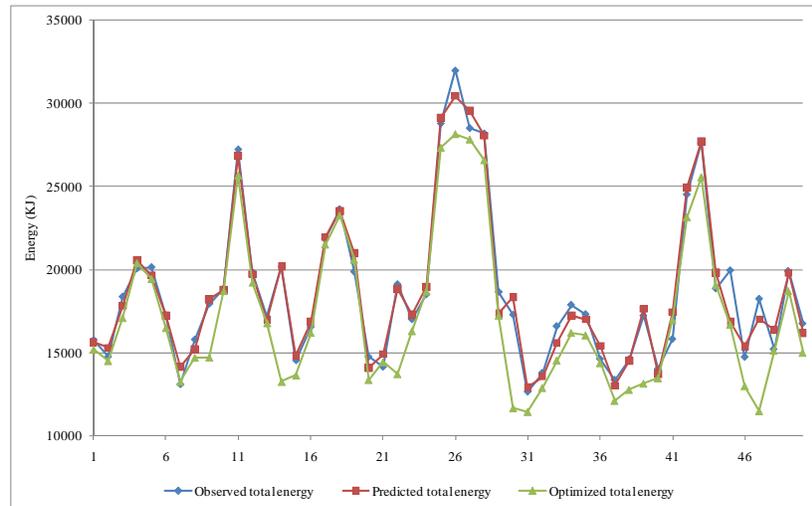


Figure 4.5. The total energy before and after optimization.

The confidence in the optimization results depends to a large degree on the accuracy of the models built. In Figure 4.5, the predicted total energy suitably matches the observed total energy. The optimization approach presented in this paper reduces the total energy consumption by 7.66%. Since the percentage error of the total energy consumption of the simulated points is 0.04%, the model prediction error does not diminish the quality of the optimization results. The peak energy usage corresponds to the high stage of the simulated internal load. It is understandable from the heat balance perspective that the more internal heat generated the more energy required to remove this extra heat. The control settings of the air temperature and the supply air static pressure are shown in Figures 4.6 and 4.7, respectively.

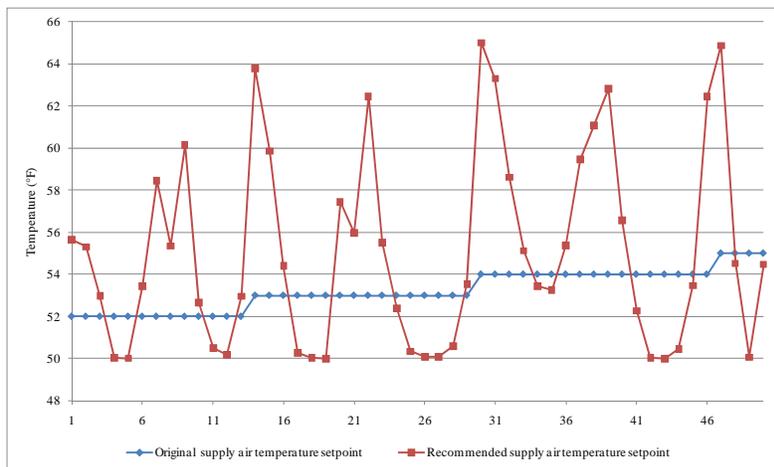


Figure 4.6. The air temperature setpoint before and after optimization.

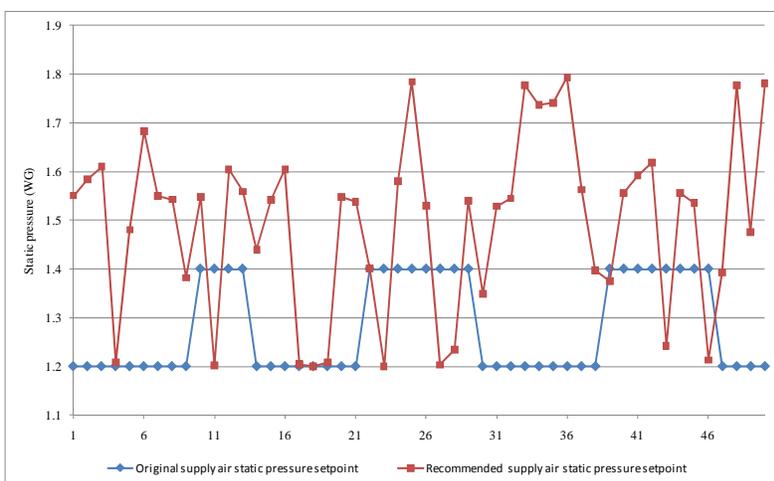


Figure 4.7. The supply air static pressure setpoint before and after optimization.

Compared to the original temperature setpoint, the optimized setpoint has a higher value when the internal load is low, and lower when the internal load is high. A higher setpoint involves less cooling output in the air handling unit and less energy usage from the chiller side. For the static pressure setpoint, most of the optimized points are higher than the original ones. A combination of the recommended setpoints results in minimization of the total energy. However,

the change of energy consumption in each sector differs. Figures 4.8 to 4.11 express the energy usage of the fan, pump, reheat device, and chiller before and after optimization.

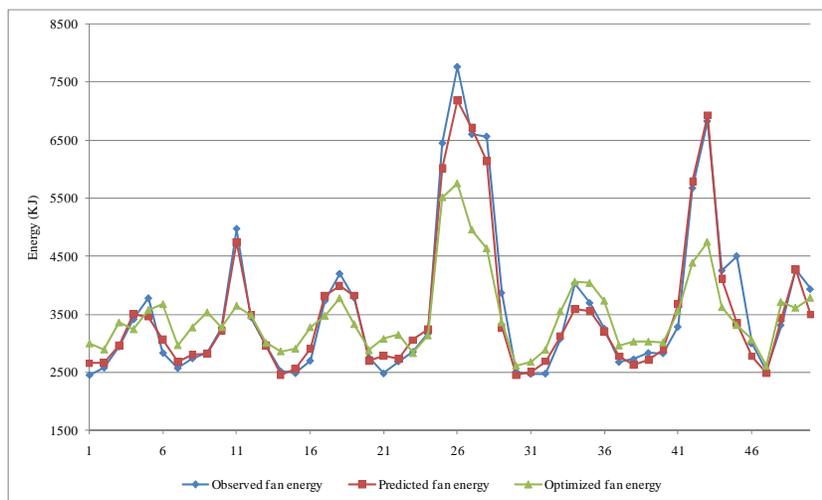


Figure 4.8. The fan energy before and after optimization.

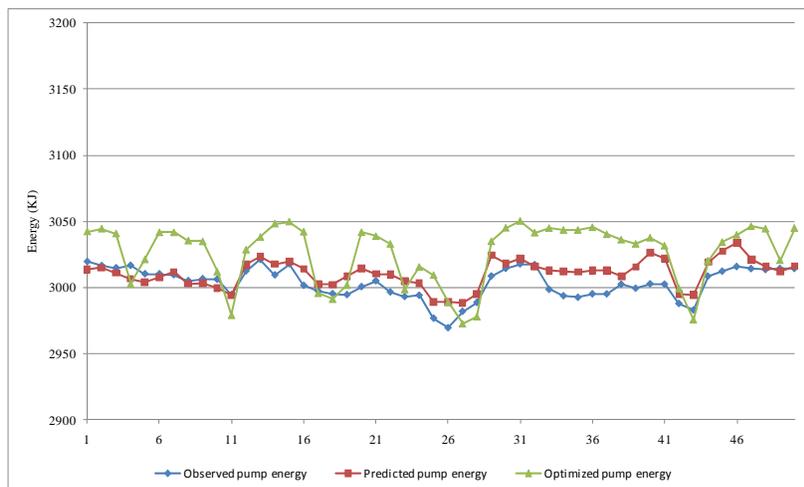


Figure 4.9. The pump energy before and after optimization.

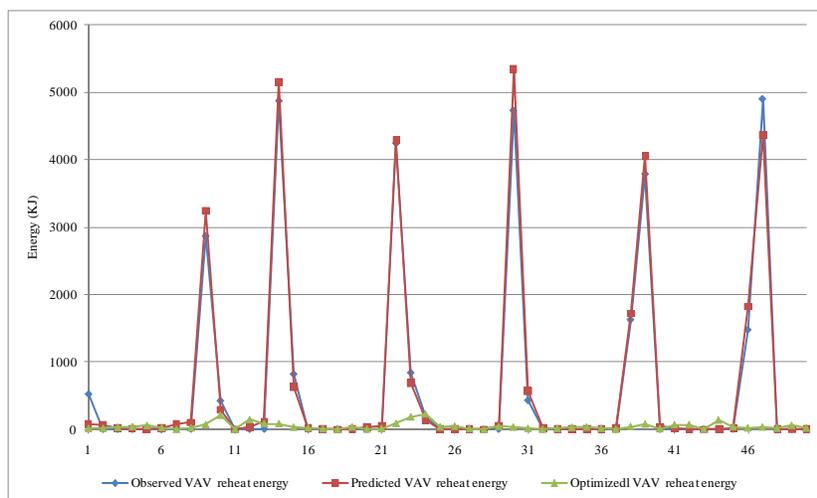


Figure 4.10. The reheat device energy before and after optimization.

