Improving building energy efficiency through novel hybrid models and control approaches including a data center case study

Tao Lu

Improving building energy efficiency through novel hybrid models and control approaches including a data center case study

Tao Lu

Aalto University
School of Engineering
Department of Civil Engineering

Supervising professor
Professor Jari Puttonen

Thesis advisor
Professor Xiaoshu Lü

Preliminary examiners
Professor Shao Li, University of Reading, UK
Professor Carey Simonson, University of Saskatchewan, Canada

Opponent
Professor Xiulan Huai, Institute of Engineering Thermophysics, Chinese Academy of Science, P.R.China

NORDIC ECOLABEL

441   697
Printed matter

Author
Tao Lu

Name of the doctoral dissertation
Improving building energy efficiency through novel hybrid models and control approaches including a data center case study

Publisher School of Engineering

Unit Department of Civil Engineering

Series Aalto University publication series DOCTORAL DISSERTATIONS 182/2016

Field of research Building Physics

Manuscript submitted 17 March 2016   **Date of the defence** 7 October 2016

Permission to publish granted (date) 8 August 2016   **Language** English

☐ **Monograph**   ☒ **Article dissertation**   ☐ **Essay dissertation**

Abstract

The building sector consumes the most energy and emits the greatest quantity of greenhouse gases of any sector. Energy savings in this sector can make a major contribution to tackling the threat of climate change. Research has produced a variety of solutions, for example, net-zero and net-positive energy buildings. Consequently, both models and controls are being challenged by increasingly complex buildings equipped with advanced information and communications technologies (ICT).

This dissertation addresses these challenges by proposing a multidisciplinary, wide-ranging modeling methodology that enables new strategies for saving building energy. The core methodology utilizes novel modeling approaches to improve predictive models and produce innovative energy solutions. Models are validated and investigated using a variety of buildings and controls. Data centers and demand-controlled ventilation (DCV) are the focus because they represent both "multifunctional buildings" and general energy system controls. This dissertation makes the following seven original contributions: (1) The first systematic, complete case study of a data center in which infrastructure, energy and air management performance, and waste heat recovery systems were investigated, analyzed, and quantified using long-term power consumption data. (2) A novel and tuning-free DCV building control strategy that is far superior to proportional control and more competitive than proportional-integral-derivative (PID) control. (3) An artificial neural network (ANN) model for predicting the water evaporation rate in a swimming hall. (4) A new ANN model for estimating prediction intervals and accounts for the uncertainty of point estimation for indoor conditions in an office building. (5) A new Maximum Likelihood Estimation (MLE) model for predicting constant and time-varying air change rates and a coupled model for estimating the number of occupants in an office. (6) Discovery of a new physical law for run-around heat recovery systems that can be used to develop a simulation model to estimate the system performance for constant volume air (CAV) and DCV systems. This new law was verified in different sites. (7) A new hybrid numerical-ANN model for building performance simulation. The hybrid model can improve not only the model accuracy but also the generalizability of ANN.

The results demonstrate the applicability of the modeling techniques and the models, and significant energy savings in buildings. The resulting improvements in model accuracy, forecasting capability, and energy efficiency were published in eight journals. By unifying the results of eight publications, this dissertation presents a comprehensive and coherent study that advances the state-of-the-art building energy research.

Preface

This Ph.D. thesis contains the results of research undertaken at the Department of Civil and Structural Engineering, Aalto University, during the years 2012 - 2014. The research work was funded by the Academy of Finland, the National Technology Agency of Finland (TEKES) and industry. I would like to gratefully acknowledge all of the financial supporters. Without their support, the research work would not have been possible.

I wish to thank my former supervisor, Professor Martti Viljanen, for his invaluable advice and support throughout the study. When I came to the department, I was just a professional computer programmer without any background in building physics. Changing to a totally new research field is really challenging. But it becomes even more challenging to identify novel topics for research in an unfamiliar field. However, his clear vision and guidance helped me a lot in the transformation. Unfortunately, he passed away after fighting an illness with bravery and courage and without seeing the completion of my dissertation.

I wish to thank my current supervisor, Professor Jari Puttonen, for his supervision and guiding me for the last few miles of the dissertation, which is much appreciated.

I wish to thank all the co-authors for cooperation. Special thanks go to Professor Xiaoshu Lü for her careful reading of my articles and her penetrating comments, Professor Charles Kibert from University of Florida, who with patience read, commented, re-read and re-commented manuscripts until finalized, and Professor Jukka Manner at the Department of Communications and Networking, who enabled me to continue my research work on data center. In addition, I would also like to thank all my colleagues for their constructive discussions, assist and pleasant working atmosphere.

I would like to dedicate this dissertation to the memory of my father, who inspired scientific thinking in my life. I can't express my gratitude for my mother in words, whose unconditional love has been my greatest strength.

Last, but not least, I would like to thank my wife for her understanding and standing beside me throughout my career. The cover image by Emma and Sonja is appreciated.

Espoo, August 2016

Tao Lu

Contents

List of the important original publications

This dissertation consists of an overview of the following publications, which are referred to in the text by their Roman numerals. The publications are presented in their published format in the following order at the end of the doctoral dissertation.

I. Lu, T., Lü, X., Remes, M. & Viljanen, M. (2011). Investigation of air management and energy performance in a data center in Finland: Case study. *Energy and Buildings,* 43(12), 3360–3372.

II. Lu, T., Lü, X. & Viljanen, M. (2014). Prediction of water evaporation rate for indoor swimming hall using neural networks. *Energy and Buildings* 81, 268-280.

III. Lu, T., Lü, X. & Viljanen, M. (2011). A novel and dynamic demand-controlled ventilation strategy for CO_2 control and energy saving in buildings. *Energy and Buildings* 43(9), 2499-2508.

IV. Lü, X., Lu, T., Viljanen, M. & Kibert, C. (2013). A new method for controlling CO_2 in buildings with unscheduled opening hours. *Energy and Buildings* 59, 161-170.

V. Lu, T., Lü, X. & Kibert, C. (2015). A hybrid numerical-neural-network model for building simulation: A case study for the simulation of unheated and uncooled indoor temperature. *Energy and Buildings* 86, 723-734.

VI. Lu, T. & Viljanen, M. (2009). Prediction of indoor temperature and relative humidity using neural network models: model comparison. *Neural Computing and Applications* 18(4), 345-357.

VII. Lu, T., Knuutila, A., Viljanen, M. & Lü, X. (2010). A novel methodology for estimating space air change rates and occupant CO_2 generation rates from measurements in mechanically-ventilated buildings. *Building and Environment* 45(5), 1161–1172.

VIII. Lu, T., Lü, X., Kibert, C. & Puttonen, J. (2016). The application of linear regression and the power law relationship of air-side heat transfer with field

measurements to model the performance of run-around heat recovery systems. *Energy and Buildings* 110, 453-467.

Author's contribution

The thesis author, Tao Lu, is the principal author of eight publications I-VIII (and the corresponding author of all publications). All ideas, simulations and analysis were carried out by Tao Lu. In (IV), the author contributed the whole idea, analysis work and a big part of simulation work. The other authors in articles I-VIII, Professor Xiaoshu Lü and Professor Martti Viljanen have acted as the supervisors of the research projects and supplemented the articles of their experiences of building physics and building simulations. Professor Charles Kibert and Professor Jari Puttonen provided comments and advice on the articles and manuscript preparation. The experiment works in (I) and (VII) were carried out by the author and co-authors Matias Remes and Anssi Knuutila.

Abbreviations

ANN	artificial neural network
APE	absolute percentage error
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
CAV	constant air volume
CFD	computational fluid dynamics
CO_2	carbon dioxide
CRAC	computer room air conditioning
DCV	demand-controlled ventilation
EU	European Union
HACA	hot aisle/cold aisle
HCHO	formaldehyde
HVAC	heating, ventilating, and air conditioning
IAQ	indoor air quality
IR	infrared
MAE	the mean of absolute errors
MAPE	the mean of absolute percentage errors
MLE	Maximum Likelihood Estimation
MSE	the mean of the sum of square errors
NNARX	nonlinear neural network autoregressive with exogenous inputs
Nu	Nusselt number
PI	proportional-integral
PID	proportional-integral-derivative
Pr	Prandtl number
PUE	power use effectiveness
Re	Reynolds number
RH	relative humidity
ROC	rate of change
RTI	return temperature index
SHI	supply heat index
UPS	uninterruptible power supply
VAV	variable air volume
VFD	variable frequency drive

Nomenclature

A	area of pool surface (m^2)
b, B	bias
c	specific heat capacity of the air ($J/kg.^oC$)
C	CO_2 concentration (ppm)
e	error
f, F	function
G	the amount of CO_2 emission per person (L/h).
h	heat transfer coefficients ($W/m^2.^oC$)
I	air change rate (1/s; 1/min; 1/h)
k	delay
K_d	the differential parameter (derivative term)
K_i	the integral parameter
K_p	the proportional parameter
$\dot{m}$	moisture generation rate (g/s)
n	sample number
n_a, n_b	order
P	pressure (Pa), number of occupants, cooling power (W)
P_a	saturation pressure at room air dew point (Pa)
P_w	saturation vapor pressure taken at surface water temperature (Pa)
$Q, V, \dot{V}$	volumetric airflow rate (m^3/s)
$\dot{Q}$	heat gain (W)
RH	relative humidity (%)
t	time (s; min; hr.)
T	temperature (^oC; K)
u	input
V	air velocity (m/s)
V, Vol	volume (m^3)
V_{DCV}	outdoor air ventilation rate (m^3/s)
$V_{ot-design}$	design ventilation rate (m^3/s)
V_{ot-min}	base ventilation rate (m^3/s)
w, W	ANN weight
W	moisture content (g/m^3)
y	output
$\hat{y}$	predicted output

Subscripts

ea	exhaust air after passing through the exhaust air heat exchanger
eb	exhaust air before passing through the exhaust air heat exchanger

enve	envelops
eq	the equilibrium value
exf	exfiltration
exh	exhaust
i	indoor
i	rack inlet
increment	temperature increment across the supply air heat exchanger
int	internal
max	maximum temperature difference between exhaust and supply airstreams
o	outdoor, rack outlet
R	at the breaching level in the space
RA, ret	return
sa	supply air after passing through the supply air heat exchanger
sb	supply air before passing through the supply air heat exchanger
si	surface
sol	solar
sou	sources
sup	supply, supply air
v	the adjacent plenum vent

Greek letters

α	signature power
η	system efficiency
Δ	difference
ρ	density of the air (kg/m^3)

List of figures

Figure 18. Post-regression results of predicted water evaporation rates (output) and measurements (target). (a) Neural network B (one-step ahead prediction). (b) Neural network A (one-step ahead prediction). (c) Neural network B (two-step ahead prediction). (d) Neural network A (two-step ahead prediction) (Original publication II).

Figure 19. The estimated and recorded numbers of occupants vs. measured indoor CO_2 concentrations on: (a) 22.9.2008, (b) 23.9.2008 and (c) 24.9.2008 (Original publication VII).

Figure 20. Supply air temperature efficiency curves for Site B (airflow rate range: 20% to 100% of design air flow rate) (a) α (signature power) = 0.3. (b) α (signature power) = 0.39.

Figure 21. Comparisons of measured indoor temperature and simulated temperatures by the testing numerical model, the numerical model of Hybrid model A and Hybrid model B, and Hybrid model A (Original publication V).

List of tables

1. Introduction

1.1 Background - energy efficient buildings and data centers

Buildings are the fastest growing energy consuming sector and, as a result, they have a significant impact on the environment. In the European Union (EU), 40% of total energy consumption and 36% of total CO_2 emissions are produced as a result of building construction and operation (European Commission 2013). Furthermore, growing demands for information technology (IT) continue to accelerate the growth of data centers and their energy use, which makes data centers the fastest-growing energy-intensive sector of all contributors in the world. More than 3% of US electricity and approximately 2% of global electricity, which has an annual growth rate of 12%, are attributed to data centers (Koomey 2011). In addition, the demand for thermal comfort and indoor air quality (IAQ) continues to increase. It is well known that people spend 80–90% of their life time indoors. The indoor air pollution levels can be much higher than outdoor levels, which could cause health problems for the occupants (Koomey 2011, Redlich et al. 1997, Akimenko et al. 1986). Therefore, efforts have been focused on energy efficient buildings that can provide a comfortable indoor environment with minimum energy cost. Currently, many new types of buildings, for example, sustainable and net zero energy buildings, which refer to a type of energy efficient building with minimum adverse impacts on the built and natural environment, are being promoted (John et al. 2005).

Scientific research on energy efficient buildings covers the areas of modeling and simulation, construction, retrofitting, and ongoing operations of buildings. The topics of retrofitting and ongoing operations are attracting the attention of the research community because they can potentially help existing buildings operate in an energy-efficient manner. An efficient control strategy and accurate building simulation are two crucial factors for successful retrofitting and operating energy efficient buildings. With efficient control strategies buildings can be operated intelligently and simulation can provide a realistic representation of the building's responses to retrofitting and optimized operation.

1.2 Scope and structure of the thesis

This thesis concentrates on the energy efficiency of large public and energy-intensive buildings, such as office buildings, sports halls, swimming pools, ice rinks and data centers. These buildings represent a great potential for energy

saving, while at the same time providing a diverse and functional complexity that makes them extremely difficult to control and operate. Interpreted as "multifunctional buildings", data centers accommodate multiple functional networks and multiscale energy infrastructure systems, for example complex air patterns, that operate across diverse domains (Barroso et al. 2009). Data centers, therefore, were chosen as a special case study that offers a wide range of building properties and allows to engage with a variety of application areas and a careful assessment of general building performance. Furthermore, demand-controlled ventilation (DCV) was especially selected as a representative for innovative control approaches because both important perspectives, energy and IAQ, are covered in DCV applications. Figure 1 schematically presents the scope and structure of this thesis.
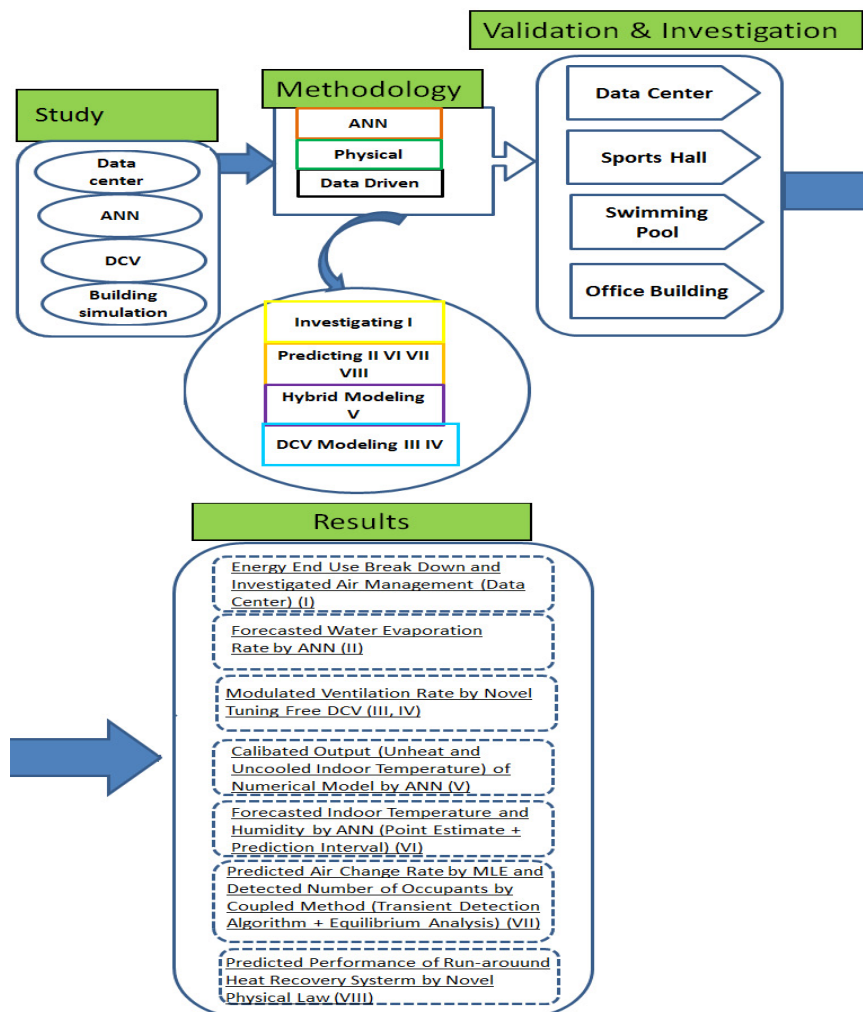
Figure 1. Schematic structure of the dissertation work, issues, and technologies related to the topic (roman numbers I-VIII indicate published journal papers contributing to the dissertation work).

- Finland has a cold climate. July temperatures in Finland average 13 to 17°C. February is usually Finland's coldest month, with temperatures averaging from - 22 to -3°C. Thus, energy consumption, particularly for space heating, is always highlighted in the building sector. According to the national statistics of Finland, in 2013 heating of residential buildings consumed 87% of total household energy (Statistics Finland 2014).
- Finland has the highest percentage of mechanical supply and exhaust ventilation and air conditioning for buildings in the EU (Kurnitski and Seppänen 2008). Before 1980, all EU countries used mainly natural ventilation. Finland was one of the first countries, i.e. before 1959, to introduce mechanical supply and/or exhaust ventilation systems. Finnish buildings constructed after 2004 have only mechanical supply and exhaust ventilation systems (Litiu 2012). Nowadays, Finland has more than half of the houses and almost half of the apartments equipped with mechanical supply and exhaust ventilation, which is the highest in the EU (Litiu 2012). And Finland has an advanced indoor climate classification in the developed world (Kurnitski and Seppänen 2008).
- However, energy performance is seldom optimized for buildings in Finland (Kurnitski and Seppänen 2008). The majority of air conditioning and ventilation systems are still constant air volume (CAV) systems. Demand-controlled ventilation is not commonly used, with the exception of meeting and conference rooms (Kurnitski and Seppänen 2008).
- With one of Europe's lowest energy prices, cold weather, and plenty of water sources, Finland has become globally attractive for data centers (Oxford research 2014).). The exponential increase in data centers world-wide is also fueling a data center boom in the Nordic countries, especially Sweden, e.g. Facebook built the first data center outside the USA in 2011 in Luleå , Sweden, and in 2014 decided to build a second data center in Luleå. Inevitably, Finland is facing increasingly global competition in the data center market. However, research on data centers is scarce in Finland.

The dissertation work consists of four broad components (Figure 1). The Study includes studies of important issues and technologies related to energy efficient buildings. These issues and technologies are presented in literature review (Ch. 2). The Methodology focuses on the development of novel modeling techniques and hybrid models to address current challenges faced in building energy research. Further validation and investigation into the detailed behavior of the models are provided in Validation and Investigation. Finally, the resulting improvements in model accuracy, model forecasting capability, and DCV control algorithm are described in Results, along with all the papers (I, II, III, IV, V, VI, VII, and VIII). In this way, the original eight publications are fully homogenized as a new package presented in this dissertation.

2. Literature review

To make it easy to read and understand, literature review is presented based on the research themes with headings. The relevant methodology is described separately after this chapter.

2.1 Data center

A data center (or datacom facility) is a facility housing high-performance computers, storage servers, computer servers, networking or other IT equipment. In typical data centers, IT equipment converts over 99% of its power into heat, leading to over 70% of the total heat load which needs to be removed in order to ensure acceptable environments for both indoor and the equipment. Therefore, cooling accounts for a large portion of those energy costs, consuming 25% or more of the total power in data centers (Kant 2009). Increasing the efficiency of the cooling system is the key to energy efficient data centers. In addition, the huge amount of removed heat from a data center could be reused for other purposes, such as heating the remainder of the building, supplying hot water, or even selling in the energy markets. These facts have brought data centers into focus in the research on energy efficient buildings.

   In general, cooling systems in data centers can be classified into two categories: forced-air cooling and liquid cooling. Air cooling is still the dominant approach used for data center cooling systems. In this technique, heat is removed by pushing cold air through racks, containing IT equipment. The aim is to keep the rack (IT equipment) inlet temperature and humidity within an acceptable range for reliable operation of equipment in data centers (Schmidt et al. 2005, ASHRAE 2009a). The complexity of a rather high internal heat dissipation with cold and warm outdoor environments requires a smart solution, capable of providing acceptable indoor climate management (Karlsson and Moshfegh 2005). IBM introduced the Hot Aisle/Cold Aisle (HACA) protocol for air cooling in 1992 (ASHRAE 2009a). This protocol became the most popular cooling technique to date for data centers. The majority of modern data centers are still using HACA. Figure 2 depicts a typical data center with Hot Aisle/Cold Aisle layout.
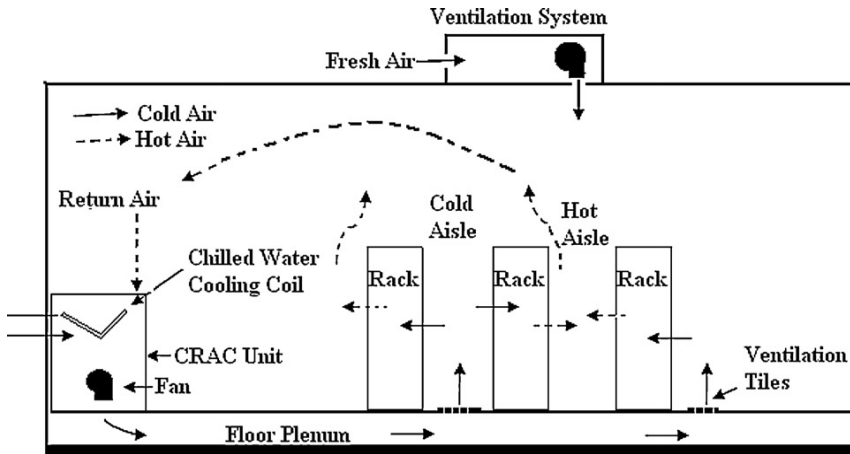
Figure 2. The cooling infrastructure with Hot Aisle/Cold Aisle arrangement (Original publication I).