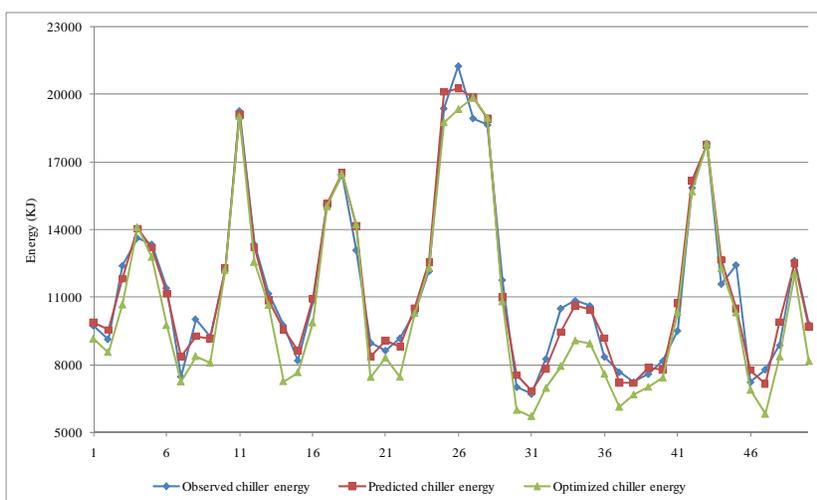


Figure 4.11. The chiller energy before and after optimization.

The optimized fan energy in Figure 4.8 is generally higher than the observed (measured) one because the optimized setpoint of the static pressure for the corresponding points is higher. The higher static pressure in the duct is due to the increased flow rate of the supply air. The

pump energy in Figure 4.9 shows minor changes before and after optimization. The reheat device energy illustrated in Figure 4.9 is minimized to 0 because a higher supply air temperature in the air handling unit decreases the chance that the room temperature drops below the heating setpoint. Thus, the reheat does not take place frequently compared to the original case. The total energy reduction is attributed mainly to the chiller energy reduction. As shown in Figures 4.5 and 4.11, energy changes are similar due to the correlation coefficient of 0.95 between observed chiller energy and total energy. Since the chiller's energy accounts for about 60% of the total energy, reducing the chiller's energy contributes to most of the total energy savings in spite of the increase in energy consumption by the pump and the fan. To investigate further the reason for the chiller energy reduction after adjusting the AHU control settings, Figure 4.12 presents the cooling output of the air handling unit before and after optimization.

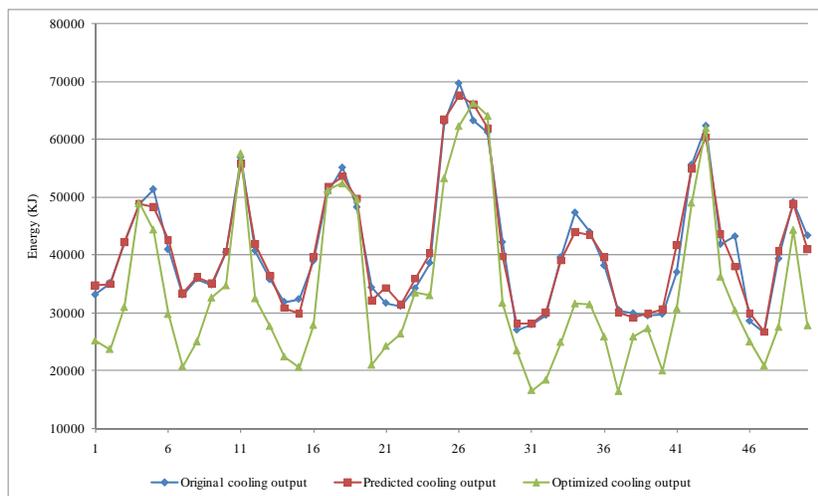


Figure 4.12. The cooling output before and after optimization.

The data set used in this paper has been collected in an experiment performed during the cooling season, and the average outside air temperature was higher than the supply air temperature. Higher values of the recommended supply air temperature setpoints reduced the

heat exchange between the chilled water side and the supply air side. Therefore, the cooling output of the air handling unit was reduced. Less cooling output decreased the chiller energy consumption due to the high linearity between these two parameters. For example, the correlation coefficient between the observed cooling output and the chiller energy for the selected 50 training points (50-h period) is 0.96. A decrease in cooling output results in energy reduction in the chiller energy consumption and further minimization of the total energy. Figure 4.13 illustrates the corresponding indoor air quality indexes in the thermal zone. Three indoor air quality models are built in the same fashion as described in Section 4.4. Average values of the indoor air quality (IAQ) metrics, i.e., the room temperature, the relative humidity, and the CO<sub>2</sub> concentration, computed from 50 training points, are displayed in Figure 4.13.

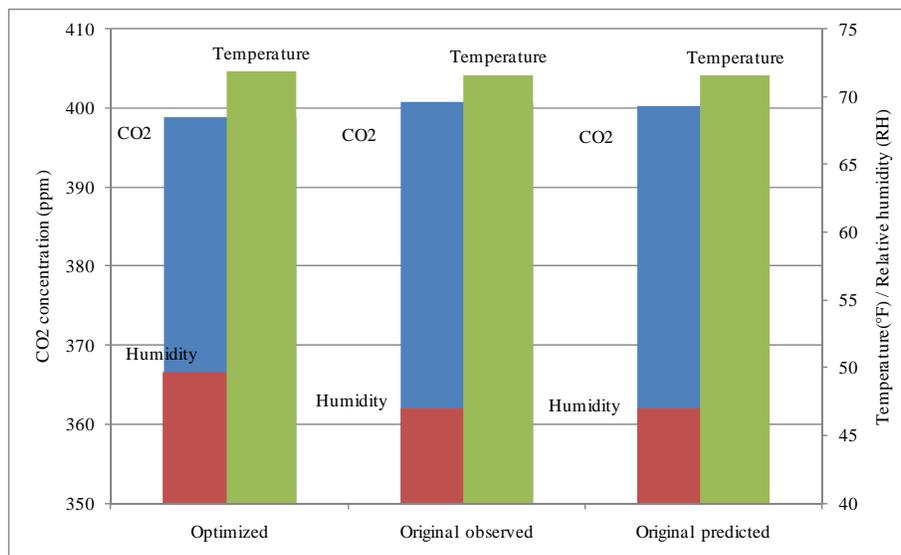


Figure 4.13. Comparison of the IAQ metrics before and after optimization.

The observed (original) and predicted values prove the three IAQ models are accurate. By modifying the control settings of the air handling unit, the room temperature remains about the same. A slight raise in humidity is observed, while CO<sub>2</sub> slightly decreases. The overall indoor

air quality is maintained at an acceptable level given the recommended control settings to save total energy.

The results discussed above represent optimal combinations of the supply air temperature and the static pressure setpoints for data points updated every hour. These setpoints can be adjusted at less frequent intervals, producing fewer energy savings, as shown in Table 4.7. Table 4.7 presents the total energy savings for four different setpoint adjustment scenarios, ranging from every hour change (1 period) to one change per 50 hours (50 periods).

Table 4.7. Energy savings with varied adjustment length of setpoints.

Setpoint Adjustment Time Period	Reduction of Chiller Energy	Reduction of Fan Energy	Reduction of Pump Energy	Reduction of Reheat Device Energy	Reduction of Total Energy
1 period	6.58%	3.12%	-0.77%	92.55%	7.66%
2 periods	5.52%	-0.20%	-0.75%	80.18%	5.95%
5 periods	4.22%	-2.88%	-0.73%	35.52%	3.12%
50 periods	3.75%	-7.50%	-0.91%	3.58%	0.83%

As the setpoints are adjusted less frequently, more energy is consumed. In an extreme case, even when the optimal combination of setpoints is kept for 50 h (50 periods), less than 1% of total energy is saved. The increased energy savings are due to the timely optimization of control settings based on the changing internal load, the outside weather temperature, and other uncontrollable parameters.

#### 4.7. Summary

In this chapter, a data-driven approach to deriving energy models has been presented. Data-mining algorithms were employed to select significant parameters and construct four models of energy consumption. To minimize the total energy consumption by a facility, a single-objective optimization model was formulated and solved by the particle swarm optimization

algorithm. The particle swarm optimization algorithm searches the near optimal solutions of supply air temperature and static pressure setpoints in the air handling unit. Optimal control settings of the air handling unit were generated every hour to minimize the total energy usage in response to different patterns of the internal load and other uncontrollable parameters. The optimization results demonstrated a 7.66% savings of the total energy in spite of an energy increase in certain individual components (pumps or fans). The analysis and discussion of IAQ metrics and the setpoint adjustment time frequency were also included.

## CHAPTER 5

### MULTI-OBJECTIVE OPTIMIZATION OF HVAC SYSTEM ENERGY MANAGEMENT

#### 5.1. Introduction

In this chapter, data-driven models of energy consumption are proposed. Data mining algorithms establish mappings between input and output variables without requiring detailed prior knowledge of the modeled process. Data collected from experiments conducted in an energy research facility has been used to investigate the relationships between control settings and energy consumption as well as facility indoor air quality index (AQI). Rather than minimizing the energy in a single objective, a trade-off between the energy consumption and AQI is considered. The total energy model and the AQI models built by data-mining algorithms are transformed into a multiple objective optimization model. To solve this four objective optimization problem, a particle swarm optimization (PSO) algorithm based on two-level non-dominated solutions is used. The PSO algorithm generates control settings in response to the changing different internal load and uncontrollable variables, including the energy consumption of the total system and the facility. The air quality metrics are assigned different weights reflecting preferences of the occupants.