The computer room air-conditioning unit (CRAC) cools the exhaust hot air from racks and pushes the chilled-air supply air (i.e. cold air) into the floor plenum. Cold air enters the cold aisle through perforated floor tiles (i.e. ventilation tiles) to cool IT equipment in the racks. Exhaust air from the racks enters the hot aisle and finally migrates back to CRACs. A chilled water system is often employed to back the above structure in cooling (Figure 3).
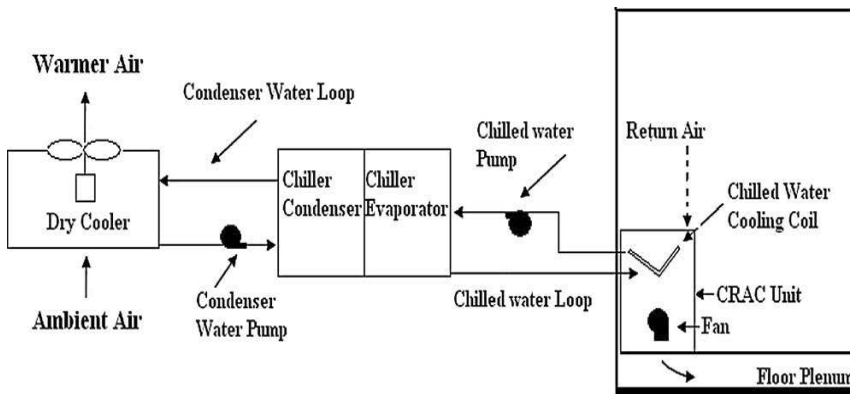
Figure 3. Chilled water system (Original publication I).

Most chilled water systems have economizers (air side or water side) installed for free cooling (ASHRAE 2009). Free cooling utilizes low temperature natural resources (e.g. outdoor air, water) to minimize costly compressor use whenever these natural resources are cool enough. Zhang et al. (2014) gives a detailed literature review on free cooling in data centers showing that free cooling generally can save 30% to 50% energy depending on geographical locations and free cooling types. This result was based on a study of 22 data center systems.

Some other alternative cooling solutions include In Row Cooling with Hot Aisle Containment (Rasmussen (2008), Cold Aisle Containment (Emerson Electric Co. 2008), Overhead Cooling (Emerson Electric Co. 2005), Kool-IT

16

(Fulton 2007) and Evaporative Cooling (e.g. Cool3 IDEC Unit by Nortek, http://datacenters.nortekair.com/Products/Cool3.aspx). However, with the growth of heat dissipation in data centers, many problems have occurred for the use of air cooling and HACA, such as hot spots and oversized cooling equipment. A study, based on more than 15,000 individual measurements in 19 computer rooms ranging in size from 230 m² to 2,400 m² and totaling 19,000 m² of raised floor, found 10% of racks had air intake conditions outside the environmental parameters recommended by hardware manufacturers for maximum reliability and performance (Sullivan et al. 2007). The percentage of racks with hot spots will increase beyond this value because data center rack power densities increase to an average of 5 kW/rack or greater (Lin 2014). Practices that reduce cold air leakage, bypass, and hot air recirculation can help to eliminate hot spots (Lin 2014). An efficient solution to eliminate hot spots and improve energy efficiency is liquid cooling. It is considered to be more efficient since liquid can carry much more heat than air. Typically, the closer to the IT rack that heat can be transferred to a liquid loop, the greater the efficiency is (Geng 2015). In fact, liquid cooling is not new in data centers, but its acceptance has been and continues to be difficult (Rakesh 2009).

The performance of a cooling system highly depends on geographical location. There is no "best" cooling solution for data centers. For example, direct evaporative cooling uses low-cost fans to draw outdoor air through a falling stream of water. Incoming air causes evaporation of the water, and the resulting cooler air is then blown into the data center (Longbottom 2012). Direct evaporative cooling eliminates costly compressors, improves energy efficiency, and lowers both capital and operating costs. However, a recent report (Manousakis et al. 2016) shows that high humidity is a greater threat to hard drive reliability than the high temperatures for free-cooled (direct evaporative cooling) data centers. The relative humidity of a free-cooled data center can sometimes be as high as 90%. The mechanical systems of the data center then must remove much of humidity, which is costly. When the humidity of the incoming air cannot increase further, the direct evaporative cooling will cease. Therefore, a hybrid system that includes various data center cooling techniques may provide substantial energy savings (Longbottom 2012).

In order to evaluate air cooling performance in data centers, the Green Grid released the first version of the power use effectiveness metric (PUE) in 2007 (The Green Grid 2007). PUE is the ratio of total amount of energy used by a computer data center facility to the energy delivered to computing equipment. Nowadays PUE is the most important metric for the efficiency of the data center infrastructure. Based on a survey for 1000 data centers (46% from the U.S. and Canada, 21% from Europe, 21% from Asia Pacific, 12% from Latin America, 6% from Africa and Middle East and 2% from Russia and Commonwealth of Independent States) conducted by Uptime Institute, average PUE was 1.7 in 2014, and from 2011 to 2014 the average self-reported PUE has only improved from 1.89 to 1.7 (Uptime 2014). In 2015, Allied Control claimed PUE ratio of 1.02 through the use of Two-Phase Immersion Cooling (3M 2015), which is the lowest by far. Other metrics, such as the Return Temperature Index (RTI)

(Herrlin 2007), and the Supply Heat Index (SHI) (Bash et al. 2003), are also used sometimes.

The case study approach dominates research on data centers because it offers direct observation of air movement and energy performance due to the complex multilevel and multiscale nature of data centers. Karlsson et al. (2005) surveyed the power consumption and indoor climate in a data center by using an IR camera for visualization of airflow and temperature patterns, aiming at identifying problems such as improper usage of the chilled air. They found (1) the power requirement within the data center was considerably high due to an oversized air conditioning system; (2) cool air did not reach the upper levels of the racks, despite a very high air exchange rate; (3) point measurements of temperatures in a rack showed that recirculation cells were present, causing accumulation of heat and improper cooling of electronic equipment. Their conclusion was that chilled air was not distributed properly and consequently the cooling energy was not used effectively. Sun et al. (2006) examined energy use of two data centers in commercial office buildings in Singapore. Results showed that data centers are high energy consuming areas in commercial office buildings—energy consumptions of approximately 3000 kWh/ (m^2 year) and 2000 kWh/ (m^2 year), respectively, were observed in the case studies. Power demands were found to be often grossly over-provided in these facilities. Potentially they estimated that in one data center approximately 56% (1.2 GWh/year) of energy consumption could be conserved through sizing and reconfiguring the equipment (i.e. HVAC and UPS) to meet the actual loads, optimizing space utilization, zoning control of HVAC and lighting, raising room temperature (e.g. from 21 ^{o}C to 25 ^{o}C), and adjusting operating schedule of lighting, leading to more than US$ 80,000 saving per year.

Some studies use a numerical model to simulate air movement and energy performance in data centers. Karki et al. (2006) simulated airflow distribution through perforated tiles in raised-floor data centers using an idealized one-dimensional computational model, in which the airflow distribution is governed by two dimensionless parameters: one related to the pressure variation in the plenum and the other to the frictional resistance. They also used a one-dimensional model to calculate flow rates for two possible arrangements of the CRAC units, and these results were compared with those given by a three-dimensional model. Cho et al. (2009) studied the design parameters and IT environmental aspects of the cooling system with a high heat density data center. They applied CFD simulation analysis to compare the heat removal efficiencies of various air distribution systems. CFD results were validated by measurements of the IT environment of an actual operating data center. A method for planning and design of an appropriate air distribution system was proposed. Based on previous works, they also proposed an energy efficiency and low carbon IT framework for data centers considering green metrics. Some research concentrates on making design guides and polices for energy efficient and green data centers. These works can be found in Uddin et al. (2012).

2.2 Artificial neural network (ANN) in buildings

Modeling and control strategies often aim to optimize operations of buildings, in which sometimes require a priori estimation of the performance of building systems and the flow of energy is needed. In many applications, however, physical and analytical modeling on performance estimation is not always feasible. Then data-driven offers an alternative approach.

Data-driven model vs. physical model

Physical model is a model which expresses the systems as a series of mathematical equations based on physical laws. It is also called numerical model if it is solved numerically. Physical models are often limited to simple building systems. Buildings, especially large or complex ones, are inherently complex and nonlinear because of the multiple interconnections among their diverse systems. Simplifications of the model equations and lack of knowledge of the physical mechanisms may lead to a lack of precision or incorrect results. In contrast to physical models, the main concept of data-driven model is to apply data analysis to find relationships between the system state variables (input and output) without explicit knowledge of the physical behavior of the system (Solomatine et al. 2008). It is to predict, not explain, the system. Samples of data-driven model are regression (identification of a mathematical expression or equation that models the data with the least error, e.g. linear regression, maximum likelihood estimation) and machine learning.

Machine learning theory is related to pattern recognition and statistical inference wherein a model is capable of learning to improve its performance of a task on the basis of its own previous experience (Mjolsness and DeCoste 2001). ANN is one of the most popular machine learning methods. ANNs are able to learn the key information patterns within a multi-dimensional information domain (Kalogirou 2000). In addition, ANN is fast, fault-tolerant (the tolerance for imprecision and uncertainty), robust, and noise-immune (Rumelhart et al. 1986).

Lee and Ho (1994) suggested the robustness to noise contamination in inputs and/or weights for ANNs can be explained in two ways. First, the orthogonal property among the output values of the hidden nodes reduces the noise effect, i.e. the hidden weights tend to be near orthogonal through learning procedure for efficient feature extraction of input patterns (Xue et al. 1990). Thus, after successful learning, the weighted sums to hidden nodes are much less correlated even when a pattern with correlated noise is presented to the input layer. In addition, magnitude of correlation coefficient between the weighted sums decreases under the sigmoidal transformations (Oh and Lee 1994). Therefore, the correlations among hidden nodes should be very small. As a result, the noise effects are averaged out when the hidden output values are summed through output weights. Second, noise immunity of ANNs can be explained in the information theoretic point of view (Abu-Mostafa 1989). It was reported that ANNs have hierarchical information extraction capabilities acquired through learning (Lee and Song 1993). It was argued that the input pattern set has the

inter-class information as well as the intra-class information. The inter-class information is the information content that an input pattern belongs to a specific class, and the intra-class variations is a measure of the average variations within the classes including noise contaminations. After learning, each layer of ANNs tries to keep the inter-class information and to minimize the intra-class variations as much as possible. In order to investigate the noise immunity of ANN, Derakhshani and Schuckers (2004) conducted the onset detection test of a distorted waveform using a continuous time delay neural network. The distorted waveform was a combination of a distorted square wave (50% duty cycle, zero phase, unit amplitude, frequency of 1 Hz, and an assigned target of D(t)=0) and a distorted sine wave (switched at t=4 sec, zero phase, unit amplitude, frequency of 1 Hz, and an assigned target of D(t)=1). A band limited noise signal (power=0.0015, sample time=0.04 sec) was constantly being added to the composite waveform. The neural network showed a very good noise immunity and robust behavior by correctly detecting both the noisy square wave (output=0) and then the following noisy sine wave (output=1).

Data from building systems, being inherently noisy, are suitable to be handled with ANNs. Therefore, ANNs are widely applied for forecast and prediction.

2.2.1 The structure of ANNs

ANNs may be defined as structures comprised of densely interconnected adaptive simple processing elements (called artificial neurons or nodes) that are capable of performing massively parallel computations for data processing and knowledge representation (Basheer and Hajmeer 2000). They are organized in layers which are made up of a number of nodes (i.e. neurons). A node contains an activation function, which takes input data and performs simple operations on the data, and selectively passes the results on to other nodes. ANN usually consists of (1) an input layer, which accepts variables from the environment, (2) an output layer, which shows response with regard to environmental inputs, and (3) some hidden layers, which do not directly interact with the environment, but have the primary function of relating the input to the output. Each single node is connected to other nodes of a previous layer through weights. Weights associated with individual nodes are also known as biases. Figure 4 illustrates a three-layer neural network with two inputs (input layer), two hidden units (i.e. nodes, hidden layer) and one output (output layer), and its mathematical formula is (II):
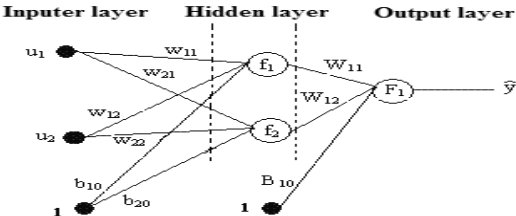
Figure 4. Three-layer feedforward neural network with two inputs (input layer), two hidden units (hidden layer) and one output (output layer).

$$\hat{y}(k) = F_1\left[\sum_{j=1}^{n_h} W_{1j} \, f_j\left(\sum_{i=1}^{n_u} w_{ji} \, u_i + b_{j0}\right) + B_{10}\right] \qquad (1)$$

where n_h is the number of hidden units; n_u the number of inputs; k sampling number; F_1 and f_j activation functions. The knowledge is acquired by ANN through a learning process, and interneuron connections (i.e. weights) are used to store the knowledge. Like the learning process of a human brain, ANN operates like a "black box" model, and does not require detailed information about the system (Kalogirou 2000). The training process of an ANN involves adjusting weights until the predicted outputs are in close agreement with the actual outputs. Because of this, ANN is good for tasks involving incomplete data sets, fuzzy or incomplete information, and for highly complex and ill-defined problems. In addition, ANN is able to handle large and complex systems with many interrelated parameters (Kalogirou 2000). They seem to simply ignore excess data that are of minimal significance, and concentrate instead on the more important inputs. There are many types of ANNs, but the most common ANN is the three-layer feed-forward neural network (one input layer, one hidden layer and one output layer, Figure 4). It has been established that a three-layer feed-forward neural network can approximate any function of interest provided that a sufficient number of neurons are used (Kalogirou 2000).

2.2.2 Predictive modeling

The most frequently used application of ANN is forecast and prediction, i.e. predictive modeling. Predictive modeling leverages statistics to predict outcomes (Geisser 1993). It can be applied to any type of unknown event, regardless of when it occurred. Forecast includes training of an ANN on samples and then using the trained ANN to predict (forecast) the behavior at subsequent times from one or more previously known historical observations. Prediction is to use the real data to train ANNs to better match the real behavior of building systems. Hence, unlike forecast, prediction estimates not only future value but also historical value. The purpose of forecast and prediction by ANN is to provide prior knowledge on building parameters, which are factors with influences on indoor environment and building energy performance, to optimize the operations of building systems.

Indoor thermal conditions depend on indoor temperature and relative humidity values. The forecasting and prediction of indoor temperature and humidity become very important because they can be used to optimize cooling and heating systems. Sigumonrong et al. (2001) developed an intelligent air handler for controlling the indoor temperature and also limiting indoor relative humidity in an industry plant. The intelligent air handler is comprised of two ANNs: an emulator and a controller. The emulator was first developed to predict future indoor temperature and humidity of the plant using the historical data set of indoor temperature and humidity as well as the current input to operate the chilled water valve. The controller was then constructed to determine the current input to open the chilled water valve by using the desired future

temperature, the current relative humidity and the historical data set of them. The aim was to regulate both the room temperature and relative humidity via the control of the supply air temperature through the chilled water flow. The details about the method or algorithm of ANN for predicting indoor humidity was not given. Similar work was done by Zhang et al. (2005). Concerning indoor temperature prediction, many publications exist. Ruano et al. (2006) applied Radial basis function (RBF) neural networks to predict indoor temperature for a school building using the current and historical data set of indoor air temperature, outside solar radiation, outside air temperature and outside relative humidity. A genetic algorithm was also employed for searching optimal structure for neural networks. Results showed that a neural network model can achieve better results than state-of-the-art physical models. Thomas and Soleimani-Mosheni (2007) used feedforward neural networks to predict indoor temperature for a small building and a factory building. In the small building, they used the current indoor temperature and indoor humidity as inputs. The current solar-air temperature, radiator power and time of day were used as optional inputs. The results showed that there were no significant differences between the different input combinations although the best testing results came from the neural network with the current indoor temperature, indoor humidity, solar-air temperature, radiator power and time of day as optional inputs. In the factory building, they used average indoor temperature, total internal heat power, temperature of water flowing in pipes behind one wall, temperature of water flowing in pipes behind the other wall and ventilation flow rate as inputs. Other works can be found in Mechaqrane and Zouak (2003), Ferreira and Ruano (2001), Gouda et al. (2002), Frausto and Peters (2004) and Ferreira et al. (2002).

ANN is also applied in other building predictions. Nassif (2012) proposed an intelligent energy management and control system (EMCS) for heating, ventilating, and air conditioning (HVAC). The EMCS collected the measured data (real data) from components or subsystems to train ANN models to predict fan power, cooling load for the cooling coil and the compressor power for the chiller. Yang et al. (2003) applied ANN to determine the optimal start time of the heating system. They took not only room air temperature and outdoor air temperature but also their variation rates as inputs in order to predict future variations. The output is the time elapsed from the start of the HVAC equipment to the time when the desired temperature is reached. Sofuoglu (2008) used six pollutants (CO_2, particulate matter 2.5 micrometers (PM2.5), formaldehyde (HCHO), total volatile organic compound (TVOCs), bacteria, fungi) and four comfort variables (temperature, RH, light and noise levels) as input variables to predict an index of occupant symptom prevalence (POPS2). Building energy consumption prediction is another hot topic concerning ANN. The details can be found in (Kalogirou, 2000). Other ANN applications include modeling of solar domestic water heating (SDHW) systems (Kalogirou et al. 1999), estimating heating loads of the building (Kalogirou et al. 1997), modeling storage-heaters of the room (Roberge et al. 1997), optimizing the energy

consumption of a commercial-scale HVAC system (Curtiss et al. 1995) and others (Kalogirou, 2000).

2.3   Demand-controlled ventilation (DCV)

Demand-controlled ventilation is a control strategy used to automatically modulate the volume of fresh air taken into a building or other occupied space by mechanical air conditioning equipment to meet occupant demand. Ventilation for DCV is often adjusted based on some building parameters (Figure 5). For instance, humidity based DCV can be set to alter ventilation rate based on whether the absolute humidity difference between the exhaust air and outdoor air is below threshold value (Nielsen et al. 2010).

   Because a data center is unoccupied at most time, electronic equipment becomes a major concern. Therefore, a DCV strategy can be applied to optimize cold airflow rate based on rack inlet temperature and humidity, for example, in a chilled-air cooled data center (Figure 1), the speed of a CRAC fan can be optimized to keep rack inlet temperature and humidity (IT equipment) close to maximum allowable operation temperature and humidity for maximum energy saving.
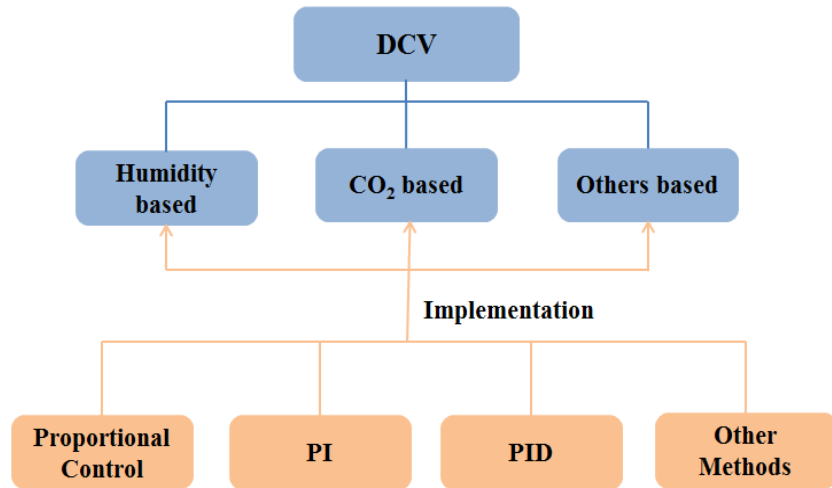
Figure 5. Graphical illustration of demand-controlled ventilation.

One of the most important control strategies is CO_2 based DCV. In this strategy, calculations of the fresh outside air (m³/s per person) provided to a space are related to indoor CO_2 level. Indoor CO_2 level is generally used as an indicator of the number of occupants and CO_2 itself is not considered as a dangerous contaminant. By doing this, the outdoor air supply rate per person recommended by the industrial standards, such as ASHRAE 62-2016 (2016), can be met for better IAQ.

   Control algorithms for CO_2-based DCV have been researched for a long time, among which proportional and exponential controls are the most popular ones.

Proportional control modulates ventilation proportionally between a lower set point of indoor CO_2 and an upper set point that represents the equilibrium concentration of CO_2 corresponding to the target per-person ventilation rate of a space (Schell et al. 1998). Exponential control is often implemented using a standard proportional-plus-integral (PI) or proportional integral- derivative (PID) control algorithm (Schell et al. 1998). In PI control, the longer the indoor CO_2 concentration varies from the set point (e.g. 1000 ppm), the more the controller output will change in order to drive the indoor CO_2 closer to the set point (Wang 2010). PID control adds a derivative term to PI to sense CO_2 approaching to or departing from the set point, thus reducing overshoot for certain conditions (CIBSE Guide 2000). Hence, the indoor CO_2 concentration can be maintained much closer to the set point with PI or PID control than with proportional control, resulting in great energy savings.

In industry standards or building codes (e.g. ASHRAE 62-2010 2010), the minimum requirement for the outdoor air ventilation rate is suggested as:

$$DVR = R_p P + BASE\_R \qquad (2)$$

where DVR is the demanded (minimum) ventilation flow rate, R_p is the fresh air requirement per person, and $BASE\_R$ is the base ventilation for unoccupied hours. $BASE\_R$ is applied to dilute non-occupant- generated pollutants. CO_2-based DCV strategy keeps ventilation rate close to minimum requirement (i.e. Eq. (2)) and maintain indoor CO_2 concentration near the set point.