#### 5.2. Data description and parameter selection

The data set 1 in Table 2.1 is used for this research. The experiment period was from August 1 2009 to August 16 2009. Since each day of the experiment covered a specific combination of setpoints, an arbitrary partitioning of the data into training and testing parts based on time could not produce a valid model. Therefore, the sampled data is used for parameter selection, algorithm selection, and model construction. The data collected in the ERS experiment is described in Table 5.1.

Table 5.1. Data description

Data Set	Description	No. of Instances
1	Parameter selection and algorithm selection, random sampled from original data set	129
2	Modeling training, random sampling; 85% of original data set	658
3	Modeling testing; The data (15%) excluded from model training	116
4	Optimization data set	77

Based on the domain knowledge and parameter selection algorithms like boosting tree and wrapper, eleven parameters have been selected for building the energy and AQI models.

Table 5.2 lists the parameters selected for building the AQI models and the total energy model.

Table 5.2. Parameter description

Parameter	Description	Unit
$x_1(t)$	Internal load at current time interval	Discrete
$x_1(t-1)$	Internal load at previous time interval	Discrete
$x_2(t)$	Supply air temperature setpoint at current time interval	°F
$x_3(t)$	Supply air static pressure setpoint at current time interval	WG
$v_1(t)$	Chilled water entering temperature (mean) at current time interval	°F
$v_2(t)$	Chilled water entering temperature (standard deviation) at current time interval	°F
$v_3(t)$	Outside air temperature (mean) at current time interval	°F
$v_4(t)$	Solar normal flux (mean) at current time interval	B/HFt2
$v_5(t)$	Solar normal flux (standard deviation) at current time interval	B/HFt3
$v_6(t)$	Outside air relative humidity (mean) at current time interval	RH
$v_7(t)$	Outside air CO <sub>2</sub> concentration (mean) at current time interval	ppm

### 5.3. Algorithm selection

The parameters listed in Table 5.1 are used to build the total energy model and the AQI models expressed in (5.1) to (5.4).

$$y_1(t) = f_1(x_1(t), x_1(t-1), x_2(t), x_3(t), v_1(t), v_2(t), v_3(t), v_4(t), v_5(t)) \quad (5.1)$$

$$y_2(t) = f_2(x_1(t), x_1(t-1), x_2(t), x_3(t), v_1(t), v_2(t), v_3(t), v_4(t), v_5(t)) \quad (5.2)$$

$$y_3(t) = f_3(x_1(t), x_1(t-1), x_2(t), x_3(t), v_1(t), v_2(t), v_3(t), v_4(t), v_5(t), v_6(t)) \quad (5.3)$$

$$y_4(t) = f_4(x_1(t), x_1(t-1), x_2(t), x_3(t), v_1(t), v_2(t), v_3(t), v_4(t), v_5(t), v_7(t)) \quad (5.4)$$

where  $y_1(t)$ ,  $y_2(t)$ ,  $y_3(t)$ ,  $y_4(t)$  denote the total energy consumption, average facility temperature, average facility humidity, and the average facility CO<sub>2</sub> concentration during 1 hour time period, respectively.

To extract the mapping among the variables involved in models (5.1) – (5.4), several data-mining algorithms are used, namely Chi-squared Automatic Interaction Detector (CHAID), Boosting tree, Random Forest, Multi-layer Perceptron (MLP), MLP Ensemble, Multivariate Adaptive Regression Splines (MARSplines), and Support Vector Machine (SVM).

As indicated in Table 4.1, data set 2 is used to build a model, while the data set 3 is used to validate it. Table 5.3 presents the training and test accuracy results of models built with eight different data-mining algorithms.

Table 5.3. Training and testing accuracy results for models extracted with different data-mining algorithms

Algorithm	Data Set	Energy Model				Temperature Model			
		MAE	Std of AE	MAPE	Std of APE	MAE	Std of AE	MAPE	Std of APE
MLP Ensemble	Training	758.28	712.03	3.73%	3.73%	0.19	0.18	0.27%	0.25%
	Testing	967.19	904.01	4.98%	5.35%	0.23	0.21	0.32%	0.29%
MLP	Training	827.55	709.88	4.06%	3.81%	0.23	0.23	0.32%	0.32%
	Testing	1098.07	965.86	5.66%	5.85%	0.28	0.22	0.39%	0.30%
Standard CHAID	Training	2147.70	1994.14	10.53%	9.53%	0.32	0.41	0.44%	0.56%
	Testing	2549.66	2316.78	12.17%	10.16%	0.29	0.33	0.40%	0.46%
Boosting Trees	Training	2249.46	2018.83	10.52%	9.06%	0.48	0.38	0.67%	0.52%
	Testing	2650.68	2111.67	13.11%	10.83%	0.45	0.35	0.63%	0.49%
Random Forest	Training	2387.45	1831.75	11.33%	8.49%	0.48	0.51	0.67%	0.70%
	Testing	2399.21	1892.64	12.28%	11.01%	0.47	0.45	0.65%	0.62%
MARSplines	Training	2309.03	2090.62	11.05%	10.81%	0.48	0.39	0.67%	0.53%
	Testing	2508.17	2003.36	13.09%	12.98%	0.48	0.36	0.66%	0.50%
SVM	Training	1290.65	1096.04	6.13%	5.13%	0.34	0.29	0.46%	0.39%
	Testing	1363.01	1138.10	6.88%	6.65%	0.35	0.25	0.48%	0.35%
Algorithm	Data set	Humidity Model				CO <sub>2</sub> Model			
		MAE	Std of AE	MAPE	Std of APE	MAE	Std of AE	MAPE	Std of APE
MLP Ensemble	Training	0.59	0.58	1.10%	1.02%	6.56	5.41	1.64%	1.34%
	Testing	0.65	0.76	1.15%	1.24%	7.34	6.33	1.81%	1.51%
MLP	Training	0.63	0.63	1.16%	1.11%	8.82	6.61	2.21%	1.63%
	Testing	0.73	0.87	1.28%	1.42%	8.98	7.74	2.22%	1.87%
Standard CHAID	Training	1.71	1.57	3.15%	2.93%	11.31	9.82	2.81%	2.32%
	Testing	2.20	2.08	3.93%	3.87%	11.95	11.91	2.91%	2.69%
Boosting Trees	Training	1.24	1.05	2.30%	1.93%	9.44	7.89	2.34%	1.83%
	Testing	1.26	1.06	2.22%	1.79%	11.56	10.61	2.82%	2.36%
Random Forest	Training	3.25	2.35	6.22%	4.66%	12.60	10.94	3.14%	2.59%
	Testing	4.09	2.79	7.38%	4.91%	14.55	14.54	3.56%	3.25%
MARSplines	Training	1.23	1.08	2.28%	1.89%	10.27	9.04	2.55%	2.15%
	Testing	1.33	1.08	2.38%	1.84%	11.41	11.12	2.79%	2.52%
SVM	Training	0.95	0.82	1.77%	1.44%	10.87	8.90	2.71%	2.10%
	Testing	1.02	1.03	1.86%	1.77%	12.08	10.37	2.97%	2.32%

Based on the training and testing error, the computational results reported in Table 4.3 show that the MLP Ensemble with 5 MLPs performs best on the MAE and MAPE metrics for

four models. Therefore, it is selected as the algorithm for building the total energy model and AQI models discussed in the section 5.4.

#### 5.4. Modeling building and validation

In developing the neural network models, the total squared error can as the cost function, and the weights can be modified to minimize that cost function accordingly. For each model 100 neural networks are trained, each with 10 to 40 hidden neurons. The best five networks with lowest squared error are selected for testing. The observed and predicted values obtained from the four models are shown in Figure 5.1 through Figure 5.4.

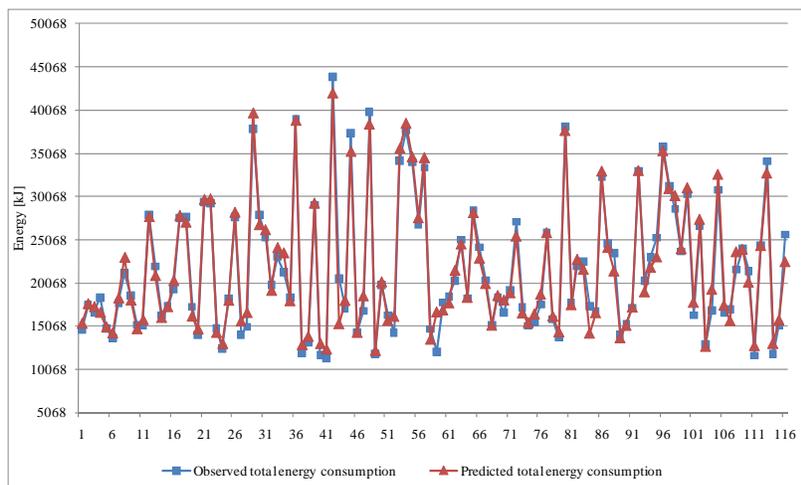


Figure 5.1. Test results from the total energy model

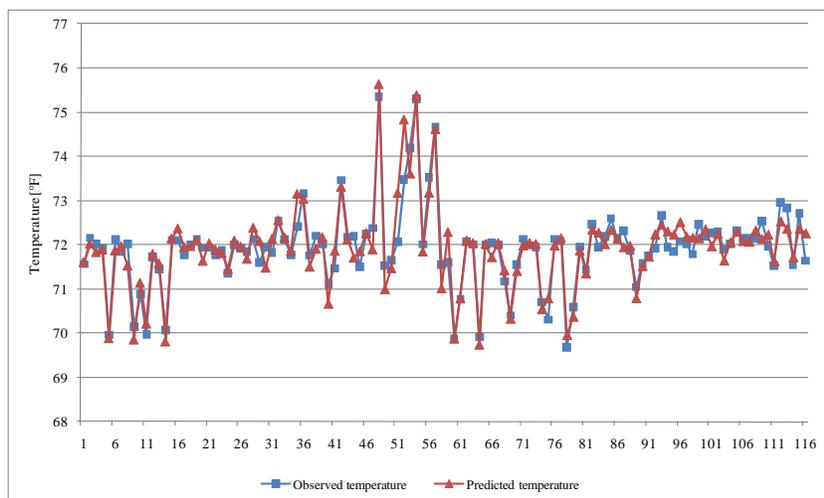


Figure 5.2. Test results from the facility temperature model

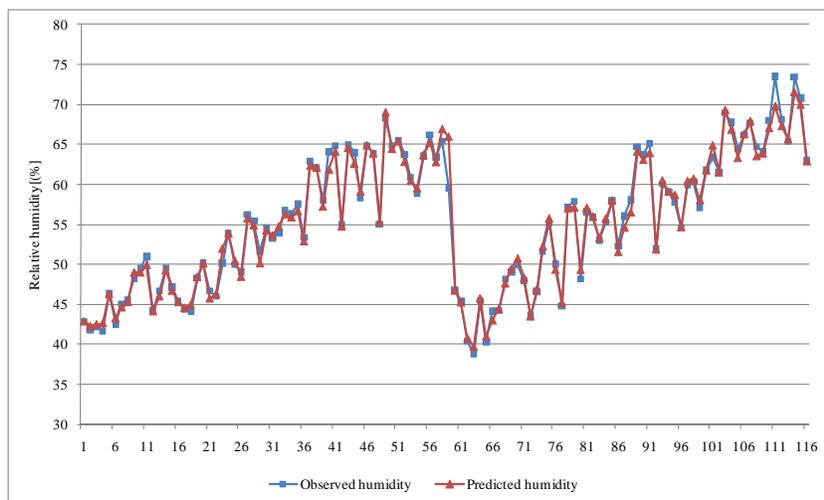


Figure 5.3. Test results from the facility relative humidity model

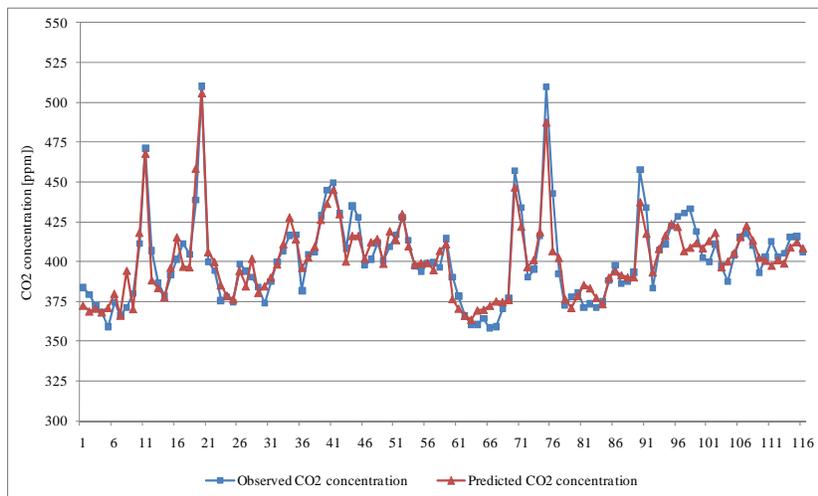


Figure 5.4. Test results from the facility CO<sub>2</sub> concentration model

As shown in the Figure 5.1 through 5.4, the predicted values follow relatively close the observed ones. Table 5.4 shows the correlation coefficients between the observed and predicted values. Therefore, the four models are used to formulate the total energy optimization model.

Table 5.4. Correlation coefficient of the observed and predicted values of the four models

Variable	Means	Std. Dev.	Correlation Coefficient	
Observed Total Energy	21559.53	7543.02	1.00	0.98
Predicted Total Energy	21647.74	7466.96	0.98	1.00
Observed Room Temp	71.93	0.91	1.00	0.94
Predicted Room Temp	71.93	0.95	0.94	1.00
Observed Room Humd	55.41	8.51	1.00	0.99
Predicted Room Humd	55.33	8.32	0.99	1.00
Observed Room CO2	401.54	27.32	1.00	0.94
Predicted Room CO2	401.41	24.12	0.94	1.00

## 5.5. Model formulation and solving

### 5.5.1 Single-objective optimization model formulation and solving

A single-objective optimization model minimizing the total energy is formulated in (5.5).

The decision variables are supply air temperature setpoint and static pressure setpoint.

$$\begin{aligned}
 & \min_{x_2(t), x_3(t)} \text{Obj} \\
 & \text{subject to:} \\
 & 50 \leq x_2(t) \leq 65 \\
 & 1.2 \leq x_3(t) \leq 1.8
 \end{aligned} \tag{5.5}$$

where  $\text{Obj} = y_1(t)$ ,  $y_1(t)$  is the total energy model built in the Section 5.4. The canonical particle swarm optimization (PSO) algorithm mentioned in Chapter 4.5 is conducted. Table 5.5 illustrates an instance of the single-objective optimization model.

Table 5.5. Instance used in single-objective optimization

Parameter	$x_1(t-1)$	$x_1(t)$	$x_2(t)$	$x_3(t)$	$v_1(t)$	$v_2(t)$	$v_3(t)$	$v_4(t)$	$v_5(t)$	$v_6(t)$	$v_7(t)$
Value	Stage 1	Stage 4	63.00	1.40	42.20	2.65	81.47	204.01	13.77	66.89	308.11

To solve this single-objective model, the initial population size is set at 100 and the maximum number of iterations at 50, which is sufficient to converge. The optimal values of  $x_2(t)$  and  $x_3(t)$  in this instance are 54.65 and 1.20, respectively. The minimal total energy is 17167.44 kJ. Since the models of facility temperature, relative humidity, and CO<sub>2</sub> concentration have been constructed in Section 5.4, the corresponding AQI indexes can be obtained by replacing the original control settings with the optimal ones. The facility temperature is 77.85 °F (25.47°C), facility relative humidity is 46.29%, and CO<sub>2</sub> concentration is 454.06 ppm. In the experiment, the facility temperature cooling setpoint was 72 °F (22.22°C), and heating setpoint was 70 °F (21.11°C). Using the optimized control settings, the total energy is minimized while the facility temperature is outside the acceptable level. Therefore, a trade-off between the energy

consumption and the AQI index needs to be accomplished by constructing a multiple objective optimization model.

### 5.5.2 Formulation and solving of the quad-objective optimization model

The generalized multi-objective optimization model is presented in (5.6) [57]:

$$\begin{aligned} & \min f_i(x) \quad i = 1, \dots, N_{obj} \\ \text{subject to: } & \begin{cases} g_j(x) = 0 & j = 1, \dots, M \\ h_k(x) \leq 0 & k = 1, \dots, K \end{cases} \end{aligned} \quad (5.6)$$

where  $f_i$  refers to the  $i^{\text{th}}$  objective function,  $x$  is the decision variable, and  $N_{obj}$  indicates the number of objectives. Several objectives are optimized simultaneously while decision variables should satisfy the equality and inequality constraints.

In the multi-objective optimization model constructed in this section, the AQI constraints are considered and then the original non-constrained optimization model is transformed into a constrained optimization model in (5.7).

$$\begin{aligned} & \min_{x_2(t), x_3(t)} y_1(t) \\ \text{subject to: } & \\ & y_1(t) = f_1(x_1(t), x_1(t-1), x_2(t), x_3(t), v_1(t), v_2(t), v_3(t), v_4(t), v_5(t)) \\ & y_2(t) = f_2(x_1(t), x_1(t-1), x_2(t), x_3(t), v_1(t), v_2(t), v_3(t), v_4(t), v_5(t)) \\ & y_3(t) = f_3(x_1(t), x_1(t-1), x_2(t), x_3(t), v_1(t), v_2(t), v_3(t), v_4(t), v_5(t), v_6(t)) \\ & y_4(t) = f_4(x_1(t), x_1(t-1), x_2(t), x_3(t), v_1(t), v_2(t), v_3(t), v_4(t), v_5(t), v_7(t)) \\ & 50 \leq x_2(t) \leq 65 \\ & 1.2 \leq x_3(t) \leq 1.8 \\ & 70.5 \leq y_2(t) \leq 71.5 \\ & 49 \leq y_3(t) \leq 51 \\ & 395 \leq y_4(t) \leq 405 \end{aligned} \quad (5.7)$$

Let  $Obj1 = y_1(t)$ ,  $Obj2 = \max[0, 70.5 - y_2(t)] + \max[0, y_2(t) - 71.5]$ ,  
 $Obj3 = \max[0, 49 - y_3(t)] + \max[0, y_3(t) - 51]$  and  $Obj4 = \max[0, 395 - y_4(t)] + \max[0, y_4(t) - 405]$ , the constrained optimization model (13) is transformed into an unconstrained quad-objective optimization model (5.8).