The fundamental concept of proportional control is the utilization of the CO_2 equilibrium value, which can be expressed as:

$$C_{eq} = \frac{G \cdot P}{Q} \qquad (3)$$

where Q is volumetric airflow rate (fresh air, m³/s) into (and out of), C_o outdoor CO_2 concentration (ppm), C_{eq} the equilibrium CO_2 concentration, P the number of occupants in the space, and G the amount of CO_2 emission per person (L/h). The equilibrium value is the maximum value of indoor CO_2 can possibly reach when C_o, Q and $G \cdot P$ are kept constant. In proportional control, C_{eq} is often taken from a preset CO_2 level, also called the CO_2 set point. The indoor CO_2 concentration should be kept below C_{eq}. $G \cdot P$ is the design CO_2 generation rate, namely the generation rate per person multiplied by the maximum number of occupants. The resultant Q from Eq. (3) is the required design ventilation rate for the space. Therefore, a proportional control can be possibly implemented like (ASHRAE 2007):

$$V_{DCV} = \frac{(C_{s-actual} - C_{s-min}) \times (V_{ot-design} - V_{ot-min})}{(C_{s-design} - C_{s-min})} + V_{ot-min} \qquad (4)$$

where V_{DCV} is ventilation rate provided by a proportional control algorithm (m³/s), V_{ot-min} base ventilation rate for non-occupant-related pollutants (m³/s), $V_{ot-design}$, design ventilation rate (maximum occupancy multiplied by design per person ventilation rate, m³/s), C_{s-min} the target indoor CO_2 concentration at the

base ventilation rate (ppm), $C_{s\text{-}design}$ the target indoor CO_2 concentration at the design ventilation rate (indoor CO_2 set point, ppm) and $C_{s\text{-}actual}$ actual indoor CO_2 concentration (ppm). Because the CO_2 set point is equilibrium value of the design ventilation rate and generation rate (Eq. (3)), indoor CO_2 concentration cannot exceed the set point even the ventilation rate is set as the design value. But obviously proportional control over-ventilates the space a bit because in practice the equilibrium concentration can hardly be reached. Exponential control can be implemented by PI and PID controls. PID can be express in a discrete form as (CIBSE Guide 2000):

$$V_{DCV}(n) = K_p . \Delta C_R(n) + K_d . \left[\Delta C_R(n) - \Delta C_R(n-1) \right] + K_i . \sum_{j=1}^{n} \Delta C_R(j)$$

(5)

where $V_{DCV(n)}$ is the outdoor air ventilation rate (m³/s), K_p the proportional parameter, K_d the differential parameter (derivative term), K_i the integral parameter (integral term) and $\Delta C_R(n)$ the difference between the CO_2 set point and the nth sampled CO_2 level. K_d is zero for PI. The integral term (K_i) has the effect of minimizing or eliminating offset while the derivative term (K_d) reduces overshooting and has a very fast response (Montgomery and McDowall 2008). Therefore, PID control is considered to have the best control performance among proportional control, PI and PID control algorithms (Chao and Hu 2004). However, obtaining optimal K_p, K_d and K_i is difficult and very much depends on engineers' experiences no matter what kind of tuning method is used (Eder 2010a). All tuning methods can give only good parameters (i.e. K_p, K_d and K_i) for PID, but never real truly optimal ones (Eder 2010a). And in practice K_p, K_d and K_i need to be updated. Details about PID and tuning can be found in numerous references (CIBSE Guide 2000).

Research regarding DCV is quite common at present. It generally falls into two categories: investigation and development (or implement) of DCV. For investigation, Pavlovas (2004) provided a case study about a Swedish multifamily apartment aimed at evaluating the DCV system with different strategies. Four strategies (i.e. reference system with constant air volume flow; demand-controlled ventilation system: carbon dioxide control; demand-controlled ventilation system: humidity control; demand-controlled ventilation system: occupancy control) were evaluated. Simulation results showed that both the CO_2 and the RH control strategy may result in >50% reduction in energy consumption, while the occupancy control strategy results in about 20% reduction in the annual heat demand for ventilation, if the indoor climate is regarded as acceptable. Mysen et al. (2005) inspected 157 Norwegian classrooms to analyze the energy use over three different ventilation systems: CAV, CO_2 sensor based demand-controlled system (DCV-CO_2) and infrared occupancy sensor based demand-controlled system (DCV-IR). Their results showed that DCV-CO_2 and DCV-IR reduce the energy use due to ventilation in the average classroom by 38% and 51%, respectively, compared to the corresponding energy for a CAV system. Wachenfeldt et al. (2007) investigated two schools for displacement ventilation (DCDV) and CO_2-sensor based

demand-controlled displacement ventilation (DCDV-CO_2). They used both measurement and a calibrated simulation tool for analysis. During daytime operation with normal school activity, DCDV-CO_2 reduces the ventilation air volume by 65–75% in both schools compared to CAV. In one school, it was found that during the analysis period of 11–17 November, 2002, DCDV-CO_2 daytime operation reduces the total heating energy demand by 21%, the amount of unrecovered heat in the exhaust ventilation air by 54%, and the average airflow rate by 50% for weekdays. It was also found that DCDV-CO_2 daytime operation reduces the fan energy consumption by 87% for the analyzed week. Jeong et al. (2010) compared the Korean ventilation standard and ASHRAE Standard 62.1-2007 for CO_2-based DCV and radio frequency identification (RFID) based DCV. Simulation was done for a theoretical public assembly space served by a dedicated outdoor air system (DOAS) with an enthalpy recovery device. They found, compared to ASHRAE, the current occupant based ventilation standard of Korea may provide unstable ventilation control especially in CO_2–DCV because of frequent and large scale changes of supply air amount. The reason was the ventilation rate (per person) indicated in the Korean standard is the sum of the outdoor air required to remove or dilute air contaminants generated by both occupants and the buildings themselves, and not a pure function of occupant numbers.

For development (or implementation) of DCV, Nielsen et al. (2010) presented a strategy for a simple demand-controlled ventilation system for single family houses where all sensors and controls are located in the air handling unit. The ventilation strategy is to switch the air flow between two levels: a high air flow rate and a low air flow rate, based on measurements of CO_2-concentration and absolute humidity in the exhaust air and supply air. Measurements in the test house showed that the ventilation can be reduced to the low rate 37% of the time without significant changes in the CO_2-concentration and moisture level in the house. In theory this gives a 35% saving on electric energy for fans. Congradac and Kulic (2009) used genetic algorithms to optimize the return damper position in the mixing box such that indoor CO_2 concentration can be kept close to the desired level as possible and at the same time the lowest value of the commanding signal for the three-way valve (i.e. the least amount of cold water passes through the cooling coil, leading to the lowest energy use for the chiller) can be accomplished. The mathematical model of the HVAC system was realized in Matlab's tool 'Simulink. Their method was verified with EnergyPlus and showed great saving compared to the case where the genetic algorithm is not used. Xu et al. (2009) proposed a model-based optimal ventilation control strategy for multi-zone variable air volume (VAV) air-conditioning systems aiming at optimizing the total fresh air flow rate by compromising the thermal comfort, indoor air quality and total energy consumption. The strategy corrects the total fresh air flow rate dynamically by utilizing the fresh air from the over-ventilation zones based on the detected occupancy of each zone and the related measurements. In the meantime, the temperature set point is optimized for the temperature control of critical zones with the aim of reducing the variation of the required fresh air fractions among all the zones and further reducing the

total fresh air intake from outdoors for energy saving. Their strategy was evaluated in a simulated building and air-conditioning environment under various weather conditions. The results showed that their strategy can maintain acceptable thermal comfort and indoor air quality compared to the conventional DCV (CO_2-based exponential control) strategy, although slightly more energy is consumed. Sun et al. (2011) developed a CO_2-based adaptive DCV strategy. The strategy employs a dynamic multi-zone ventilation equation for multi-zone air-conditioning systems, in which a CO_2-based dynamic occupancy detection scheme is used for online occupancy detection. The strategy was implemented in an independent Intelligent Building Management and Integration platform (IBmanager). The performance of the strategy was practically tested and validated by comparing it with that of the original fixed outdoor air flow rate control strategy used on the site. Test results showed that the annual total energy consumption in one zone using the DCV strategy can be significantly reduced when compared to that using the fixed outdoor air flow rate control strategy.

2.4 Building simulation

As a separate discipline, building simulation can be traced back to the late 1970s. It often involves decision-making at the design stage, but nowadays it is increasingly used to assess the thermal response (e.g. indoor temperature, humidity) to building parameters (e.g. ventilation rate, wall structure, supply water temperature, water flow rate, and others) in existing buildings for the purpose of articulating an energy-efficiency strategy. Building simulation is normally underpinned by a numerical model because it is fast, economical, and particularly powerful for studying conditions that are too difficult and expensive for actual experimentation (Malkawi and Augenbroe 2003). In general, the results from black-box models are closer to measurements than the ones from numerical approaches (e.g.ANN, etc.). Errors result from a variety of sources, for example, incomplete information regarding material properties and types, the introduction of assumptions and the simplified modeling of (complex) physical processes, discretization and mathematical errors in the model, coding errors, inaccurate information about outdoor climate conditions and occupant behavior, and others. However, diagnosing error sources can be very difficult, time consuming, and expensive and is an area of research in itself (Judkoff and Neymark 2006). Some efforts have been made to increase the accuracy of numerical models. But, the improvement of simulation accuracy is often based on experiment and building audit (Malkawi and Augenbroe 2003), requiring significant effort and high costs.

The accuracy improvement of building simulation is a complex and difficult process. Judkoff and Neymark (2006) suggested that possible sources of errors in a simulation consist of seven types divided into two groups: external error types and internal error types. External error types are those caused by the user, such as (1) differences between the actual microclimate that affects the building versus input used by the program; (2) differences between actual schedules,

control strategies, effects of occupant behavior, and other effects from the real building versus those assumed by program user; (3) user error in deriving building input files; (4) differences between the actual physical properties of the building versus those input by the user. Internal error types include (1) differences between the actual thermal transfer mechanisms in the real building and its HVAC systems versus the simplified model of those processes in the simulation; (2) errors and inaccuracies in the mathematical solution of the models and code errors. They recommended that the validation of building simulation should be a process combining analytical verification, empirical validation, and intermodal comparisons. In literature, external errors are commonly reduced by means of experiment, audit, survey, specifications and documents (i.e. model calibration). For example, Pan et al. (2007) calibrated DOE-2 energy model with the detailed data of buildings and systems that were collected on as-built drawings, specifications, operating records and site surveys. Raftery et al. (2011a, 2011b) applied a systematic, evidence-based methodology to calibrate whole building energy models. A version control repository was created to store each revision of the model. An initial model was built first according to design documentation and program defaults where design information was unavailable. The model was then revised using the information of standards and guide-lines, benchmark or best practice models, as-built documentation, operation and maintenance manuals, material data sheets, interviews, surveys and physical verification, spot measured data and logged measured data. The final calibrated model had excellent correlation with the measured HVAC consumption data for the analyzed year of 2007. Borg and Kelly (2012) designed a generic model for a single effect lithium bromide–water absorption chiller. The model was calibrated by deriving component coefficients from measurement data.

2.5 Purpose of the study

The objective of this thesis is to provide modeling methodology, algorithms, information, and control strategies for improving the operations of building systems and the accuracy of building simulation models. The ultimate goal is to make buildings more energy-efficient. Based on different methodological perspectives, four sub-objectives of this study are (Figure 1):

- Methodology 1 (physical modeling approach): (1) to investigate the data center's air management, energy performance and the opportunities for improving energy efficiency in cooling and waste heat reuses (I). (2) to develop control method to implement DCV strategy for CO_2 and energy savings, which can be an alternative for conventional proportional, PI and PID controls (III and IV)
- Methodology 2 (data-driven modeling approach): (3) to promote the application of ANN and regression in predictive modeling for buildings (II, VI, VII and VIII);

- Methodology 3 (hybrid modeling): (4) to improve the accuracy of numerical models without tedious and difficult error analysis to minimize the time and cost of building simulation calibration (V).

As will be seen in the Method chapter, this thesis draws attention to the innovative use of these modeling methods presented in eight papers (I) to (VIII) (see the second column of Figure 1 for details)

Energy application in data centers is inherently site specific, depending heavily upon geography, climate and local environmental conditions. To date, even though a number of case studies have been published, most of them were conducted in the US and other areas. As for Nordic countries, the leading area in high technology, relevant research papers are still lacking. Publication (I) focuses on the investigation of the overall aspects of energy management.

Although PI and PID are considered as efficient methods for control and energy saving, their tuning is problematic and sometimes it is quite difficult to achieve the needed performance (Eder 2010b) and it demands special skills and experiences for engineers, even for autotuning (Eder 2010b). In order to work effectively for PI or PID, common control parameters must be updated online through ANN, which makes the tuning even more complicated (Roberge et al. 1997). Furthermore, existing DCV strategies other than PI and PID are either for a specific system (e.g. VAV system) or too complicated to use. Publication (III) develops a novel and simple method, based on the CO_2 mass balance equation, to implement CO_2-based demand-controlled ventilation strategy for hourly scheduled buildings. Because the method is physically based, it is universal and does not need to be updated. Publication (IV) is a follow-up of (III) by extending the method of (III) to buildings with unscheduled opening hours.

ANN and regression are widely applied data-driven approaches in predictive modeling. But still modeling challenges exist, which were dealt in Publications (II, VI, VII and VIII). Publication (II) develops an ANN model to predict water evaporation rate for indoor swimming pool. The water evaporation rate is one of the key design and operational variables in the study of various energy performance systems. In spite of this importance, its stochastic nature and complexity makes it nearly impossible to predict for indoor swimming pools. Its forecasting is rare in the literature. Publication (VI) deals with the prediction of indoor temperature and humidity with ANN. Prior to (VI), the prediction of indoor humidity by ANN was barely detailed. In addition to conventional point estimate, publication (VI) also offers interval estimate for indoor temperature and humidity, which adds novelty value to the application of ANN. In the literature, there lacks a simple and handy method for estimating space air change rates, which compares the supply airflow rate (i.e. recirculated as well as outside air in the supply air) to the volume of that space, and CO_2 generation rates for a particular space with field measured indoor CO_2 concentrations. Knowing the space air change rate is important for evaluating the ventilation system, for example, whether the desired space air change rate is met. Estimation of the number of occupants can help CO_2 based DCV meet industry

standards or building code requirements. Publication (VII) develops a novel and simple method for accurately quantifying ventilation rates using Maximum Likelihood Estimation (MLE) and CO_2 generation rates from measured CO_2 concentrations for individual spaces. Because most ventilation systems are CAV in Finland, and even a DCV system cannot ensure sufficient information on ventilation rate, data-driven model like ANN could have poor generalization when using ventilation rate to predict supply air temperature efficiency in a run-around heat recovery system. The evaluation of supply air temperature efficiency is essential for a run-around heat recovery because in Finland many run-around heat recovery systems often run with the supply air temperature efficiency less than the desired value, i.e. 50%. Publication (VIII) develops a simple and novel field measurement based methodology, supported by the power law relationship of air-side heat transfer, to evaluate the performance of run-around heat recovery systems using linear regression. The methodology (VIII) can work for CAV and demand-controlled systems with great accuracies, and greatly simplifies the assessment of air-to-air heat recovery systems.

Publication (V) proposes a novel hybrid numerical-neural-network model based on both physical and data-driven modeling techniques to improve the accuracy of numerical model without going through difficult and costly error analysis.

A summary of the research methods for eight papers and their research goals is provided in Table 1.

Table 1. Summary of the research methods and research goals for eight papers.

Research goal*	Paper	Research method
(1)	I	Methodology 1 (physical modeling approach): energy balance equation, performance metric equations, and cooling power equation for investigating air management and making energy end use breakdown in a data center
(3)	II	Methodology 2 (data-driven modeling approach): ANN model for predicting water evaporation rate
(2)	III, IV	Methodology 1 (physical modeling approach): CO_2 mass balance equation for developing novel and tuning free DCV
(4)	V	Methodology 3 (hybrid modeling): hybrid numerical-neural-network model for building simulation
(3)	VI	Methodology 2 (data-driven modeling approach): ANN model for predicting indoor temperature and humidity
(3)	VII	Methodology 2 (data-driven modeling approach): Maximum Likelihood Estimation model for predicting space air change rate; Coupled model (transient detect algorithm and equilibrium analysis) for predicting number of occupants
(3)	VIII	Methodology 2 (data-driven modeling approach): linear regression model, developed by a new physical law, for predicting supply air temperature increment

* the number indicates corresponding sub-objective of this thesis (see the bullets in Ch. 2.5).

3. Method

The methodology adopted in this thesis is composed of three broad modeling categories: physical modeling, data-driven modeling and hybrid modeling as shown in Figure 6. Figure 6 provides an overview of the methodology and lists the location of each component in this thesis. The modeling techniques and the developed models are detailed and example applications are provided. The case study will illustrate the feasibility and effectiveness of the proposed modeling methods and developed models.
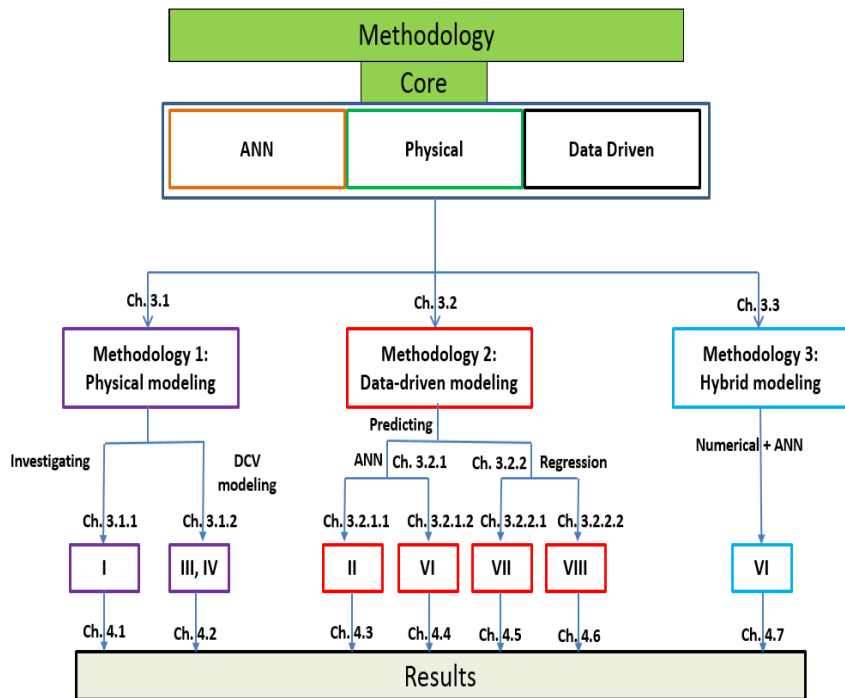
Figure 6. Overview of methodology, related methods with their associated publications and their locations in the thesis.

3.1 Methodology 1: Physical modeling

Physical models describe physical processes observed in a building. The processes depending on the state variables are dictated by the laws of conservation of mass, momentum and energy. Models can represent the

physical interrelationships between the subsystems and their flow paths. Physical models can be used to investigate various scenarios, compare them, and optimize building energy systems without the need for extensive measurement. In addition, a physical model aids in the analysis of system energy performance and in the development of control algorithms. For example, the energy balance equation can be used for energy end use breakdown (Ch. 3.1.1). Similarly, the CO_2 mass balance equation can assist DCV in the development of a control algorithm (Ch. 3.1.2).

3.1.1 Investigating cooling performance of a data center: case study (I)

This case study was conducted in a data center (built with raised floor and racks arranged in Hot Aisle /Cold Aisle layout, Figure 7) located in a southern city, Espoo, in Finland (I). The data center incorporates a fluid-side economizer with the chilled water system for free cooling (Figure 3). The design IT power is near 1 MW, but currently the data center is operating with less than half the designed power.
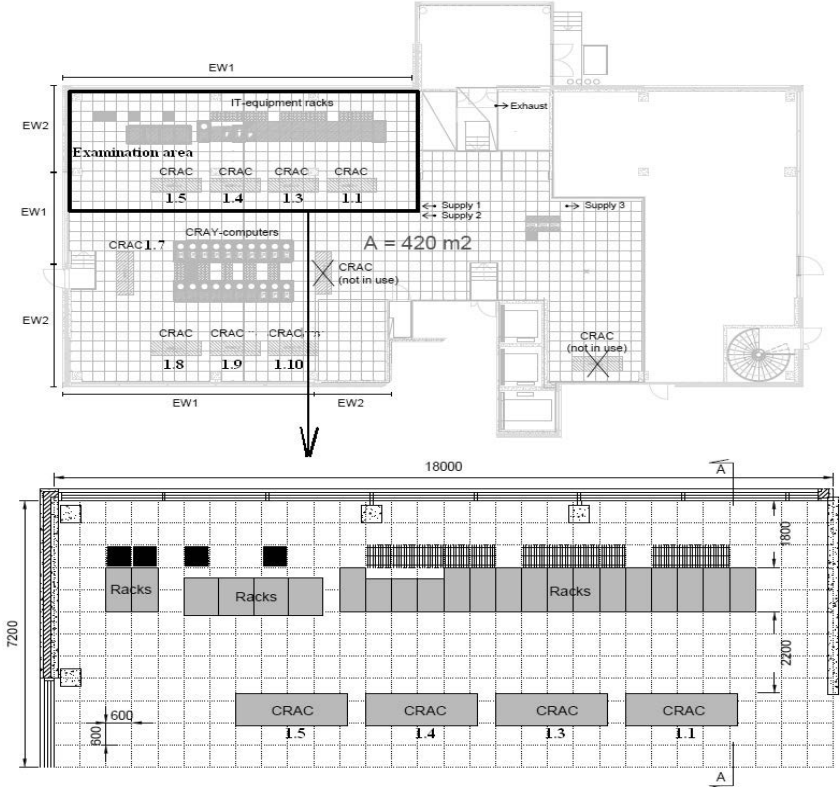
Figure 7. Plan view of the data center. The black arrows denote the approximate locations of the supply and exhaust ventilation ducts. The dimensions in the lower figure are approximated in mm (Original publication I).

The center is divided into two parts: Main cluster and Examination area. Examination area utilizes the conventional Hot Aisle /Cold Aisle protocol with locally ducted supply and flooded return (Figure 2, twenty-two racks, about 240 kW, mainly HP and Dell products) while Main cluster is made up of air-cooled supercomputers (Cray XT4 and XT5, about 490 kW, 20 x 11 kW and 30 x 9 kW units) where chilled air from CRACs is directly ducted in and hot air exits at the top (see Fig. 7 in I). This research study includes (1) an investigation of air management and (2) energy end use breakdown.

The investigation of air management was performed in Examination area by measurements (Table 2 and Figure 8, temperature and humidity were not measured for empty and almost empty racks).
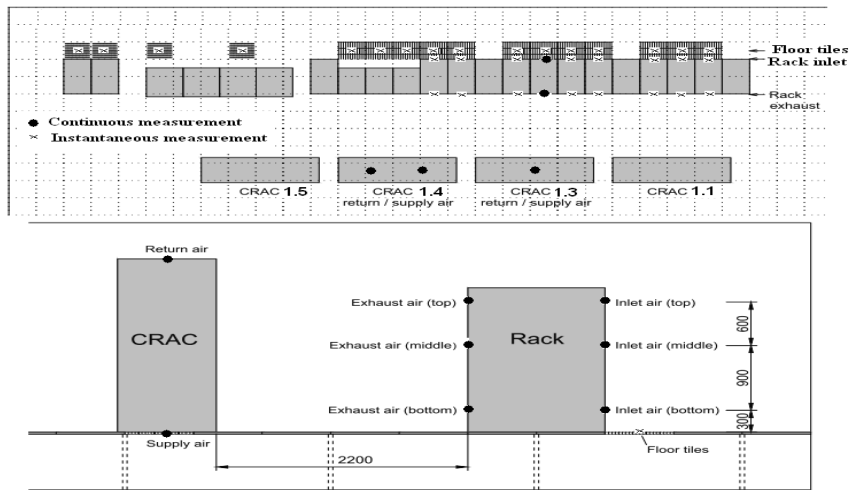
Figure 8. The measurement points of CRACs and IT equipment racks. CRAC1.4 was measured from two points on the center axis and CRAC1.3 from one point. The rack inlet and exhaust air was measured at three different heights, at the centerline of the rack (mm) (Original publication I).