$$\begin{aligned}
& \min_{x_2(t), x_3(t)} (Obj1, Obj2, Obj3, Obj4) \\
& \text{where:} \\
& Obj1 = y_1(t) \\
& Obj2 = \max[0, 70.5 - y_2(t)] + \max[0, y_2(t) - 71.5] \\
& Obj3 = \max[0, 49 - y_3(t)] + \max[0, y_3(t) - 51] \\
& Obj4 = \max[0, 395 - y_4(t)] + \max[0, y_4(t) - 405] \\
& 50 \leq x_2(t) \leq 65 \\
& 1.2 \leq x_3(t) \leq 1.8
\end{aligned} \tag{5.8}$$

Evolutionary computation algorithms are natural candidates for solving multi-objective optimization models [58]. Fonseca et al. [59] classified them into three categories: aggregation-based, non-Pareto, and Pareto-based approaches. Unlike other evolutionary computational algorithms, Particle Swarm Optimization (PSO) algorithm is a stochastic optimization technique inspired by bird flocks. As one of the swarm intelligence algorithms, PSO is able to search for global and local solutions. The canonical PSO is usually applied for solving single-objective optimization models. In this paper, a modified PSO based on two levels of non-dominated solutions is used. In contrast to the canonical PSO usually used for solving the single-objective models, the Pareto-optimality concept is incorporated into the algorithm to expand its capability to handle models with several conflicting objectives.

Pareto optimality is defined next [25]. Assume  $F(x) = (f_1(x), \dots, f_{N_{obj}}(x))$ . A decision variable  $x_u$  is said to be Pareto-optimal, if and only if, there is no  $x_v$  for which  $v = F(x_v) = (v_1, \dots, v_{N_{obj}})$  dominates  $u = F(x_u) = (u_1, \dots, u_{N_{obj}})$ . In other words, there is no  $x_v$  such that

$$\forall i \in \{1, \dots, N_{obj}\}, v_i < u_i \tag{5.9}$$

The set of all Pareto-optimal decision vectors is called the Pareto-optimal, while the corresponding set of objective vectors is called the non-dominated set.

The canonical PSO algorithm has been modified to make it applicable for solving the multi-objective optimization model.

### **Modification I: Generation of non-dominated solutions**

Establish a storage set  $S_i$  for each particle  $x_i$  to store the non-dominated solutions at every iteration. Establish a storage set  $G$  to store the non-dominated solutions of  $S_i$  produced at every iteration. Create an external set  $E$  to store the non-dominated solutions of  $G$  produced at every iteration.

### **Modification II: Update of non-dominated solutions**

At each iteration, for every local non-dominated set  $S_i$ , compare the current iteration particle solution and the solutions from in the previous iterations. Remove the dominated solutions and store the non-dominated. At each iteration, for the global non-dominated set  $G$ , copy all local non-dominated sets into the global non-dominated set. Select and keep the non-dominated solutions only. At each iteration, for external non-dominated set  $E$ , copy the global non-dominated set  $G$  to  $E$ . Compare the solutions and keep the non-dominated ones only.

### **Modification III: Generation of local and global best solutions**

At each iteration, for each particle, compute the Euclidean distance between every pair of local non-dominated and global non-dominated solution. The pair with minimum distance in the objective value space is selected as the local and global best. For the corresponding particle velocity and position are updated.

The MOPSO algorithm determines a set of non-dominated solutions of the quad-objective model (14). Figure 5.5 illustrates the solution generation process in two dimensions (the objective values), rather than the original four dimensional space which is not able to be graphically interpreted. The solutions can be interpreted from the plots in Figure 5.5. For example, in the upper left corner of Figure 5, solutions having both small values of  $Obj1$  and  $Obj2$  are included. However, in the  $Obj1$  and  $Obj3$  space, the corresponding value of  $Obj3$  is much larger.

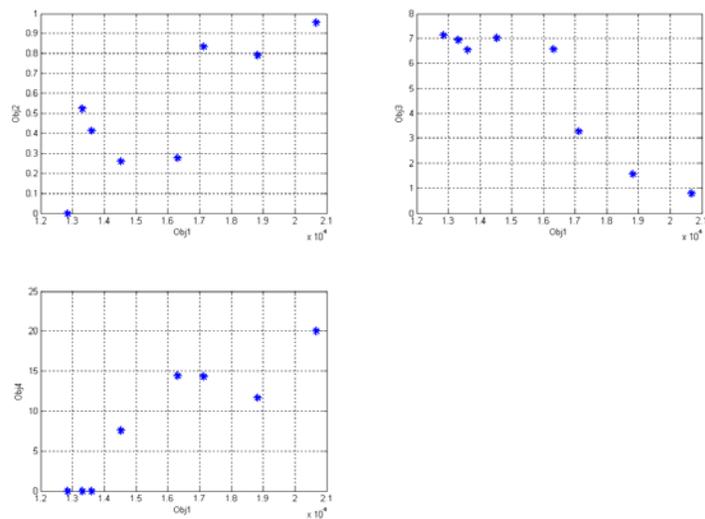


Figure 5.5. Two dimensional solution spaces

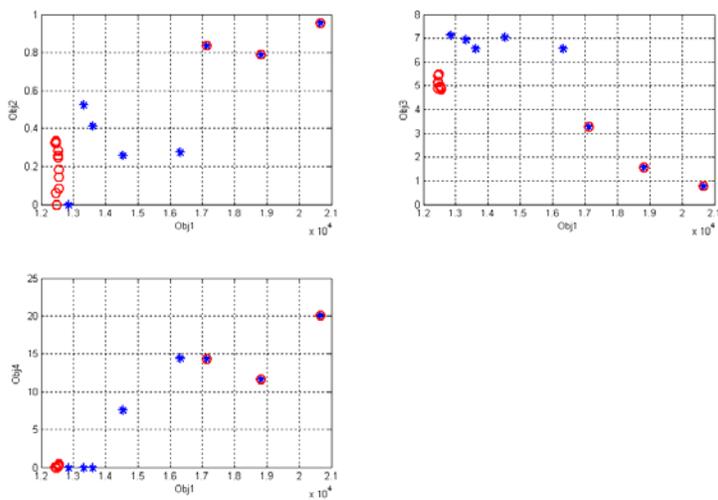


Figure 5.6. The solution process

Figure 5.6 illustrates the search process. The asterisks are the elite set solutions at early iterations. The circles are the elite set solutions at a later stage of the search process. As the number of iterations has increased, most asterisks and dots tend to move towards the bottom-left

corner. This indicates that the many newly generated solutions dominate the old ones in the space of objective values.

### 5.5.3 Weight assignment and solution selection

The optimal solution is selected from the final elite set by the weighted normalized objective function (5.10).

$$Obj = w_1 \frac{Obj1 - Obj1_{\min}}{Obj1_{\max} - Obj1_{\min}} + w_2 \frac{Obj2 - Obj2_{\min}}{Obj2_{\max} - Obj2_{\min}} + w_3 \frac{Obj3 - Obj3_{\min}}{Obj3_{\max} - Obj3_{\min}} + w_4 \frac{Obj4 - Obj4_{\min}}{Obj4_{\max} - Obj4_{\min}} \quad (5.10)$$

where  $w_1, w_2, w_3, w_4$  are the user-defined weights indicating the importance of the corresponding objective, and  $Obj1_{\max}$  and  $Obj1_{\min}$  are the maximum and the minimum values of  $Obj1$  in the final elite set. Similar notation is used for  $Obj2_{\max}$ ,  $Obj2_{\min}$ ,  $Obj3_{\max}$ ,  $Obj3_{\min}$ ,  $Obj4_{\max}$  and  $Obj4_{\min}$ . Note that  $\sum_{m=1}^4 w_m = 1$ , with  $w_1, w_2, w_3, w_4$  being either constants or functions of other objectives.

Table 5.6 presents eight scenarios representing different assignments of weights to the objectives. Weights vary within different preference bounds associated with the AQI indexes,  $\Theta_A = [70.5, 71.5]$ ,  $\Theta_B = [49, 51]$ ,  $\Theta_C = [395, 405]$ . Scenario 1 is equivalent to the single objective optimization problem without considering the AQI constraints. Scenario 2-4 represent that each factor in the AQI constraints has equal importance of energy. The original problem is transformed into a bi-objective optimization problem. Scenario 5-7 assigns equal weights to the two components of the AQI constraints and the energy, which result in a three-objective optimization model. In Scenario 8, a four-objective optimization model is constructed by assigning equal weights to each objective.

Table 5.6. Description of the eight weight assignment scenarios.

Scenario	Weights of four objectives	Description
1	$w_1 = 1, w_2 = 0, w_3 = 0, w_4 = 0$	No AQI constraints
2	$w_1 = \begin{cases} 1 & \text{obj2} \in \Theta_A \\ 0.5 & \text{obj2} \in \Theta_A^c \end{cases}, w_2 = \begin{cases} 0 & \text{obj2} \in \Theta_A \\ 0.5 & \text{obj2} \in \Theta_A^c \end{cases}, w_3 = 0, w_4 = 0$	Preference bound for facility temperature
3	$w_1 = \begin{cases} 1 & \text{obj3} \in \Theta_B \\ 0.5 & \text{obj3} \in \Theta_B^c \end{cases}, w_2 = 0, w_3 = \begin{cases} 0 & \text{obj3} \in \Theta_B \\ 0.5 & \text{obj3} \in \Theta_B^c \end{cases}, w_4 = 0$	Preference bound for facility humidity
4	$w_1 = \begin{cases} 1 & \text{obj4} \in \Theta_C \\ 0.5 & \text{obj4} \in \Theta_C^c \end{cases}, w_2 = 0, w_3 = 0, w_4 = \begin{cases} 0 & \text{obj4} \in \Theta_C \\ 0.5 & \text{obj4} \in \Theta_C^c \end{cases}$	Preference bound for facility CO <sub>2</sub>
5	$w_1 = \begin{cases} 1 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B \\ 0.5 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B^c \\ 0.5 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B^c \end{cases}, w_2 = \begin{cases} 0 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B \\ 0 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B^c \\ 0.5 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B^c \end{cases}$ $w_3 = \begin{cases} 0 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B \\ 0.5 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B^c \\ 0 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B^c \end{cases}, w_4 = 0$	Preference bound for facility temperature and preference bound for facility humidity
6	$w_1 = \begin{cases} 1 & \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C \\ 0.5 & \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C^c \\ 0.5 & \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C \\ 0.33 & \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C^c \end{cases}, w_2 = 0$ $w_3 = \begin{cases} 0 & \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C \\ 0 & \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C^c \\ 0.5 & \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C \\ 0.33 & \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C^c \end{cases}, w_4 = \begin{cases} 0 & \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C \\ 0.5 & \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C^c \\ 0 & \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C \\ 0.33 & \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C^c \end{cases}$	Preference bound for facility temperature and preference bound for facility CO <sub>2</sub>

Table 5.6. --continued

7	$w_1 = \begin{cases} 1 & \text{obj2} \in \Theta_A, \text{obj4} \in \Theta_C \\ 0.5 & \text{obj2} \in \Theta_A, \text{obj4} \in \Theta_C^c \\ 0.5 & \text{obj2} \in \Theta_A^c, \text{obj4} \in \Theta_C \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj4} \in \Theta_C^c \end{cases}, w_2 = \begin{cases} 0 & \text{obj2} \in \Theta_A, \text{obj4} \in \Theta_C \\ 0 & \text{obj2} \in \Theta_A, \text{obj4} \in \Theta_C^c \\ 0.5 & \text{obj2} \in \Theta_A^c, \text{obj4} \in \Theta_C \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj4} \in \Theta_C^c \end{cases}$ $w_3 = 0, w_4 = \begin{cases} 0 & \text{obj2} \in \Theta_A, \text{obj4} \in \Theta_C \\ 0.5 & \text{obj2} \in \Theta_A, \text{obj4} \in \Theta_C^c \\ 0 & \text{obj2} \in \Theta_A^c, \text{obj4} \in \Theta_C \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj4} \in \Theta_C^c \end{cases}$	Preference bound for facility humidity and preference bound for facility CO <sub>2</sub>
8	$w_1 = \begin{cases} 1 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C \\ 0.33 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C^c \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C^c \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C \\ 0.5 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C^c \\ 0.5 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C \\ 0.5 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C \\ 0.25 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C^c \end{cases}, w_2 = \begin{cases} 0 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C \\ 0 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C^c \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C^c \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C \\ 0 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C^c \\ 0 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C \\ 0.5 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C \\ 0.25 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C^c \end{cases}$ $w_3 = \begin{cases} 0 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C \\ 0.33 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C^c \\ 0 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C^c \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C \\ 0 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C^c \\ 0.5 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C \\ 0 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C \\ 0.25 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C^c \end{cases}, w_4 = \begin{cases} 0 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C \\ 0.33 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C^c \\ 0.33 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C^c \\ 0 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C \\ 0.5 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C^c \\ 0 & \text{obj2} \in \Theta_A, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C \\ 0 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B, \text{obj4} \in \Theta_C \\ 0.25 & \text{obj2} \in \Theta_A^c, \text{obj3} \in \Theta_B^c, \text{obj4} \in \Theta_C^c \end{cases}$	Preference bound for facility temperature, preference bound for facility humidity and preference bound for facility CO <sub>2</sub>

The solutions at some time stamp for each of the eight scenarios are shown in the Table 5.7. After assigning different weights to the objectives, the multi-objective optimization model is transformed into a single objective model to be minimized with the objective function shown in (5.10). Scenario 1 has the lowest total energy consumption as the AQI constraints are not considered. As additional constraints are included, the total energy consumption is affected.

Table 5.7. Solutions for the eight scenarios at some time stamp

Scenario	1	2	3	4	5	6	7	8
Total energy	25449.47	26053.00	26822.61	25865.57	26833.86	26822.61	26391.62	26833.86
Recommended SA settings	51.21	50.67	54.56	52.59	52.99	54.56	52.47	52.99
Recommended SP settings	1.77	1.55	1.75	1.78	1.56	1.75	1.61	1.56
Corresponding facility temperature	72.12	72.01	72.10	72.12	72.02	72.10	72.03	72.02
Corresponding facility humidity	43.74	42.84	48.46	45.39	45.82	48.46	45.27	45.82
Corresponding facility CO <sub>2</sub> concentration	375.72	364.89	406.45	401.85	402.85	406.45	395.35	402.85

### 5.6. Optimization results and discussion

Data set 4 in Table 5.1 is used for optimization since most of the points are out of control. The room temperature and room humidity in the optimization data set can be maintained in the restricted range to meet the system load. The MOPSO algorithm of Section 5.5.2 is employed to solve the quad-objective optimization model (5.8). The initial population size is set to 100 while the maximum number of iterations is set at 50. The recommended control settings are Scenario 8 described in Table 5.7. Figures 5.7 and 5.8 compare the original and recommended control settings of the supply air temperature and the static pressure.

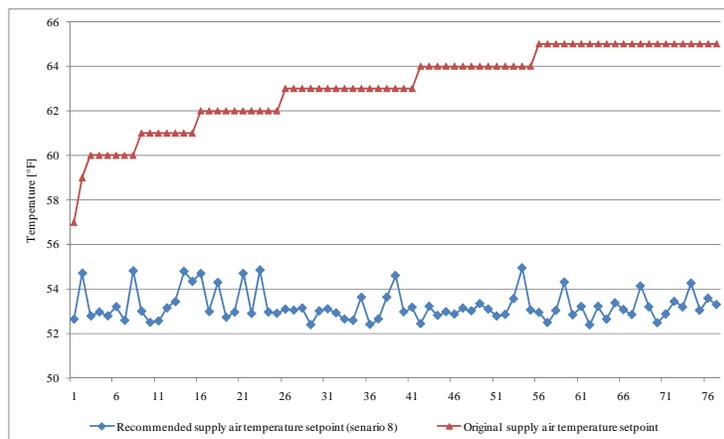


Figure 5.7. The original and recommended setpoints of the supply air temperature

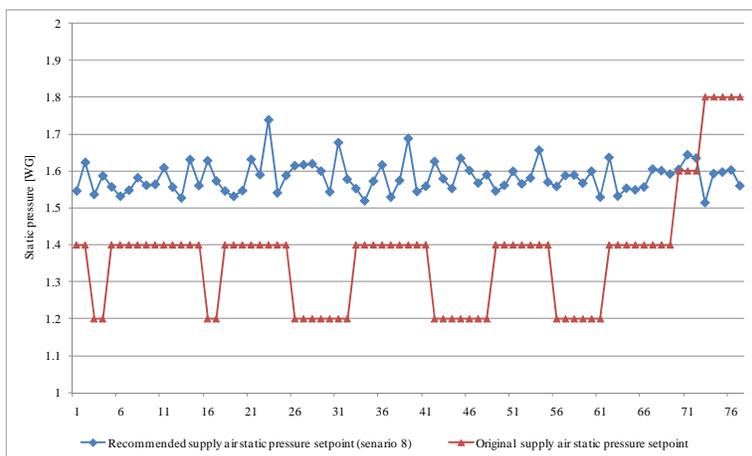


Figure 5.8. The original and recommended setpoints of the supply air static pressure

Based on the optimal control settings, the corresponding total energy consumption and the facility air quality metrics are estimated using the energy and the AQI models of Section 5.4. Figures 5.9 to 5.12 compare the original and optimized energy as well as the AQI indexes.