Table 2. Measurements conducted in the data center.

Subject	Measured quantity	Measurement type	Equipment
Computer Room Air Conditioner (CRAC)	T, RH (return and supply air ducts)	Continuous	T/RH logger ST-171 (Clas Ohlson) (Accuracy: ± 0.2 °C and ± 0.2% RH)
IT-equipment racks	T, RH (inlet and exhaust air)	Continuous, instantaneous	ST-171,Vaisala HMI41 humidity and temperature meter (Accuracy: ± 0.1 °C and ± 0.1% RH)
Perforated floor tiles	T, RH, air velocity	Instantaneous	Vaisala HMI41, Alnor GGA-65P Thermo-Anemometer
Ventilation	T, RH (supply and exhaust air)	Continuous	ST-171

Continuous measurement means the sensor continuously reads the data on a specified time interval while instantaneous measurement indicates that the measurement was taken several times at a specified location and that finally

these measurement values were averaged. Instantaneous measurements were taken at the middle height (1.2 m from the floor) of racks. The airflow rates from the perforated floor titles were measured with an airflow meter. This airflow meter measured air velocities from the center and some corner holes. Airflow rates were calculated based on open areas and average measured air velocities. Temperature and relative humidity were measured at all duct openings (supply and exhaust) for outdoor air ventilation system and these values were averaged. These measurements were used to (1) examine rack inlet air conditions; (2) evaluate performance metrics RTI (%) (*measuring the actual utilization of the available airflow, above 100% suggests mainly recirculation air while below 100% mainly bypass air*, see Eq. (1) in I) and the SHI (*measuring the local magnitude of hot and cold air mixing, typical SHI value is less than 0.4, and the smaller the better*, see Eq. (2) in I); (3) analyze outdoor air ventilation.

The center monitoring system provided one-year IT equipment and total facility powers (1.11.2009-1.11.2010) and the facility information for 10:00-14:00 1.11.2010 including (1) IT, pump, dry cooler and CRAC fan powers; (2) chilled supply and return water temperatures and water flow rate. The average IT power for the period of 10:00-14:00 1.11.2010 (494 kW) is near average IT power from 1.11.2009 to 1.11.2010 (492 kW). Therefore, this facility information for 10:00-14:00 1.11.2010 is suitable for calculating energy end use breakdown by applying the energy balance equation (*total heat load from IT equipment and others = total cooling power from CRACs*) and cooling power equation ($P=Q.\rho.c.\Delta T$, *where P is the cooling power, W; Q chilled water flow rate, m^3/s; ρ the density of air, kg/m^3; c the specific heat capacity of air, $J/kg\ ^oC$; ΔT the temperature difference between return and supply chilled water, ^oC*) (Ch. 5.3 in I). When estimating energy end use breakdown, only average values (e.g. the average power of dry coolers estimated by the site is 17.4 kW for 10:00-14:00 1.11.2010) were used. A PUE ($=\frac{Total\ facility\ power}{IT\ power}$) analysis was also made.

3.1.2 Developing novel DCV control strategy (III and IV)

Sports arenas are selected as the base case to develop a novel CO_2-based DCV control strategy. In Finland, there are over 2,000 indoor sports facilities and many of these are sports training arenas, and represent a potential sector of energy saving. In addition, control strategies for sports training arenas are certainly applicable to a large range of buildings/spaces such as classrooms, theatres, conference rooms and so on. These buildings/spaces have a common feature that their opening hours are dominated by schedules, that is, occupied and unoccupied hours are well-scheduled and known in advance.

For a well-mixed and mechanically-ventilated space, the mass balance of CO_2 concentration can be expressed as:

$$V\frac{dC}{dt} = Q(C_o - C) + G.P \tag{6}$$

where V= space volume, C = indoor CO_2 concentration, Q = volumetric airflow rate into (and out of) the space, C_o = supply CO_2 concentration. The number of

occupants can be solved from Eq. (6) using the finite difference approach for the derivative term as:

$$P_n = \frac{1}{G}\left[V\frac{(C_n - C_{n-1})}{\Delta t} + Q_n(C_n - C_{o,n})\right] \tag{7}$$

where n and $n-1$ are the current and previous time steps, respectively. Eq. (7) is called the *transient detection algorithm*. Assuming Q, C_o and $G \cdot P$ are constant, Eq. (6) can be solved as follows:

$$C(t) = C_o + \frac{G.P}{Q} + (C(0) - C_o - \frac{G.P}{Q})e^{-It} \tag{8}$$

where $C(0)$ = indoor CO_2 concentration at time 0, $I = Q/V$, space air change rate.

Return CO_2 concentration (return duct) is generally quite different from the CO_2 concentration in the breathing zone for a large space like sports hall (Heiselberg et al. 1998). If the return CO_2 concentration is the concern, the mass balance equation of CO_2 concentration will be the same as Eq. (6) with C= return CO_2 concentration because the CO_2 concentration in return airstreams may approximately represent average CO_2 concentration in the section of building from which the return air is drawn (International Performance Measurement & Verification Protocol Committee 2001). Eq. (8) then can be used as an essential tool to calculate the next time step's ventilation rate Q for developing novel DCV control strategy. A novel DCV control strategy (III and IV) was developed for sports halls with scheduled and unscheduled opening hours.

Sports hall with scheduled opening hours
Opening hours are divided into training sessions (e.g. team training or match) and breaks. The developed novel DCV control strategy is summarized as follows (Ch. 2.2 and Fig. 4 in III):

1. If the building is in a break, set the base ventilation as the ventilation rate.
2. If the building is in a training session, the strategy proceeds through the following steps:
Step 1: Estimate the number of occupants by directly solving Eq. (7).
Step 2: Calculate the minimum requirement for the outdoor air ventilation rate based on Eq. (2) and local building code.
Step 3: Calculate the ventilation rate via Eq. (8) by setting: C(0)=measured indoor CO_2 concentration at the time, $G \cdot P$ =estimated from Eq.(7), t = the_remaining_time_for_the_training_session + time_supplement, and C(t)= CO_2_set_ point – CO_2_supplement.
Step 4: Select the maximum value from the calculated ventilation rates in Steps 2 and 3 as the building/space ventilation rate.

The time_supplement and the CO_2_supplement are positive values, providing a more secure way to ensure the indoor CO_2 concentration is below the CO_2 set point at the end of a training session.

Sports hall with unscheduled opening hours

Opening hours are divided into pseudo training sessions and breaks. For each pseudo training session, the outdoor airflow rate is calculated in the same way as the novel DCV control strategy does in each training session for sports hall with scheduled opening hours. However, in each pseudo break, if indoor CO_2 concentration is above the preset *pseudo break CO_2 threshold*, set the design ventilation rate as the required ventilation rate. If the indoor CO_2 concentration is below (or equal to) the preset pseudo break CO_2 threshold, the novel DCV control strategy includes the following two steps (Ch. 2.1, IV):

Step 1: Estimate the number of occupants using Eq. (7);
Step 2: Compute the minimum outdoor air ventilation requirement based on local building code using Eq. (2) and then set it as the required ventilation rate.

Experimental study for the developed novel DCV strategy

An experimental study was carried out in a sports hall in Finland equipped with a 100% outdoor air system. The CO_2 generation rates are results of simulation and real measurements. The measurement of CO_2 generation rates was for a vocational school gym supplied with 100% outdoor air with a throughout constant ventilation rate. The proportional control algorithm (Eq. (4)) and the novel DCV control strategy were implemented and compared for the case of scheduled opening hours (III). For the case of unscheduled opening hours, the novel DCV control strategy were compared with the proportional control algorithm and the PID control algorithm (Eq.(5)). The CO_2 set point was set 800 ppm.

3.2   Methodology 2: Data-driven modeling

The data-driven method often employs ANN and regression to perform prediction and forecast, i.e. predictive modeling.

3.2.1   ANN modeling

ANN can be divided into two categories – static and dynamic. In a static network, the output at a time step depends only on the inputs at the same time instant. Therefore, static networks can be applied for static nonlinear system modeling but cannot be used to predict future. For "black box" problems, the important task is to build models for dynamic systems. In a dynamic system the output depends not only on its current inputs but also on the previous behavior of the system. Dynamic networks have memories and are designed for modeling dynamic systems. Dynamic networks can "forecast" and one of most used dynamic networks is nonlinear neural network autoregressive with exogenous inputs (NNARX), which has the following form (VI):

$$y(t)=f[y(t-1),...,y(t-n_a),u(t-k),....,u(t-k-n_b+1)]+e(t) \tag{9}$$

where y and u represent observed outputs and independent inputs respectively, n_a and n_b the orders and k the delay. Eq. (9) is also called serial-parallel model (II) if feed-forward network is adopted without feedback from output (the true output is used as inputs instead of feeding back the estimated output). In this dissertation, the serial-parallel model (Eq. (9)) is employed to predict the water evaporation rate in an indoor swimming hall (Ch. 3.2.1.1) and indoor temperature and humidity in an office building (Ch. 3.2.1.2). There are several methods for selecting structures (e.g. n_a, n_b, k and the size of hidden units) for the serial-parallel model, including exhaust search, generic algorithm and other optimization methods (Nørgaard et al. 2000). The genetic algorithm is a method for solving optimization problems originally inspired by biological evolution. The algorithm encodes a potential solution to a specific problem to a chromosome-like structure and applies recombination operators to these structures in order to preserve critical information. The genetic algorithm starts with an initial population and then selects parents to produce the next generation using specific rules. For example, a genetic algorithm can select optimal structures for ANN by giving each combination of input variables a binary representation (i.e. 1 represents an input variable "taken" as input for ANN while 0 "not taken") while the number of hidden units is a base ten number (e.g. 4 means the number of hidden units is 4).

3.2.1.1 Predicting water evaporation rate for indoor swimming hall (II)

Water evaporation from free surfaces is a function of water temperature, air temperature, relative humidity, and air velocity, as well as the number and type of activity of the occupants (Asdrubali 2009). Occupants can increase water surface area to enhance water evaporation by causing a wet deck, waves, sprays and the additional area of contact between air and the wet bodies of occupants exposed to air. Hence, evaporation is much higher if pools are occupied. This type of information, such as wet deck, waves, spray and wet bodies, is not possible to obtain. But it is related to schedules and time to some degree, particularly for a group activity (e.g. swimming club training). Therefore, time is chosen as an independent input variable to account for these types of impacts. Indoor temperature, relative humidity and pool water temperature are measurable. Pool water temperatures can be considered as constant and are neglected along with indoor air velocity, which has the least impact on water evaporation rate (ASHRAE 1999). The final model variables are indoor temperature, indoor relative humidity and time. Exhaust search was used to select the optimal structures (Fig. 3 in II).

Experiment

The measurement data were collected from an indoor swimming hall with five pools located in the central part of Finland. The test swimming hall is equipped with a 100% outdoor ventilation system, providing about 80% of its design flow rate during the operating hours (6 a.m. – 10 p.m. and 9 a.m. – 10 p.m.), and about 50% of its design flow during the non-operating hours. The building automation system for the test indoor swimming pool provided the ratios of

actual airflow rate to design air flow rate for supply and exhaust fans. Figure 9 depicts air handling systems and measurement locations.
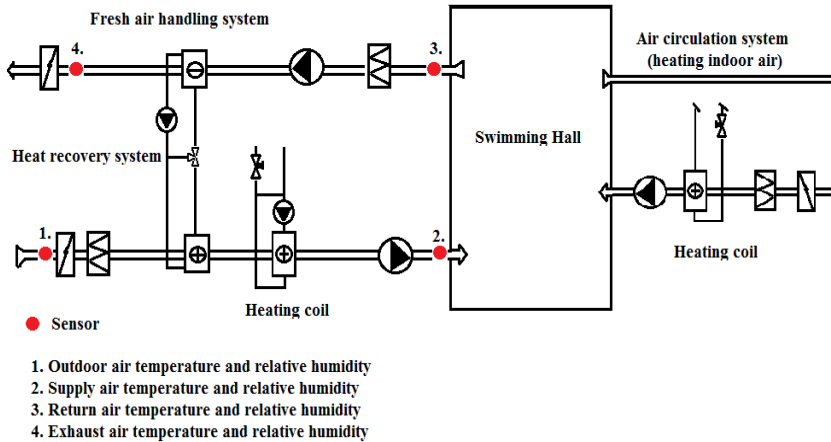
Figure 9. Air handling units of the swimming hall and sensor locations.

Sensors are Clas Ohson ST-171 temperature and moisture loggers with 10-minute interval and accuracies of ± 0.2 °C and ± 0.2% RH. The water evaporation rate was estimated by the mass balance equation ($Vol \frac{dW_{hall}}{dt} = V_{sup}W_{sup} - V_{exh}W_{ret} - V_{exf}W_{hall} + \dot{m}_{sou} + \dot{m}_{enve}$, Eq. (7) in II, the detail of calculating water evaporation rate is in Ch. 4.2 of II) using measured temperature and humidity data. The total size of the measured data is 4300 (Feb. 9 to Mar. 9 2012). The time was coded as a binary format (i.e. 1 indicates operating hours and 0 non-operating hours) and an hour format (e.g. 11:30 p.m. is coded as 23.5 =23+30/60).

3.2.1.2 Predicting indoor temperature and relative humidity (VI)

The aim of this research is to examine whether indoor temperature and humidity can be accurately predicted using the ANN model (Eq. (9)) and with simple measurements of indoor/outdoor temperature and humidity for a location with complicated indoor conditions.

In this research, two different approaches were used to implement ANNs: (1) using a mathematical method described and implemented in (Nørgaard 2000) to determine model orders and delays (i.e. n_a, n_b and k, Eq. (9)); the size of the hidden units were set the same as the number of input variables; (2) using a genetic algorithm to determine model input variables and the size of hidden units. For the convenience of model comparison, the first approach is referred to as NNARX while the second is referred to as the genetic algorithm.

In addition to MSE and the MAE, two other model validation methods were also presented:

- Prediction interval: A prediction interval is an estimate of an interval in which future observations will fall, with a certain probability

(normally 95%), given what has already been observed. An interval estimate expands on point estimate by incorporating the uncertainty of the point estimate which is preferable. In addition to supporting prediction, a prediction interval can also show the reliability of a specific input to the network. There are two major methods for estimating a prediction interval: delta method and bootstrap (Tibshiran 1996). The bootstrap method is more accurate. However, its limitation is heavy computation. For the network model, bootstrap requires hundreds of repetitions of retraining. The delta method instead gives a fast calculation but is less accurate and requires a Hessian matrix calculation. An alternative, more practical method was derived from the delta method to estimate prediction interval (De Veanux et al. 1998). However, this alternative method is also quite complicated. The details can be found from (VI) and (De Veanux et al. 1998).

- K-step prediction: The k-step prediction is able to reveal whether important information is captured by the model or not. It is often taken as an auxiliary tool to detect underfitting and overfitting problems. A four-step ahead prediction was adopted in this study.

Experiment

The dataset was obtained from the weather station located inside a central ventilation control room (Figure 10).
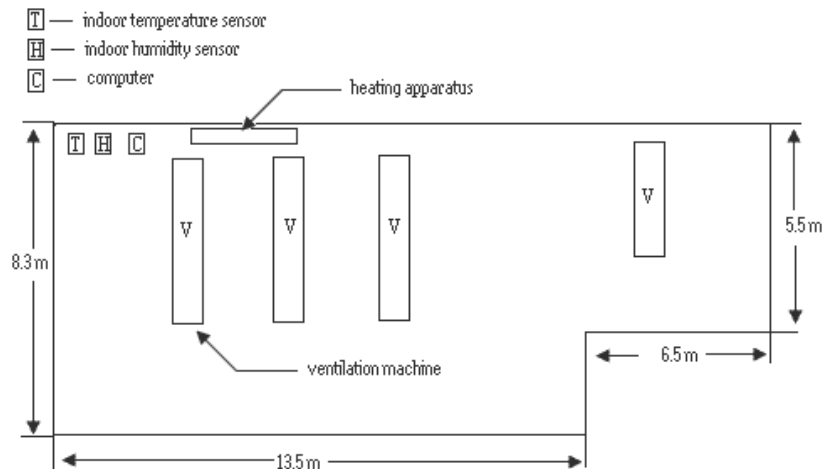
Figure 10. Central ventilation control room (Original publication VI).

The ventilation control room is on top of the department store building with outdoor temperature/humidity sensors mounted on the roof. The room has three large and one small ventilation machines which are responsible for ventilating half of the building. These machines constantly generated heat which was difficult to measure. There were ventilation ducts all over the room. It is easy to see that such a test house presents complicated indoor characteristics that are difficult to describe physically.

The experiment was carried out during January, 2007 for 30 days. All variables (temperatures and relative humidity indoors and outdoors) were measured within a 15 min interval. The total sample size was 2930.

3.2.2 Regression modeling

One type of regression is to predict systems parameter by applying the least squares or other optimization methods to the system governing equation. For instance, when CO_2 generation rate (i.e. G) is zero, Eq. (8) can be expressed as:

$$C(t) = C_o + (C(0) - C_o)e^{-It} \qquad (10)$$

With this simple governing equation, the least squares method can be used to predict space air change rate (i.e. I) by minimizing the sum of the squared indoor CO_2 concentration residuals between Eq. (10) and measurement. The least squares method can also be replaced by other methods, e.g. by the Maximum Likelihood Estimation (Ch. 3.2.2.1). The big advantage of this type of regression is that the methodology developed in one system is applicable to other systems because the physically based governing equation is universal (e.g. Eq. (9)). Therefore, when applying the regression method, the first and most important job is to try to identify and develop a simple equation which can physically interpret the system. In this dissertation, a new and novel physical law was developed to predict the performance of a run-around heat recovery system with linear regression (Ch. 3.2.2.2).

3.2.2.1 Predicting space air change rates and occupant CO_2 generation rates (VII)
In order to best apply the CO_2 mass balance equation (Eq. (10)) to predict space air exchange rate, the occupied period of a working day was split into (1) the occupied working period when the staff is present and (2) the unoccupied working period when the staff has left for home with the ventilation system on. The prediction was made for unoccupied working period. MLE was adopted. Supposing α is a vector of parameters to be estimated and $\{d_n\}$ is a set of sample or experimental data points, Bayes theorem gives

$$p(\alpha \mid \{d_n\}) = \frac{p(\{d_n\} \mid \alpha)p(\alpha)}{p(\{d_n\})} \qquad (11)$$

What MLE tries to do is to maximize $p(\alpha|\{d_i\})$ to get the best estimation of parameters (i.e. α) from $\{d_n\}$. Through some expansions and substitutions (e.g. substituting Eq.(10) for $d_{n,}$ see Eqs. (6)-(13) in VII), two MLE equations are obtained as:

$$\frac{\partial logp(\{d_n\}|\alpha)}{\partial \alpha_0} = \frac{1}{\delta^2}\sum_n (1 - \exp(-\alpha_1 n\Delta t))[d_n - (C(0) - \alpha_0)\exp(-\alpha_1 n\Delta t) - \alpha_0$$
$$]=0 \qquad (12)$$

$$\frac{\partial logp(\{d_n\}|\alpha)}{\partial \alpha_0} = \frac{(C(0) - \alpha_0)\Delta t}{\delta^2} \sum_n nexp(-\alpha_1 n\Delta t)[d_n - (C(0) - \alpha_0)exp($$

$$-\alpha_1 n\Delta t) - \alpha_0] = 0 \hspace{2cm} (13)$$

where δ^2 is the variance of the measurement errors and assumed to be independent of the time. Solving these two equations, Eqs. (12) and (13), simultaneously allows for the prediction of the space air change rate (i.e.α_1) and supply CO_2 concentration (i.e. α_0) for a working day. Unlike the least square method, the error analysis (δ^2) can be possibly done by MLE for the results.

The finite difference approximation for differentiation in Eq. (7) is unstable due to the growth of round-off error especially for the noise contaminated data which further amplify the measurement errors (Anderssen & Bloomfield, 1974). Instead of directly estimating the CO_2 generation rates, the number of occupants is evaluated because: (1) almost all ventilation regulations were stipulated based on the number of occupants; (2) knowing the number of occupants can somehow compensate for the losses from calculation errors (i.e. different CO_2 generation rates could result in the same number of occupants). The algorithm is described as followed (The details and examples are in Ch. 3.2.2 in VI):

Step 1: Compute a range of the numbers of occupants (Eq. (7), using a range of outdoor CO_2 concentrations to target uncertainties because normally outdoor CO_2 concentration isn't measured) for each measured CO_2 concentration to locate actual CO_2 generation rate.
Step 2: Identify significant jumps and drops from the measured CO_2 concentrations (e.g. 10 plus ppm jump or drop as significant change). Because one significant jump/drop does not mean a change in the number of occupants, further analysis is needed, followed by Step 3.
Step 3: Analyze the number of occupants at the jumped or dropped point as well as subsequent points.
Step 4: Finally, further confirm the obtained possible numbers of occupants by computing the value of the equilibrium CO_2 concentration (Eq.(3)). This step mainly targets the complex in estimating the number of occupants described in Step 2.

Experiment
The field measurement was set up in an office (27.45 x2.93 m^3, on the third floor) of a three-story school building (Figure 11). The mechanical ventilation is supplied (100% outdoor air) during the daytime on working day from 6:10 a.m. to 8:00 p.m. and shut down during nights, weekends, and public holidays. Three persons, two males and one female, work at the office regularly and the design airflow rate is around 200 m^3/h (2.5 ach). In addition to indoor CO_2 concentrations, the pressure differences between the return air vent and room were also measured.
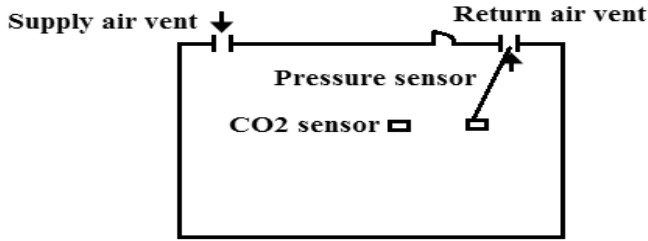
Figure 11. The layout of the office and sensor locations.

The measurement was categorized based on two stages. At the first stage (22.9.2008– 28.9.2008), the existing ventilation system was examined. At the second stage (13.10.2008–19.10.2008), the ventilation system was reconfigured by blocking some outlets at the supply and return air vents, aiming at reducing airflow rates. Finally, five-day data (with 5 min interval) were obtained. The measurements show that the pressure differences between the return air vent and room, important indicators of the airflow rate, were almost constant for all working hours each day despite small fluctuations. This implies that space air change rates on each working day are nearly constant.