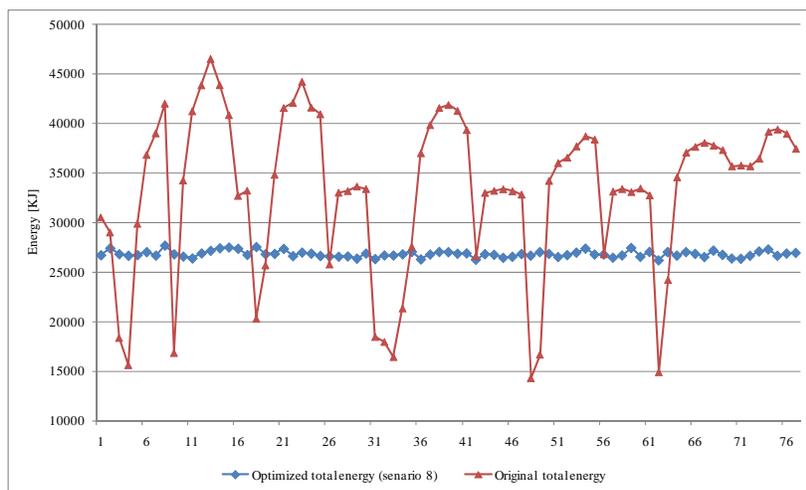


Figure 5.9. The original and the optimized total energy

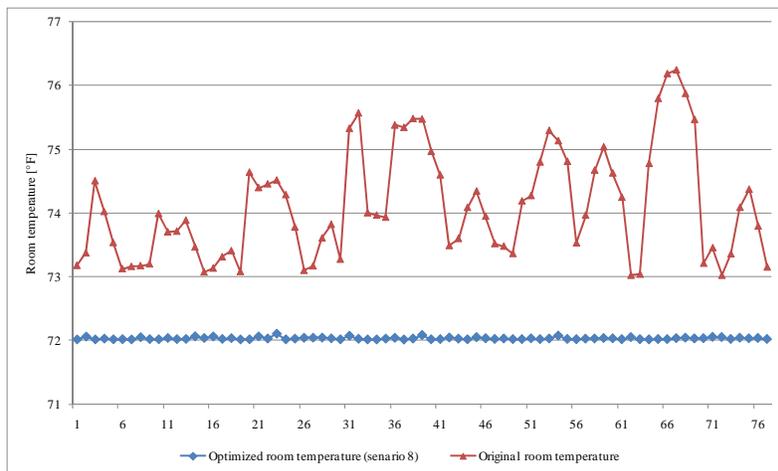


Figure 5.10. The original and the optimized facility temperature

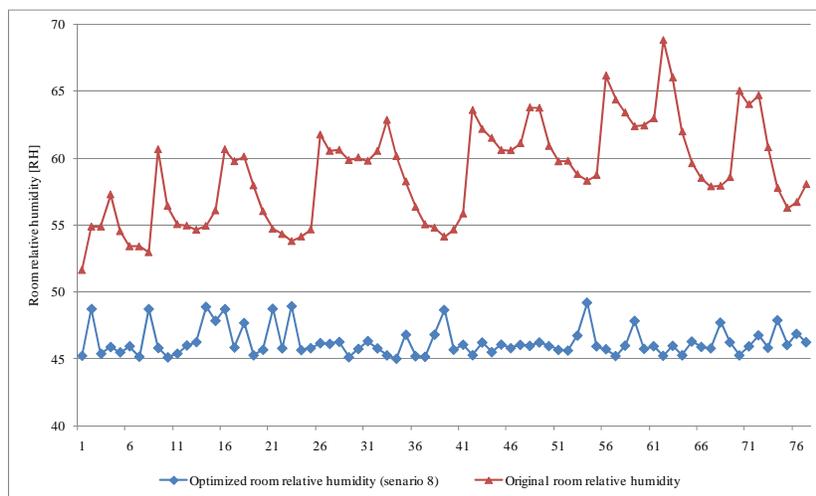


Figure 5.11. The original and the optimized facility relative humidity

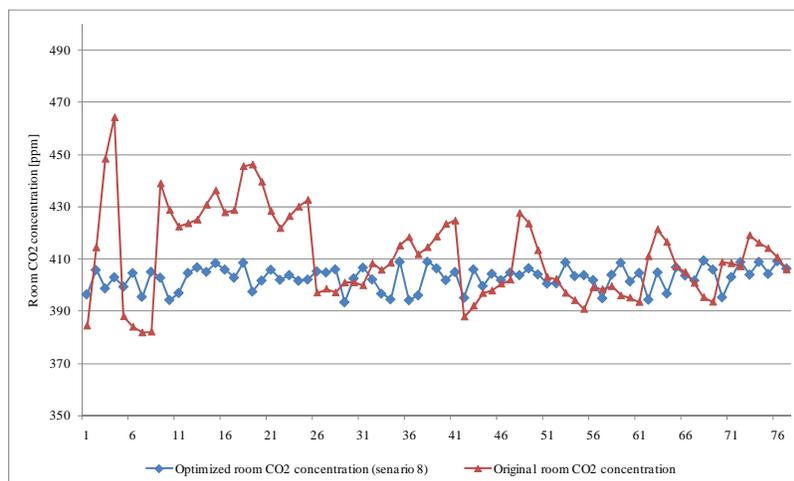


Figure 5.12. The original and the optimized facility CO<sub>2</sub> concentration

Figure 5.9 demonstrates that the optimized total energy consumption is relatively stable. It is not uniformly smaller than original energy due to the fact that the total energy is minimized in one objective. The corresponding values of the AQI indexes (within the allowable ranges) are presented in Figure 5.10 through 5.12. At times, the energy minimization objective is compromised in order to maintain the required values of AQI. Consider all three AQI constraints, the total energy saved for the 77 out-of-control points is 12.4% compared to 17.4% with no AQI constraints included.

### 5.7. Summary

Data-mining algorithms were applied to model the nonlinear relationship among the energy consumption, AQI indexes, control settings, and uncontrollable variables. The Multilayer Perceptron (MLP) ensemble algorithm performed better than any of the seven algorithms and it was selected to build the total energy model and three AQI models. To minimize the total energy consumption while maintaining AQI indexes within acceptable ranges, a single-objective optimization model was generalized to a quad-objective optimization model. A modified Multiple Objective Particle Swarm Optimization (MOPSO) algorithm was employed to optimize

control settings of the supply air temperature and the static pressure. A trade-off between the total energy consumption and preserving AQI indexes was optimized in response to different patterns of the internal load and uncontrollable variables. The total energy savings for the data set considered in this paper was 17.4% in the case where the AQI constraints were not considered and 12.4% when the AQI constraints were applied for one of the eight user preference scenarios (Scenario 8). Optimization results demonstrated that that a proper balance between the energy savings and the AQI indexes can be accomplished in by representing the preference of occupants with the weights used in the objective function.

## CHAPTER 6

### CONCLUSION

The thesis is focused on the evolutionary computation in HVAC system modeling and optimization. Chapter 2 mainly introduces a clustering-based model to construct the predictive models. Neural network is applied to establish the relation between inputs and outputs. The proposed model is demonstrated to be both effective in modeling and short-term prediction. The computational cost is reduced by applying the clustering method.

Chapter 3 applies the evolutionary computation on a multi-objective optimization problem. A bi-objective problem is transformed from a three-objective function and solved by the SPEA. The results show that the HVAC system can be optimized by implementing the optimal control settings and a large amount of energy can be saved with the IAQ maintained at an acceptable level.

In Chapter 4, a single objective optimization algorithm is employed to optimize the HVAC components separately. The energy consumption of each component is shown. Although the energy use increases for some components (fan, pump), the overall performance is improved by using the optimized settings. The analysis and discussion of IAQ metrics and the setpoint adjustment time frequency are also discussed.

Chapter 5 employs a multi-objective particle swarm optimization algorithm to optimize the overall performance of the existing HVAC system. A trade-off between energy saving and AQI maintenance is addressed in detail. A linear operation strategy is presented for the management of the system. It is demonstrated that a proper balance between the energy savings and the AQI indexes can be accomplished by representing the preference of occupants with the weights used in the objective function.

The future research will try to construct an adaptive dynamic model of the HVAC system. Some statistical components will involve in the integration of existing data-driven models. The practical application, computation cost, and lean management will be also considered.

## REFERENCES

- [1] L. Pérez-lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and Buildings*, Vol. 40, No. 3, pp. 394-398, 2008.
- [2] W. Huang, M. Zaheeruddin and S. H. Cho, "Dynamic simulation of energy management control functions for HVAC systems in buildings," *Energy Conversion and Management*, Vol. 47, No. 7-8, pp. 926-943, 2006.
- [3] Y. W. Wang, W. J. Cai, Y. C. Soh, S. J. Li, L. Lu and L. Xie, "A simplified modeling of cooling coils for control and optimization of HVAC systems," *Energy Conversion and Management*, Vol. 45, No. 18-19, pp. 2915-2930, 2004.
- [4] X. Yu, J. Wen and T. F. Smith, "A model for the dynamic response of a cooling coil," *Energy and Buildings*, Vol. 37, No. 12, pp. 1278-1289, 2005.
- [5] Y. Yao, Z. Lian, W. Liu, Z. Hou, M. Wu, "Evaluation program for the energy-saving of variable-air-volume systems," *Applied Thermal Engineering*, Vol. 24, No. 7, pp. 1037-1050, 2004.
- [6] S. Wang and Z. Ma, "Supervisory and optimal control of building HVAC systems: A review," *HVAC&R Research*, Vol. 14, No. 1, pp. 3-32, 2008.
- [7] K. Lomas, H. Eppel, C. Martin, D. Bloomfield, "Empirical validation of thermal building simulation programs using test room data," Final report: Vol. 1, Leicester: De Montfort University, 1994.
- [8] M. Khemani, "Energy audit software directory," *Ottawa: Natural Resources Canada*, September 1997.
- [9] J.P. Waltz, "Computerized building energy simulation handbook," *The Fairmont Press*, Lilburn, Georgia, 2000.
- [10] P. Jacobs, H. Henderson, "State-of-the-art review of whole building, building envelope, and HVAC component and system simulation and design tools," *Final report ARTI-21CR/30010-01*, Arlington: Air-Conditioning and Refrigeration Technology Institute, February 2002.
- [11] S.W. Wang, "Dynamic simulation of a building central chilling system and evaluation of EMCS on-line control strategies," *Building and Environment*, Vol. 33 No. 1, pp. 1-20, 1998.
- [12] D. B. Crawley, J. W. Hand, M. Kummert, and B. T. Griffith, "Contrasting the capabilities of building energy performance simulation programs," *Building and Environment*, Vol. 43, No. 4, pp. 661-673, 2008.
- [13] M. M. Gouda, C. P. Underwood and S. Danaher, "Modeling the robustness properties of HVAC plant under feedback control," *Building Services Engineering Research and Technology*, Vol. 24, No. 4, pp. 271-280, 2003.

- [14] J. Sun, and A. Reddy, "Optimal control of building HVAC&R systems using complete simulation-based sequential quadratic programming (CSB-SQP)," *Building and Environment*, Vol. 40, No. 5, pp. 657–69, 2005.
- [15] R. E. Rink and N. Li, "Aggregation / disaggregation method for optimal control of multi-zone HVAC systems ," *Energy Conversion and Management*, Vol. 36, No. 2, pp 79-86, 1995.
- [16] N. N. Kota, J. M. House, J. S. Arora and T. F. Smith, "Optimal control of HVAC systems using DDP and NLP techniques," *Optimal Control Applications & Methods*, Vol. 17, pp. 71-78, 1996.
- [17] B.B. Ekici, U.T. Aksoy, "Prediction of building energy consumption by using artificial neural networks," *Advances in Engineering Software*, Vol. 40, No. 5, pp. 356-362, 2009.
- [18] S. Karatasou, M. Santamouris, V. Geros, "Modeling and predicting building's energy use with artificial neural networks : Methods and results," *Energy and Buildings*, Vol. 38, No. 8, pp. 949-958, 2006
- [19] A. Kusiak, M.Y. Li, Z.J. Zhang, "A data-driven approach for steam load prediction in buildings," *Applied Energy*. Vol. 87, No. 3, pp. 925-933, 2010.
- [20] O.A. Dombayci, "The prediction of heating energy consumption in a model house by using artificial neural networks in Denizli-Turkey," *Advances in Engineering Software*, Vol. 41 No. 2, pp. 141-147, 2010.
- [21] S.A. Kalogirou, "Applications of artificial neural-networks for energy system," *Applied Energy*, Vol. 67 No. 1-2, pp. 17-35, 2000.
- [22] S. Kalogirou, S. Panteliou, A. Dentsoras, "Modeling of solar domestic water heating systems using artificial neural networks," *Solar Energy*. Vol. 68, No. 6, pp. 335-42, 1999.
- [23] S.A. Kalogirou, "Artificial neural networks in renewable energy systems applications: A review," *Renewable and Sustainable Energy Reviews*, Vol. 5, No. 4, pp. 373-401, 2000.
- [24] J.F. Kreider, X.A. Wang, "Artificial neural network demonstration for automated generation of energy use predictors for commercial buildings," *ASHRAE*, pp.193-8, 1995.
- [25] Y. Ke, S. Mumma, "Optimized supply air temperature in a variables air volume systems," *Energy*, Vol. 22, No. 6, pp. 601-614, 1997.
- [26] S. Wang, and X. Jin, "Model-based optimal control of VAV air-conditioning system using genetic algorithm," *Building and Environment*. Vol. 35, No. 6, pp. 471-487, 2000.
- [27] N. Nassif, S. Kajl, and R. Sabourin, "Evolutionary algorithms for multi-objective optimization in HVAC system control strategy," *Proceedings of NAFIPS*, North American Fuzzy Information Processing Society, Alberta, Canada; 2004.
- [28] N. Nassif, S. Kajl, and R. Sabourin, "Optimization of HVAC control system strategy using two-objective genetic algorithm," *HVAC&R Researc*, Vol. 11, No. 3, pp. 459-486, 2005.
- [29] M. Mossolly, K.Ghali, N.Ghaddar, "Optimal control strategy for a multi-zone air-conditioning system using a genetic algorithm," *Energy*, Vol. 34, No. 1, pp. 58-66, 2009.

- [30] L. Magnier, F. Haghghat, "Multi-objective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network," *Building and Environment*. Vol. 45, No. 3, pp. 739-746, 2010.
- [31] A. Sfetsos, "Short-term load forecasting with a hybrid clustering algorithm," *IEE Proceedings: Generation, Transmission and Distribution*, Vol. 150, No. 3, pp. 257-262, 2003.
- [32] H. Mori and A. Yuihara, "Deterministic annealing clustering for ANN-based short-term load forecasting," *IEEE Transactions on Power Systems*, Vol. 16, No. 3, pp. 545-551, 2001.
- [33] H. Mori and A. Yuihara, "Normalized RBFN with Hierarchical Deterministic Annealing Clustering for Electricity Price Forecasting," *2007 IEEE Power Engineering Society General Meeting*.
- [34] A. Kusiak, W.Y. Li, "Short-term prediction of wind power with a clustering approach," *Renewable Energy*, Vol. 35, No. 10, pp. 2362-2369, 2010.
- [35] J. Wang, "Data mining: opportunities and challenges," Hershey, PA: Idea Group Pub.; 2003.
- [36] J. Friedman, "Stochastic gradient boosting," *Stanford University, Statistics Department*, 1999.
- [37] L. Breiman, "Random forests," *Machine Learning*, Vol. 45, No. 1, pp. 5-32, 2001.
- [38] J.A. Herz, A. Krogh, and R.G. Palmer, "Introduction to the theory of neural computation," Boulder, CO: Westview Press; 1999.
- [39] T. Hastie, R. Tibshirani, and J.H. Friedman, "The elements of statistical learning," New York: Springer; 2001.
- [40] P. N. Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining," New York: Addison Wesley; 2005.
- [41] G. Casella, R. Berger, "Statistical inference, 2<sup>nd</sup> edition," Pacific Grove, CA: Duxbury Press; 1990.
- [42] G.R. Zheng and M. Zaheer-Uddin, "Optimization of thermal processes in a variable air volume HVAC system," *Energy*. Vol. 21, No. 5, pp. 407-420, 1996.
- [43] K. F. Fong, V. I. Hanby, and T.T. Chow, "HVAC system optimization for energy management by evolutionary programming," *Energy and Buildings*, Vol. 38, No. 3, pp. 220-231, 2006.
- [44] A. Kusiak and M.Y. Li, "Cooling output optimization of an air handling unit," *Applied Energy*, Vol. 87, pp. 901-909, 2010.
- [45] R. J. Dossat, "Principles of Refrigeration," Wiley; 1991.
- [46] E. Zitzler, and L. Thiele, "An evolutionary algorithm for multi-objective optimization: The strength pareto approach," *Technical Report 43, Computer Engineering and Networks*

Laboratory (TIK), Swiss Federal Institute of Technology (ETH) Zurich, Gloriastrasse 35, CH-8092 Zurich, Switzerland.

- [47] R. Kohavi and G.H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, Vol. 97, No. 1-2, pp. 273-324, 1997.
- [48] T. Nakashima, T. Morisawa, H. Ishibuchi, "Input selection in fuzzy rule-based classification systems," *Proceedings of IEEE International Conference on Fuzzy Systems*, Vol. 3, pp. 1457-1462, 1997.
- [49] G.V. Kass, "An exploratory technique for investing large quantities of categorical data," *Applied Stat*, Vol. 29, No. 2, pp. 119-27, 1987.
- [50] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, Vol. 19, No. 1, pp. 1-67, 1991.
- [51] Z. Ma and S. Wang, "Energy efficient control of variable speed pumps in complex building central air-conditioning systems", *Energy and Buildings*, Vol. 41, No. 2, pp. 197-205, 2009.
- [52] TRNSYS: A transient system simulation program, Volume 1 (reference manual). *Solar Energy Laboratory*, University of Wisconsin-Madison, WI, 1996.
- [53] L. Lu, W. Cai, L. Xie, S. Li and Y. C. Soh, "HVAC system optimization—in-building section," *Energy and Buildings*, Vol. 37, No.1, pp. 11-22, 2005.
- [54] J. Kennedy and R.C. Eberhart, "Particle swarm optimization," *Proceedings of International Conference on Neural Networks*, pp. 1942-1948, IV (Perth, Australia), Piscataway, NJ: IEEE, 1995.
- [55] M.A. Abido, "Optimal design of power system stabilizers: Using particle swarm optimization," *IEEE Transactions on Energy Conversion*, Vol. 17, No. 3, pp. 406-413, 2002.
- [56] M.P. Wachowiak, R. Smolíková, Y. Zheng, J. M. Zurada, and A.S. Elmaghraby, "An Approach to Multimodal Biomedical Image Registration Utilizing Particle Swarm Optimization," *IEEE Transactions on Evolutionary Computation*, Vol. 8, No. 3, pp. 289-301, 2004.
- [57] M. A. Abido, "Two-level of nondominated solutions approach to multiobjective particle swarm optimization," *Genetic and Evolutionary Computation Conference*, pp. 726-733, 2007.
- [58] M.A. Abido, "Multiobjective evolutionary algorithms for electric power dispatch problem," *IEEE Transactions on Evolutionary Computation*, Vol. 10, No. 3, pp. 315-329, 2006.
- [59] C. M. Fonseca and P. J. Fleming, "An overview of evolutionary algorithms in multi-objective optimization," *Evolutionary Computation*, Vol. 3, No. 1, pp. 1-16, 1995.