Most literatures summarize the relationship between airflow rate and pressure difference across an opening as the following empirical formula (VII):

$$Q = C(\Delta P)^n \tag{14}$$

where Q is airflow rate, ΔP pressure difference across the opening, C a constant value depending on the geometry effects of the opening, n flow exponent (between 0.5 and 1.0. it is close to 0.5 for large openings and near 0.65 for small crack-like openings). Eq. (14) is called powerlaw relationship for opening also. If the pressure differences between the return air vent and space are measured, Eqs. (12) and (13) can be modified by Eq. (14) for time-varying ventilation system, such as DCV systems or variable air volume (VAV) systems, e.g. $\alpha_1 n$ in Eqs. (12) and (13) can be expressed as $\alpha_{ref}\sqrt{\frac{P_n}{P_{ref}}}$ by Eq. (14), where P_{ref} is a random pressure difference between the return air vent and space, such as 8, 9 or 10 Pa, α_{ref} corresponding space air change rate for this random pressure difference and P_n measured pressure difference, see Ch. 4.1.1 in VII.

3.2.2.2 Predicting the performance of run-around heat recovery systems (VIII)
The goal of this research is to develop a simple and useful physical law to simplify field measurement based modeling and to avoid multiple solutions.

A typical run-around heat recovery system is comprised of two finned-tube water coils (i.e. liquid-to-air heat exchangers, normally in cross-flow arrangement), connecting pipes, a three-way temperature valve and a pump as shown in Figure 12.
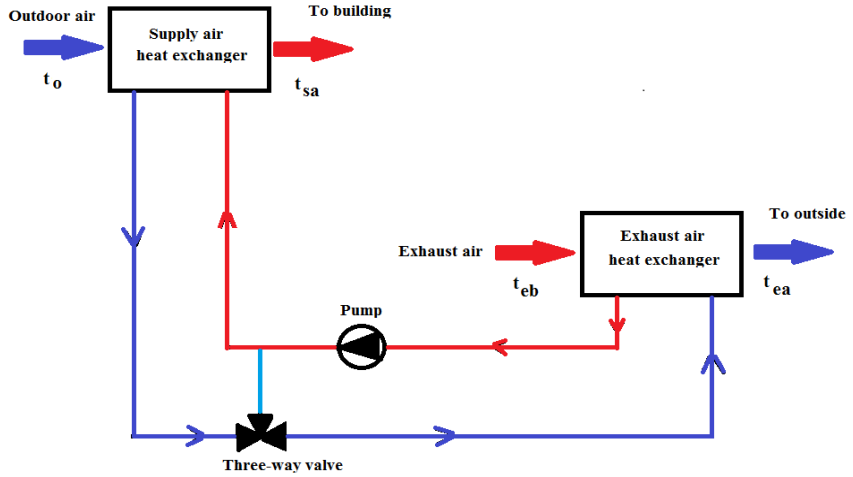
Figure 12. Run-around heat recovery system (Original publication VIII).

The supply air temperature efficiency is defined as (Figure 12):

$$\eta_{sup}(\%) = \frac{Supply\ air\ temperature\ increment}{Maximum\ temperature\ difference} = \frac{t_{increment}}{t_{max}} = \frac{(t_{sa}-t_o)}{(t_{eb}-t_o)} \tag{15}$$

where η_{sup} is the supply air temperature efficiency, t_{max} maximum temperature difference ($t_{eb}-t_o$) and $t_{increment}$ supply air temperature increment($t_{sa}-t_o$). Without loss of generality, the exhaust airflow rate ($\dot{V}_{ex}$) is assumed to be equal to supply airflow rate ($\dot{V}_{sup}$). Because the heat transfer coefficient on the liquid-side is usually more than an order of magnitude larger than the heat transfer coefficient on the air-side, air-side heat transfer dominates the overall heat transfer (i.e. a smaller heat transfer coefficient creates a bigger heat transfer resistance on the path of heat flow and seriously impedes heat transfer for the system even with the additional surface area added by the high-efficiency fins, Nellis&Klein, 2009). External forced convection over tubes in cross-flow can be experimentally expressed with reasonable accuracy by a simple power-law relationship of the form (Nellis&Klein, 2009):

$$Nu = CRe^m Pr^n \tag{16}$$

where Nu is the Nusselt number (proportional to convective heat transfer coefficient), Re is the Reynolds number (proportional to air velocity) and Pr is the Prandtl number. C, m (less than 1), and n are constant numbers depending on the system configuration and air flow. As the air thermal properties involved in Eq. (16) (e.g. kinematic viscosity and Pr) do not change much for normal air operating conditions in Finland (i.e. air temperature from -15 $^{\circ}$C to +21 $^{\circ}$C), the air-side heat transfer coefficient for the supply air or exhaust air heat exchanger (i.e. over a finned tube bank in cross-flow) is proportional to a power of the airflow rate.

By applying Eq. (16) and some empirical formulas (run-around heat recovery systems can be considered as air-to-air cross flow heat exchangers with both fluids unmixed), the following important equation is obtained:

$$\eta_{sup} \approx k \; \frac{1}{(\dot{V}_{sup})^\alpha} \tag{17}$$

where k is constant and $\alpha < 1$. By Eq. (15), there exists a linear relationship between the supply air temperature increment and the maximum temperature difference divided by a power of airflow rate:

$$t_{increment=} \; \eta_{sup} \, t_{max} \approx k \; \frac{t_{max}}{(\dot{V}_{sup})^\alpha}, \text{ namely, } t_{increment} \propto \frac{t_{max}}{(\dot{V}_{sup})^\alpha} \tag{18}$$

α is important system characteristic value, called a *signature power*. Eq. (18) can be considered as a new physical law governing run-around heat recovery system. In practice, Eq. (18) is adapted as:

$$t_{increment} \approx a \; \frac{t_{max}}{(V_{sup})^\alpha} + b \tag{19}$$

where $t_{increment}$ is measured temperature increment for supply air (i.e. dependent variable), $\frac{t_{max}}{(V_{sup})^\alpha}$ regressor and b error term, which accounts for the influence on the supply air temperature increment from all sources other than the regressor $\frac{t_{max}}{(V_{sup})^\alpha}$. It is important to determine the signature power α in Eq. (19) because, if α is known, *a* and b in Eq. (19) can be easily obtained by linear regression. As the majority of ventilation systems are CAV systems with two ventilation rates in Finland, data can be naturally divided into two very different groups based on ventilation rate. As for a correct signature power (i.e. α in Eq. (19)), resultant linear regression equations (i.e. Eq. (19)) from two groups should produce very near supply air temperature increment ($t_{increment}$) for the same $\frac{t_{max}}{(V_{sup})^\alpha}$, resulting in two very close and similar supply air temperature efficiency curves. These curves are 2-dimensional plots displaying the trend of supply air temperature efficiencies (y-axis) over airflow rates (x-axis). A tuning algorithm is then developed for CAV system with two ventilation rates to search the correct signature power (i.e. α) by linear regression:

Step One: Randomly set the signature power as a value (<1).
Step Two: Manually tune the signature power after step one so that the two supply air temperature efficiency curves from two groups are near enough when the maximum temperature difference is kept constant.

With little modification, the process described works for demand-controlled ventilation systems but not work for CAV systems with a single ventilation rate.

Experiment

The measurement data were collected from four sites (Site A, Site B, Site C and Site D, Tables 1 and 2 in VIII), which utilize a 100% outdoor air ventilation system and are equipped with a run-around heat recovery system. These data were obtained from the building automation systems of the sites and our own measurement systems. The building automation systems provide ratios of actual airflow rate to design air flow rate for supply and exhaust air fans (Note: Site D converted ratios of actual airflow rate to design air flow rate into actual airflow rates). The measurement systems provide (Figure 12): (1) outdoor temperature (t_o); (2) supply air temperature after the supply air heat exchanger (t_{sa}); (3) exhaust air temperature before the exhaust air heat exchanger (t_{eb}); (4) exhaust air temperature after the exhaust air heat exchanger (t_{ea}). The air temperatures were continuously monitored by four Clas Ohson ST-171 temperature and moisture loggers (located in the supply and exhaust air handing machines for each site) with 5 minute sample interval (30 minute sample interval for Site D) and accuracies of ± 0.2 $^\circ$C.

Data from Sites A, B and D were classified into three groups: (1) Group A: data with the base airflow rate (15276 samples for Site A and 26367 samples for Site B) and data with the airflow rate around 2.3 m3/s (only Site D, 1873 samples); (2) Group B: data with the design airflow rate (1255 samples for Site A and 2059 samples for Site B) and data with the airflow rate around 3.8 m3/s (only Site D, 719 samples); (3) Group C: data with the airflow rates between the base and design airflow rates (3200 samples for Site A and 1573 samples for Site B) and data with the airflow rates between 2.3 m3/s and 3.8 m3/s (only Site D, 1175 samples).

Data were divided into two groups for Site C: Group A (data with 80% of the design airflow rate, 13573 samples) and Group B (data with 90% of the design airflow rate, 10721 samples). Whole measurement in this study denotes entire measurement for a site: (1) all data from Group A, Group B and Group C (Sites A, B and D). (2) all data from Group A and Group B plus some unseen data, which should belong to Groups A and B but are used just for testing (See Table 3 in VIII) because site C does not have Group C (Site C).

The two-step strategy previously described was used to determine the signature powers by first guessing a value (<1) and then tuning it until the supply air temperature efficiency curves from Groups A and B (each site) were close enough by keeping the maximum temperature difference to 36 $^\circ$C.

3.3 Methodology 3: Hybrid modeling

An important challenge in building energy research is to model the evolving complex building systems that contain a large number of components with unknown multiple physical mechanisms. Both data-driven and physical modeling approaches haven been applied to demonstrate their advantages. A crucial limitation of both modeling methods is their generalization capability. Large quantities of data are needed for data-driven models and the models often lack interpretation. Detailed physical properties are not always available for

physical models. A contemporary trend is to couple them and this type of model is generally called hybrid model.

Improving accuracy of building simulation (V)

In this research study, a hybrid numerical-ANN model was developed, comprised of two components: a numerical model, serving as an estimator to predict the output of a building system with building parameters, and an ANN, acting as a calibrator to calibrate the output of the numerical model with the measured output of the building system. Figure 13 illustrates the hybrid numerical-ANN model.
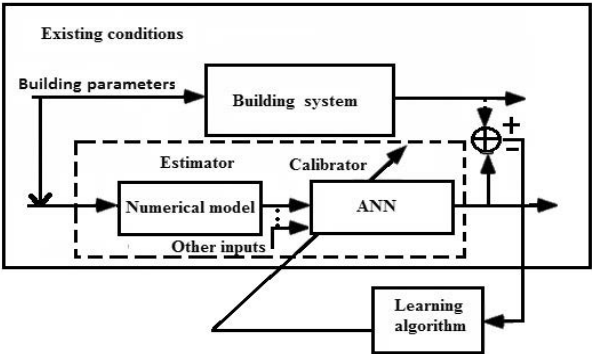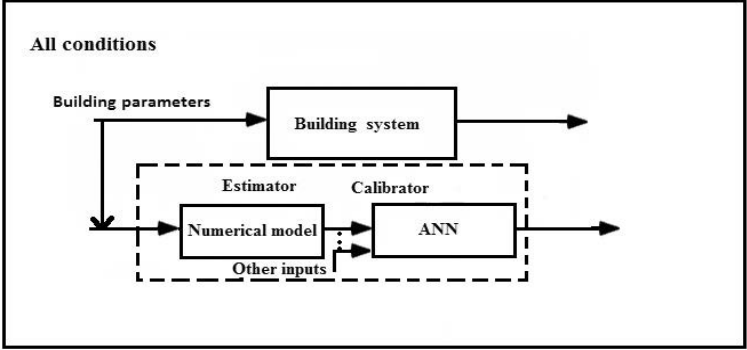
Figure 13. The hybrid numerical-neural-network model for building simulation (Original publication V).

The big advantage of the hybrid numerical-ANN model is that it plays the same role as a numerical model but the accuracy is much improved by importing ANN as a calibrator. As we know, no matter how carefully we model physical processes and how well we know the system, errors are inevitable for a numerical model (Malkawi and Augenbroe 2003). In fact, some error sources

46

for numerical modeling are quite difficult to identify. However, ANN is good at handling the uncertainty problems. By comparing simulated values from the numerical model with measured ones, the ANN can learn and simulate the error patterns between the two without knowing the actual causes. In addition, the ANN can also facilitate simulation processes in the numerical model. For instance, a building physical process can be ignored in the numerical mode if it is too difficult to model. The error caused by the exclusion of this physical process in the numerical model is expected to be corrected by the ANN.

By acting like a numerical model, the hybrid numerical-ANN model should be capable of analyzing the impact of a building parameter on the output of building system even when the building parameter is constant in practice. For example, for a CAV ventilation system, the ventilation rate remains constant, but the hybrid numerical-ANN model can show how an unheated and uncooled indoor temperature is influenced by the ventilation rate of this CAV system. As is well known, a single ANN generally cannot generalize about a constant parameter. However, in the hybrid numerical-ANN model, the parameter, which is constant in practice, is moved from the ANN to the numerical model as input. If the impact of the building parameter on the output of building system is modelled correctly and accurately in the numerical model, relevant numerical errors will become small compared to other numerical errors. Because the ANN is supposed to predict the error patterns between the numerical model and measurement, the ANN can be assumed to be independent of the parameter being constant, namely, relevant numerical errors are small compared to other numerical errors. Therefore the building parameter, constant in practice, can be assumed as having negligible impact on the error patterns between the numerical model and measurement. In this way, the ANN in the hybrid model can generalize about the building parameter (constant in practice) as long as the ANN can generalize about the output of building system. More details regarding this can be found from (V).

Experimental study

The indoor temperature simulation for an unheated and uncooled single-zone building was taken as a case study. The study was conducted in a sports center situated in the south of Finland, employing a 100% outdoor air CAV system to provide two different ventilation rates, one for non-operating hours (10 p.m.–9 a.m.) and the other for operating hours (9 a.m.–10 p.m.), respectively. Data for the return air temperature (return duct) of the sports hall and outdoor temperature were collected from the building automation system at an interval of 20 min (2357 samples). During the period of data collection (18.7.2007–20.8.2007), the rotary air-to-air heat recovery system was switched off, meaning that there was no heat exchange between return air and supply air, and the sports hall was unheated and uncooled. In the case study, two hybrid numerical-ANN models, Hybrid model A and Hybrid model B, and a testing numerical model were developed. The testing numerical model is not a part of any hybrid numerical-ANN model (i.e. Hybrid model A or Hybrid model B) and its results (i.e. simulated unheated and uncooled indoor temperature) were

compared with those from the two hybrid numerical-ANN models. The differences between the two hybrid numerical-ANN models are as follows: Hybrid model A uses mixed measurement data from both the non-operating and operating hours to train and test the ANN while the training data for the ANN in Hybrid model B are all from the non-operating.

The heat balance for the sports hall is shown in Eq. (21) (ENERGYPLUS 2009):

$$\rho.c.V.\frac{dT_{ret}}{dt} = \sum_{i=1}^{N_{surfaces}} h_i A_i (T_{si} - T_{ret}) + \dot{Q}_{int} + \dot{V}_{sup}.\rho.c.(T_{sup} - T_{ret}) + \dot{Q}_{sol}$$

(20)

Eq. (20) uses the return air temperature to represent the average indoor temperature leading to an approximation of a uniform temperature condition for the hall. The difference between the testing numerical model and the numerical model in Hybrid model A or Hybrid model B is that in the latter $\dot{Q}_{sol}$ and $\dot{Q}_{int}$ are treated as zero. The reason is to show that the hybrid model can enhance the accuracy of a numerical model (i.e. the testing numerical model) even with simplified simulation processes. $\dot{Q}_{int}$ includes heat gains from occupants, lights, supply air fan, exercise equipment (e.g. jogging machine), and unknown sources. The heat gains from occupants and exercise equipment are difficult to estimate and depend on occupant profiles and time. Because occupant profiles are not available, they were ignored along with the heat gain from unknown sources in the testing numerical model.

Hybrid model A
For the ANN of Hybrid model A, the training and testing data included both non-operating hours and operating hours. Because lights, sports and excises are related to schedules and time in some degree, time (e.g. a time, hh.mm.ss, can be converted to hours) was chosen as another input for the ANN to account for the simulation errors caused by excluding them from Eq. (20). Therefore, the inputs to the ANN of Hybrid model A are the simulated indoor temperature from the numerical model and time.

Hybrid model B
All training data for the ANN of Hybrid model B are from non-operating hours, but testing data are a mixture of non-operating hour and operating hour (9 a.m. to 12 a.m.) data. Only the part of testing data from non-operating hours was used to decide the optimal structure indicating that the ANN of Hybrid model B was constructed purely from non-operating hour data. The aim of Hybrid model B is to evaluate the generalization ability of the hybrid numerical-ANN model for ventilation rate. The ANN of Hybrid model B, built from non-operating hour data, is expected to be used directly to operating hour data, where the ventilation rate is much larger than in non-operating hours. Because there is no internal heat gain from occupants or lights in non-operating hours, time (e.g. a time, hh.mm.ss, can be converted to hours) was not taken as an input for the

ANN of Hybrid model B, leading to only one input – simulated indoor temperature from the numerical model.

4. Results

4.1 Air management and energy performance in data center in Finland (I)

Air management

The measured rack inlet temperatures and humilities (including continuous and instantaneous measurements) were in the ranges of 13 – 15 °C (average values from three different heights for continuous measurements, Figure 8) and 7 – 10 °C (dew point). The rack inlet dew point temperatures were within the recommended moisture level, and the inlet temperatures fell within the allowable level but outside the recommended level (18 °C to 27 °C) from ASHRAE (2009b). Low inlet temperatures stem from the operating requirement from CRAY supercomputers at Main cluster (Figure 7), which require cold supply air within 10 °C – 16 °C. Low inlet temperatures often lead to inefficient cooling performance, such as low chiller Coefficient of Performance (COP) of the chiller and excess cold air (Moore et al. 2005). Reported cold aisle temperatures from recent literature are mostly between 13 and 32 °C (Ebrahimi et al. 2014). The measured rack inlet temperatures in this study fall in this range.

All the SHI-values were near zero because the measured inlet air temperatures of the racks were either lower than or near the measured air temperatures from the adjacent plenum vents. This indicates that the recirculation of hot air was negligible and that the hot and cold air streams were perfectly separated. The RTI was estimated as 41%, suggesting extreme bypass air or oversupplied cold air. In fact, the CRAC airflow could ultimately be reduced by 59 % (=1-41 %). Due to the cubical relationship between the fan power and fan flow, the fan energy could be reduced by 93% if Variable Frequency Drive (VFD) is installed. Overall, calculated performance metrics suggest that enough cooling was provided for Examination area but cold air was oversupplied. RTIs from another study, which investigated cooling performance in 13 computer rooms (total area=15,800 m²) in USA, were found to be in the range of 34-79% with the average value=59% (Sullivan et al. 2007). Bypass air does occur in air-cooled data centers.

The supply and exhaust air flow rates were about 0.9 and 0.8 m3/s (from the center's monitoring system), respectively, namely the center was over pressurized (positive pressure to outside). The air change rate was then 2.66 1/h (=0.9*60*60/1218). ASHRAE recommends a minimum 0.25 (1/h) air change rate for data centers so as to dilute contaminants. Compared to this guideline value, the air change rate for the data center was too large. The ventilation rate

should be minimized in order to decrease the humidification and dehumidification load.

Energy end use breakdown
The PUE value was slightly above 1.2 in winter time and about 1.5 in summer (Figure 14). The difference was likely caused by the free cooling, which was activated when outside temperature was below 8 ℃, according to the information received from the site.

Figure 15 shows energy end use breakdown of electricity with the percentage information for a typical year (having average outdoor temperatures from measurements of 2000 to 2010).
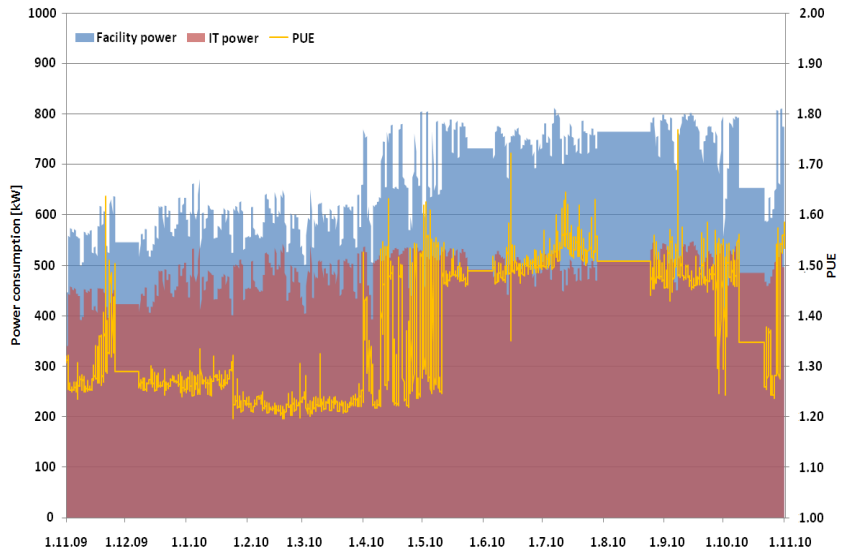
Figure 14. One year's (1st November 2009–1st November 2010) facility and IT power consumption (measurement interval = 1 h) (Original publication I).
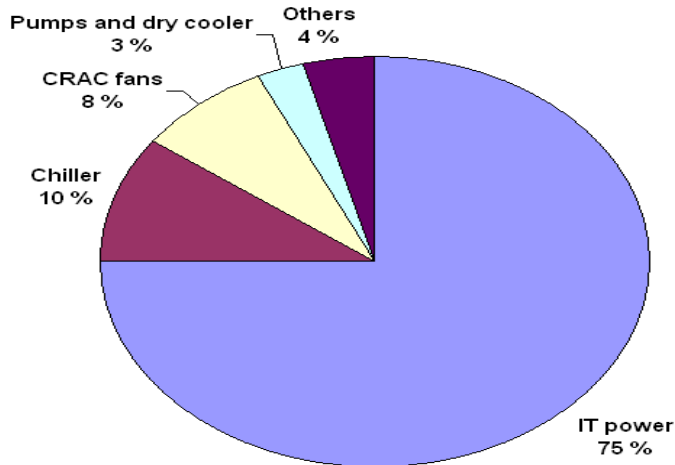
Figure 15. Electrical end use breakdown for a typical year (Original publication I).

The total power consumption of the CRAC fans is close to that of the chiller which did not match our original expectation (Figure 15). The main reason is the cold climate in Finland which gives the center more time for free cooling, but CRAC fans have to be in operation at all times. Therefore, CRAC fans have a great potential for energy saving. The use of VFD (i.e. adopting DCV strategy) will lower energy use of CRAC fans since currently the CRAC fan speed is fixed at full speed. The average PUE value (one year) is about: 1.33=100/75, which is very close to calculated average PUE value from the measurement – 1.35. The data center is an energy efficient one. The cooling system (chiller + CRAC fans + pumps + coolers) consumes 21% of total. About 97% of total electrical power (chiller + CRAC fans + IT power + Others) is converted to waste heat and then rejected to outdoor each year. This part of the waste heat (approximate 5627 MWh, one year) actually can be reused. In Finland, the average energy demand for space heating and water heating is about 182 kWh/m² for non-domestic buildings (one year), meaning the reuse heat can support 30916 m² (=5627*1000/182) non-domestic building for yearly heating (space heating + water heating).

4.2 Novel DCV control strategy (III and IV)

Scheduled opening hours
The controlled CO_2 concentrations by the novel DCV control strategy (Ch. 3.1.2) are much nearer to the CO_2 set point (i.e. 800 ppm) compared to the ones by a proportional control approach. Figure 16 displays simulated CO_2 concentrations using simulated CO_2 generation rates for the proportional control approach and the novel DCV control strategy when opening hours are scheduled. Table 3 shows comparison results.

Table 3. Comparisons of average ventilation rates between the novel DCV control strategy and proportional control (simulated CO_2 generation rates) for scheduled opening hours (Original publication III).

Control approach	Average ventilation rate (m³/s)	Average CO_2 concentration during opening hours (7 a.m. – 23 p.m.)
Proportional control (1.92 m³/s)[a]	0.72	628 ppm
Proportional control (1.52 m³/s)[a]	0.65	648 ppm
Novel DCV control strategy	0.43	749 ppm

[a] The number in parentheses is design ventilation rate

Because the proportional control is based on the assumption of an equilibrium condition, which seldom occurs in practice, the space is often over-ventilated although the minimum requirement for the amount of outdoor air can be satisfied. In contrast, the developed novel DCV control strategy takes the dynamic effects (i.e. CO_2 changes in both time and quantity dimensions) into account to calculate the instantaneous ventilation rate by solving the CO_2 mass balance equation, resulting in the controlled CO_2 concentration much closer to the CO_2 set point while the minimum requirement of outdoor air is met (Figure 16). This phenomenon is more strongly manifested in training (e.g. ice hockey club training) dense periods (after 5 p.m.) where the indoor CO_2 concentration is mostly controlled between 770 ppm and 800 ppm at each training session

(Figure 16). The novel DCV control strategy can respectively save +40% ($\approx$ (0.72-0.43)/0.72) (vs. 1.92 m³/s) and +34% ($\approx$ (0.65-0.43)/0.65) (vs. 1.52 m³/s) the energy required to condition outdoor air (heat recovery is not included. Therefore, this part of energy is directly proportional to outdoor air intake) compared to the proportional control approach. The smaller ventilation rate also increases the efficiency of air-to-air heat recovery system (Table 3). Similar results can be seen if CO_2 generation rates were obtained experimentally (e.g. Table 2 and Fig. 9 in III).
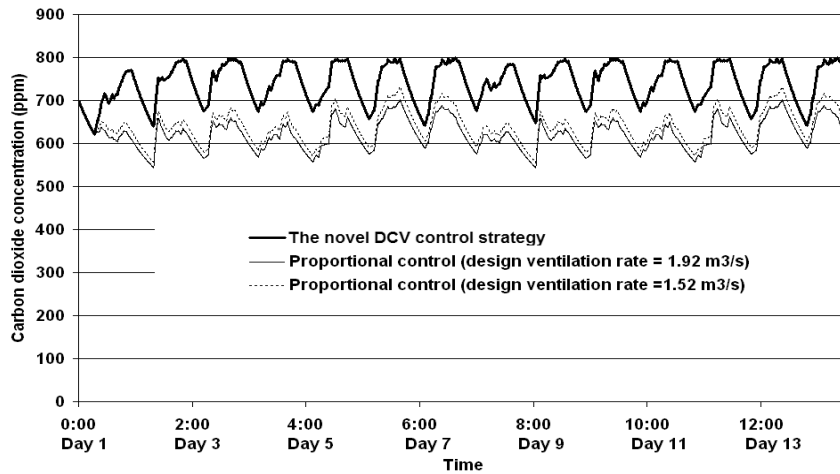
Figure 16. Comparison of simulated indoor CO_2 concentrations between the novel DCV control strategy and the proportional control in the sports training arena for scheduled opening hours (simulated CO_2 generation rates, 14 days) (Original publication III).

Unscheduled opening hours

The performance for unscheduled opening hours is shown in Figure 17 and Table 4.
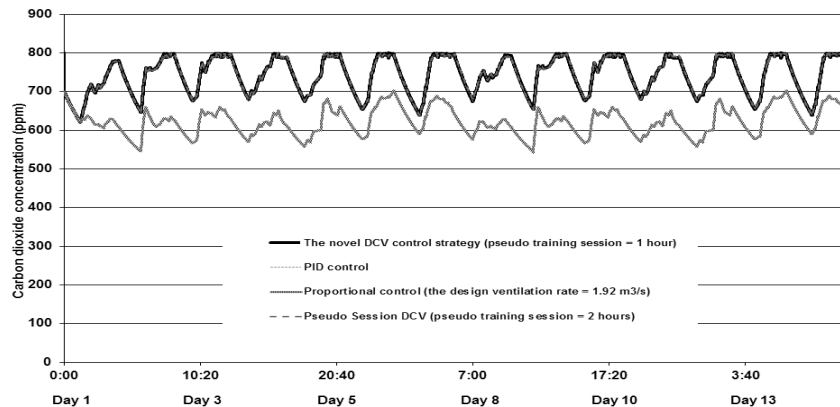
Figure 17. Comparison of simulated CO_2 concentrations between the novel DCV control strategy, proportional control and PID control in the sports training arena for unscheduled opening hours (14 days, real schedules with simulated CO_2 generation rates per person) (Original publication IV).

Table 4. Comparison of performance between proportional control, PID control and the novel DCV control strategy for unscheduled opening hours (real schedules with simulated CO_2 generation rates per person) (Original publication IV).

Control approach	Average ventilation rate (m³/s)	Average cubic ventilation rate (m³/s)³	Average CO_2 concentration during opening hours (7 a.m. – 23:20 p.m.)
Proportional control (1.92 m³/s)[a]	0.72	0.8	629 ppm
PID control	0.45	0.18	753 ppm
Novel DCV control strategy (pseudo training session = 1 hour)	0.45	0.24	754 ppm
Novel DCV control strategy (pseudo training session = 2 hours)	0.45	0.17	754 ppm

The simulation results in Tables 4 show that in all the cases (also see Tables 1 and 3 for other two cases in IV) using the novel DCV control strategy, the same amount of the energy required to condition outdoor air (heat recovery is not included) is consumed as is the case with PID control (i.e. their average ventilation rates are near). Changing settings for the novel DCV control strategy (e.g. pseudo training session length, the design ventilation rate, and pseudo break CO_2 threshold) does not have much effect on the energy required to condition the outdoor air (see Table 4 in IV). This interesting phenomenon is another strong manifestation of the novel DCV control strategy's quick response to indoor CO_2 changes and its ability to reduce the actual indoor CO_2 discrepancy from the set point.

The results of fan energy vary for different settings for the novel DCV control strategy. According to the fan laws, the input energy required for a fan (i.e. energy use of fans) is proportional to cubic outdoor airflow rate (cubic fan law) (Lonnberg 2007). With the pseudo training session of two hours, the novel DCV control strategy consumes the equivalent amount of energy as PID but with the pseudo training session of one hour, 11-33% more fan energy is consumed compared to PID control (the third column of Table 4). The design ventilation rate and pseudo break CO_2 threshold affect fan energy consumption (see Table 4 in IV). A smaller design ventilation rate or higher pseudo break CO_2 threshold would help the system save more fan energy (see Table 4 in IV). Both the novel DCV control strategy and PID save significant amounts of fan energy, ranging from 61% to 79%, compared to proportional control. The developed novel DCV control strategy achieves the same level of energy consumption as PID control for both the energy required to condition outdoor air and the energy to power fans.

4.3 Prediction of water evaporation rate for indoor swimming hall (II)

The best performance in predicting water evaporation rate is the ANN model with water evaporation rate and the dichotomized time variable in binary

format as inputs (Ch. 3.2.1.1). Measurement data further suggest that there is no clear correlation between water evaporation rate and indoor moisture content (determined by indoor temperature and indoor relative humidity). Occupant influences are probably one reason and another is the moisture content in return air stream, roughly representing the average hall moisture content which may be different than those above the pools.

"Neural network A" is used to denote ANNs with only water evaporation rate as inputs while "Neural network B" for ANNs with water evaporation rate and time as inputs. Their performance is presented in Figure 18 and Table 5.
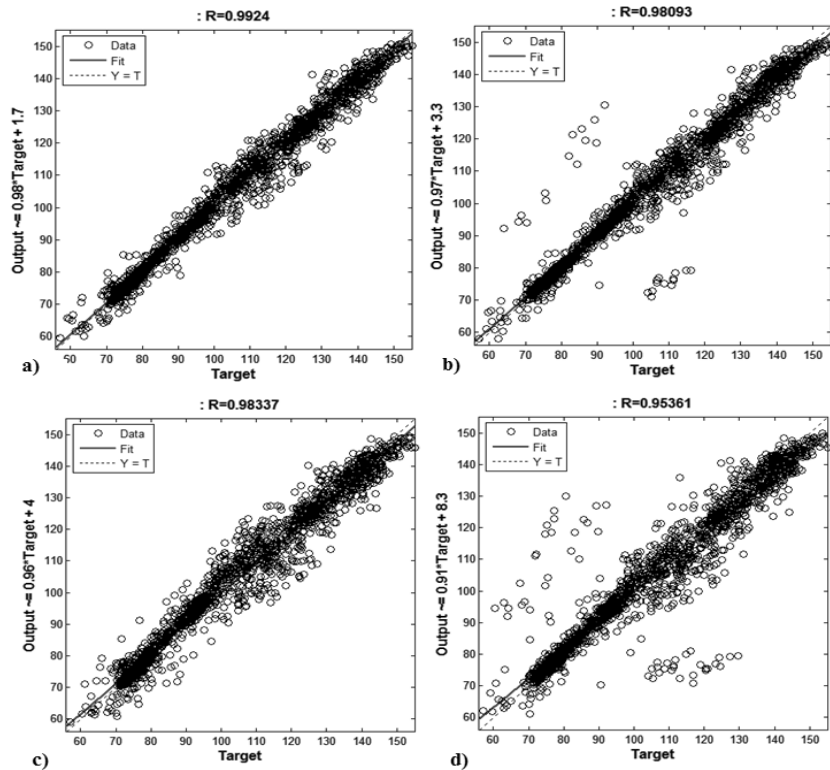
Figure 18. Post-regression results of predicted water evaporation rates (output) and measurements (target). (a) Neural network B (one-step ahead prediction). (b) Neural network A (one-step ahead prediction). (c) Neural network B (two-step ahead prediction). (d) Neural network A (two-step ahead prediction) (Original publication II).

Outliers arise from Neural network A (Figure 18), implying that Neural network A did not capture some extreme fluctuations characterized by Rate Of Change (ROC). ROC indicates the degree of fluctuations of water evaporation rate, which is defined as the percentage of mass change between the current time and the previous time ($=|E(k)-E(k-1)|/E(k-1)$). The larger the ROC is, the more difficult the forecast is.

Outliers of ROC generally occur either at 6:10 a.m. (or 9:10 a.m.) or 10:10 p.m. (i.e. one interval after opening or closing time) due to the reservation of training sessions. Neural network A suffers from one-interval-delay (10 min) in predicting the jump at 6:10 a.m. (or 9:10 a.m.) The same happens with the drop

at 10:10 p.m. More detailed studies, results not shown here, reveal two-interval-delay prediction with two-step ahead prediction of Neural network A.

However, by introducing time into the inputs, Neural network B can accurately detect the big jump at 6:10 a.m. (or 9:10 a.m.) and drop at 10:10 p.m. without time delay for one- and two-step ahead predictions. As for one-step ahead prediction, the biggest difference between predicted and measured at 6:10 a.m. (or 9:10 a.m.) and 10:10 p.m. is only 8.5 kg/h for Neural network B while up to 38.62 kg/h for Neural network A. Consequently MSE decreases from 23.45 to 9.388 (Table 5) and maximum APE from 44.98% (Neural network A) to 12.65% (Neural network B). For two-step ahead prediction, even though forecast qualities decline for both as expected, Neural network B still shows superiority over Neural network A (see maximum APE 21.45% for Neural network B vs. 60.96% for Neural network A). Neural network B is much more efficient and effective than Neural network A in predicting the extremes of water evaporation rate.

Table 5. Comparisons between Neural network A (only water evaporation rate as inputs) and Neural network B (water evaporation rate and a binary format of time as inputs) (Original publication II).

Time step	MSE (Mean Square Error)	MAE (Mean Absolute Error)	MAPE (%) (Mean Absolute Percentage Error)
Neural network A (only water evaporation rate as inputs)			
One-step ahead	20.770 (training) 23.446 (testing)	2.35 (training) 2.392 (testing)	2.37 (training) 2.298 (testing)
Two-step ahead	46.275 (training) 56.424 (testing)	3.6 (training) 3.806 (testing)	3.7 (training) 3.71 (testing)
Neural network B (water evaporation rate and a binary format of time as inputs)			
One-step ahead	9.314 (training) 9.388 (testing)	1.937 (training) 2.0389 (testing)	1.912 (training) 1.949 (testing)
Two-step ahead	17.64 (training) 20.49 (testing)	2.6 (training) 3.079 (testing)	2.56 (training) 2.96 (testing)

Considering the fact that water evaporation rate fluctuates much more than indoor temperature and relative humidity during operating hours (revealed by measurements), the results of Neural network B are good with a one-step ahead prediction and acceptable with two-step ahead prediction. Neural network A gives unacceptable results during shifts between non-operating and operating. Strong overfitting (i.e. the error on the training set is very small, but when new data is presented to the network the error is large) is not observed from Neural network A and Neural network B for one- and two- ahead predictions, and the performance of Neural network B are very near between training and testing for one-step ahead prediction (Table 4), indicating that Neural network B is capable to generalize.

Results also show that Neural network B performs excellent for non-operating hours (see Table 4 in II). Due to high fluctuation of water evaporation rate and

lack of occupant information, prediction accuracies in operating hours are in general worse than those in non-operating hours. But Neural network B are still able to achieve good forecasts for one-step ahead prediction and acceptable forecasts for two-step ahead predictions in operating hours (see Table 4 in II).

4.4 Prediction of indoor temperature and relative humidity (VI)

When using NNARX to construct ANN, only $T_i(t-1)$ and $T_i(t-2)$ were finally identified as significant for indoor temperature prediction while $RH_i(t-1)$, $T_i(t-2)$, $T_o(t-2)$ and $RH_o(t-1)$ were significant for indoor relative humidity prediction (Ch. 3.2.1.2).

When using the genetic algorithm to search optimal ANN models, the five best models were obtained after running 200 generations for indoor relative humidity (the best one is $RH_i(t-1)RH_i(t-2)$ $RH_i(t-3)RH_i(t-5)$ with the number of hidden units=6) while the two best models for indoor temperature were obtained after 100 generations (the best one is $T_i(t-1)T_i(t-3)T_i(t-5)$ with the number of hidden units=6) (Tables 1 and 2 in VI).

Tables 6 and 7 show the comparisons of one-, two-, three- and four-step ahead indoor temperature and relative humidity predictions for NNARX and the genetic algorithm.

Table 6. Performance comparison of indoor temperature prediction between networks (Original publication VI).

Input signals	One-step ahead prediction	Two-step ahead prediction	Three-step ahead prediction	Four-step ahead prediction
$T_i(t-1)$, $T_i(t-2)$ (NNARX)	Testing: MSE=0.0062 MAE=0.048 Training: MSE=0.0185 MAE=0.047	Testing: MSE=0.0162 MAE=0.082	Testing: MSE=0.0313 MAE=0.1138	Testing: MSE=0.0533 MAE=0.1461
$T_i(t-1)$, $T_i(t-3)$, $T_i(t-5)$ (Genetic algorithm)	Testing: MSE= 0.0080 MAE=0.056 Training: MSE=0.023 MAE=0.0608	Testing: MSE=0.029 MAE=0.110	Testing: MSE=0.057 MAE=0.2137	Testing: MSE=0.0909 MAE=0.250

Table 7. Performance comparison of indoor relative humidity prediction between networks (Original publication VI).

Input signals	One-step ahead prediction	Two-step ahead prediction	Three-step ahead prediction	Four-step ahead prediction
$RH_i(t-1), T_i(t-2)$, $T_o(t-2), RH_o(t-1)$ (NNARX)	Testing: MSE=0.239 MAE=0.14 Training: MSE=0.0805 MAE=0.0974	Testing: MSE=0.7331 MAE=0.2568	Testing: MSE=1.3309 MAE=0.3562	Testing: MSE=1.9242 MAE=0.4445
$RH_i(t-1), RH_i(t-2)$, $RH_i(t-3), RH_i(t-5)$ (Genetic algorithm)	Testing: MSE= 0.1584 MAE=0.1717 Training: MSE=0.0446 MAE=0.0624	Testing: MSE=0.3617 MAE=0.4262	Testing: MSE=0.822 MAE=0.7255	Testing: MSE=1.5021M AE=0.9996

Indoor temperature predictions perform much better than those of indoor relative humidity. Theoretically, relative humidity is governed by the nonlinear diffusion equation while indoor temperature is governed by a linear diffusion equation, e.g. thermal conductivity is assumed to be constant or linear in almost all buildings while the moisture diffusivity coefficient is a highly nonlinear function of moisture or humidity contents for most materials. Indoor relative humidity is more difficult to predict. Both methods (i.e. NNARX and genetic algorithm) can provide four-step ahead indoor temperature predictions with great accuracies. However only with the genetic algorithm, a satisfactory accuracy for a two-step ahead indoor relative humidity prediction can be obtained because, after two-step ahead prediction, MAE is smaller than MSE, indicating that some errors are getting much greater than one (>1). These facts also indicate that the indoor relative humidity may be influenced by more factors than the indoor temperature. Indoor temperature predictions by both methods perform better in testing stage than in training stage, indicating that obtained networks are capable to generalize. The prediction intervals are displayed in Table 8.

Table 8. Results for 95% prediction intervals (Original publication VI).

	T_i (NNARX)	T_i (Genetic algorithm)	RH_i (NNARX)	RH_i (Genetic algorithm)
The number of testing Points	1228	1228	1118	1118
Average interval width (full)	0.5945	0.4797	1.8677	1.1844
The number of targets which don't fall in the prediction interval	6	22	55	76

Table 8 shows that both models for indoor temperature prediction possess very good prediction intervals, in which only few targets are not in intervals. On the other hand for indoor relative humidity prediction, many targets are out of prediction intervals. Probably this is because indoor humidity is more dynamic and uncertain as previously described.

However, the exact binomial distribution shows that there will be probably (92.6% probability) 45-80 targets falling outside the prediction interval. Therefore, the prediction interval still works well for indoor relative humidity prediction. Moreover, these results also provide more evidence that the prediction of indoor relative humidity is more difficult as large prediction intervals were obtained when predicting indoor relative humidity.

4.5 Prediction of space air exchange rates and CO_2 generation rates (VII)

Although the pressure difference between the return air vent and space was measured, there was no direct measurement available for airflow rate due to technical difficulties (Ch. 3.2.3). Therefore, an extra equilibrium analysis, Eq. (3), was adopted as a supplement tool to analyze results. Eq. (3) is used to estimate airflow rate (i.e. space air change rate) if an equilibrium of CO_2 concentration is reached. Fortunately, on 22.9.2008 and 13.10.2008, one person was present in the office for a long time which allowed reaching near-equilibrium. On 22.9.2008, only one person worked in the office for nearly the whole afternoon (15:40–17:20) with a number of visitors for less-than-five minute visits. The measured indoor CO_2 concentration is near constant (from 488 ppm to 479 ppm). The average value, 481 ppm, was then taken as the equilibrium concentration value. The CO_2 generation rate for this person was estimated as 0.0052 L/s based on his size. Similarly, the near-equilibrium was observed at an even longer period of 14:25–17:00 on 13.10.2008. Again the average value, 681 ppm, was used as the equilibrium concentration value. Tables 9 and 10 show the estimated space air change rates from the equilibrium analysis and MLE on 22.9.2008 and 13.10.2008 as well as comparisons with other days.

On 22.9.2008, 23.9.2008 and 24.9.2008, all periods have close pressure differences, indicating the space air change rate is nearly constant. The results (Table 9) verifies this and the space air change rates estimated by MLE coincide with those from the equilibrium analysis. Table 10 shows similar results for 13.10.2008 and 15.10.2008.

Table 9. Ventilation rates estimated from equilibrium analysis and MLE on 22.9.2008, 23.9.2008 and 24.9.2008 (Original publication VII).

Date	Method	Ventilation rate (α_1, ach)	Supply CO_2 concentration (α_0, ppm)	Pressure difference (Pa)[a]
22.9.2008	Equilibrium analysis	2.93	401[b]	58
	MLE	2.92	401[b]	56
23.9.2008	MLE	2.94	378	56
24.9.2008	MLE	2.92	370	55

[a] This is average pressure difference between the room and return air vent for the estimated period
[b] Supply CO_2 concentration estimated by MLE

Table 10. Ventilation rates estimated from equilibrium analysis and MLE on 13.10.2008 and 15.10.2008 (Original publication VII)

Date	Method	Ventilation rate (α_1, ach)	Supply CO_2 concentration (α_0, ppm)	Pressure difference (Pa)
13.10.2008	Equilibrium analysis	0.77	378	92
	MLE	0.74	378	91
15.10.2008	MLE	0.76	387	90

The four-step method the method described in Ch. 3.2.2.1 was employed to estimate the number of occupants, and then compared the results with records from diaries (Figure 19).
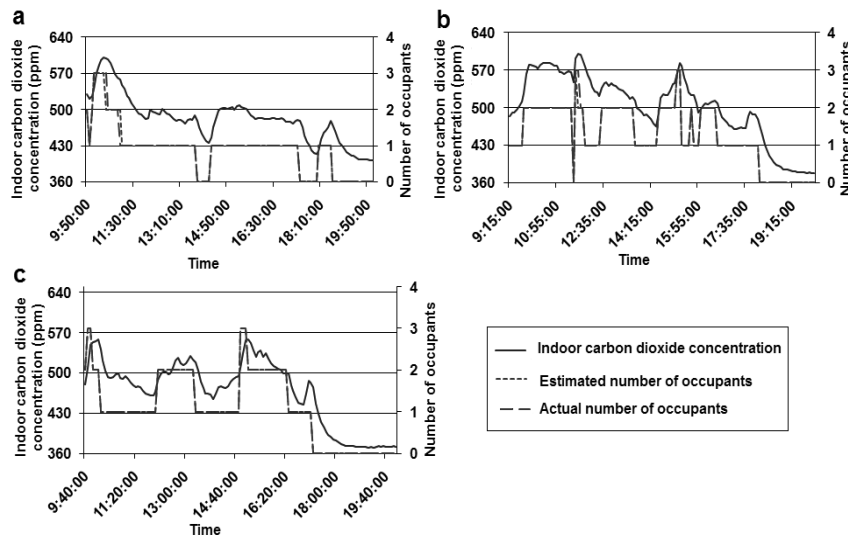
Figure 19. The estimated and recorded numbers of occupants vs. measured indoor CO_2 concentrations on: (a) 22.9.2008, (b) 23.9.2008 and (c) 24.9.2008 (Original publication VII).

A significant change on indoor CO_2 concentrations does not necessarily mean a change of the number of occupants. For example, on 23.9.2008, one significant change from 461 ppm to 484.2 ppm was observed, which seemed to be associated with the change of occupants. However, using the developed model, the number of the occupants was found to be the same as before. An informal

request from the person revealed that he actually did a little excise so as to alleviate the tiredness from the long-time work. Human activity has significant impact on CO_2 generation rate, and sometimes misleads judgments on the number of occupants. The developed model can correctly estimate the number of the occupants in such complicated case.

4.6 Prediction of supply air temperature efficiency for run-around heat recovery systems (VIII)

Tuned signature powers (by observing the nearness of two supply air temperature curves for Sites A, B and D and by comparing the sum of the square difference between two curves for Site C) are listed in Table 11 (Ch. 3.2.4). Table 11 also shows the performance of the estimated supply air temperature increment for sites A, B, C and D. The estimation accuracy of supply air temperature increment is very good for all the sites.

Figure 20 shows an example of how to use tuning algorithm to search the correct signature power α from Groups A and B (Site B).
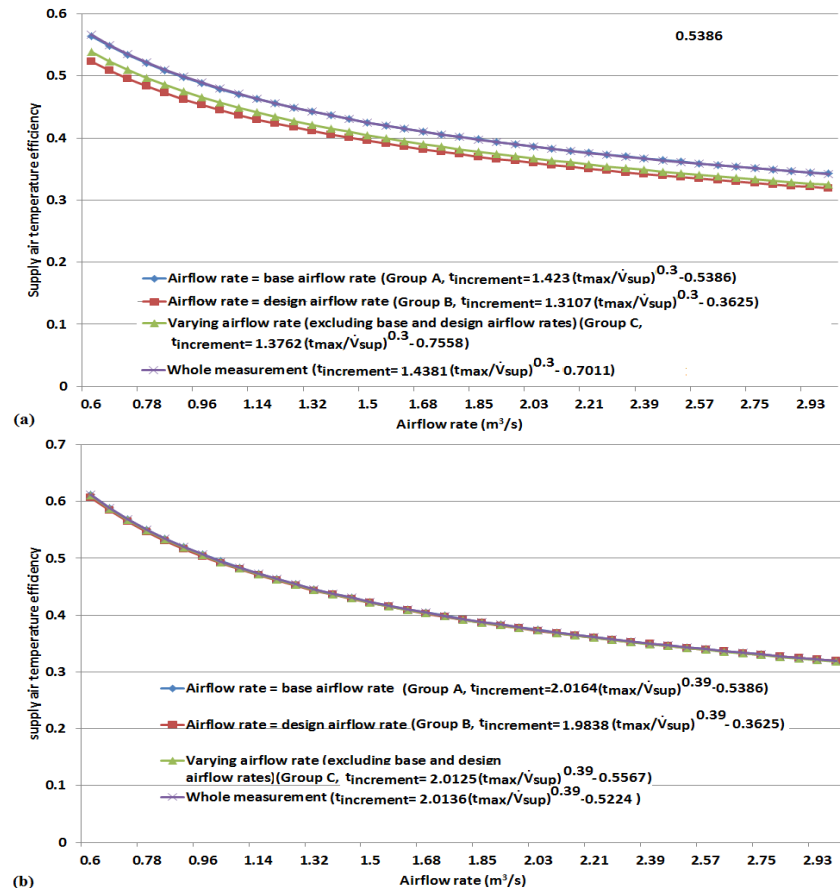
Figure 20. Supply air temperature efficiency curves for Site B (airflow rate range: 20% to 100% of design air flow rate) (a) α (signature power) = 0.3. (b) α (signature power) = 0.39.

When α is set as 0.3, the supply air temperature efficiency curves generated by Groups A (the base ventilation rate) and B (the design ventilation rate) have clear differences. When α is tuned to 0.39, supply air temperature efficiency curves from Groups A and B are very close. Thus the correct signature power should be around 0.39. Although signature power is determined by Groups A and B, the curve from Group C is nearer to those from Groups A and B when signature power is tuned (Figure 20 (b)). This indicates the correct signature power should give similar supply air temperature efficiency curves regardless of airflow rate. Similar results can be seen from other sites. For examples, the maximum mean of the sum of the square difference between two supply air temperature efficiency curves (i.e. $\frac{\sum_{k=1}^{N}[y1(k)-y2(k)]^2}{N}$) among all groups is: 1.3e-05 (Site A, α=0.4), 5.8e-06 (Site B, α=0.39), 7.3e-07 (Site C, α=0.69) and 1.9e-05 (Site D, α=0.67).

Table 11. Tuned signature powers (α) and corresponding performance of supply air temperature increment for sites (Original publication VIII).

	Site A* (α=0.4) (Group C)	Site B* (α=0.39) (Group C)	Site C** (α=0.69) (Groups A B, and unseen data)	Site D* (α=0.67) (Group C)
MSE (Mean square error)	0.189	0.15	0.12	0.09
MAE (Mean Absolute Error)	0.33	0.28	0.26	0.24
MAPE (%) (Mean Absolute Percentage Error)	4.6	3.82	6.65	2.11
R^2 (Coefficient of Determination)	0.983	0.982	0.93	0.97

*Group C was used as testing. The linear equation of supply air temperature increment, Eq. (19), was obtained from Groups A and B.

** unseen data (about 4000 samples) of Groups A and B were used as testing. The linear equation of supply air temperature increment, Eq. (19), was obtained from Groups A and B.

As can be seen in Figs. 20, no matter what signature power is used, two supply air temperature efficiency curves from Group A and the whole measurement are always very close. The reason is that measurement data are dominated by Group A for Site B (88% for Site B, similar case for Site A, about 77%, see Fig. 3 in VIII). Thus, the signature power obtained from the whole measurement (using the least squares to find α in Eq. (19)) is biased and must be tuned. Ideally two groups should have two very different ventilation rates. In this way, a small deviation from the correct signature power could give an obvious difference between two supply air temperature efficiency curves so that the signature power can be approximated by tuning and observation. Otherwise manual tuning would be difficult.

The inversely proportional relationship between supply air temperature efficiency and a power of airflow rate (Eq. (17)) was also used to directly estimate supply air temperature efficiency in this research. The accuracy is reasonable (using the least squares to search optimal power α and factor k), for example, R^2 is about 0.88 for sites A and B (sites C and D are CAV systems) although the newly developed physical law in this study (Eq. (19)) can improve accuracy (R^2 are about 0.93 for sites A and B). Because Eq. (18) is a one-dimensional model (airflow rate is only variable) and in practice there lacks information on airflow rate even for a demanded-control system, multiple results are unavoidable.

Most importantly, Eq. (19) shows maximum temperature difference has the same impact on supply air temperature increment as a power (α) of airflow rate does. Hence, as a 2-dimentional model, the new physical law (Eq.(19)) takes advantage of sufficient information of the maximum temperature difference to compensate for the lack of information of airflow rate in order to find the "exact" signature power. The new physical law is more applicable than one-dimensional model, i.e. Eq. (17)

4.7 Accuracy improvement of building simulation (V)

The results of Hybrid model A are given in Table 12 and Figure 21 (Ch. 3.3)

Table 12. Performance for the hybrid model (Combination of numerical model and neural network) and numerical models (Original publication V).

	Numerical model of Hybrid model A and Hybrid model B	Testing numerical model	Hybrid model A
MSE (Mean Square Error)	4.98	0.97	0.43
MAE (Mean Absolute Error)	2.03	0.77	0.5
MAPE (%) (Mean Absolute Percentage Error)	8.66	3.4	2.2
IA (Index of Agreement)			0.95
R^2 (Coefficient of Determination)			0.87
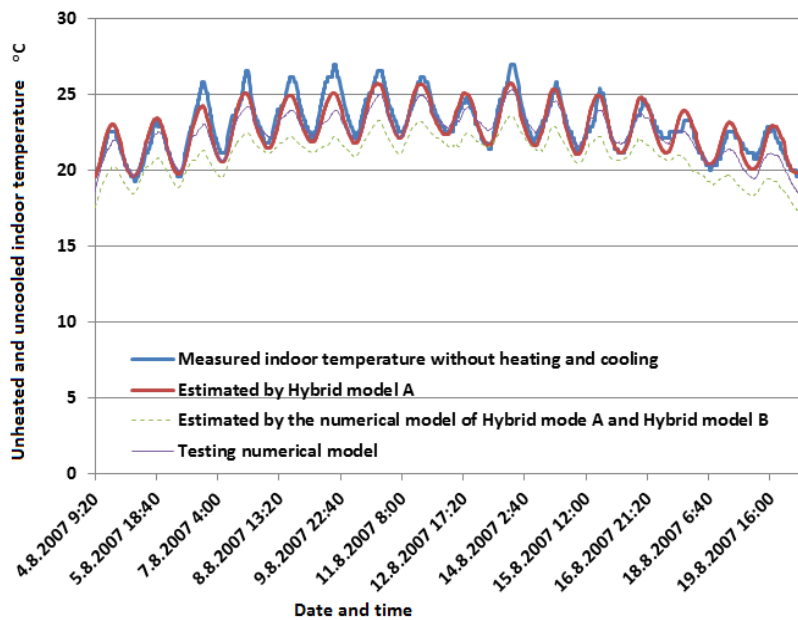
Figure 21. Comparisons of measured indoor temperature and simulated temperatures by the testing numerical model, the numerical model of Hybrid model A and Hybrid model B, and Hybrid model A (Original publication V).

Due to the inclusion of internal gains (i.e. lighting and supply air fan), the accuracy is improved greatly by the testing numerical model in comparison to the numerical model of Hybrid model A or Hybrid model B. A further improvement is achieved by Hybrid model A by decreasing the maximum APE

from 11.5% (the testing numerical model) to 6.5% (the hybrid numerical-ANN model).

Performance gaps between operating hours and non-operating hours are much larger in the numerical model than in Hybrid model A, indicating the ANN (in Hybrid model A) effectively captures errors that are related to operating hours such as lighting, people, solar radiation through exterior walls and roof, and others (see Table 3 in V). Some errors from the numerical model cannot be removed by the ANN even if in non-operating hours because there is no lighting, people, solar radiation through exterior walls and roof, equipment etc. These errors may be due to algorithms, assumptions, and the simplified physical process (Malkawi and Augenbroe 2003). Nevertheless the accuracy is much improved during non-operating hours, where the biggest difference between the simulated indoor temperature (by Hybrid model A) and the measured one is only 1.2 ^{o}C (4.2 ^{o}C for the numerical model of Hybrid model A or Hybrid model B, and 2.4 ^{o}C for the testing numerical model).

Error analysis for the numerical model is extremely difficult and requires expertise in the fields of building systems, building physics, simulation techniques, validation techniques, uncertainty analysis techniques, statistics, and probability theory. As can be seen in Figure 21, the simulated indoor temperatures from the numerical model of Hybrid model A or Hybrid model B, and the testing numerical model are obviously declining during non-operating hours from 16.8.2007 to 19.8, 2007 due to decreasing outdoor temperature. But the measured data set does not behave in the same fashion. During non-operating hours (22:00 p.m. – 9 a.m.) from 18.8.2007 to 19.8.2007, the measured temperature is actually higher than the one from 17.8.2007 to 18.8.2007, leading to worse numerical results (i.e. the numerical model of Hybrid model A or Hybrid model B: MAE=2.7 and MSE=7.6, 22:00 18.8.2007 – 8:40 19.8.2007. The testing numerical model: MAE=1.44 and MSE=2.1, 22:00 18.8.2007 – 8:40 19.8.2007). The physical explanation for such a phenomenon is difficult to make because the internal heat gains, e.g. occupants, lighting, solar radiation through exterior walls and roof, are very small and can be ignored during non-operating hours. However, the result from Hybrid model A basically follows the trend of measurement, giving MAE =0.53 and MSE =0.43 (22:00 18.8.2007 – 8:40 19.8.2007). The performance does not change much (see Table 3 in V). This strongly indicates the ability of ANN to handle highly nonlinear problems with uncertainties.

The results of Hybrid model B for a part of operating hour data (9 – 12 a.m.) are given in Table 13.

Table 13. Performance of the hybrid model (Combination of numerical model and neural network, trained by data only from non-operating hours) for a part of operating hours (9 a.m. – 12 a.m.) vs. numerical models (Original publication V).

	Numerical model of Hybrid model A and Hybrid model B	Testing numerical model	Hybrid model B
MSE (Mean square error)	3.77	0.57	0.28
MAE (Mean Absolute Error)	1.8	0.59	0.4
MAPE (%) (Mean Absolute Percentage Error)	8.67	2.85	1.95
IA (Index of Agreement)			0.95
R^2 (Coefficient of Determination)			0.87

The ANN of Hybrid model B was built from non-operating hours but its performance for operating hours (9-12 a.m.) are even better than those for non-operating hours for Hybrid model A. This reveals one very important fact that the hybrid numerical-ANN model can realize the generalization for a building parameter, which is constant in operation, with enhanced accuracy in comparison to a numerical model. Such building parameter cannot be generalized by ANN alone. In this study, Hybrid model B first models the influence of ventilation rate on indoor temperature in the numerical model (the second term on the right of Eq. (20)), and then uses the output of the numerical model (i.e. simulated indoor temperature) as the only input for the ANN. Unlike ventilation rate (constant in non-operating hours), the output of the numerical model is ranging from 15 ℃ to 23 ℃ in non-operating hours, implying that the ANN can possibly generalize about the output of the numerical model. As for a new ventilation rate value, its impact is reflected in the output of the numerical model. Due to the generalization for the output of the numerical model, the exact indoor temperature can be predicted by the ANN.

5. General Discussion

The principle aim of this study is to improve the energy efficiency of buildings through modeling, simulation and control approaches. Four sub-objectives corresponding to the three methodological perspectives were previously presented in Ch. 2.5. The following discussion follows along the same lines. Finally, the main contributions of the study are discussed. A summary of the novelty values of this dissertation work is presented in Table 14.

5.1 Data centers in Finland

Cold weather gives data centers more free cooling in Finland, resulting in low and satisfactory PUE. The study (I) indicates that the indoor air management of data center is not optimized, for example, cold air is oversupplied in data centers, not only in this study but also in other data centers investigated in Finland. This also hints that, through improving air management, it is possible to get more energy savings from CRAC fans than from compressors due to the cold weather conditions in Finland (e.g. using DCV). Another big issue is that waste heat recovery systems are commonly lacking in Finnish data centers.

There are two common ways to reuse waste heat. One is to supply waste heat to local buildings. This method is relatively simple and does not create a high demand for supply water temperature. However, the disadvantage is that waste heat cannot be completely used in summer. It is technically difficult to store waste heat from the summer for use during the winter. This problem can be partially solved by supplying more buildings, but it makes piping more expensive. The other solution is to sell heat to energy companies, i.e. joining the district heating network. The advantage is that waste heat can be reused totally, but the disadvantage is that energy companies often have restrictions on supply water temperature (e.g. 60^+ ^{o}C for supply water temperature in Finland), which could reduce the efficiency of compressors. There may also exist an ideal solution to combine both approaches, e.g. selling waste heat to the energy company in summer and providing heat to local buildings for the rest of the time to utilize waste heat efficiently. Therefore it is still challenging to choose a proper method for reusing waste heat from data centers.

5.2 Predictive modeling

ANN is a powerful "black box" model and is very efficient at forecasting and predicting a building parameter where the governing equation is not available

and relevant information is lacking. For example, although occupants have great impact on the water evaporation rate, it is too complex to model physically. However ANN is capable of forecasting the water evaporation rate in a swimming hall without exact information about the occupants (II). This capability is also manifested in predicting indoor humidity using only outdoor and indoor humidity as inputs in a machine room with complicated thermal situations (VI).

However, if the governing equation is well-defined, regression sometimes performs better in predictive modeling than ANN does. For example, Publication (VIII) suggests a linear relationship between the supply air temperature increment (across the supply air heat exchanger in the air supply duct) and the maximum temperature difference (between exhaust and supply airstreams) divided by a power of the supply airflow rate in run-around heat recovery systems. This relationship is designed by linear regression to be able to predict the supply air temperature efficiency for CAV systems with two ventilation rates and also for demand-controlled ventilation systems. Excellent results (Table 11 and Ch. 4.6) were obtained when testing this relationship. In a similar fashion, ANN has been used to predict the supply air temperature increment using the supply airflow rate and the maximum temperature difference as inputs to estimate supply air temperature efficiency for Sites A and B, which are both demand-controlled systems (not a part of this dissertation work). Unfortunately, the performance was unsatisfactory for the ranges of airflow rate that are not in the measurement data, e.g. estimated supply air temperature efficiency jumped to near 100%, which is unrealistic. The reason is that even a demand-controlled ventilation system cannot normally give adequate information on the supply airflow rate to allow ANN to make generalization. Regression, on the other hand, utilizes this relationship as well as the fact that (1) the information on maximum temperature difference is often adequate and (2) the maximum temperature difference has the same impact on supply air temperature increment as the power of airflow rate (VIII).

5.3 DCV control strategies

In Finland, fans in ventilation systems are normally equipped with a frequency inverter, but run mostly in a two speed model (one ventilation rate for daytime and another one for nighttime). This provides an opportunity for a DCV strategy similar to the one developed in (III and IV) to optimize the system.

Even though the novel DCV control strategy (III and IV) would seem to be very different from PID, they share some common features. The both offer two-dimensional capability for modeling both temporal and quantity results, which is significantly different from proportional control. They can be referred to as a 2-dimensional (2D) DCV model. Proportional control adjusts ventilation rates based only on the indoor CO_2 concentration and thus it is a one-dimensional (1D) DCV model. The novel DCV control strategy improves the calculation based upon the indoor CO_2 concentration, the quantity difference in the time dimension, i.e. the difference between indoor CO_2 concentration and CO_2 set

point in the training (or pseudo training) remaining time, and the quantity change rate related to the CO_2 generation rate. In a similar fashion, PID control determines the ventilation rate based on the difference between the indoor CO_2 concentration and the CO_2 set point, the quantity accumulation in time dimension i.e. integral term, and the quantity change rate i.e. derivative term (Eq. (5)). Compared with one-dimensional DCV proportional control, the two-dimensional control strategies have obvious advantages in their quick response to indoor CO_2 changes and reducing the difference between indoor CO_2 and CO_2 set points.

5.4   Hybrid numerical-ANN model

In the case of unheated and uncooled indoor temperature simulation, the estimated indoor temperature of the numerical model can be considered as a historical value of (true) indoor temperature (Ch. 3.3). They (the estimated indoor temperature from a numerical model and the measured one) are linked by the physical governing equation of indoor temperature (Eq. (20)). This is the reason why the estimated indoor temperature of numerical model can be used as input to predict measured indoor temperature by ANN in hybrid numerical-ANN model (e.g. Hybrid model A and Hybrid model B). Therefore, the key to (1) successfully calibrating a numerical model by ANN and (2) enhancing the generalization of ANN is the numerical model which must be physically correct. Assumptions, simplifications of the physical process, and discretization must be carefully considered. Material properties and types must be correctly selected. It is important to point out that, in order to study a parameter (i.e. generalization), its impacts on the output must be modeled correctly by the numerical model. Using unheated and uncooled indoor air temperature simulation as an example in studying the ventilation rate, its (ventilation rate) impacts include: (1) adding or removing heat to/from the indoors by the outdoor air and (2) adding heat to the indoors by the supply air fan. The first impact is significant, and it must be considered and modelled as accurately as possible in the numerical model. The second part is less significant and can be ignored. However if the first part is not modelled in the numerical model but is included in the ANN model by adding the ventilation rate as input, the accuracy may be increased but the generalization may be lost, particularly for a CAV system, which cannot provide adequate ventilation rate information for this purpose.

5.5   New contributions of the study

This thesis presents a multidisciplinary and wide-ranging modeling methodology and novel control approaches for enabling new possibilities for improving both simulation accuracy and the energy efficiency of building systems. Specifically, data centers and DCV were the focus because they represent "multifunctional buildings" and require a more general control strategy applied to both energy and indoor environmental issues. The study is

not limited to conventional outdoor air systems but also includes the air delivery systems of data centers. It is the first systematic and complete case study of a data center in Finland where the infrastructure, energy and air management performance, and waste heat recovery system of a data center were investigated, quantified and discussed. Particularly, long term power consumption measurement data (one year) make the investigation and energy end use breakdown more convincing compared with existing research studies that contain only short term power measurement data. Potential of waste heat reuse was quantified with the help of energy end use breakdown analysis, which is generally lacking in existing researches on data center. Several technical gaps are filled:

- A novel DCV control strategy was developed to achieve the same performance as PID does. Compared with earlier DCV methods, the novel DCV control strategy is much easier to implement, does not need tuning and it is universal. It is particularly suitable for system retrofitting. In addition, a 2-dimensional new theory was recognized first time, that is, an effective DCV should adjust ventilation rate in time and quantity (e.g. CO_2 concentration) dimensions. This provides a guideline for developing effective DCV control strategy.
- For the first time, ANN was applied to forecast the water evaporation rate for indoor swimming pools without the knowledge of occupant information. Most existing studies concentrate on developing empirical equations derived from experiments to estimate the water evaporation rate for indoor swimming pools. These equations are for the purposes of sizing the air conditioning equipment as well as energy consumption calculations (not for control). They cannot be easily applied to occupied swimming pools because occupant influences on water evaporation rate are significant, quite random, and very difficult to estimate.
- An innovative hybrid numerical-neural-network framework was developed to maximize the benefits of both the numerical model and ANN so that (1) the accuracy of the numerical model is enhanced by ANN without error analysis (e.g. building audit, survey, documentation, specification and so on), which greatly simplifies the calibration process; (2) the generalizability of ANN is improved by the numerical model (e.g. building parameters, constant in practice, can be analyzed). To the best of our knowledge, this is the first study to apply this hybrid framework in building simulation.
- A new field measurement based physical law for run-around heat recovery systems was proposed by suggesting that a simple linear relationship can be set up between the supply air temperature increment across the supply air heat exchanger in the air supply duct and a parameter. This parameter is equal to the temperature difference between the exhaust and supply airstreams divided by a power (i.e. signature power) of supply airflow rate. Most relevant research focus on either the numerical model, which is computationally intensive, or ANN,

whose training data are either generated from laboratory experiment or a numerical model in order to achieve better generalization. Laboratory experimentation is difficult, costly, and does not accurately represent the actual conditions. Field measurement based studies for run-around heat recovery systems, which can show the actual behavior of a run-around heat recovery system, are lacking in the literature. The new physical law will likely be valuable to industry. For example, if manufacturers can provide the system characteristic value proposed by the new physical law (i.e. signature power) with their products, performance assessment will become very easy, even for CAV systems with only one ventilation rate for the run-around heat recovery systems.

Some improvements were made to existing techniques and research:

- The prediction interval was first introduced for the prediction of indoor temperature and humidity by ANN for uncertainty analysis.
- Transient detection algorithm and equilibrium analysis were coupled to detect the number of occupants.
- The Maximum Likelihood Estimation were combined with the power law relationship for openings to determine time varying space air change rates.

A summary of major findings and novelty values of this thesis work is presented in Table 14.

Table 14. A summary of the novelty values of the research works.

Paper	Finding	Novelty value
I	– inlet conditions (racks) were all within the ASHRAE recommended or allowable ranges – notable recirculation air was not observed – bypass air existed – the total power consumption of CRAC fans was close to the chiller – PUE between 1.2 and 1.5 depending on whether free cooling was on or not – waste heat recovery system was lacking	– the first systematic and complete case study of a data center in Finland – long term measurements for IT and facility powers (one year) – quantification of air management and waste heat reuse
II	– the inclusion of indoor temperature and indoor relative humidity (in the return duct) in inputs did not improve the accuracy of ANN – using only water evaporation rate as ANN inputs gave unacceptable results during shifts between non-operating and operating hours – the best performance came from the ANN model with water evaporation rate and the dichotomized time variable in binary format as inputs – strong overfitting was not observed for all ANN models	– first ANN model for forecasting of the water evaporation rate in indoor swimming halls
III, IV	– CO_2 mass balance equation (Eq. (8)) can offer great help for developing effective DCV algorithms – for the case of unscheduled opening hours, settings (e.g. the lengths of the pseudo training session and pseudo break) affect fan energy consumption for the developed novel DCV	– a novel, physically based (CO_2 mass balance equation) and tuning free DCV model, which is as effective as PID – identification of 2-dimensional new theory for effective DCV strategy
V	– ANN can calibrate the numerical model without error analysis – the numerical model can improve the generalizability of ANN – the numerical model can be simplified in a hybrid numerical-neural-network model – The output from a numerical model can be considered as historical value of the true result	– innovative hybrid numerical-neural-network model for building simulation, where the numerical model serves as an estimator and ANN acts as a calibrator
VI	– it is more difficult to predict indoor humidity than indoor temperature for ANN – ANN models constructed by both NNARX and the genetic algorithm perform better in testing than in training for predicting indoor temperature	– prediction intervals for the predictions of indoor temperature and humidity
VII	– the power law relationship for openings can assist in predicting space air change rate for time-varying ventilation system – A significant change on indoor CO_2 concentrations does not necessarily indicate a change of the number of occupants	– new MLE model for predicting constant and time-varying space air change rates – the coupled model by transient detection algorithm and equilibrium analysis for detecting the number of occupants in offices
VIII	– one-dimensional model (Eq. (17)) leads to multiple solutions for predicting supply air temperature efficiency – ANN often results in unrealistic results for unseen range of airflow rates when predicting supply air temperature efficiency – supply air temperature efficiency curves generated by different groups of data are close for the correct signature power	– new physical law for supply air temperature increment (Eq. (18)) – tuning algorithm for determining signature power

6. Conclusions and further research

This thesis is composed of several studies which address and discuss issues, methods, algorithms, and possibilities for improving system performance and the accuracy of simulation in buildings. The details, overview, and contributions of these studies are well presented in original papers and summarized in this thesis. From these studies, some implications can be stated as follows:

- Data centers cover almost all issues of buildings, such as indoor environment control, cooling, air movement management, smart control, sustainability, and waste heat reuse. Due to the situation with data centers and weather conditions in Finland, there is an urgent need to reuse waste heat from data centers to benefit other buildings, particularly in winter
- ANN is excellent at sorting out complex patterns from measurement data with limited knowledge. This feature makes ANN very powerful in forecasting and predicting not only the behavior of dynamic systems but also that of static systems. It is also easy to implement ANN. Generally we just need to decide on the inputs for an ANN.
- If the governing equation is well-established, regression can achieve better generalization in predictive modeling than ANN.
- An efficient DCV algorithm should have the capability of updating ventilation rates based on time and quantity dimensions, referred to as the 2-dimensional DCV model. The one-dimensional DCV model (i.e. updates results based on the quantity dimension, such as proportional control) often over-ventilates the space.
- Finally, methodologically, each of physical, data-driven and their combined models has advantages in certain applications. Each of these models has advantages in certain flow categories and application limits. Hybridization of different modeling techniques shows promise for handling various types of uncertainty in complex building systems.

More research is clearly needed. One issue for future research is to find a better way to account for occupant impacts on the water evaporation rate for ANN. CO_2 is a good indicator of occupant number and activity. However, evaporation may not be in direct proportion to CO_2 concentration because an occupant could be a spectator, thus making no or little contribution to evaporation. It is also important in the future to extend the novel CO_2-based DCV developed in (III) and (IV) to other types of DCVs, such as humidity- and non-occupant-related-

pollutant-based DCVs. The prediction interval will be further applied in the hybrid numerical-neural-network model (V) for uncertainty analysis so that the hybrid-neural-network model can give not only point estimate but also interval estimate. The concept of a "green data center" will be another important issue of future research. This includes issues of smart cooling control, using waste heat effectively, and utilizing natural resources (e.g. solar or wind) to generate electricity.

References

3M (2015). Two-Phase Immersion Cooling: A revolution in data center efficiency.

Abu-Mostafa, Y.S. (1989). Information theory, complexity, and neural networks. IEEE communication magazine, 22-28.

Akimenko, V.V., Anderson, I., Lebovitze, M.D. & Lindvall, T. (1986). The sick building syndrome. In: Berglund B, et al., editors. Indoor air. Evaluations and conclusions for health sciences and technology, vol. 6. Stockholm: Swedish Council for Building Research, 87–97.

Anderssen, R.S., Bloomfield, P. (1974). Numerical differentiation procedures for non-exact data. Numerische Mathematik 22, 157–82.

Asdrubali, F. (2009). A scale model to evaluate water evaporation from indoor swimming pools, Energy and Buildings 41, 311–319.

ASHRAE. (1999). ASHRAE Handbook HVAC Applications. Atlanta, USA.

ASHRAE. (2007). Standard 62.1-2007 User's Manual. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta.

ASHRAE Standard 62.1-2007 (2007). Ventilation for Acceptable Indoor Air Conditioning Engineers. Atlanta, GA.

ASHRAE. (2009). Design Considerations for Datacom Equipment Centers. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta.

ASHRAE. (2009b). Thermal Guidelines for Data Processing Environments. Refrigerating and Air-Conditioning Engineers, Inc., Atlanta.

ASHRAE Standard 62.1-2010. (2010). Ventilation for Acceptable Indoor Air Quality. American Society of Heating, Refrigeration, and Air Conditioning Engineers, Atlanta, Georgia.

Bash, C.E., Patel, C.D. & Sharma, R.K. (2003). Efficient Thermal Management of Data Centers- Immediate and Long-Term Research Needs, Intl. J. HVAC&R Res. 9, 137-152.

Basheer, I.A. & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. Journal of Microbiological Methods 43, 3–31.

Borg, S.P. & Kelly, N.J. (2012). The development and calibration of a generic dynamic absorption chiller model. Energy and Buildings 55, 533–544.

Cengel, Y.A. (2003). Heat Transfer: A Practical Approach. McGraw-Hill.

Chao, C.Y.H. & Hu, J.S. (2004). Development of a dual-mode demand control ventilation strategy for indoor air quality control and energy saving. Building and Environment 39, 385–397.

Chen, J. & Huang, T. (2004). Applying neural networks to on-line updated PID controllers for nonlinear process control. Journal of Process Control 4, 211–230.

Cho, J., Lim, T. & Kim, B. (2009). Measurements and predictions of the air distribution systems in high compute density (Internet) data centers. Energy and Buildings 41, 1107–1115.

CIBSE Guide H. (2000). Building Control Systems. Butterworth-Heinemann, UK.

Congradac, V. & Kulic, F. (2009). HVAC system optimization with CO_2 concentration control using genetic algorithms, Energy and Buildings 41, 571–577.

Curtiss, P.S., Brandemuehl, M.J. & Kreider, J.F. (1995). Energy management in central HVAC plants using neural networks. In: Haberl JS, Nelson RM, Culp CC. editors. The use of artificial intelligence in building systems. ASHRAE, 199-216.

Derakhshani, R., Schuckers, S.A.C. (2004). Continuous time delay neural networks for detection of temporal patterns in signals. In: Proceedings of 2004 IEEE International Joint Conference on Neural Networks, Vol. 4, Budapest (Hungary), 2723 – 2728.

De Veanux, R.D., Schumi, J., Schweinsberg, J. & Ungar, L.H. (1998). Prediction intervals for neural networks via nonlinear regression. Technometrics 40(4), 273-282.

Eder, H.H. (2010a). PID tuning – science, art or both? Act-control.

Eder, H.H. (2010b). The PID: Hommage to a controller – with some critical remarks. Act-control.

Emerson Electric Co. (2005). Liebert XD, Cooling Solutions for High Heat Density Applications. Liebert Technical Document.

Emerson Electric Co. (2008). Knurr CoolFlex. Knurr Technical Document.

ENERGYPLUS. (2009). Energy Plus Engineering Reference. EERE.

European Commission. (2013). ENERGY-EFFICIENT BUILDINGS: MULTI-ANNUAL ROADMAP FOR THE CONTRACTUAL PPP UNDER HORIZON 2020. European Union.

Ferreira, P.M. & Ruano, A.E. (2001). Predicting the Greenhouse inside Air Temperature with RBF Neural Networks. 2nd IFAC/CIGR Workshop on Intelligent Control for Agricultural Applications, Bali Indonesia.

Ferreira, P.M., Faria, E.A. & Ruano, A.E. (2002). Neural Network Models in Greenhouse Air Temperature Prediction. Neurocomputing 43, 51–75.

Fisk, W.J., Faulkner, D., Prill, R.J. (1991). Air exchange effectiveness of conventional and task ventilation for offices. Report LBL_31652, Berkeley, CA: Lawrence Berkeley Laboratory.

Frausto, H.U. & Peters, J.G. (2004). Modelling Greenhouse Temperature Using System Identification by Means of Neural Networks. Neurocomputing 56, 423–428.

Fulton, J. (2007). A Strategic Approach to Datacenter Cooling. Afco Systems, White Paper.

Geisser, S. (1993). Predictive Inference: An Introduction. New York: Chapman & Hall.

Geng, H. (2015). DATA CENTER HANDBOOK. WILEY.

Gouda, M.M., Danaher, S. & Underwood, C.P. (2002). Application of an artificial neural network for modelling the thermal dynamics of a building's space and its heating system. Mathematical and Computer Modelling of Dynamic Systems 8, 333–344.

Heiselberg, P., Murakami, S. & Roulet, C.A. (1998). Ventilation of Large Spaces in Buildings: Analysis and Predictions Techniques. IEA.

Herrlin, M.K. (2007). Improved Data Center Energy Efficiency and Thermal Performance by Advanced Airflow Analysis. Digital Power Forum 2007, San Francisco, CA, USA.

Herrlin,M.K. (2010). Air Management Research, Application Assessment Report #0912. ANCIS Incorporated.

International Performance Measurement & Verification Protocol Committee. (2001). International Performance Measurement & Verification Protocol: Concepts and

Practices for Improved Indoor Environmental Quality Volume II. OFFICE OF ENERGY EFFICIENCY AND RENEWABLE ENERGY U.S. DEPARTMENT OF ENERGY.

Janczak, A. (2005). Identification of Nonlinear Systems Using Neural Networks and Polynomial Models: A Block-Oriented Approach. Springer.

Jeong , J., Choi, A. &No, S. (2010). Improvement in demand-controlled ventilation simulation on multi-purposed facilities under an occupant based ventilation standard, Simulation Modelling Practice and Theory 18, 51–62.

Johansson, L. & Westerlund, L. (2001). Energy savings in indoor swimming-pools: comparison between different heat-recovery systems. Applied Energy 70, 281–303.

John, G., Clements-Croome, D. & Jeronimidis, G (2005). Sustainable building solutions: a review of lessons from the natural world. Building and Environment 40, 319–328.

Judkoff, R. & Neymark, J. (2006). Model validation and testing the methodological foundation of ASHRAE standard 140. ASHRAE Transactions 112, 367–376.

Kalogirou, S.A., Neocleous, C.C. & Schizas, C.N. (1997). Heating load estimation using artificial neural networks. In: Proc. of CLIMA 2000 Conf., Brussels (Belgium).

Kalogirou, S., Eftekhari, M. & Pinnock, D. (1999a). Prediction of air flow in a single-sided naturally ventilated test room using artificial neural networks. In: Proceedings of Indoor Air' 99, vol. 2, The 8th International Conference on Indoor Air Quality and Climate, Edinburgh, (Scotland), 975-980.

Kalogirou, S., Panteliou, S. & Dentsoras, A. (1999b). Modelling of solar domestic water heating systems using artificial neural networks. Solar Energy 65(6), 335-342.

Kalogeria, S.A. (2000). Applications of artificial neural-networks for energy systems. Applied Energy 67, 17-35.

Kant, K. (2009). Data center evolution: a tutorial on state of the art, issues, and challenges. Computer Networks 53, 2939–2965.

Karlsson, J.F. & Moshfegh, B. (2005). Investigation of indoor climate and power usage in a data center. Energy and Buildings 37, 1075–1083.

Koomey, J.G. (2011). GROWTH IN DATA CENTER ELECTRICITY USE 2005 TO 2010. Analytics Press.

Kurnitski, J. & Seppänen, O. (2008). Trends and drivers in the Finnish ventilation and AC market. In: AIVC.

Lin, P. (2014). How to Fix Hot Spots in the Data Center. APC.

Lee, Y. & Oh, S.H. (1994), Input Noise Immunity of Multilayer Perceptrons, ETRI Journal 16, 35-43.

Lee, Y. & Song, H.K. (1993). Analysis on the efficiency of pattern recognition layers using information measures. Pro IJCNN'93 Nagoya, 2129-2132.

Litiu, A. (2012). Ventilation system types in some EU countries. REHVA Journal.

Longbottom, C. (2012). Evaporative cooling cuts costs in the data centre. ComputerWeekly.

Lonnberg, M. (2007). Variable Speed Drives for energy savings in hospitals. World Pumps 494, 20-24.

Lü, X., Lu, T., & Viljanen, M. (2011). A Case Study in Analysis and Improvement of Energy Efficiency in Data Center. 1st Annual International conference on Construction, Athens, Greece, 20-23.6.2011. pp. 1-1.

Malkawi, A.M. & Augenbroe, G. (2003). Advanced Building Simulation. Spon Press, London, UK.

Manousakis, L., Sankar, S., McKnight, Gregg. Nguyen, T.D. & Bianchini, R. (2016). Environmental Conditions and Disk Reliability in Free-cooled Datacenters. In: the Proceedings of the 14th USENIX Conference on File and Storage Technologies, Santa Clara, CA, USA, 53-65.

Mechaqrane, M. & Zouak, M. (2003). Evolutionary neural network in prediction of indoor temperature in buildings. AMSE Modeling Periodicals 46(7), 11–24.

Mjolsness E., DeCoste D. (2001). Machine learning for science: state of the art and future prospects. Science 293:2051–2055.

Montgomery, R. & McDowall, R. (2008). Fundamentals of HVAC Control Systems. Elsevier, USA.

Moore, J., Chase, J., Ranganathan, P& Sharma, R. (2005). Making Scheduling "Cool": Temperature-Aware Workload Placement in Data Centers. In: Proceedings of the annual conference on USENIX Annual Technical Conference, USENIX Association Berkeley, CA, USA, 61-75.

Mysen, M., Berntsen, S., Nafstad, P. & Schild, P.G. (2005). Occupancy density and benefits of demand-controlled ventilation in Norwegian primary schools. Energy and Buildings 37, 1234–1240.

Nassif, N. (2012). Modeling and optimization of HVAC systems using artificial intelligence approaches. Paper presented in ASHRAE Annual Conference, San Antonio, TX, USA.

Nellis, G., Klein, S. (2009). Heat Transfer. Cambridge University Press.

Nielsen, T.R. & Drivsholm, C. (2010). Energy efficient demand controlled ventilation in single family houses. Energy and Buildings 42, 1995–1998.

Nourani, V. & Fard, M.S. (2012). Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. Advances in Engineering Software 47, 127–146.

Nørgaard, M. (2000). Neural network based system identification toolbox. Tech. Report. 00-e 891, Department of Automation Technical University of Denmark.

Nørgaard, M., Rvan, O., Poulsen, N.K. & Hansen, L.K. (2000). Neural networks for modelling and control of dynamic systems. Springer, London.

Oh, S.H. & Lee, Y. (1994). Effect of nonlinear functions on correlation between weighted sums in multilayer perceptions. IEEE Trans. Neural Networks 5, 508-510.

Oxford research. (2014). Finland's Giant Data Center Opportunity From the Industrial Heartland to Digital Age. Report.

Pan, Y., Huang, Z. & Wu, G. (2007). Calibrated building energy simulation and its application in a high-rise commercial building in Shanghai. Energy and Buildings 39,651–657.

Pavlovas,V. (2004). Demand controlled ventilation a case study for existing Swedish multifamily buildings. Energy and Buildings 36, 1029–1034.

Raftery, P., Keane, M. & Costa, A. (2011b). Calibrating whole building energy models: detailed case study using hourly measured data, Energy and Buildings 43, 3666–3679.

Raftery, P., Keane, M. & O'Donnell, J. (2011a). Calibrating whole building energy models: an evidence-based methodology. Energy and Buildings 43, 2356–2364.

Rakesh,D. (2009). Choices in liquid cooling for your data center. The Data Center Journal.

Rasmussen, N. (2008). An Improved Architecture for High-Density Data Centers. APC White Paper 126.

Redlich, C.A., Sparer, J. & Cullen, M.R. (1997). Sick-building syndrome. Lancet 349(9057), 1013–6.

Roberge, M.A., Lamarche, L., Karjl, S. & Moreau, A. (1997). Model of room storage heater and system identification using neural networks. In: Proc. of CLIMA 2000 Conf., Brussels (Belgium).

Ruano, A.E., Crispim, E.M., Conceicao, E.Z.E. & Lucio, M.M.J.R. (2006). Predictions of building's temperature using neural networks models. Energy Build 38, 682–694.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning internal representations by error propagation. In: Parallel distributed processing: explorations in the microstructure of cognition, Vol. 1. Cambridge (MA), MIT Press (chapter 8).

Schell, M.B., Turner, S.C. & Shim, R.O. (1998). Application of CO2-based demand-controlled ventilation using ASHRAE Standard 62: optimizing energy use and ventilation. ASHRAE Transactions 104, 1213–1225.

Schmidt, P.R., Cruz, E.E. & Iyenar, M.K. (2005). Challenges of data centers thermal management. IBM Journal of Research and Development 49,709–723.

Solomatine, D.P., Abrahart, R., See, L. (2008). Data-driven modelling: concept, approaches, experiences. In: Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications, Springer-Verlag.

Sigumonrong, A.P., Bong, T.Y., Fok, S.C. & Wong, Y.W. (2001). Self-learning neurocontroller for maintaining indoor relative humidity. In: Proceedings of the International Joint Conference on Neural Networks v2, IEEE, Washington, DC, USA, 1297– 1301.

Sofuoglu, S.C. (2008). Application of artificial neural networks to predict prevalence of building-related symptoms in office buildings. Building and Environment 43, 1121–1126.

Statistics Finland (2014). Energy consumption in households.

Sun, H.S. & Lee, S.E. (2006). Case study of data centers energy performance, Energy and Buildings 38, 522–533.

Sun, Z., Wang, S. & Ma, Z. (2011). In-situ implementation and validation of a CO2-based adaptive demand-controlled ventilation strategy in a multi-zone office building. Building and Environment 46, 124-133.

Sullivan, R.F., Strong, L. & Brill, K.G. (2007). Reducing Bypass Airflow Is Essential for Eliminating Computer Room Hot Spots. The Uptime Institute, Inc.

The Green Grid. (2007). The Green Grid data Center Power Efficiency Metrics: PUE and DCiE. The Green Grid.

Thomas, B. & Soleimani-Mosheni, M. (2007). Artificial neural network models for indoor temperature prediction: investigations in two buildings. Neural Comput Appl 16, 81–89.

Tibshiran, R. (1996). A comparison of some error estimates for neural network models. Neural Computation 8,152-163.

Uddin, M. & Rahman, A.A. (2012). Energy efficiency and low carbon enabler green IT framework for data centers considering green metrics. Renewable and Sustainable Energy Reviews 16, 4078– 4094.

Uptime 2014. 2014 Data Center Industry Survey. https://journal.uptimeinstitute.com/2014-data-center-industry-survey/.

Wachenfeldt, B. J., Mysen, M. & Schild, P.G. (2007). Air flow rates and energy saving potential in schools with demand-controlled displacement ventilation, Energy and Buildings 39, 1073–1079.

Wang, S.W. & Jin, X.Q. (1998). CO2-based occupancy detection for on-line outdoor air flow control, Indoor Built Environment 7(3), 165-81.

Wang, S. (2010). Intelligent Buildings and Building Automation. Spon Press, USA.

Xu, X., Wang, S., Sun, Z. & Xiao, F. (2009). A model-based optimal ventilation control strategy of multi-zone VAV air-conditioning systems. Applied Thermal Engineering 29, 91–104.

Xue, Q., Hu, Y. & Tompkins, W.J. (1990). Analyses of the hidden units of backpropagation model by singular value decomposition. Pro.IJCNN'90 San Diego, 739-742.

Yang, I.H, Yeo, M.S. & Kim, K.W. (2003). Application of artificial neural network to predict the optimal start time for heating system in building. Energy Conversion and Management 44, 2791–2809.

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

Zhang, Q., Wong, Y.W., Fok, S.C. & Bong, T.Y. (2005). Neural-based air handling unit for indoor relative humidity and temperature control. In: ASHRAE Transactions v 111 PART 1—Technical and Symposium Papers presented at the 2005 Winter Meeting of the American Society of Heating, Refrigerating and Air-Conditioning Engineers, ASHRAE, Orlando, FL, USA, 63–70.

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS