New methods and models for the ongoing commissioning of HVAC systems in commercial and institutional buildings

Nunzio Cotrufo

A Thesis
In the Department
of
Building, Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Building Engineering)
Concordia University
Montreal, Quebec, Canada

November 2017

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Nunzio Cotrufo

Entitled: New Methods and Models for the Ongoing Commissioning of HVAC Systems in Commercial and Institutional Buildings

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Building Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.
Signed by the final examining committee:

_____ Chair
Dr. Jia Yuan Yu

_____ External Examiner
Dr. Stanislaw Kajl

_____ External to Program
Dr. Amin Hammad

_____ Examiner
Dr. Andreas Athienitis

_____ Examiner
Dr. Zhenhua Zhu

_____ Thesis Supervisor
Dr. Radu Zmeureanu

Approved by _____
Dr. Fariborz Haghighat, Graduate Program Director

Thursday, December 14, 2017

_____
Dr. Amir Asif, Dean
Faculty of Engineering and Computer Science

Abstract

New methods and models for the ongoing commissioning of HVAC systems in commercial and institutional buildings.

Nunzio Cotrufo, Ph.D.

Concordia University, 2017

The performance of the HVAC systems in buildings tends to decrease after few years of operation. Equipment and sensors degradation lead to remarkable wastes of energy and money, as well as to the increase of building occupants thermal discomfort. HVAC ongoing commissioning (OCx), the continuation of HVAC commissioning well into the occupancy and operation phase of a building life, has been recognized as a cost-effective strategy to reduce energy wastes, equipment degradation and thermal discomfort. Building Automation Systems (BAS) collect and store huge amount of data for the purpose of building systems control. Those data represent a golden mine of information that can be used for the OCx of the building HVAC systems.

This research work develops and validates new methods and models to be used for the OCx of HVAC systems using BAS measurements from commonly installed sensors. A Fault Detection and Identification (FD&I) method for chillers operation, and several virtual sensor models for variables of interest in Air Handling Units (AHUs) are presented.

A FD&I method based on Principal Components Analysis (PCA) has been developed and used to detect abnormal operation conditions in an existing chiller operation and identify the responsible variables. The proposed FD&I method has been trained using measurements from summer 2009, and then used to detect abnormal observations from the following seven summer seasons (2010-2016). When the detected abnormal observations were replaced with artificially generated fault-free data, the proposed FD&I method did not detect any abnormal value along those artificially faulty-free variables. In summer 2016 the building operators changed several HVAC system operation set points, the FD&I method was effective in detecting almost 100% of the observations and properly identifying those variables whose set point was changed.

For two different operation modes of an AHU several virtual outdoor air flow meters have been developed and the predictions have been compared against short-term measurements using uncertainty analysis and statistical indices. Three models have been investigated when the heat

recovery coil was off. Results showed that the model with the simplest mathematical formulation was the most accurate, with the lowest value of uncertainty. When a heat recovery coil at the fresh air intake was on, two virtual flow meters have been developed to predict the outdoor air flow rate without the need of additional sensors. Both the models predicted the outdoor air ratio with good statistical indices: the Mean Absolute Error (MAE) was 0.015 for model *a* and 0.016 for model *b*.

Three methods for the virtual measurement and/or calibration of air temperature and relative humidity have been developed for different AHU operation modes. These methods are different in terms of modelling strategy, information needed and technical knowledge required for implementation. For instance, results from the correction of the faulty measurements of the outdoor air temperature along a 24 hours period using Method A showed a high virtual calibration capability: MAE = 0.2°C and the Coefficient of Variation, CV-RMSE = 1.7%.

A new definition of *virtual sensor* is proposed at the end of this research work. From a review of publications on virtual sensors for building application, the two most recurrent reason for the implementation of virtual sensor models (costs and practical issues) have been highlighted and integrated into the proposed new definition.

Acknowledgments

I wish to thank Dr. Zmeureanu for the supervision and the support received throughout my work at Concordia. His invaluable guidance during the last four years has been meaningful, his human qualities meant a lot to me. This research work would not have been possible without him.

A special thank goes to Olga Soares, from the Department administration. Thank you Olga for the smiles and the optimistic approach to life you spread out. It helped me to give the right value to things and stay positive.

I would also like to thank all the colleagues with whom I shared my time at Concordia.

Finally, thanks to five *funny* guys met at Concordia and surroundings: Mathieu, Eric, Costa, Harry and Daniel Jose Candelario. Thank you Daniel, it was important to me to mention you here.

"How was it, that tail fish story?" [1]

[1] Samuele Olivero, *Tips for a satisfying life*, 2006.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1. Problem Statement

The energy consumption is significantly growing all around the world. The International Energy Agency (IEA) reported a continuous increase in the yearly energy consumption during the last 45 years (more than 50% since 1971), while the CO_2 doubled (IEA 2016). This trend does not seem to change in the next years. The building sector is responsible for a big portion of the world energy consumption, and the portion of energy consumption due to the building sector is growing faster than others. The Department of Energy of the United States of America estimated that in 2010 buildings were accountable for 41% of the primary energy consumption, more than industry and transportation (U.S. Department of Energy 2012). In Canada 27% of the secondary energy use and 23% of the Green House Gas (GHG) emissions in 2013 were due to buildings (Natural Resources Canada 2016). In the developed countries most of the energy used for building operation is used for heating and cooling. HVAC systems are estimated to account for most of the energy end use in the building sector, 47% in the U.S. and 59% in Canada (Natural Resources Canada 2016; U.S. Department of Energy 2012). Investments and research on the energy efficiency in buildings, thus, should be strategic to have a great impact on the buildings energy consumption trend in the next decades. Improvements in building energy efficiency would reduce costs and pollution due to improper, non-efficient building operation.

The re-commissioning of existing HVAC installations has proved to be an extremely cost-effective strategy to identify and fix energy wastes and reset the system to meet the occupant's needs. Although this practice is gaining more and more popularity, the savings from HVAC re-commissioning do not persist over time (IEA 2010). Few years after recommissioning, HVAC systems and equipment decrease their performance, faults in operation start to rise, and occupants complains augment (Roth et al. 2008). The recent, remarkable augment of technology content in HVAC systems allows for more sophisticated control strategies but, at the same time, exposes building systems to a higher level of risk to fail. Components and equipment degradation, sensors miscalibration and improper control strategies and schedules are such of failures. The occurrence of those so called *soft* failures is not easy to detect, and it may remain undetected for a long time before being fixed, producing noticeable waste of energy, increase of costs and pollution, as well as thermal discomfort for occupants (Haves 1999). A new, non-stop monitoring strategy,

implemented on a *real time* basis, arose in the last two decades. This approach is named Ongoing Commissioning (OCx), and is considered highly effective for HVAC monitoring, performance analysis and efficiency maintenance (Roth et al. 2008). OCx uses HVAC trend data from Building Automation Systems (BAS) to continuously use control algorithms, in order to detect and diagnostic faults, and monitoring equipment efficiency and degradation. Thus, large buildings with modern BAS are ideal candidates for OCx implementation. Large amount of data are available from BAS, which represent an invaluable source of information about the state and operation of building and building systems. The extraction and interpretation of those information for maintenance purpose can be tedious, time demanding, and even impossible to be performed by humans, although they may have high technical skills.

In this context, there is a need for mathematical methods to be implemented in order to: i) benchmark equipment performance levels, ii) detect anomalies in measurements and iii) assist building operators in maintaining the desired level of system performance. OCx methods should account for several issues, like the availability and the quality of data, the cost for additional devices needed for OCx implementation, and building operator's technical knowledge. This research aims to develop new methods for the Ongoing Commissioning of HVAC systems using measurements commonly available from BAS. Mathematical methods could be integrated in the HVAC control code, taking advantage of the high capability of nowadays commonly installed BASs, at no additional costs.

1.2. Scope of the thesis

This thesis focus on the use of BAS trend data to derive valuable information on the operation of HVAC systems. BAS trend data contain useful information about the state and the operation of HVAC systems. The challenge is to extract those information and use them to improve systems operation and maintenance strategy. Mathematical methods are needed to detect faults and unexpected events in equipment operation, and provide building operators with useful information to address inspection and maintenance tasks.

This thesis has four main goals towards the main topics of the Ongoing Commissioning of HVAC systems:

1. a new method for the fault detection and identification in chillers operation;
2. the development of new virtual flow meter models for the virtual measurement of the outdoor air flow rate into an AHU under different operation modes;
3. the development of a series of methods for the development of virtual sensor models for the virtual measurement of variables or the virtual re-calibration of faulty sensors of air properties at the AHU mixing box inlets and outlet;
4. a new definition of virtual sensors for building application.

1.3. Overview of the thesis

The literature review from chapter 2 introduces to the different forms of building commissioning available tools for the automatic monitoring of buildings. A review of publications on the main topics of this thesis are presented: benchmarking models in section 2.3 and virtual sensors in section 2.4. The literature review chapter concludes with the main objectives of the present research work. The methodology presented in chapter 3 starts with an overview of the Knowledge Discovery in Databases (KDD), followed by the methodology used for each objective. The chapter concludes with the uncertainty analysis strategy used in this thesis to quantify and interpret the results uncertainty. Chapter 3 is not intended to be exhaustive of the methodology used in this work. Detailed description of methods and models development for each thesis objective is given in the further chapters. The case study used for this research work is presented in chapter 4.

The fault detection and diagnosis method for chillers is presented in chapter 5, while the virtual flow meter models, and the methods for the virtual measurement and calibration air temperature and relative humidity are presented in chapter 6 and chapter 7, respectively. The discussion on a new definition of *virtual sensors* for building application is addressed in chapter 8. Results are discussed at the end of each chapter.

Finally, the conclusions from chapter 9 recall the main contributions of the present research work and propose potential further works and developments.

2. LITERATURE REVIEW

This chapter presents a review of the literature on the main topics related to the Ongoing Commissioning of HVAC systems. First, the existing forms of HVAC commissioning are presented. Following the methods for the HVAC ongoing commissioning are reviewed and needs for further research are identified.

2.1. The commissioning process

The commissioning process intends to verify that the building systems and facilities meet the Owner Project Requirement (OPR). ASHRAE published commissioning guidelines (ASHRAE 2005a) in which several forms of commissioning are identified: Initial commissioning, Re-commissioning, Retro commissioning, Continuous Commissioning® and Ongoing Commissioning. While Initial, Re- and Retro commissioning are performed once, Continuous Commissioning® and Ongoing commissioning go "*well into the Occupancy and Operation Phase to verify that a project continues to meet current and evolving Owner's Project Requirements*" (ASHRAE 2005a). Continuous Commissioning® is performed periodically, e.g. several time per year, during the entire building life. Ongoing Commissioning goes more in deep, verifying operation and performance ideally with hourly or shorter time steps.

Initial commissioning takes place at the early stage of building design, going through the construction phase and the pre-occupancy phase. Inspections and design documents reviews are performed by the commissioning team. Initial commissioning intends to verify that all the systems are properly installed and work as expected. The goal is the delivery of a building which meet the OPR.

Re-commissioning and Retro commissioning are commissioning on existing building. While Re-commissioning refers to buildings that have been already commissioned in the pass, Retro commissioning is performed on buildings never commissioned before. For both, a re-commissioning team of energy managers and experts performs in situ inspections, reviews of design documents and operation records. Re-commissioning intends to ensure that the current building occupant's needs are meet by building systems operation. Low-cost improvements are investigated in order to improve system operation and control strategies. Re-commissioning may be part of a bigger retrofit project. Retro commissioning review the whole of the installed equipment and verify the way the work together. Non optimal design and installations are

identified and resolved, along with problems raised during the building's life. Finally, similarly to Re-commissioning, retro commissioning aims to adapt systems operation to meet the current occupant's needs.

Continuous Commissioning® has been first promoted in 1993 by the Texas A&M University. It starts from the idea that a program of continuous operation and maintenance (O&M) of the building system operation would allowed to better identify wastes and develop corrective actions (Claridge et al. 2000). The main target of CC® is to maintain the desired energy performance through the time through periodical verification of the system and equipment (Liu 1999). In the A&M first case, CC® used long-term measurements of electricity, chilled water and heating water consumption at a one hour time step. Data quality was checked weekly, and data analysis was performed monthly by humans, resulting in a monthly energy consumption report, which showed trends and savings. The use of a continuously implemented approach makes correct operations to persist and, in the A&M case, it resulted in an averaging saving of 28% for cooling and 54% for heating energy consumption. The O&M procedures implemented at A&M University require technical knowledge and system experience. It is performed by engineers and experts, and the building operator participation is desired in order to make them autonomous for future (Claridge et al. 2004).

Ongoing Commissioning (OCx) is commissioning implemented on a non-stop time basis through the building useful life (ASHRAE 2005a). The measurements collected from the Building Automation System (BAS) are used to verify the compliance of the equipment with the expected performance levels. The amount of data collected from BAS is huge, and the compliance to the expected performance levels has to be verified each time new measurements come to be available (e.g. one hour, or few minutes). Potentially the measurements from BAS allow for the verification of each single HVAC component. Because of the huge amount of information to be reviewed, and the analysis to be run with a high frequency, the OCx commissioning cannot be performed by humans without computational support. Measurements validation, benchmarking compliance, equipment fault detection and diagnostics, and automated reports generation can be provided by dedicated tool that helps building operators to monitor system and equipment performance levels and identify opportunities to improve saving strategies.

2.2. Analysis tools

HVAC control strategies in modern buildings are implemented through BASs, which collect the needed measurements and use them to accommodate needs according to the control codes. The BAS collects, and eventually stores, huge amount of data from building system operation. Those data represent an evaluable source of information on the system current operation. Several tools have been developed which use trend data from BAS to perform Fault Detection and Diagnosis (FDD) on HVAC operation. The most common FDD strategies are performed comparing the current values against set-points or benchmarks. A review of the available analysis tools is presented in this section.

2.2.1. Automated diagnostic tools

Automated diagnostic tools which perform Fault Detection and Diagnosis (AFDD) on BAS trend data are often developed in the form of add-on to the BAS. AFDD automate the process of collecting measurements, evaluate performance indices (PI), detecting faults and identifying their origin.

The Pacific Northwest National Laboratory, in collaboration with Honeywell and the University of Colorado, developed a **Whole Building Diagnostician** (WBD) tool to automate the process of fault detection in AHU economizers and central plants operation (Brambley et al. 1998). WBD provides automated data acquisition and achievement directly from BAS. The economizer module, named OAE, requires for measurements of the air temperatures, along with the fans modulation and the valves position signals, to detect faults in operation. Faults are detected through a decision tree which implements engineering rules of proper operation. A user interface uses color coding to display faults when detected. The tool provide additional information to help in fault diagnosis. The central plant module, WBE, uses measurements from the BAS along with neural networks to predict the whole building electrical and thermal consumption, and the HVAC equipment energy consumption. Fault diagnosis in WBE module are manual. No dedicated sensors are installed from WBD implementation, all required points are commonly recorded by BAS for control.

PACRAT (Performance and Continuous Recommissioning Analysis tool) was developed by Facility Dynamics Engineering and commercialized in 1999 (FacilityDynamics Engineering, n.d.). It is addressed to AHU, chillers, hydronic system, whole building energy and zones distribution. PACRAT allows for a large variety of raw data visualization techniques: time series, X-Y, 3D,

daily profile, load duration, aggregate. Fault detection and diagnosis is demanded to thermal and electric energy baselines (Santos and Brightbill 2000). PACRAT verifies the whole building consumption through comparison against reference buildings, based on performance metrics. At equipment level, metrics are plotted to support building operators in fault detection. The commissioning of the economizer operation is performed by PACRAT using measurements of the air temperature and humidity. Sensor faults and lack of calibration are highlighted if the tool identifies a thermodynamically inconsistent status of the system. At both building and equipment level, PACRAT use performance baselines of thermal and electric energy, binning the energy data by time of the week, air temperature and relative humidity. Energy wastes are quantified from the deviation between measurements and modelled baselines.

ABCAT (Automated Building Commissioning Analysis Tool) is a *semi*-automated tool developed by the Texas A&T University to monitor the energy consumption at the whole building level (Bynum et al. 2012). ABCAT consists of a simplified, first principle based building calibrated model, used to predict the whole building, cooling and heating energy consumption under given weather conditions. Faults detection is carried out manually by users comparing energy consumption predictions to measurements. ABCAT support the detection phase providing statistical indices and plots. If thresholds are previously defined, ABCAT performs automated fault detection looking at the magnitude and persistence of the difference between measurements and predictions.

DABO^{TM} (Diagnostic Agent for Building Operators) has been developed by the Intelligent Building Group of the CanmetENERGY Research Center. DABO^{TM} provides automated analysis and report of trend data from building BAS (Choinière 2008). Data analysis is performed by three modules: the building energy agent, the FDD agent and the Condition-based Maintenance agent. The FDD tool uses data from BAS to perform HVAC analysis at i) component level (hourly), ii) system (hourly, daily and weekly) and iii) building level. Designed for big institutional and commercial buildings, DABO^{TM} uses more than 800 rules to automate human analysis of PIs and reference values. Set-point PIs account for the difference between a sensor output and correspondent set-point.

CITE-AHU is an automated commissioning tool for air-handling units, developed in 2003 by the US *National Institute of Standard Technology* and the French *Scientifique et Technique du Bâtiment* center (Castro and Vaezi-Nejad 2005). CITE-AHU performs automated functional tests

before occupancy, and continuous commissioning during the system life. The AHU Performance Assessment Rules (APAR) identify different AHU operation modes based on 28 rules, control signal and measurements from available sensors from BAS. Five AHU operation modes are distinguished: i) heating, ii) cooling with outdoor air, iii) mechanical cooling and 100% outdoor air, iv) mechanical cooling with minimum outdoor air, and v) undefined case. A set of expert rules verify the system operation based on mass and energy conservation equations. Faults are identified from detecting unexpected operation modes. Although this tool is said to be automated, expert knowledge is required for configuration and rules customizing. After CITE-AHU configuration, functional testing and continuous commissioning can be implemented. Data can be analyzed daily, weekly, or monthly. Once a fault is detected, temperature and control signal plots are provided through the tool interface to the users for diagnosis validation.

VPACC (VAV Box Performance Assessment Control Charts) consists of a set of algorithms developed by the NIST for the FDD of VAV boxes (Schein and Bushby 2005). The algorithms are based on a statistical process control named *cumulative sum chart*. For each monitored variable, the expected value and variation range are defined. Faults are flagged when the cumulative sum of process deviations from the expected value is exceeded. A set of pre-defined process errors is then used to interpret the detected deviation and diagnose the fault. Target VAV monitored variables are, for instance, the space air temperature in relation to heating and cooling set points, the air flow rate, and the differential air temperature. VPACC uses measurements from few points: air temperature, dampers position, and air flow rate. Through the collaboration with manufacturers, VPACC has been integrated into BAS control codes, in order to execute the FDD algorithms along with the normal control logic. The AFDD tool has access to the BAS measurements, and eventual fault alarms and work orders are displayed on the BAS interface and included in reports.

VOLTTRON Lite™ has been developed by the Pacific Northwest National Laboratory to perform AFDD on AHU (Lutes et al. 2014). It includes seven proactive tests, performed on a daily or weekly frequency, to verify the correct operation of system sensors and components. Automated changes or simulated operating conditions are produced, and outputs are compared to expected values. The expected outputs derive from thermodynamic-based rules. An air temperature adjustable tolerance threshold of 2-4°C is used to evaluate measured and expected outputs during tests. Required inputs are variables commonly measured by BAS for HVAC control: air temperature at mixed, return, supply and outdoor condition, outdoor and mixed dampers signal,

heating and cooling coil valves signal. Because the implementation of proactive tests during occupancy can affect occupant's comfort, those tests should be run out of occupancy time. Tests are performed in a given sequence, and the outcome from a test may be used as hypothesis for the following one.

2.2.2. Building Automation Dashboards

Automated commissioning tools often provide user-interface to easier understand the system operation. User-interfaces, also known as building automation dashboards, aim to inform customers and make them aware of the building ongoing operation. A survey performed by Shadpour and Kilcoyne (2015) on a large number of HVAC professionals reported the features that should be integrated in a dashboard. The three main features are: i) real-time energy costs, ii) fault detection and diagnosis, and iii) facilities control. Depending on the available functions, building automation dashboards have been categorized in four levels:

i) Level 0, the simplest one, includes static values from historic data and simulations. At this level the target are the overall systems energy consumption and costs. Those information are mainly used by designers and project managers to make decisions on a master planning and life cycle level;

ii) Level 1, in addition to all the functionalities from level 0, display real time energy data. Building Automation Dashboards from level 1 intend to increase awareness in occupants, displaying the building performance level, as well as cost and energy meters. Strategies consist of comparing yearly or monthly cost and energy use values to previous records. Displayed data can concern the whole building as well as systems and equipment energy use;

iii) A level 2 dashboard, in addition to the previously listed features, allows to customize performance indicators and functions for specific equipment and components, as well as display energy use and trend plots. The target is support users in fault detection and diagnosis;

iv) Level 3 is the most advanced version of Building Automation Dashboard. At this level dashboards integrate software for AFDD, which use trend data from BAS collected at a short time step (e.g. hourly or each few minutes). Integrated AFDD software automatically detect and diagnosis the origin of an abnormal operation, monitor

performance indices trends, compare measured values to benchmarks, prioritize maintenance interventions and quantify potential savings.

EnteliWEB is a facility management tool intended to support owners and energy managers to analysis and interpret data from BAS (Deltacontrols n.d.). Visualization tools aim to support the analysis of the system performance, helping in identifying low performance situations and opportunities for operation optimization. EnteliWEB is produced and commercialized by Delta™ Controls, and allows to visualize energy profile and consumption data from HVAC and lighting. Visualization techniques include bar and pie charts, as well as plot line graphs.

A contribute to HVAC performance visualization is given by Abdelalim et al. (2017). The authors proposed the use of Sankey diagrams to visualize energy, mass and costs flows through HVAC systems in order to support the building operators in better understating the system performance and identifying opportunities for improvement. The proposed approach consists of using trend past data from BAS along with first principles to derive energy, mass and cost values. Those values are then used to build the Sankey diagram. As an example, the Sankey based approach was applied to the HVAC system from a university campus, and allowed to effectively identify the opportunity to optimize the heat recovery system set-point.

2.3. Benchmarking models

Many monitoring strategies and fault detection methods have been proposed in the last decades for the purpose of the commissioning of buildings. Although they varies in terms of application level (whole building, building systems, equipment, components, etc.) and source of information (questionnaires, utility bills, trend data from BAS, etc.) all of them implement some sort of comparison between i) an observed value which is representative of the actual behavior of the building/system, and ii) a benchmark which represents the expected performance of the building/system.

Existing models can be classified within few categories: grey and black box models (also known as data-driven methods), and white box models (also known as first principle based models).

Black box models refer to a category of models developed through learning machine approaches. Correlation functions between regressors and predictions are not known. *Black box* models work

primarily on experience, on how the system performed in the past. No engineering knowledge is used in model architecture.

White box models are based on the knowledge of the physical principles involved in the observed phenomena. White box models usually implement mass and energy balance equations.

Grey box models combine data fitting techniques with physical knowledge (Li and Wen 2014).

A great variety of data-driven models have been developed for the benchmarking of HVAC equipment. Data-driven models require for a reduced amount of information, time and technical knowledge compared to *white box* models. Supervised data-driven models use available measurements to predict some indices representative of the performance of the equipment to be benchmarked (e.g. COP for chillers). Non-supervised data-driven models explore data and benchmark the equipment through an *internal* representation of the performance pattern. In this latter case the model does not predict any variable, the benchmark consists of the representation of the performance pattern discovered within the data set. The most popular supervised models are classification models, regression models, artificial neural networks, and fuzzy logic. Examples of non-supervised models are clustering models. Once the benchmarking model is developed, its prediction (if the model is supervised), or the data internal patter (if the model is non-supervised), is compared to the actual values from measurements. If the comparison does not satisfy some threshold condition, an abnormal operation is flagged and a fault is detected.

In the following sections a literature review on the existing benchmarking models is presented, for both system and equipment levels, focusing in particular on the inverse (data-driven) models.

2.3.1. Building signature

When all the building systems are working as expected, for instance just after Initial, Re or Retro commissioning, a building energy signature can be used to benchmark the building energy consumption with respect to some parameter (e.g. outdoor air temperature, square meter of floor, etc.). Example of building signature are the Energy Use Indices (EUI) or Utility Cost Indices (UCI), used to compare a building to a similar sample of other buildings (Thompson and Moore 2011). At the building level, the utility bills are a valid source of information to build the energy consumption profiles through the previous years, and thus quantify the energy save due to the building Re or Retro commissioning.

The use of *calibrated models* is another strategy of building benchmarking (Efficiency Valuation Organization. 2007). Building and/or building system models are calibrated in order to minimize the difference between predicted and measured performance over a given time period (e.g. daily, weekly, monthly or yearly). Calibrated simulations can reach high levels of accuracy, but they are time demanding and require technical knowledge of building systems and physical principles (e.g. thermodynamics). For this reason simplified approaches are often preferred (Haves et al. 2001). The Lawrence Berkeley National Laboratory produced a guideline for energy performance signature in small commercial buildings (Granderson and Lin 2016). The guideline proposes a *Top-Down* approach in which the energy use per square meter (EUI) and the Energy Star Rating are used as reference to calibrate the building energy model. Monthly bills are collected and used to estimate the monthly energy consumption. Trend data from the BAS were used by Mihai (2014) to calibrate the model of a research center in Montreal, Qc. A *bottom-up* approach is proposed by Zibin which uses trend data from the BAS to calibrate building models (Zibin 2014). Calibration starts at the HVAC component level, moving up to equipment, system, and building levels progressively. Trend data are pre-processed and used as input to the model, while the model outputs are compared to the actual system outputs which are derived from trend measurements.

Data-driven models for building signature have the advantage to require for an extremely reduced amount of information and technical knowledge if compared to calibrated models. At the same time data driven models provide more reliable information than index based building signature methods as the EUI or the UCI. The most wide spread data-driven modelling approach for building signature use regression models to link the building energy consumption or cost to input variables (e.g. weather conditions). Example of regression based building signature models are given in Crawford et al. (1991), Hadley (1993), and Fels (1986). These studies will be introduced in detail in the next section.

Other examples of building signature strategies based on data-driven models are given in Dong et al. (2005), Chung et al. (2006), and Lee and Rajagopalan (2008). In Dong et al. (2005) a support vector machine (SVM) based model is used to take into account the non-linear correlation between weather data and building energy consumption. Four years monthly bills are used along with weather data (outdoor dry-bulb temperature, relative humidity and global radiation) to benchmark the energy performance of four buildings. In Chung et al. (2006) the EUI is evaluated

from nine inputs: building age, indoor air temperature set-point, type of equipment and lighting system details, and occupancy related parameters. Given those inputs, the authors presented a table of EUI benchmarking values. A similar approach is proposed in Lee and Rajagopalan (2008): a labelling program was developed in Singapore with the intent of ranking buildings based on the annual energy performance. The benchmarking strategy used data from surveys on a great number of commercial buildings. A new index was proposed for the energy efficiency of each building system, and the whole building index was finally calculated as the sum of every system index.

In Yan et al. (2017) a whole building benchmarking approach based on Gaussian Process (GP) regression is presented. The GP regression model predicts the whole building energy consumption accounting for the uncertainty of the output. Larger inputs uncertainties would lead to a larger output uncertainty (in this case the whole building energy consumption). The whole building benchmark is used for the fault detection and for the evaluation of the impact of each single input variable on the whole building energy consumption. Results from both fault detection and estimation of variables impact are affected by the considered uncertainty: the larger is the benchmarking uncertainty, the less sensitive would be the fault detection strategy, and the lower would be the impact of each single considered input (the variables) on the whole building energy consumption.

2.3.2. Inverse models for HVAC systems and equipment

A wide range of inverse (data-driven) models has been proposed in the last decades for the benchmarking of HVAC systems and equipment (Katipamula and Brambley 2005a and 2005b). A study on models development for HVAC equipment benchmarking is presented in Reddy et al. (2003). The author presented four models of the centrifugal chiller COP profile. In this study the author investigated few key aspects to be considered in HVAC equipment modelling: i) the model inputs, ii) the eventual physical meaning of the model coefficients, iii) the size and type of the data set used for training the model (training data set), and, for the case of regression models, iv) the regression technique.

Fault Detection and Diagnosis (FDD) strategies have been developed for HVAC components, mainly chillers, coupling data-driven models and with identification procedures (e.g. classification rules). In Comstock, Braun, and Groll (2001) the authors presented a FDD strategy for chillers based on few data-driven models which predict a set of five Performance Indices (PIs). Predicted PIs are compared to the observed ones to detect eventual faults. The residual between

the predicted and observed PIs values is used with a set of classification rules to diagnose the fault. A similar FDD strategy is presented in Cui and Wang (2005). Based on five chiller components PIs, the authors presented an effective fault classifier which consists of a set of rules linking residuals from predicted PIs to pre-identified common faults in chillers.

From the different data-driven models available in the literature, four of them are here presented as the most common and effective for HVAC equipment benchmarking: i) regression models, ii) Artificial Neural Network (ANN), iii) Support Vector Machine (SVM), and models based on the Principal Component Analysis (PCA-based models).

2.3.2.1. Regression models

Regression models are the most investigated models for HVAC benchmarking. The literature proposes a wide variety of regression models aimed to characterize systems and equipment energy performance under observed conditions. Fels (1986) presented PRISM, a univariate linear regression model which links the energy consumption to the outdoor air temperature. The main issue in modelling HVAC systems and equipment is the non-linear correlation among the parameters involved. A modelling procedure was proposed by Crawford et al. (1991) which consists of a segmented linear model. The entire range of variation of the input was split in several intervals in which the system response could be approximated by a linear function. The extreme point of the intervals in which the input values are split are known as change points. Hadley (1993) showed that including a day-type index improves the benchmarking capability. The day-type index is a factor which gives information on the general weather condition of the considered day.

In Katipamula et al. (1998) the authors investigate the effectiveness of Multiple Linear Regression (MLR) models for benchmarking the energy consumption of HVAC systems. MLR are remarkably better than single variable regression models because the thermal load is function of weather conditions, system characteristics, building usage, etc. The authors also investigated the most appropriate time resolution for HVAC thermal load prediction (monthly, daily, hourly, and sub-hourly). Reddy and Claridge (1994) observed that the prediction capability of MLR for HVAC benchmarking can be improved using Principal Components (PCs) based regression models when the inter-correlation among regressors is high.

The use of non-linear regression models for the benchmarking of HVAC equipment has been investigated by Monfet and Zmeureanu (2012). Four multivariate models, two linear (ML)

and two non-linear (MP), were proposed which predict the electrical power input (E) and the Coefficient Of Performance (COP) of a chiller (eqs. 2.1-2.4).

$$E = \alpha_1 \cdot \frac{Q_E}{Q_{Edesign}} + \alpha_2 \cdot T_{CNDS} + \alpha_3 \cdot T_{OA} \quad \text{eq. 2.1}$$

$$COP = \gamma_1 \cdot \frac{1}{\frac{Q_E}{Q_{Edesign}}} + \gamma_2 \cdot T_{CNDS} + \gamma_3 \cdot T_{OA} \quad \text{eq. 2.2}$$

$$E = \beta_0 + \beta_1 \cdot \left(\frac{Q_E}{Q_{Edesign}}\right)^2 + \beta_2 \cdot \frac{Q_E}{Q_{Edesign}} + \beta_3 \cdot (T_{CNDS})^2 + \beta_4 \cdot T_{CNDS} + \beta_5 \cdot T_{OA} \quad \text{eq. 2.3}$$

$$COP = \delta_1 \cdot \left(\frac{Q_E}{Q_{Edesign}}\right)^2 + \delta_2 \cdot \frac{Q_E}{Q_{Edesign}} + \delta_3 \cdot (T_{CNDS})^2 + \delta_4 \cdot T_{CNDS} + \delta_5 \cdot T_{OA} \quad \text{eq. 2.4}$$

where α_i , γ_i , β_i , and δ_i are the regression coefficients, Q_E is the thermal load at the evaporator, $Q_{Edesign}$ is the design thermal load, T_{CNDS} is the supply condenser water temperature, and T_{OA} is the outdoor air temperature.

Although both models predicting the electrical power input (eqs. 2.1 and 2.3) showed good accuracy, the non-linear model (eq. 2.3) predictions were slightly more accurate (CV-RMSE = 4.0% against CV-RMSE = 6.5% from the linear model). For the COP, the non-linear model (eq. 2.4) showed again good prediction capability (CV-RMSE = 4%), while accuracy for the linear model (eq. 2.2) varied (CV-RMSE between 2.8% and 11.5%) depending on the training data set. The superiority of non-linear regression models over the linear ones, thus, is not obvious. Monfet and Zmeureanu (2011) investigated the impact of the training data set on the model accuracy. Along with the development of a linear model for chiller benchmarking, the authors investigated the impact of the training data set size on the model prediction accuracy. Moreover, two different training techniques were considered: i) the *augmented window*, and ii) the *sliding window*. When the augmented window technique was used, models showed good prediction accuracy if trained along data sets of at least 14 days (CV-RMES < 6%). In the case of the sliding window technique, results showed that large training data sets decrease the prediction accuracy of the model. The models trained with a 14 days sliding window data set (CV-RMSE between 4.8% and 8.8%) performed slightly better than those models trained with a 21 days slightly window data set (CV-RMSE between 4.5% and 30.7%).

2.3.2.2. Artificial Neural Networks

Artificial Neural Networks (ANNs) model the relationship between inputs and outputs using compositions of basic functions. Because of their high capability in modelling non-linear correlations, in the last decades ANNs have been investigated for HVAC modelling (Teeter and Chow 1998). There are two main techniques to identify an ANN: i) *forward modeling* and ii) *inverse modeling*.

The *forward modelling* technique is a classical supervised learning problem, in which the system to be modelled and the ANN are *in parallel*. The error between the ANN model outputs and the observed actual values are used to refine the model coefficients and thus train the model. In the *inverse modelling* approach the real system to be modelled and the ANN model are in series. The goal is to map the inverse of the system to be modelled, thus the system output is used as input to the ANN model. The ANN model output is compared to the system input, and the error is used to refine the ANN model.

ANN based models were used to predict the cooling and heating load to estimate the energy savings due to a building retrofit (Krarti et al. 1998 and Yokoyama et al. 2009). Results showed good prediction capabilities. Zmeureanu (2002) developed and compared three different models to predict the COP of an existing rooftop units: a General Regression Neural Network (GRNN), a Back-propagation Neural Network (BNN), and a Multiple Linear Regression analysis (MLR). The GRNN model showed to perform better than the others. In Kusiak and Xu (2012) a dynamic neural network was used for predictive control of an HVAC system. The predictive control strategy was feed by the ANN outputs, having two objective functions: the minimization of energy consumption and the room temperature set-point. The proposed model allowed for up to 30% of energy savings.

2.3.2.3. Support Vector Machines

Support Vector Machines (SVMs) are machine learning models used for supervised learning (classification and regression) (Cortes and Vapnik 1995). SVMs identify a discriminant function to split data in categories, performing a linear classification. SVM performing non-linear classification are named *Kernel trick*. SVM based models (hyperplanes) are then used to classify new data in the same categories. If data are not labelled, SVM models implement unsupervised learning, looking for natural clustering in data. Introduced by Vapnik in the late ’60, SVMs have been recently used for prediction and forecasting in HVAC systems.

A SVM based model was used to predict the thermal load and power input of a chiller (Kusiak and Li 2010). Compared against three other prediction models based on multilayer perceptron, random forest and boosting tree, SVM based model showed a higher prediction accuracy. Compared to ANNs, SVM based models proved to perform better when used to predict the cooling thermal load (Li et al. 2009 and 2009b). Also, SVM algorithms require less parameters than ANNs, and thus are easier to design. SVM enable to model non-linear relationships, and thus recently become a topic of interest in HVAC modelling. For instance, SVM was used to model the non-linear correlation among air temperature, humidity and CO_2 concentration in an AHU served environmental (Kusiak, Li, and Zheng 2010). Prediction accuracy was improved coupling SVMs with ARIMA based time series models (Nie et al. 2012 and Kavousi-Fard and Kavousi-Fard 2013) and evolutionary algorithms (Hong 2009). SVM have been used for the prediction of the energy consumption in HVAC systems. For instance, Le Cam et al. (2017) used SVM to forecast the electrical consumption of the supply and return fans at the AHU of an existing building up to six hours in the future.

Although SVM based models are promising in HVAC benchmarking, this family of algorithms has not been exploited enough yet (Le Cam 2016).

2.3.2.4. PCA-based models

The Principal Components Analysis (PCA) is a multivariate technique which transforms a data set of j inter-correlated variables into another data set of k independent new variables, the Principal Components (PCs), where $j > k$. Draper and Smith (1981) introduced the use of PCA for the identification of new candidate regressors, other than the variables from an initial data set, to predict a dependent variable. PCA is an effective technique to overcome the presence of high inter-correlation along variables. Also, PCA is useful for the reduction of number of variables, as only the first few PCs can explain most of the variance present in original data set. Using synthetic data, Reddy and Claridge (1994) found that the PCA-based prediction models had better prediction performance than the Multivariate Regression (MLR) model, when the correlation between regressors and predictions is high. Lam et al. (2008) applied PCA to five weather variables and reduced the initial data set to two PCs, which were used as regressors in a MLR model for prediction of yearly and monthly electricity use in 20 office building in Hong Kong. In the past, several studies focused on PCA-based methods for the monitoring and Fault Detection and Diagnosis (FDD) of systems. Some representative works are presented herein. An introduction to

PCA-based techniques for data quality control of processes was given by Jackson (1991). An ellipse formulation was used to define the border of a control region, out of which the measurements are considered out of control. The ellipse was constructed by using Hotelling's T^2 statistic, which is a multivariate statistical generalization of the Student t-test.

Many studies investigated the use of PCA for HVAC equipment benchmarking and fault detection (Luo et al. 1999, Pranatyasto and Qin 2001, Lennox and Rosen 2002, Dunia et al. 1996, Wang and Xiao 2004a and 2004b.

Wang and Cui (2005) used PCA for sensor fault detection and identification in centrifugal chillers. They developed one PCA model for the energy performance indices of chillers, and another PCA model for the energy balance of the chillers. The models were trained by using measurements from centrifugal chillers in a large commercial building. The first three PCs were used for the first PCA model that explains 95.62% of the total variance, and the four first PCs for the second model. When the Q-statistic exceeded the threshold limit, which was calculated in terms of confidence level, a fault was detected. The faulty sensors were detected by using the Q-contribution plot. Multi-level PCA models were developed by Du and Jin (2007), at the system and local levels, to detect the multiple faults that might occur simultaneously in a Variable Air Volume system. The PCA models were developed from the energy and mass balance equations, and trained by using simulation results. There was an indication of the occurrence of faults when the SPE of measurements exceed the threshold of corresponding model. The fault signatures and expert rules were integrated for the fault identification. In Xu et al. (2008) PCA was coupled with wavelet transform analysis to exclude noise and dynamics from measurements, and to enhance the effectiveness of PCA based fault detection, diagnosis and estimation methods. The PCA models were trained with measurements from a chiller plant serving a high-rising building. A similar approach was used by Li and Wen (2014).

2.4. Virtual sensors

A *virtual sensor*, also referred to as *soft*, *smart* or *inferential sensor*, uses measured variables and other available information to predict the value of a variable of interest. In the last decades the implementation of virtual sensor models progressively spread out in several fields, as for instance process control, automobiles, wireless communication, robotic, traffic and building monitoring (Li et al. 2011). As an example, virtual sensors have been developed for application on

automobiles which estimate the air pressure in tires, the vehicle road friction, the vehicle velocity and motor combustion time.

The adoption of virtual sensors in buildings is slower than for other fields, and the interest on its implementation in buildings only raised in the last decade. This delay is mainly due to the uniqueness of each building. Differently from production in series, buildings are individually engineered. Costs for integrated virtual sensors is more relevant on each individually engineered products than on mass produced (Li et al. 2011). Nowadays, the implementation of virtual sensing technology in building systems is recognized to enable more effective on-line monitoring of equipment and components performance without increasing costs (Li and Braun 2009 ; Ploennigs et al. 2011 ; Hjortland and Braun 2016). In the last few years a growing number of virtual sensors has been proposed for HVAC monitoring. For instance, virtual sensor models have been developed for monitoring and fault detection of chillers (Li and Braun 2007 ; Mcdonald and Zmeureanu 2014), reliable measurements of pipes internal fluid temperature (Gorman et al. 2013), prediction and correction of air temperature measurements in AHU ducts (Lee and Dexter 2005 ; Brambley et al. 2011 ; Yu et al. 2011) and in occupied spaces (Alhashme and Ashgriz 2016), and prediction of the air flow rate (Tan and Dexter 2006 ; Yu et al. 2011).

Wichman and Braun (2009) proposed a model to correct the measurements of the mixed air temperature in Roof Top Units (RTUs) from an embedded single sensor. A set of previous tests allowed to develop the correlation between the measured mixed air temperature (T_{ma}) and the sensor error. Further prediction of the sensor error from the developed correlation were then used to correct the faulty measurements of T_{ma}. Fernandez et al. (2009) proposed a set of algorithms to correct faults from air temperature and relative humidity sensors, and from signals of dampers position in AHUs. Yu et al. (2011a) developed a regression model to correct faulty measurements of the supply air temperature (T_{sa}) in a RTU. The model predicts the T_{sa} correction term, at each new time step, as function of the number of heating stages (N_{Hst}) and the outdoor dampers position. An improved version of this model was proposed by Yu et al. (2011b), in which measurements of the outdoor air temperature (T_{oa}) were included in the list of regressors. Tan and Dexter (2006) developed data driven models to predict: (1) the outdoor air flow rate from the control signal to the inlet damper of a VAV system, and (2) the supply and extract airflow rates from the control signals to the fans and dampers. The Virtual Flow Meter (VFM) produced relatively small errors

of 8%, 2% and 3% for the recirculated, supply and outdoor air flow rates, respectively, when compared with direct measurements.

Wang et al. (2014) developed a VFM that uses measurements of the fan power input, along with the motor and fan efficiency, to predict the supply air flow rate from an AHU. In Hjortland and Braun (2016) the supply air flow rate of a RTU was predicted from the fans VFD signal and the pressure difference across the supply fan. The same study proposed linear correlation models to correct: (1) the outdoor air temperature from the dampers position signal, the faulty outdoor air temperature and the return air temperatures as measured by the embedded sensors, and (2) the mixed air temperature from the supply mass air flow rate along with the dampers position signal, and the faulty mixed air temperature, and the correct outdoor and return air temperatures as measured by the embedded sensors. Finally, a third order correlation model was proposed to predict the outdoor air fraction, the factor α (eq. 1), from the damper actuator control signal.

$$\alpha = \frac{T_{ma} - T_{ra}}{T_{oa} - T_{ra}} \qquad \text{eq. 1}$$

where T_{ma}, T_{ra} and T_{oa} are the mixed, return, and outdoor air temperatures, respectively.

Padilla et al. (2015) presented a model for the virtual measurement of the supply air temperature in AHU as part of a FDD strategy. Starting with the measurement of the mixed air temperature, the supply air temperature is calculated by adding: (1) the temperature change due to heating coil, (2) the temperature change due to mechanical cooling coil, and (3) the temperature change due to the supply fans. The contribution of each term was correlated to measurements from the BAS through genetic algorithms (GA). The GA were used to estimate the model coefficients which minimize the difference between measurements and model predictions.

Several studies focused on the development of VFM for the virtual measurement of water flow rate in cooling plants (Song et al. 2012 ; Swamy et al. 2012 ; Zhao et al. 2012 ; Mcdonald and Zmeureanu 2014 ; Andiroglu et al. 2016). Mcdonald and Zmeureanu (2014) developed two, first principle based, models to estimate the water mass flow rate at the chiller evaporator and condenser. Five different scenario were considered, corresponding to different number of available sensors.

A complete list of the reviewed articles on the topic of virtual sensors for building applications is given in Table 2.1. The virtual sensor models are reported along with the correspondent modelling method, type of application and required measurements.

2.4.1. The need for a new definition of virtual sensor

According to Li et al. (2011), a widely accepted definition for virtual sensors does not exist yet. The most recurrent definition from scientific literature on HVAC states that a virtual sensor consists of a mathematical model which use measurements from other measured variables to predict (virtually measure) a variable of interest (e.g. Li et al. 2011 ; Mcdonald et al. 2014 ; Hjortland and Braun 2016). Someone may argue that, according to this definition, most of the existing physical sensors actually work as virtual sensors, as they do not *directly* measure the physical variable they are used for, but they derive their output from another measured parameter: for instance an ultrasonic flow meter does not *directly* measure the liquid flow rate, but derive it from the delay in time of an ultrasound traveling through the fluid, between two transducers, in both directions; a thermocouple does not *directly* measure the air temperature, but derive it from a temperature-dependent voltage. Eventually, according to the above given definition, all the formulas to derive performance parameters, as for instance the building EUI (e.g. Lee and Rajagopalan 2008), or any chiller's characteristic physical parameters (e.g. Jia and Reddy 2003), may be referred to as a *virtual sensor*. The notion of *virtual sensor*, thus, risks to be assimilated to a generic process of deriving a parameter (a medium physical property, an index, a probability, etc.) from other values. In order to make the subject clearer, within the field of HVAC monitoring, the present research work proposes a new formulation of the definition of *virtual sensor*.

Table 2.1 – Virtual sensor model from literature review.

System / equipment	Virtual sensor	Modeling method	App	Required measurements	Reference
Vapor compression AC	Power consumption	Grey box	*Obs*	$T_{suc,d}$; $T_{dis,d}$	Li and Braun 2007
	Refrigerant flow rate	First principle	*Obs*	$\dot{W}$; Q_{loss} ; $T_{suc,d}$; $T_{dis,d}$	
	Volumetric efficiency	Grey box	*Obs*	P_{suc} ; P_{dis} ; T_{amb} ; T_{suc}	
Chillers	Refrigerant charge level	First principle	*Obs*	*Tsc,1 ; Tsh,1 ; Tsc,2 ; Tsh,2*	Li and Braun 2009
	Refrigerant pressure	Grey box	*Obs*	$\dot{m}_{ref}$	Li and Braun 2009b
	Condenser fouling	First principle	*Obs*	T_{CNDS} ; T_{CNDR} ; $T_{CND,ref}$	Zhao et al. 2012a
	Water flow rate	First principle *(five scenarios from 8 to 6 points)*	*Obs*	T_{CHWS} ; T_{CHWR} ; T_{CNDS} ; T_{CNDS} ; T_{DIS} ; T_{SUC} ; T_{II} ; T_{EV} ; P_{EV} ; T_{CN} ; P_{CN} ; $\dot{W}$; *manufacture specification*	McDonald and Zmeureanu 2014
		First principle	*Obs*		Zhao 2012b
	Condenser heat loss	First principle	*Obs*	$\dot{m}_{cdS}$; $T_{cnd,ref}$; T_{CNDS} ; T_{CNDR}	Reddy 2007
	Evaporator heat loss	First principle	*Obs*	$\dot{m}_{cHW}$; $T_{chw,ref}$; T_{CHWS} ; T_{CHWR}	
	Polytropic efficiency of the compressor	First principle	*Obs*	T_{SUC} ; P_{SUC} ; T_{DIS} ; P_{DIS}	
Pumps	Water flow rate	Grey box	*Obs*	ΔP_{pump} ; ω	Song et al. 2012
		Grey box	*Obs*	H ; W_{VFD} ; V ; f ; Q	Andiroglu et al. 2016
Pipes	Fluid temperature	Grey box	Obs	T_{air} Pipe dimensions	Gorman et al. 2013
AHU / RTU	Water flow rate	First principle/ Grey box	*Obs*	ΔP_L ; *x ; design and manuf. specifications*	Swamy et al. 2012

AHU / RTU	Supply, Mixed and Return air temperature	First principle	*B-up*	T_{oa} ; T_{ma} ; T_{sa} ; OAD_{st}	Fernandez et al. 2009
	Supply air static pressure	Black box	*B-up*	x_{fan} ; N_{fan} ; x_{VAV} ; Q_{SA}	Padilla et al. 2015
	Outdoor air temperature	Grey box	*B-up*	*OADst* ; T_{oa} ; T_{ra}	Hjortland and Braun 2016
	Supply air temperature	Grey box	*B-up*	$N_{Hstages}$; *OADst* ; T_{oa} ; *along with other short-term measurements*	Yu et al. 2011a
		Grey box	*B-up*	$N_{Hstages}$; *OADst* ; *along with other short-term measurements*	Yu et al. 2011c
		Black box	*B-up*	T_{ma} ; T_{hw} ; x_{hc} ; T_{chw} ; x_{cc} ; ΔP_{fan} ; x_{fan}	Padilla et al. 2015
	Mixed air temperature	Black box	*B-up*	T_{ma} ; CFD model of the AHU	Lee and Dexter 2005
		Grey box	*Obs*	T_{oa} ; T_{ra} ; T_{sa}	Wichman and Braun 2009
		Grey box	*B-up*	*OADst* ; T_{oa} ; T_{ra} ; T_{ma} ; $\dot{m}_{sa}$	Hjortland and Braun 2016
	Outdoor air flow rate	Grey box	*Obs*		Tan and Dexter 2006
		First principle	*Obs*	T_{oa} ; T_{ra} ; T_{sa} ; *OADst*	Yan et al. 2017
	Air flow rate	Grey box	*Obs*	T_{oa} ; T_{sa} ; OAD_{st} ; N_{Hstg}	Yu et al. 2011b
		Grey box	*Obs*	V_g ; T_{ma} ; $T_{ma,wb}$; T_{sa} ; ΔT_{fan}	Yu et al. 2011d
		First principle		T_{sa} ; T_{ra} ; *design specifications*	Mishukov and Horyna 2015
		First principle/ Grey box	*Obs*	$\dot{W}_{motor}$; H	Wang et al. 2014

		Grey box	*Obs*	VFD ; ΔP_{fan}	Hjortland and Braun 2016
		First principle	*Obs*	T_{oa} ; T_{ma} ; T_r	Zmeureanu and Vanderbrook 2015
Fans	Electrical power input	Grey box	*Obs*	VFD ; ΔP_{fan}	Hjortland and Braun 2016
	Fan air temp. increase	First principle	*Obs*	$\dot{W}_{fan}$; V_{air} (air velocity)	
Zone	Room temperature	First principle	*Obs*	T_{room}	Alhashme and Ashgriz 2016
		Black box		V_{in} ; $\dot{m}_{chw}$	Reppa et al. 2014
	Energy consumption	First principle		$\dot{m}_{pipe,w}$	Ploennings et al. 2011
	IAQ (T)	Black box	*Obs*	IAQ(CO_2) ; IAQ(RH) ; T_{SET} ; Ic_{room} ; P_{atm} ; RH_{oa} ; T_{oa} ; $I_{sol,beam}$; $I_{sol,horz}$; W_{dir} ; W_{spd}	Kusiak et al. 2010
	IAQ (RH)	Black box	*Obs*	IAQ(CO_2) ; IAQ(T) ; T_{SET} ; Ic_{room} ; P_{atm} ; RH_{oa} ; T_{oa} ; $I_{sol,beam}$; $I_{sol,horz}$; W_{dir} ; W_{spd}	Kusiak et al. 2010
	IAQ (CO_2)	Black box	*Obs*	IAQ(T) ; IAQ(RH) ; T_{SET} ; Ic_{room} ; P_{atm} ; RH_{oa} ; T_{oa} ; $I_{sol,beam}$; $I_{sol,horz}$; W_{dir} ; W_{spd}	Kusiak et al. 2010

NOTE: *Obs* observation; *B-up* back-up/replacement.

2.5. Objective of the thesis

In the context of HVAC systems ongoing commissioning there is a need for more flexible and easily scalable monitoring methods. The main objectives of this work are the development of the following points:

1. *A new method for the detection and identification (FD&I) of faults in chillers operation.*
   Chillers have been observed to decrease their performance over the time. The rise of faults in operation, along with components degradation, can provoke energy waste, increase operation costs, and generate thermal discomfort for building occupants. Real-time FD&I would promptly inform the building operators about the occurrence of a fault and the need of further actions. Addressing commissioning efforts, an automated FD&I method would require less experienced knowledge from building operators, and it would reduce the time needed for inspection and maintenance. On the other hands, the implementation of FD&I on historical data can help energy managers and manufacturers to better understand the system operation and its evolution through its life-cycle. A data transformation and reduction technique, the Principal Component Analysis (PCA), is considered for the development of a data-driven model used for benchmarking the fault-free chiller operation and detect abnormal events.
2. *Two new virtual flow meter models for the virtual measurement of the outdoor air flow rate into the AHU.*
   Although the measurement of air flow rates in AHUs is of a major significance for control and performance monitoring, such a sensor is not always installed. A physical air flow meter would come at additional cost, and practical issues may occur (e.g. improper installation, miscalibration, sensor degradation, etc.). The implementation of virtual sensor can effectively overcome practical issues and, at the same time, prevent additional costs.
3. *Several methods for the development of virtual sensor models for the virtual measurement of variables or the virtual re-calibration of faulty sensors of air properties at the AHU mixing box inlets and outlet.*
   The measurements of the air temperature and relative humidity at the AHU mixing box inlets and outlet are known to face several practical issues which may affect the quality of the measurements, e.g. compact structure of the AHU, air stratification, etc. The virtual

sensor models of the air properties should be implemented to overcome these issues and avoid additional costs.

4. *A new definition of virtual sensor for building application.*

   Although slower than for other fields, the implementation of virtual sensors in buildings is progressively spreading out. Virtual sensors are used to avoid the installation of a new sensor, which would come at additional cost, or when the variable of interest is difficult to measure because of HVAC system configuration (e.g. AHUs usually have compact structure which may make hard to measure air properties at specific points). From the review of the literature on virtual sensor models for building application, the need for a new definition has been identified. The objective is to provide a clear definition which highlight the key characteristics of virtual sensors, and distinguish from other mathematical formulations, e.g. performance indices, which are already commonly used in buildings for benchmarking and signature purpose.

3. METHODOLOGY

This research work aims to develop new methods for the OCx of HVAC systems. For this purpose, four specific objectives are proposed, which consist in the development of the following:

1. a new FD&I method for chillers;
2. new VFM models for the outdoor air flow rate into AHUs;
3. new methods for the development of virtual air properties sensors in AHUs;
4. a new definition of *virtual sensor* for building application.

Objectives 1, 2 and 3 have been addressed using a real case study building. Trend data from the HVAC system operation are used which were collected from the BAS for control purpose. Field measurements, thus, have been used to develop and validate new models. The final intent is to derive *actionable information* for the ongoing commissioning of HVAC equipment. The process of using raw data to extract knowledge (e.g. operation patterns) is a multi-step process known as Knowledge Discovery in Databases (KDD) (Fayyad et al. 1996). The next section introduces the KDD main steps. Following, the methodologies used for the development and validation of the FD&I method, and the methods and models for air virtual measurements in AHUs are presented (sections 3.2, 3.3 and 3.4). A schematic of the main features from the methods and models which this research work proposes is showed in Figure 3.7. Finally the uncertainty analysis approach used to quantify the propagation of uncertainty on virtual models predictions is outlined.

The development of a new definition of *virtual* sensors (objective 4) has been addressed through an extensive literature review of publications which explicitly used the term *virtual sensors* for building application. From the literature review, the most recurrent reasons why a virtual sensor should be used have been highlighted and used to develop a new definition of *virtual sensors* for building application.

3.1. Knowledge Discovery in Databases (KDD)

An exhaustive introduction to the KDD process is presented in Fayyad et al. (1996). KDD is a multi-step process aiming to use raw data to extract useful information. Five main steps within the KDD process model and transform the original raw data set and finally extract the hidden patterns. Those five main steps are: i) Selection, ii) Pre-processing, iii) Transformation, iv) Data mining, and v) Interpretation and evaluation.

Selection is the first step in KDD. From the *initial raw data set*, the selection step aims to isolate the *target variables* at given target condition. The amount of available recorded items can be huge, compromising the effectiveness of data analysis. The selection of *target variables* is based on the overall knowledge of the domain, and applying sample techniques.

Pre-processing includes limit checks and statistical techniques intended to *prepare* data for further analysis. The quality of field measurements needs to be verified as the condition under which measurements are collected are not always known. Typical issues in data collection are noise, missing data, quasi-steady versus transient operation. Also, differently from laboratory experiments in which all the conditions are strictly controlled, field measurements may be affected by sensors miscalibration and improper installation, incorrect assumptions, etc. In order to overcome those issues, the ASHRAE Guideline on *Engineering Analysis of Experimental Data* (ASHRAE 2005b) proposes to check limits, which means measurements must comply with physical, expected and theoretical conditions. Measurements which do not comply with those conditions should be excluded by data sets. Furthermore, when the data set includes variables with different units and different range of variation. When statistical techniques are used, those variables with larger range of variation may *hide* the information from the variables with a smaller range of variation. Data normalization techniques, thus, may be necessary to normalize each variable (Reddy 2011).

Transformation consists of representing the pre-processed data set with new system of coordinates, making easier to manipulate it. Transformation techniques are also used for data reduction. For instance, Principal Component Analysis (PCA) allows to represent an initial data set through new independent variables, and reduce the number of variables to be used for information extraction.

Data mining is the core of the KDD process, and consists of extracting the needed information and patterns from available data. Different data mining techniques respond to different KDD goals. Common data mining methods are classification, clustering, and regression.

Interpretation and evaluation is the final step in KDD, when the information extracted with data mining technique are interpreted and eventually converted into useful *actionable information*. Visualization techniques and summarization can be very powerful in information interpretation.

The KDD process can include iterations on one or more of the above listed steps. The KDD multi-step process, as proposed by Fayyad et al. (1996), is resumed in Figure 3.1.

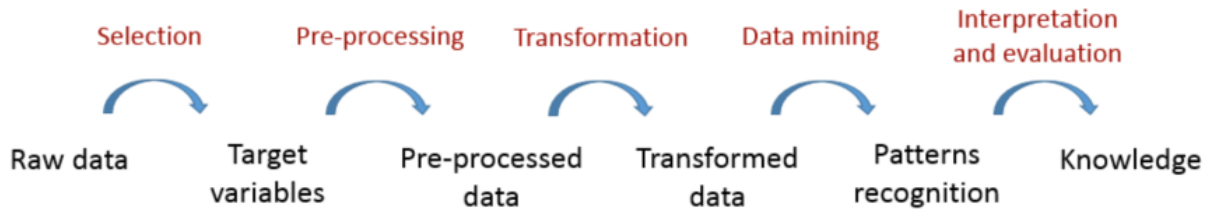


Figure 1 – KDD multi-step process by Fayyad et al. (1996).

The potential application of KDD has been used for building energy modeling. Le Cam et al. (2016) used KDD to extract insights from measurements of an existing HVAC operation with the goal of forecasting the energy consumption of the AHU fans.

3.2. Chillers benchmarking and Fault Detection and Identification (FD&I)

HVAC equipment tend to deteriorate their operation performance through time. Few months or years after building initial commissioning or re-commissioning, the occurrence of operation faults, components degradation and sensors miscalibration issues rise, affecting energy consumption, equipment life cycle and occupants comfort. Economic benefits from building re-commissioning can drop by up to 25% four years after recommissioning is completed (IEA 2010). The present research work proposes a new Fault Detection and Identification (FD&I) method for chillers, which uses trend data to benchmark the chiller fault-free operation (ideally just after initial commissioning or re-commissioning). The proposed benchmarking model is a statistical model which aims to describe the fault-free operation of a specific, as-operated, chiller. The proposed benchmarking development consists of learning from trend data of the chiller fault-free operation pattern. After initial, re- or retro commissioning HVAC systems operation is expected to be fault-free. Thus, the models for OCx should learn from data collected just after the system commissioning. The proposed model is an inverse, grey/black, model. The developed benchmarking model is than applied to further measurements in order to detect abnormal values which do not comply with the benchmarked fault-free operation condition, and identify the variable which caused the detection.

3.2.1. Principal Component Analysis (PCA) based method for FD&I

Chillers include several components, at least two heat exchangers, one compressor and one throttling valve. In order to properly describe a chiller operation, thus, several variables should be

monitored (e.g. the chiller power input, the supply chilled water temperature, etc.). At each time step, thus, chiller operation is described with a set of variable's measurements. The entire set of measurements at a given time step is here named *observation*. Chillers FD&I is a multivariate analysis problem, including inter-correlated variables. In order to reduce the number of variables, and remove the inter-correlation among them, Principal Component Analysis (PCA) is used. PCA transforms the initial data set of j inter-correlated variables into a new data set of j independent (orthogonal) variables, the Principal Components (PCs). Most of the information included into the original data set (the variance) is explained by few firsts PCs. Variables reduction is implemented retaining only few PCs and excluding the others. Chiller fault-free operation benchmarking is thus implemented into the PCs-based space.

Fault-free operation benchmarking consists of identify a boundary which delimits the region of space where the fault-free observations are expected to fall once projected into the PCs-based space. The boundary is defined using statistical indices from the distribution of the fault-free observations along each axis into the PCs-based space. This step is named *training*, and the fault-free data set used for *training* is referred to as *training data set*. The identified boundary consists of an ellipsoidal threshold condition that further observations must comply with in order to be labelled as fault free observations. Observations whose PCs-projections fall outside the identified threshold condition are taken as abnormal events, and labelled as outliers. Once the fault-free benchmarking model is trained, thus, it can be applied to further measurements in order to detect abnormal (non-fault-free) observations (*application*).

Fault identification consists of identifying which variable, from the list of variables selected to describe the chiller operation, took an abnormal value, and thus caused the detection of an abnormal projected observation. Figure 3.2 shows a schematic of the proposed PCA-based FD&I method, which is explained in detail in Chapter 5 along with method validation.

Chapter 5 gives a details explanation of the model development and application here introduced.

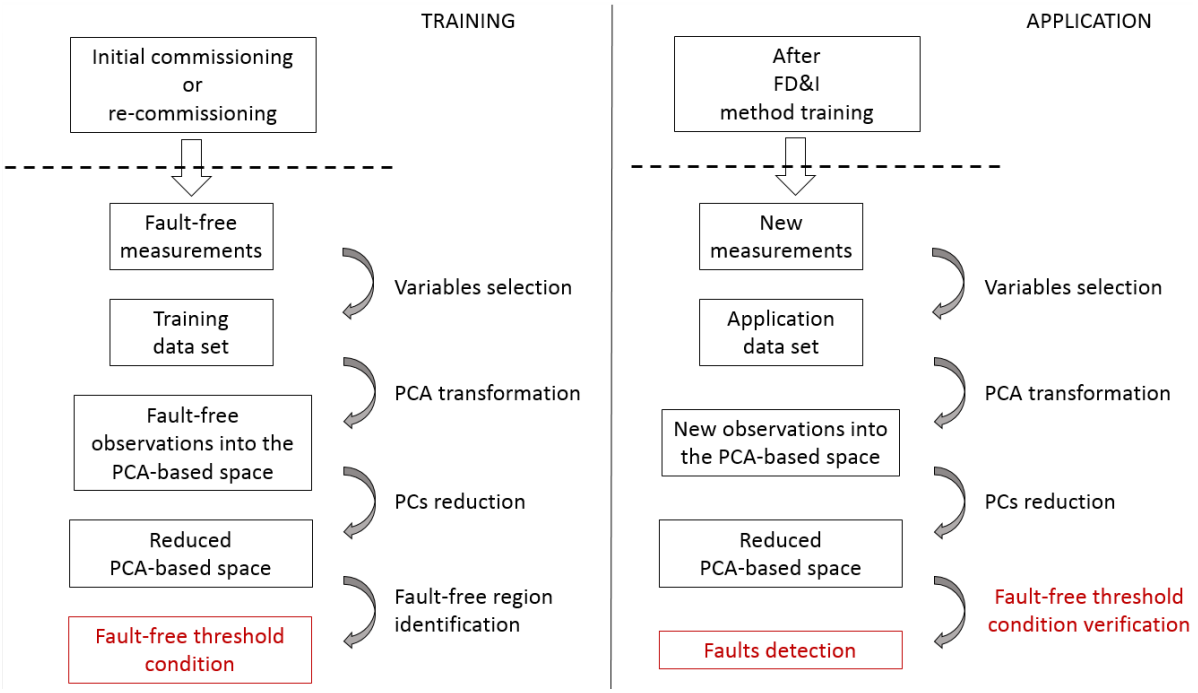


Figure 3.2 – Schematic of the proposed PCA-based FD&I method.

For validation purposes, measurements of chiller operation from eight summer season (from 2009 to 2016), is used. Measurements from the entire summer 2009 are used as training period to evaluate the fault-free operation boundary (benchmarking). The derived boundary threshold condition is then applied to measurements from the following seven summer seasons (validation) to detect and identify faults. The variables found to be responsible for the highest percentages of detected faults are investigated.

Along chiller operation, if a fault occurs, it can be detected with the proposed PCA-based FD&I method, and the measured variable which caused the fault detection be identified. These information can be used by building operators and energy managers to plan inspections and address maintenance efforts.

3.3. Virtual Flow Meters for AHUs

Reliable values of the outdoor air flow rate into the AHU economizer are needed for control schemes and performance monitoring purpose. Virtual measurements of the outdoor air flow rate can be obtained with eq. 3.2, from the supply air flow rate and the outdoor air ratio (factor α) (eq.

3.1). Combining eq. 3.1 with the energy balance equation in the mixing box (eq. 3.3), the factor α can be formulated as function of the air specific enthalpy at outdoor, return and mixed condition (eq. 3.4).

The VFM model from eq. 3.2, thus, is a first principle based model (*withe box model*) based on the energy balance across the mixing box. If the water content in the air streams entering and leaving the mixing box is neglected, the factor α can be expressed as function of air temperatures only (eq. 3.5). The factor α from eq. 3.5 is a common formulation, accepted by several authors for both HVAC ongoing commissioning and models calibration (e.g. Zibin et al. 2016 ; Yan et al. 2017).

$$\alpha = \frac{\dot{m}_{oa}}{\dot{m}_{ma}} \qquad \text{eq. 3.1}$$

$$\dot{m}_{oa} = \alpha \cdot (\rho_{air} \cdot V_{sa}) \qquad \text{eq. 3.2}$$

$$\dot{m}_{ma} \cdot h_{ma} = \dot{m}_{oa} \cdot h_{oa} + \dot{m}_{rec} \cdot h_{ra} \qquad \text{eq. 3.3}$$

$$\alpha = \frac{h_{ma} - h_{ra}}{h_{oa} - h_{ra}} \qquad \text{eq. 3.4}$$

$$\alpha_0 = \frac{T_{ma} - T_{ra}}{T_{oa} - T_{ra}} \qquad \text{eq. 3.5}$$

where $\dot{m}_{oa}$, $\dot{m}_{ma}$ and $\dot{m}_{rec}$ are the outdoor, mixed and recirculated air mass flow rate respectively, kg/s; ρ_{air} is the air density, kg/m^3; and V_{sa} is the air volume flow rate, m^3/s; h_{ma}, h_{oa} and h_{ra} are the air specific enthalpies at mixed, outdoor and return conditions respectively, kJ/kg; and T_{ma}, T_{oa} and T_{ra} are the air temperatures at mixed, outdoor and return conditions respectively, °C.

3.3.1. Operation modes

Three different AHU operation modes have been identified for this case study (section 4.2). When the mixing dampers are completely closed (operation mode #1), the outdoor air flow rate is expected to be the same as the BAS measured supply air flow rate. VFM based on eq. 3.2 are proposed for the case when the AHU is working under operation modes #2 and #3. Under operation mode #2 the measured outdoor air properties are the same as at the mixing box outdoor inlet (Figure 3.3). Under operation mode #3 a heat recovery system is used to pre-heat the outdoor air stream (Figure 3.4). Thus, during the cold season, the temperature of the outdoor air entering the

mixing box (T_{ac}) will be higher than the air temperature at outdoor conditions (T_{oa}) (Figure 3.4). In this case, the measured outdoor air temperature is not the same as at the mixing box outdoor inlet, and should not be used with eq. 3.5.

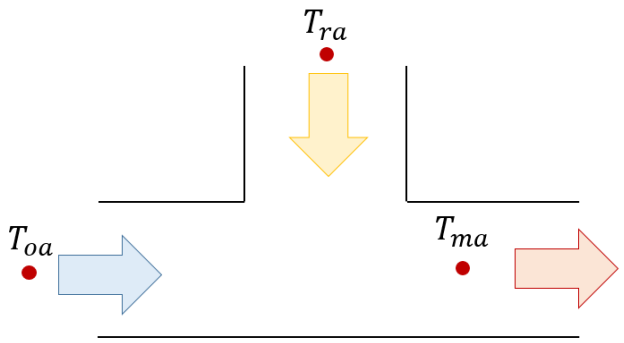


Figure 3.3 – Schematic of a mixing box with known inlets conditions.

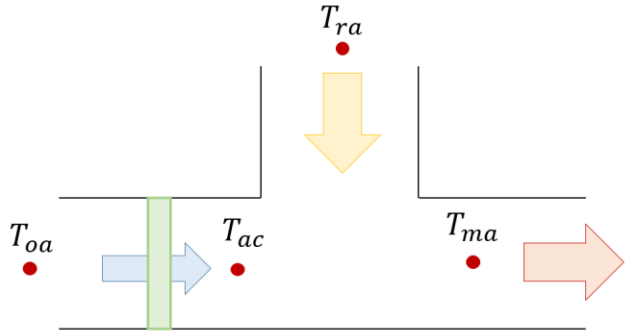


Figure 3.4 – Schematic of a mixing box, with measured and actual mixing box inlets air temperatures.

The factor α from eq. 3.5 is replaced by eq. 3.6, where the outdoor and return air temperatures (T_{oa}) is replaced by the actual air temperatures at the mixing box inlet (T_{ac}).

$$\alpha = \frac{T_{ma} - T_{ra}}{T_{ac} - T_{ra}} \qquad \text{eq. 3.6}$$

The implementation of eq. 3.6 for the evaluation of the factor α requires for values of a variable (T_{ac}) which is not always measured and collected by the BAS. Thus, one additional, dedicated air temperature sensor should be purchased and installed, which would increase costs and involve installation issues (e.g. air stratification, AHU mixing box compact structure, etc.). A new model for the prediction of the factor α is proposed which does not need for values of T_{ac}. The new model

consists of a system of two equations (the factor α from eq. 3.6 and the heat recovered through the heat recovery system from eq. 3.7) and two unknown (the factor α and the air temperature T_{ac}). The heat transfer Q_{HR} from the heat recovery coils (in green in figure 3.4) to the outdoor air stream is given (eq. 3.7):

$$Q_{HR} = \rho_{glc} \cdot V_{glc} \cdot C_{p,glc} \cdot (T_{hre} - T_{hra}) \qquad \text{eq. 3.7}$$

where ρ_{glc} is the glycol density, kg/m^3; V_{glc} is the glycol flow rate, L/s$\cdot 10^{-3}$; $C_{p,glc}$ is the glycol specific heat, $C_{p,glc}$ = 3.35 kJ/(kg K); and T_{hre} and T_{hra} are the glycol temperatures respectively before and after the heat recovery coil, °C.

3.3.2. Models development and validation

New models for the prediction of the factor α are presented for operation modes #2 and #3, and compared to the reference factor α from eqs. 3.5 and 3.6, in which the actual values of the air temperatures at the mixing box inlets and outlet are used.

Field measurements from an existing system are used. The raw data sets are pre-processed: physical, expected and theoretical limits are checked. According to ASHRAE 2005b, observations which do not comply with checked limits are discharged from data sets. Quasi-steady state equilibrium is required when using factor α based on air temperatures. The occurrence of transient phenomena in operation data has been detected and corresponding measurements are discharged from data sets. Short-term measurements (STM) from portable calibrated sensors have been collected for models training and validation only.

Models validation is performed comparing the predictions of the factor α to reference values in terms of statistical indices: the Mean Bias Error (MBE), the Mean Absolute Error (MAE), the maximum occurred Absolute Error (AE_{max}) and the Coefficient of Variance of the Root Mean Square Error (CV-RMSE).

3.3.3. Virtual sensors selection criteria

Two main aspects should be considered when developing or selecting a virtual sensor model: i) the information and measurements needed for model implementation, and; ii) the uncertainty associated to the model predictions. In order to enhance the feasibility of virtual sensors implementation, the model should require for a reduced number of measurements usually

selected from those already available from BASs. A virtual sensor model which needs measurements from several sensors not already installed in AHUs may be difficult or impossible to implement due to the lack of information. Also, costs would increase if additional physical sensors are purchased. On the other hands, a virtual sensor model which only needs for few measurements from commonly installed sensors would be easy to implement without additional costs. Furthermore, the uncertainty associated to models predictions must be considered. Too large prediction uncertainty values can make the model predictions unreliable. Those two criteria have been taken into account, in this research work, along the development and the selection of the proposed virtual sensor models.

3.4. Virtual sensors for the air properties at the AHUs mixing box

In an AHU the measurements of the air temperature and relative humidity at outdoor, recirculated and mixed conditions are needed for system control purpose. Three methods are presented in this research work for the virtual measurement or virtual calibration of faulty sensors of the air temperature and relative humidity at the mixing box inlets and outlet. The methods differ in the model strategy (first principle or data driven models), the needed information and the technical skill required to building operators. The virtual sensors developed using the proposed methods are intended to be used each time new measurements are collected by the BAS, before those measurements are used for control purpose.

The need information, the uncertainty propagation, and the required technical for each method have been assessed.

3.4.1 Method A

Method A is used for the virtual re-calibration of one sensor (e.g. the outdoor air temperature) out of the six air temperature and relative humidity sensors installed in the AHU mixing box (Figure 3.5 and Table 3.1), which has been detected to be faulty by some fault detection method or by the building operators. The faulty measurement itself are used by Method A along with the measurements from remaining fault-free five sensors. Method A uses a first principle-based model based on the energy balance of the AHU mixing box (eqs. 3.3 and 3.4), along with the selection of optimum correction values, to adjust the measurements from a faulty sensor. This section presents, as an example, the case of virtual re-calibration of the outdoor air temperature sensor. The same approach can be applied for any other of the five sensors involved.

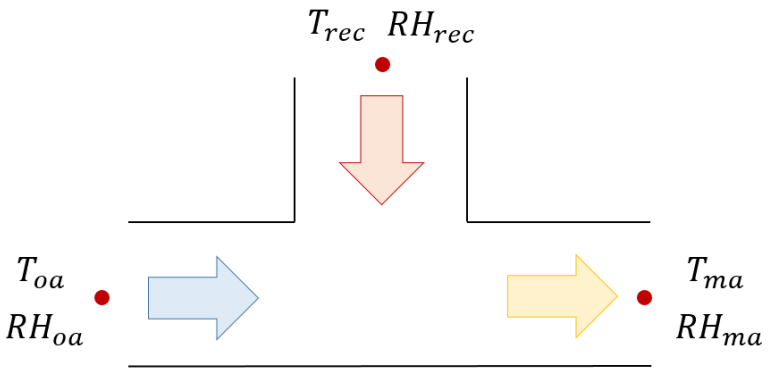


Figure 3.5 – Schematic of the AHU mixing box including temperature and relative humidity of the outdoor (oa), mixed (ma), and recirculated (rec) air flows.

Table 3.1 - List of variables required by method A.

Description	**Units**	**Variable names**
Outdoor air temperature	°C	T_{oa}
Outdoor air relative humidity	%	RH_{oa}
Recirculated air temperature	°C	T_{rec}
Recirculated air relative humidity	%	RH_{rec}
Mixed air temperature	°C	T_{ma}
Mixed air relative humidity	%	RH_{ma}

The air humidity ratio (x) and specific enthalpy (h) at outdoor, recirculated and mixed conditions are derived using the measurements of the correspondent air temperature and relative humidity. For instance, at outdoor conditions, the air humidity ratio and specific enthalpy are derived as follows (eqs. 3.8-3.11):

$$Ps_{oa} = e^{65.81 - \frac{7066.27}{T_{oa}+273.15} - 5.976 \cdot \log(T_{oa}+273.15)} \quad \text{eq. 3.8}$$

$$Pv_{oa} = \frac{RH_{oa}}{100} \cdot Ps_{oa} \quad \text{eq. 3.9}$$

$$x_{oa} = 0.62198 \cdot \frac{Pv_{oa}}{(P - Pv_{oa})} \quad \text{eq. 3.10}$$

$$h_{oa} = C_a \cdot T_{oa} + x_{oa} \cdot \left(h_{fg} + C_v \cdot T_{oa}\right) \quad \text{eq. 3.11}$$

where Ps_{oa} and Pv_{oa} are the saturation and the partial pressure of water vapor, respectively, Pa; P is the atmospheric pressure, P = 101,325 Pa; h_{fg} is the water vaporization heat, h_{fg} = 2501 kJ/kg; C_a and C_v are the dry air and water vapor specific heats at constant pressure, C_a = 1.006 kJ/(kg K) and C_v = 1.875 kJ/(kg K).

Combining the factor α definition (eq. 3.1) with the energy balance equation (eq. 3.3) at the AHU mixing box, the outdoor air fraction (factor α) is calculated with eq. 3.4 as function of the air specific enthalpy at outdoor, recirculated and mixed conditions which are derived from eqs. 3.8-3.11.

At each time step, if one of the six considered sensors is found to be faulty (e.g. the outdoor air temperature), an iterative procedure is implemented to identify the optimum correction term, from a set of candidate correction terms (eq. 3.12), which will be used to correct (virtually re-calibrate) the faulty measurement from the BAS. For instance, for the outdoor air temperature, the vector *dT* of candidate correction terms includes elements from -5.0 to 5.0, with a 0.1 step.

$$dT = [dT_1 , dT_2 , \ldots , dT_{j-1} , dT_j] = [-5.0, -4.9, -4.8, \ldots\ldots\ldots 4.8, 4.9, 5.0] \quad \text{eq. 3.12}$$

For each term of *dT* an iteration of the re-calibration procedure is performed. For instance, for a generic *j* term from *dT*, the following steps are performed:

a. The faulty measurement from the BAS (T_{oa}) is corrected with the *j* term from vector *dT* (dT_j) (eq. 3.13):

$$T_{oa}^{j} = T_{oa} + dT_j \quad \text{eq. 3.13}$$

b. The five fault-free measurements of the air temperature and relative humidity from BAS, along with the faulty measurement corrected with dT_j (T_{oa}^{j}), are used to estimate the air specific enthalpies, and the value of the factor α_j (eq. 3.4);

c. The thermodynamics properties of outdoor air are estimated from the factor α_j: humidity ratio (eq. 3.14), specific enthalpy (eq. 3.15), and temperature (eq. 3.16), which are dependent from the factor α and thus denoted with a subscript 'α'.

$$x_{oa,\alpha}^{j} = \frac{(x_{ma} - x_{rec})}{\alpha_j} + x_{rec} \qquad \text{eq. 3.14}$$

$$h_{oa,\alpha}^{j} = \frac{(h_{ma} - h_{rec})}{\alpha_j} + h_{rec} \qquad \text{eq. 3.15}$$

$$T_{oa,\alpha}^{j} = \frac{h_{oa,\alpha}^{j} - h_{fg} \cdot x_{oa,\alpha}^{j}}{C_p + C_v \cdot x_{oa,\alpha}^{j}} \qquad \text{eq. 3.16}$$

d. The outdoor air temperature derived from eq. 3.16 ($T_{oa,\alpha}^{j}$) is compared to the j corrected measurement from the BAS (T_{oa}^{j}), and their difference ($\Delta T_{BAS,\alpha}^{j}$) is retained (eq. 3.17):

$$\Delta T_{BAS,\alpha}^{j} = \left| T_{oa}^{j} - T_{oa,\alpha}^{j} \right| \qquad \text{eq. 3.17}$$

3.4.1.1 The objective function

At each time step, the objective function $\Delta T_{BAS,\alpha}^{j}$ is minimized. The optimum correction term (dT^*) is thus selected, among the j available candidates, as the one which corresponds to the minimum $\Delta T_{BAS,\alpha}^{j}$. Finally, the optimized (re-calibrated) measurement of the faulty variable (in this case the outdoor air temperature) is given by eq. 3.18.

$$T_{oa}^{*} = T_{oa} + dT^{*} \qquad \text{eq. 3.18}$$

where T_{oa}^{*} is the optimized (re-calibrated) measurement of the faulty variable, and dT^* is the selected optimum correction term.

Given a set of six, fully fault-free measurements of the air temperature and relative humidity at the mixing box inlets and outlet, the correspondent factor α (eq. 3.4) is expected to be fault-free, and so the terms derived from the fault free factor α with eqs. 3.14, 3.15 and 3.16 (if no correction terms are used to correct any initial variable, $j = 0$). In particular, if the output from eq. 3.16 ($T_{oa,\alpha}^{j=0}$) is faulty free, it must be the same as the initial measurement from the BAS (T_{oa}), thus $\Delta T_{BAS,\alpha}^{j=0} = \left| T_{oa} - T_{oa,\alpha}^{j=0} \right| = 0$.

If one over the six air properties measurements is faulty (e.g. T_{oa}), the derived factor α is expected to be faulty, and so the derived air specific humidity ($x_{oa,\alpha}^{j=0}$) from eq. 3.14 and the derived air specific enthalpy from eq. 3.15. The initial and the derived air specific enthalpies (h_{oa} and $h_{oa}^{j=0}$) have the same value as eq. 3.4 and eq. 3.15 are equivalent. On the other hands, the initial value of air specific humidity (x_{oa}) and the derived one ($x_{oa}^{j=0}$) are different because calculated through different equations (eqs. 3.8-3.10 for x_{oa} and eq. 3.14 for $x_{oa}^{j=0}$), and thus affected in two different ways by the initial error in the measurement of the outdoor air temperature. As an example figure 3.6 shows the divergence between the initial and derived outdoor air specific humidity values for a set of six faulty free measurements where the initially faulty free outdoor air temperature has been modified adding an artificial error.

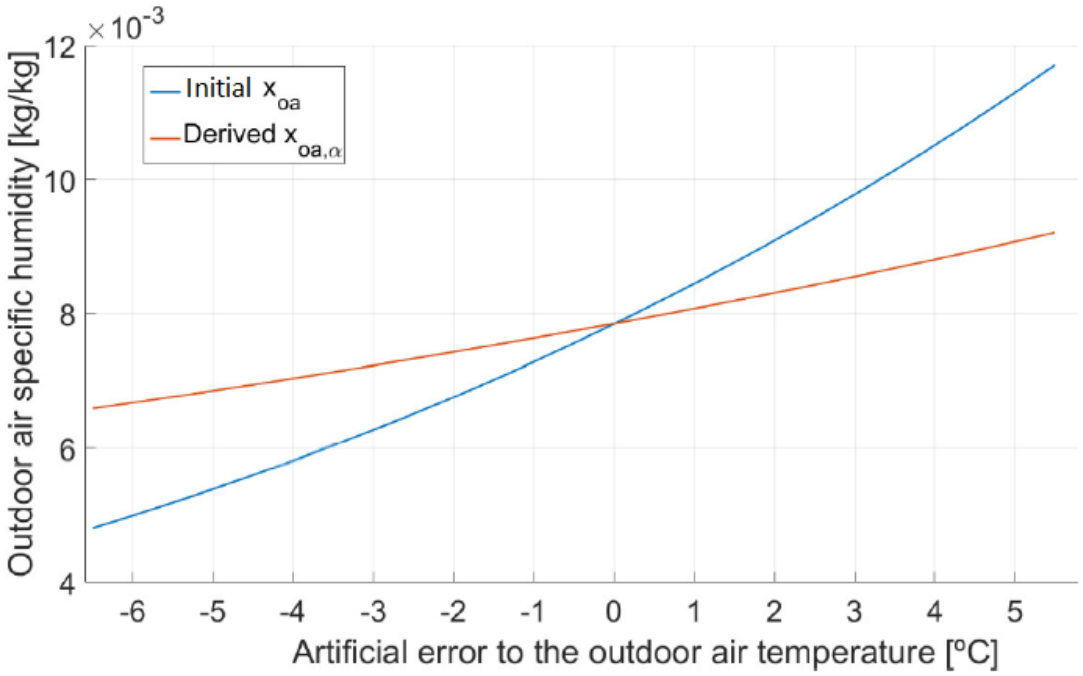


Figure 3.6 – Initial (eqs. 3.8-3.10) and derived (eq. 3.14) outdoor air specific humidity values as function of the error affecting the outdoor air temperature.

Finally, the derived faulty $T_{oa,\alpha}^{j=0}$ would not be the same as the initial faulty T_{oa} ($\Delta T_{BAS,\alpha}^{j=0} = \left|T_{oa} - T_{oa,\alpha}^{j=0}\right| > 0$), and the reason is the $T_{oa,\alpha}^{j=0}$ is derived from $x_{oa,\alpha}^{j=0}$ which is different from the one derived from T_{oa} (x_{oa}). The Initial and derived outdoor air temperatures converge to the same value when no initial faulty measurements produce a divergence between x_{oa} from eqs. 3.8-3.10

and $x_{oa,\alpha}^{j=0}$ from eq. 3.14. Thus, the re-calibration procedure consists of an iterative process which tests a number of candidate correction terms, and the select the optimum one through an objective function which minimize the term $\Delta T_{BAS,\alpha}^{j=0} = \left|T_{oa} - T_{oa,\alpha}^{j=0}\right|$.

Method A is here presented for virtual calibration (self-correction). Nevertheless, method A can be used for virtual measurement if one over the six air properties is not available. A default value should be assigned to the missing variable, which would then be corrected with method A.

3.4.2 Method B

Method B consists of developing grey box (data driven) models for the prediction of a variable of interest using other correlated variables. There are three models B for the three different operation modes of an AHU.

The variables of interest are modelled as function of other variables available from the BAS. Short-term measurements (STM) from calibrated portable sensors are collected to be used as reference for models validation.

3.4.3 Method C

Method C aims to virtually re-calibrate a faulty sensor. STM from portable calibrated sensors are used, along with correlation analysis to estimate the sensor error. The estimated sensor error is then used to correct the faulty measurements from the BAS. As an example, method C is used for the re-calibration of an outdoor air temperature faulty sensor.

STM from calibrated portable sensors are collected to derive the error *e* from comparison against the faulty measurements, and to be used as reference for models validation.

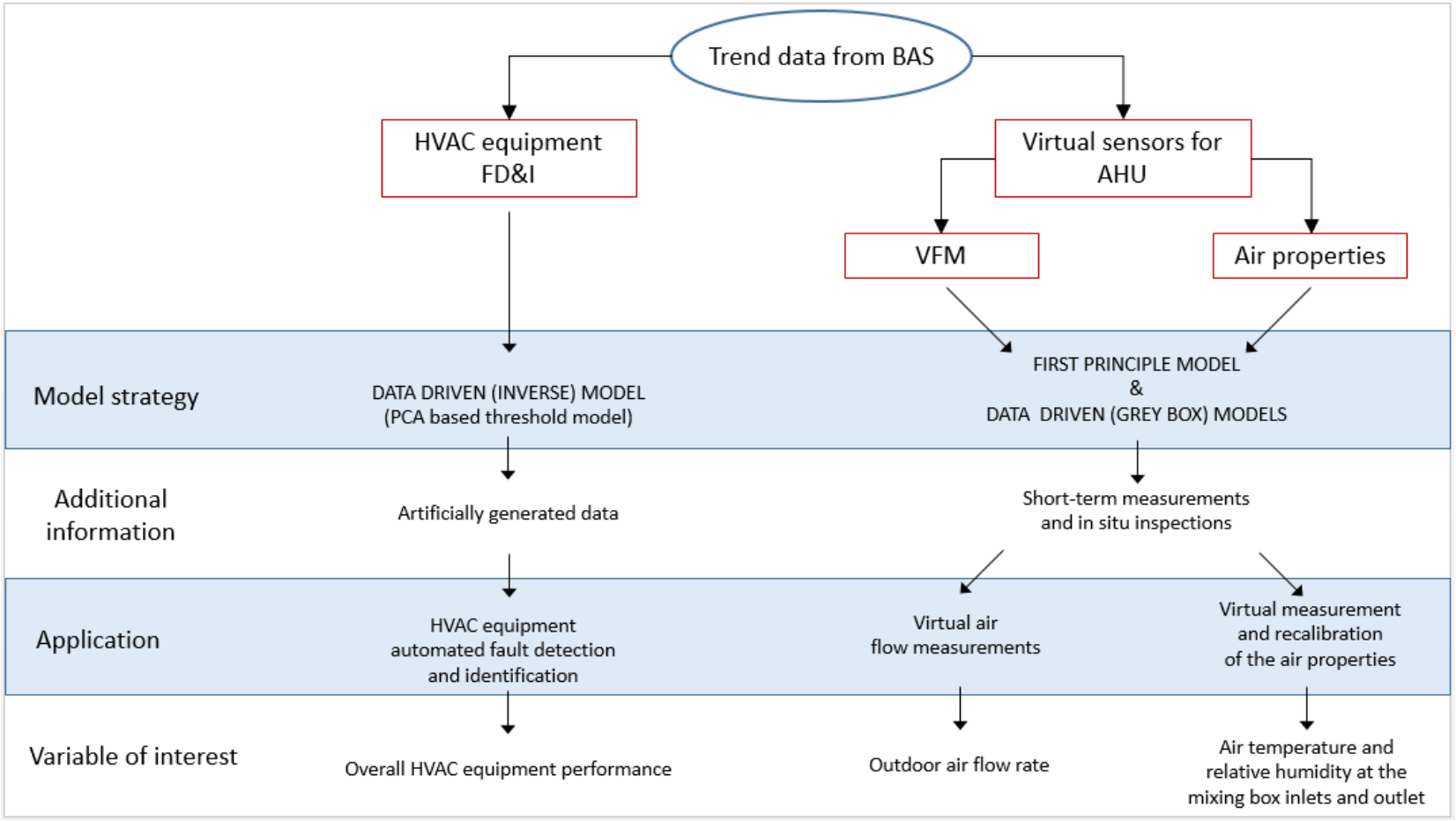


Figure 3.7 – Features of the proposed methods and models for the OCx of HVAC.

3.5. Uncertainty analysis

The measurements of a variable are always affected by some residual from the true value. The occurrence of residuals is accounted for through the uncertainty analysis. The uncertainty of a measurement is defined as the interval, around the measured value, within the true value is expected to fall with a certain level of confidence (e.g. 95%). A measured value without any statement about the correlated uncertainty has limited meaning (Reddy 2011). For engineering application purpose, two main sources of error affecting measurement are considered: the random error (R) and the bias (fixed) error (B). Those two errors contribute to define the overall uncertainty affecting the measurements (U). If a value is not directly measured, but it is derived by other measurements through some mathematical formulations, the uncertainty propagates from the measurements, through the mathematical formulation, to the derived value. According to ASHRAE (2005b), the random and bias errors are here introduced along with the overall uncertainty analysis and its propagation through mathematical formulations. Further, a procedure to remove the effect of the bias (fixed) error from measurements residual is presented.

The random error (R_x) is a deviation from the true value which varies at each observation following some probability distribution. Usually the random error distribution around the mean (the true value) is approximated to a normal (Gaussian) distribution. Thus, given a population of n measurements only affected by the random error, the true value is approximated to the mean of the measurements (eq. 3.19):

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{eq. 3.19}$$

where $\bar{X}$ is the mean value from the population of n measurements X_i.

For measurements only affected by the random error, R_x represents the amplitude of the interval, around the measured value, within which the true value is expected to fall, with a given level of confidence. For a level of confidence of 95%, the random error is given by (eq. 3.20):

$$R_x = Z \cdot S_x = 1.96 \cdot S_x \qquad \text{eq. 3.20}$$

where S_x is the standard deviation of the measurements distribution around the mean, and Z = 1.96 accounts for 95% of the measurements population around the mean.

The bias (fixed) error (B_x) is defined as a constant deviation of the measurements from the true value of an x variable. The bias error is mainly due to sensors miscalibration and installation issues. For a given level of confidence, the bias error is provided by the sensor manufacturer as the maximum deviation of measurements from the true value. The true value is identified with a reference sensing device. The given bias error is a constant value, derived from a larger distribution of possible bias errors, which is assumed to be normally distributed around the mean true value. Thus, given a population of bias errors normally distributed around the true value of a variable x, the bias error B_x with a confidence level of 95% is (eq. 3.21):

$$B_x = Z \cdot S_B = 1.96 \cdot S_B \qquad \text{eq. 3.21}$$

where S_B is the standard deviation of the bias errors distribution around the mean, and Z = 1.96 accounts for 95% of the bias error population around the mean.

The random and the bias (fix) errors of a variable x are combined together to evaluate the overall uncertainty (U_x) which affects the measurements of the variable x with a given level of confidence (eq. 3.22). When a dependent variable Y is derived from j variables x, the random and the bias (fixed) errors propagate through the function $Y = f(x_1, x_2, \dots, x_j)$ to the dependent variable Y. The propagation of the random and bias errors is evaluated separately with eq. 3.23 (which applies to both bias and random errors) and then combined through eq. 3.22.

$$U_x = \sqrt{B_x^2 + R_x^2} \qquad \text{eq. 3.22}$$

$$S_Y = \sqrt{\sum_{j=1}^{n} \left(\frac{\partial Y}{\partial x_j} S_{x_j} \right)^2} \qquad \text{eq. 3.23}$$

Figure 3.8 shows a schematic of the contribution of the bias and random errors on each measurement (or derivation) of a value.

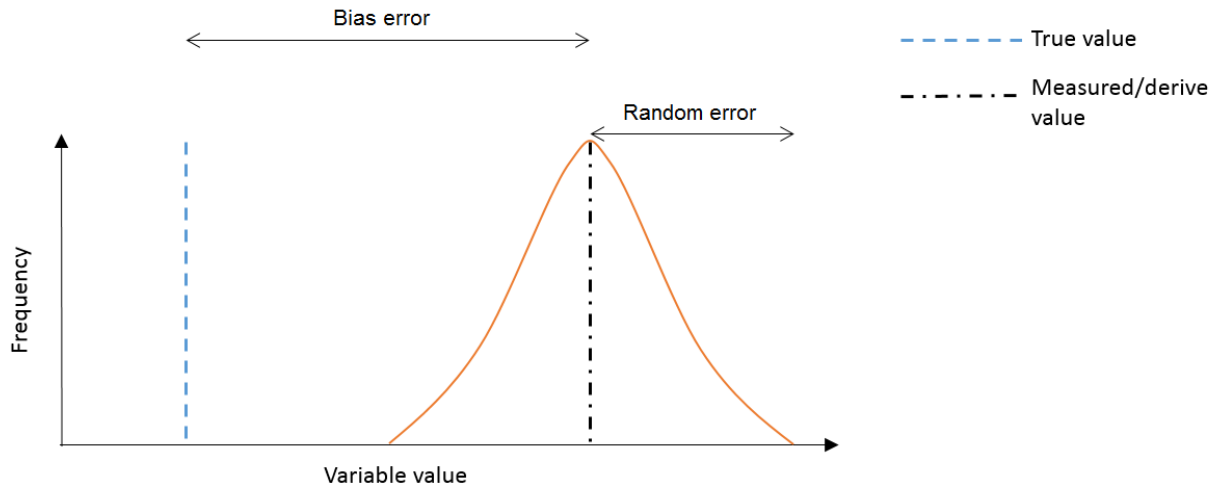


Figure 3.8 – Schematic of the uncertainty on a measured (or predicted) value, due to the bias (fixed) error and the random error.

3.5.1. Bias (fixed) error effect: estimation and removal

A procedure is here proposed to remove the effect of the bias (fixed) error (Figure 3.8) from the measurements (or predictions). As introduced above, the two main sources of error accounted for in engineering data analysis are: i) the bias (fixed) error, which is constant and it is not expected to change through time, and ii) the random error, which change at each new observation, whose distribution curve is often considered to be symmetrical around the mean. Both are associated to a certain level of confidence, usually 95%, and so it is the overall uncertainty given by the combination of those two errors (eq. 3.22). It is worth to remark that the bias error B_x, as stated by the manufacturer, is the maximum possible value of B_x with a given level of confidence. Although its actual value is unknown, the bias (fixed) error affecting a single specific physical sensor will be smaller than the maximum B_x given by the manufacturer. On the other hands, because the random error distribution is approximated to a Gaussian (symmetric) distribution (ASHRAE 2005b), the average random error on the measurements x (or predicted variables Y) is zero. Thus, averaging a statistically relevant number of residuals between measurements and true values would remove the effect of the random error on the measurements.

If reference values of the variable of interest are collected from a *calibration* data set (e.g. short term measurements with high precision sensing equipment), the average residual between reference values and measurements would give an estimation of the constant effect of the bias

error on the measurements. The constant effect of the bias error on measurements is estimated through the Mean Bias Error (MBE) between measurements and reference values (eq. 3.24). Once the constant contribution of the bias error on the measurements is estimated along the *calibration* data set, it can be removed from further measurements when the reference values are not available (eq. 3.25). The resulting, *unbiased* measurements would be affected by the random error only (Figure 3.9).

$$\text{MBE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i) \qquad \text{eq. 3.24}$$

$$x_{i,c} = \hat{x}_i \text{ - MBE} \qquad \text{eq. 3.25}$$

where $\hat{x}_i$, x_i and $x_{i,c}$ are the biased measurement, the reference value and the unbiased measurement respectively, at the *i* time step.

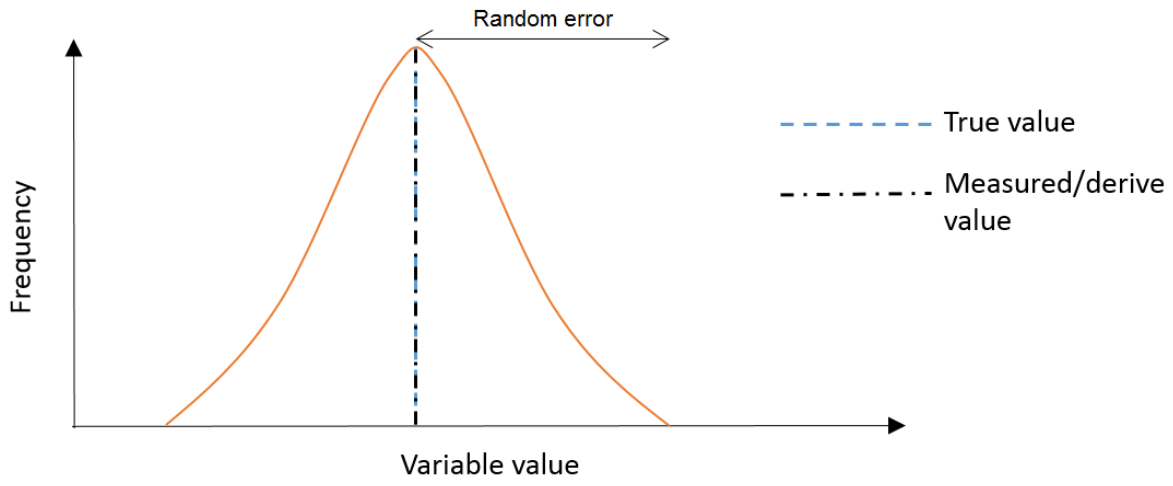


Figure 3.9 – Schematic of a measurement (or prediction) affected by the random error only (*unbiased*).

The MBE is an estimation of the effect of the bias (fixed) error on the measurements of a variable of interest. Three main issues may prevent from good estimation of the bias effect through the MBE:

i) the size *n* of the population of measurements used to average the residual (eq. 3.24): increasing the size *n* the precision of the estimation would increase;

ii) the random error may not take a Gaussian (thus symmetrical) distribution, thus the use of eq. 3.24 to estimate the bias (fixed) contribution to the overall error risks to be affected by the random component;

iii) the available *true values* of some variable of interest are the best available estimation of the considered variable, which are often measured by some high precision instrument, thus affected by some error too: eq. 3.24 will give an estimation of the effect of the bias error from the reference values, and not from the true value.

4. CASE STUDY

The case study used in this research work includes: i) the cooling plant serving several buildings at Loyola Campus, Concordia University, in Montreal, and ii) the air handling system installed In the Genomic Research building (GE building) at the same university campus (Figure 4.1).

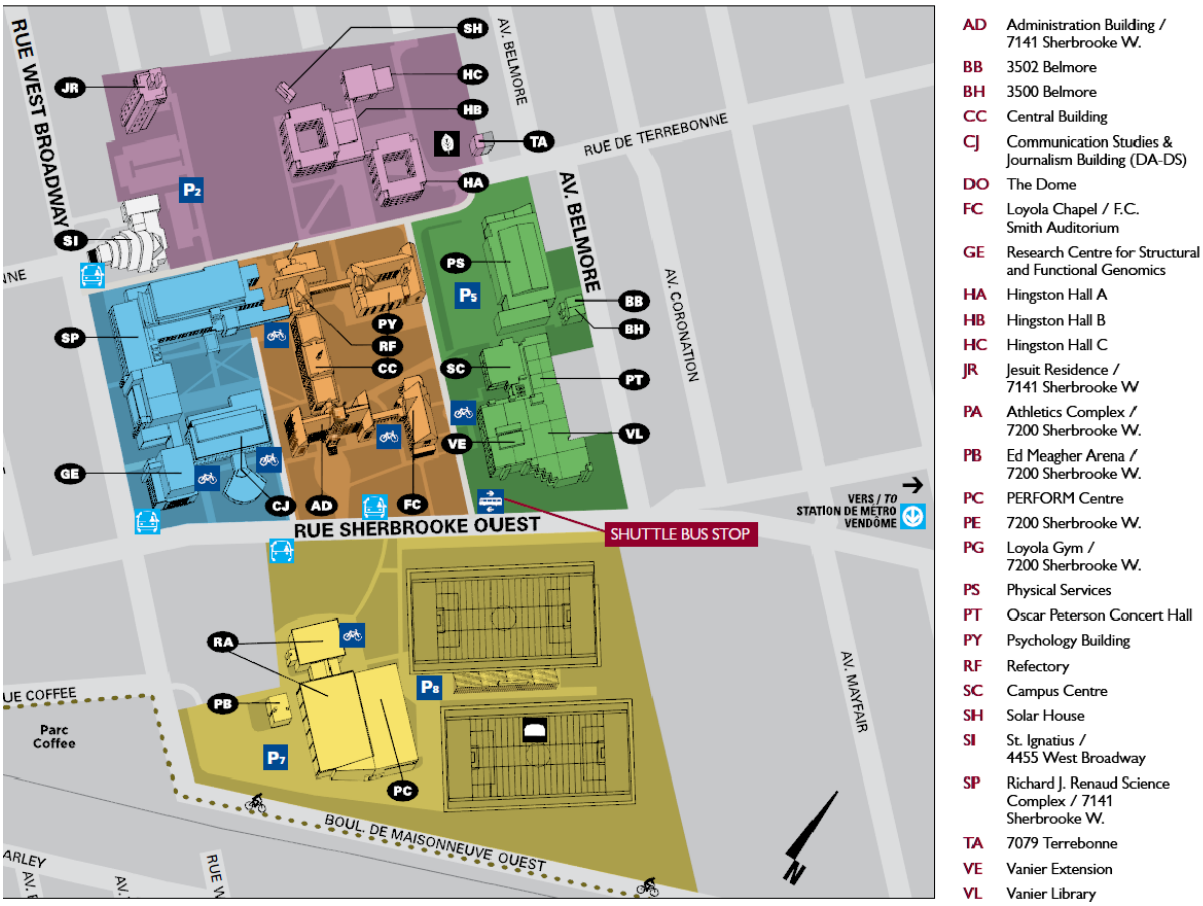


Figure 4.1 – Map of the Loyola Campus, Concordia University, Montréal, Qc, Canada (extracted from www.concordia.ca).

A Building Automation System (BAS) collects measurements of the systems operation from more than 200 points with a 15 minutes time step.

4.1. The cooling plant

The cooling plant is located in the SP building and provides chilled water to three buildings within the campus: the Science (SP) building itself, the Administrative (AD) building, and the

Genomic Research (GE) building. The cooling plant consists of two sub-systems (#1 and #2), each of them including a centrifugal chiller of 3,165 kW (900 tons) cooling capacity, two constant speed pumps, and one perpendicular flow cooling tower. When one of the chillers starts, the correspondent pumps and cooling tower are started as well. If one chiller is not sufficient to match the thermal load, the second one starts, working simultaneously with the first one. Thus, four different operation modes are identified: i) both chillers and corresponding pumps and cooling towers work; ii) only chiller CH-1 and corresponding pumps and cooling tower work; iii) only chiller CH-2 and corresponding pumps and cooling tower work; iv) the entire cooling plant is turned off. A heat exchanger (HX-3) is used to recover heat from the condenser supply water, and use it to pre-warm the heating water return. Up to 80% of the condenser supply water is directed to HX-3 and then mixed with the remaining 20% before being sent to the cooling tower. The cooling tower fans are turned off when the outdoor conditions allow it.

The as operated equipment parameters can be extracted from the BAS trend data. Measurements from the summer season in 2009 are here used to derive those parameters. For the present research work the cooling plant operation from 2009 is taken as reference for further analysis.

Figure 4.2 shows a schematic of the central plant equipment along with location of the sensors used for control purpose by the BAS.

4.1.1. The chillers

The two centrifugal chillers are rated with a 549 kW power input, and a design COP of 5.76. The supply chilled water temperature set-point is 6.7°C, while the average supply condenser water temperature is 35.0°C for both chillers (as operated). Whenever one chiller is not enough to match the cooling load (electrical power input higher than 525 kW), the second chiller starts as well. The two chillers working simultaneously operate at the same power input, of 200 and 400 kW each (Figure 4.3). The electrical power input to the chillers increases with the outdoor air temperature, and reaches its maximum value at around 22.0°C of outdoor air temperature (Figure 4.4).

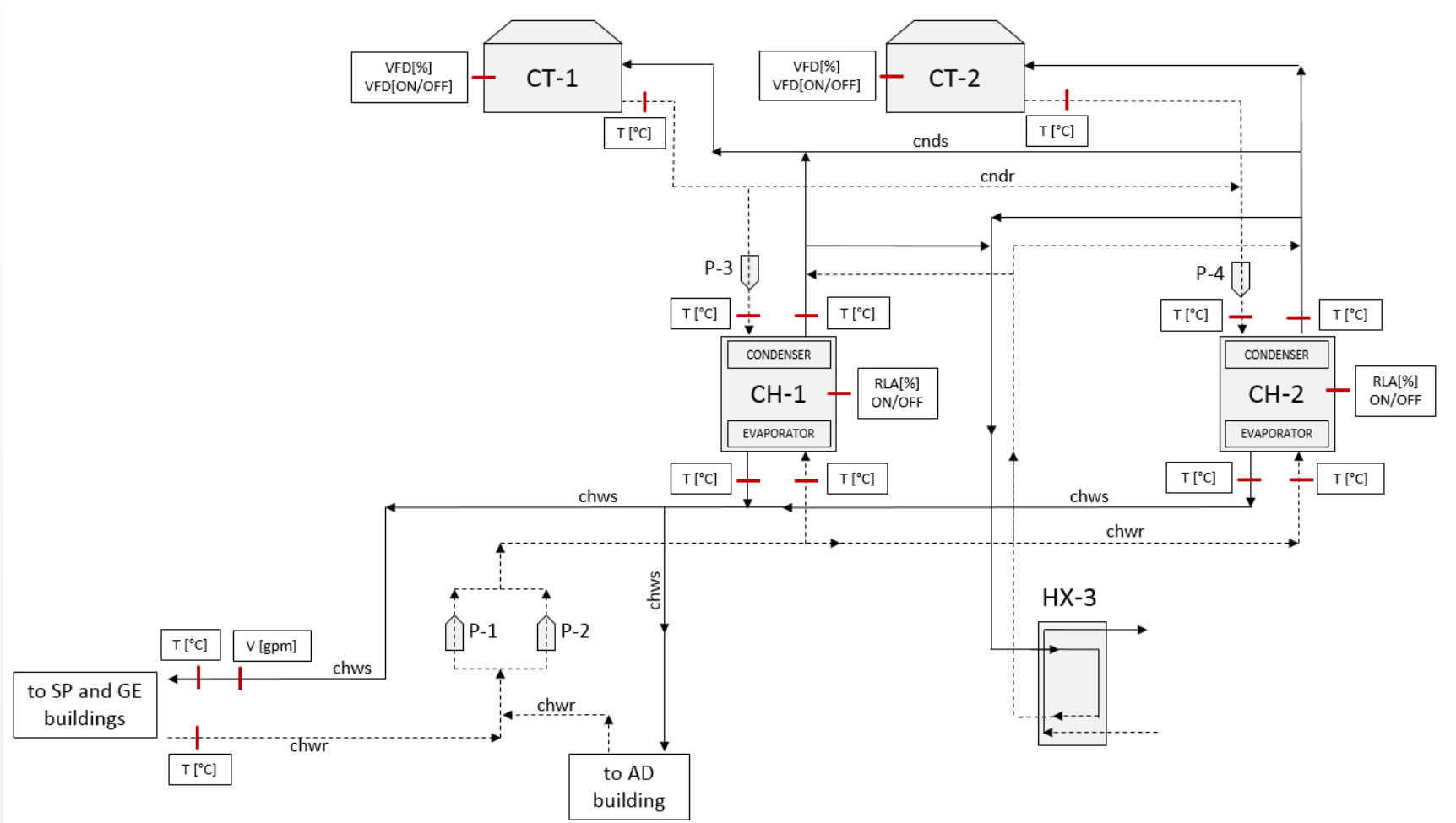


Figure 4.2 – Schematic of the cooling plant along with sensors location (red bars).

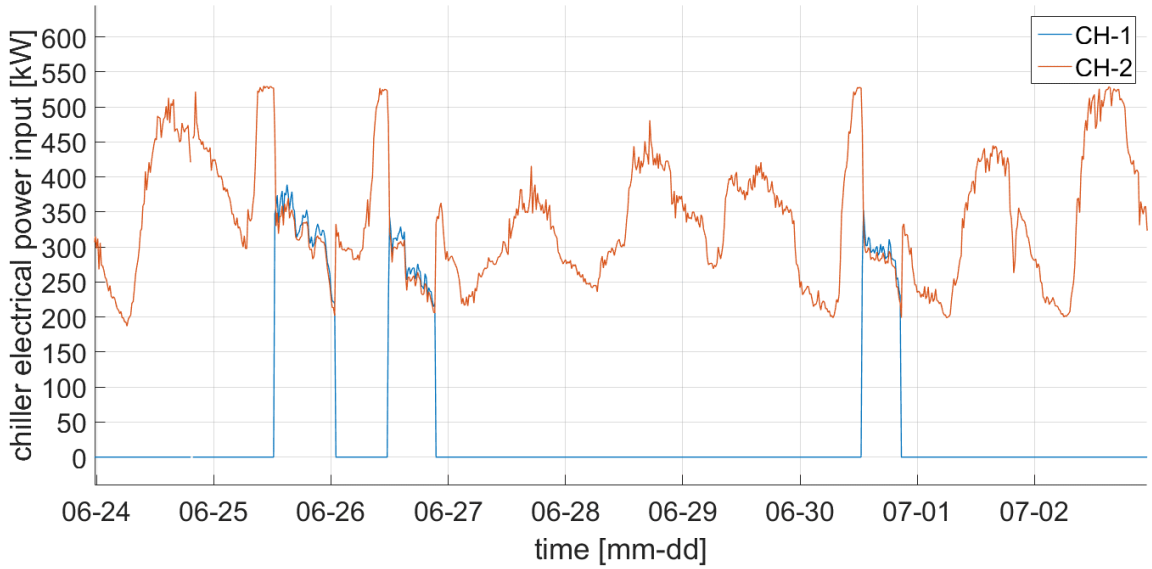


Figure 4.3 – Electrical power input to the two chillers along several days in 2009.

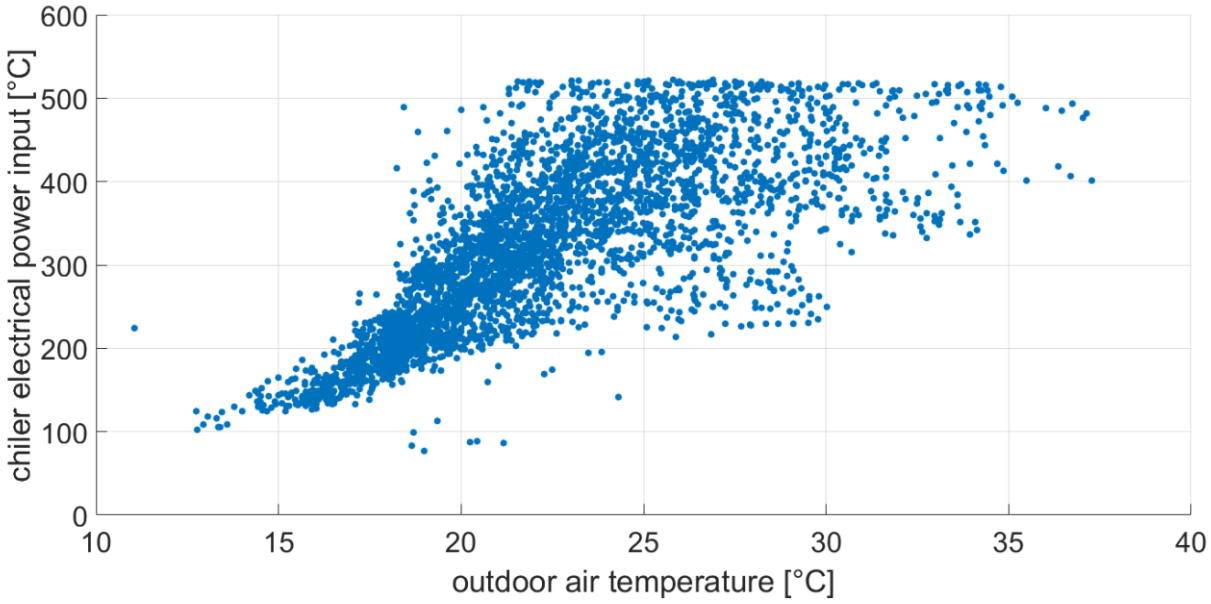


Figure 4.4 – Electrical power input to CH-1 against the outdoor air temperature when CH-1 works alone, between June 1 and August 30, in 2009.

4.1.2. The cooling towers

The cooling towers have a capacity of 4750 kW (1350 tons) at design condition. The supply cooling tower water temperature is set at 29.0°C (as from trend data). Variable Frequency Drive VFD is used for cooling towers fans control. When a chiller starts, the corresponding cooling tower starts as well. When the outdoor air temperature allows it, the cooling towers are turned off. The fans VFD signal increases with the outdoor air temperature, ranging between 30% and 100% (Figure 4.5).

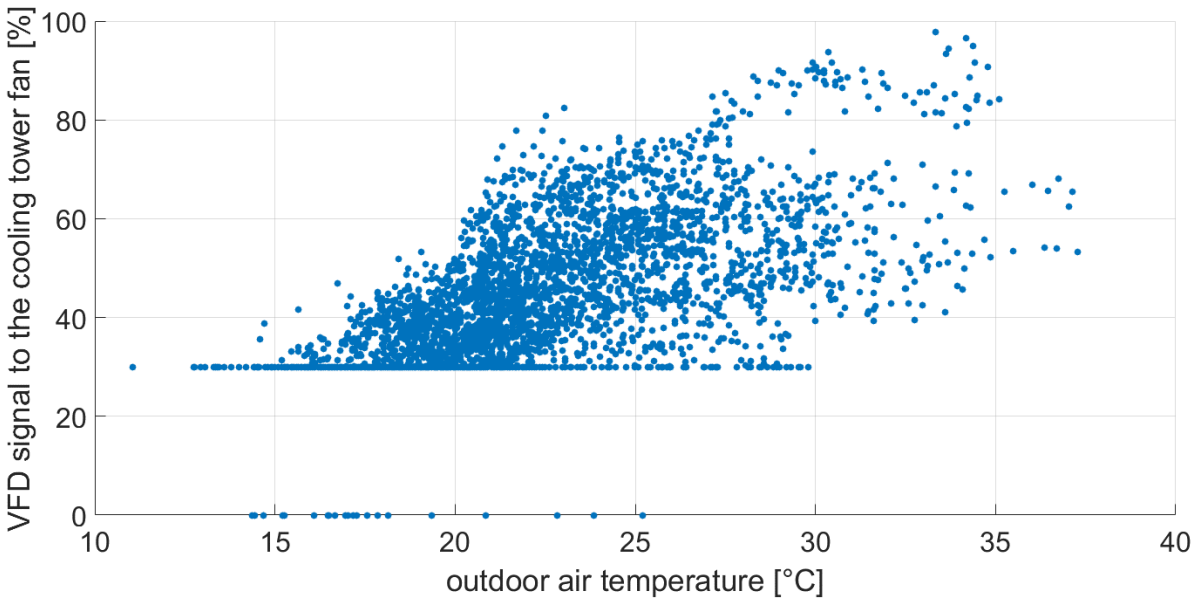


Figure 4.5 – VFD signal to CT-1 fan against the outdoor air temperature when CH-1 is the only chiller working, between June 1 and August 30, in 2009.

4.1.3. The pumps

Two pumps installed in parallel (P-1 and P-2) drive the chilled water, between the chiller evaporators and the cooling coils in the buildings (Figure 4.2). When one chiller is started, the corresponding pump is started as well. If two chillers work simultaneously, the two constant speed pumps work at the same time. Other two constant speed pumps (P-3 and P-4), one for each chiller, drive water between the chiller condensers and the cooling towers. Spot measurements of the water flow rate on the evaporator and condenser constant speed pumps have been collected in 2014, and reported in Table 4.1 (Mcdonald 2014).

Table 4.1 – Constant speed pumps spot measurements, 2014.

Chilled water flow rate [kg/s]	
P1	89.8 ± 2.7
P2	90.1 ± 2.7
P1 & P2	151.7 ± 3.5
Condenser water flow rate [kg/s]	
P3	112.4 ± 3.8
P4	-

4.2. The Air Handling Unit (AHU)

The Air Handling Unit (AHU) is installed at the Genomic (GE) research building. The GE building was completed in 2011, and was certified LEED Gold in 2013. The AHU consists of two AHU sub-systems (AHU#1 and #2) installed in parallel. Each sub-system includes a recovery coil (R-HC), a mixing box, an heating coil (HC), a humidifier (HH), a cooling coil (CC), and two supply funs in parallel. Two return fans installed in parallel extract the air from the occupied space, which is then rejected outside or recirculated by controlling the mixing and rejection dampers (Figure 4.6). Sub-system #2 handles and supplies roughly double the air mass from sub-system #1. From figure 4.7, two different daily profiles are distinguished: week days and weekends.

Three different AHU operation modes are identified:

- **Operation mode #1**: the mixing dampers are completely closed, 100% of the supply air flow comes from the outdoor air intake, and 100% of the return air flow is rejected outside the building;
- **Operation mode #2**: the mixing dampers are modulated and part of the return air flow is recirculated and mixed up in the mixing box with the outdoor air flow;
- **Operation mode #3**: the heat recovery system is turned on, the outdoor air flow is pre-heated with two coils before it reaches the mixing box and mix up with the recirculated air flow.

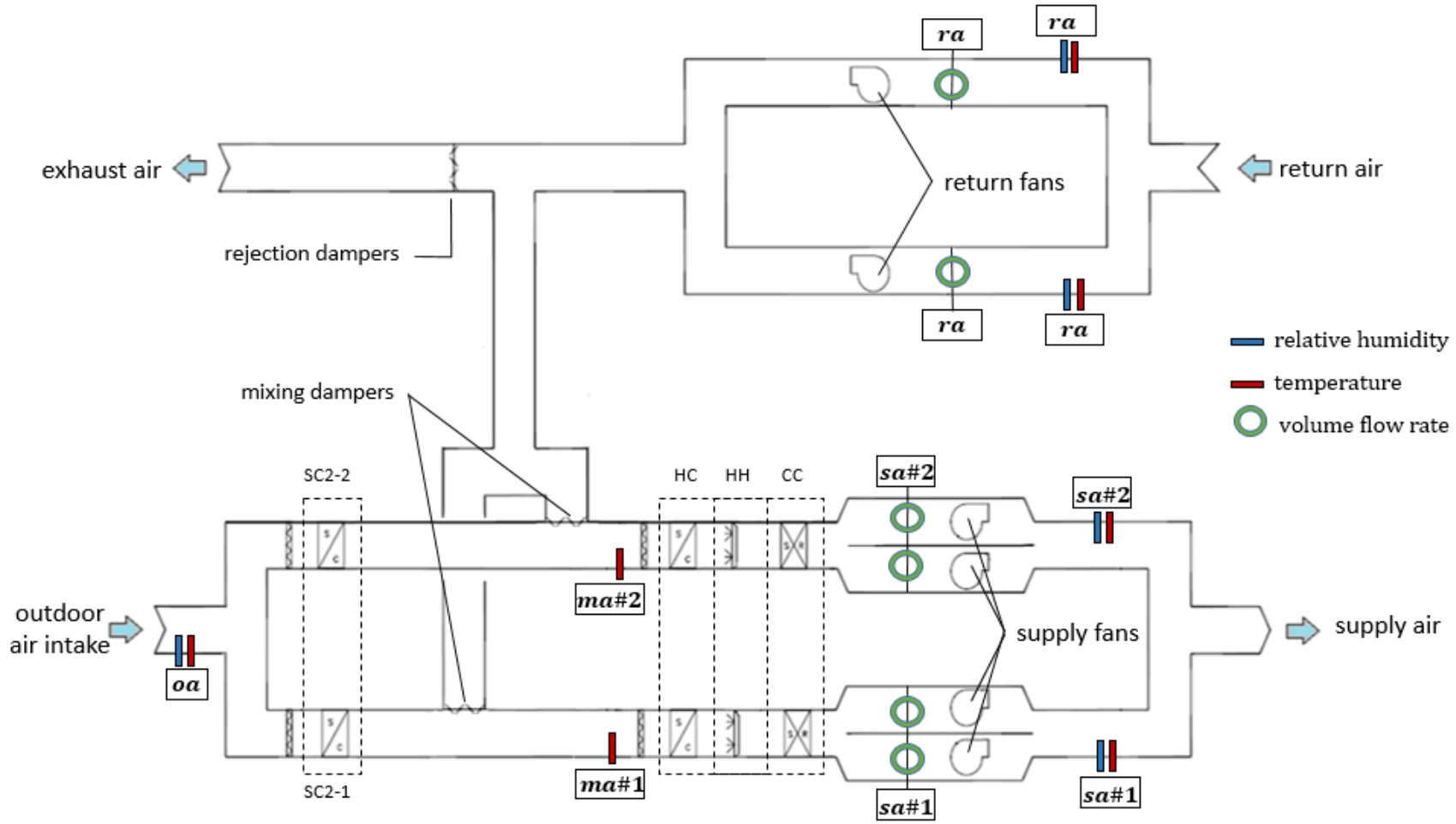


Figure 4.6 – Schematic of the case study air handling system with measurements location: outdoor (oa), return (ra), mixed (ma), supply (sa), recirculated (rec), and after the pre-heating coil (ac).

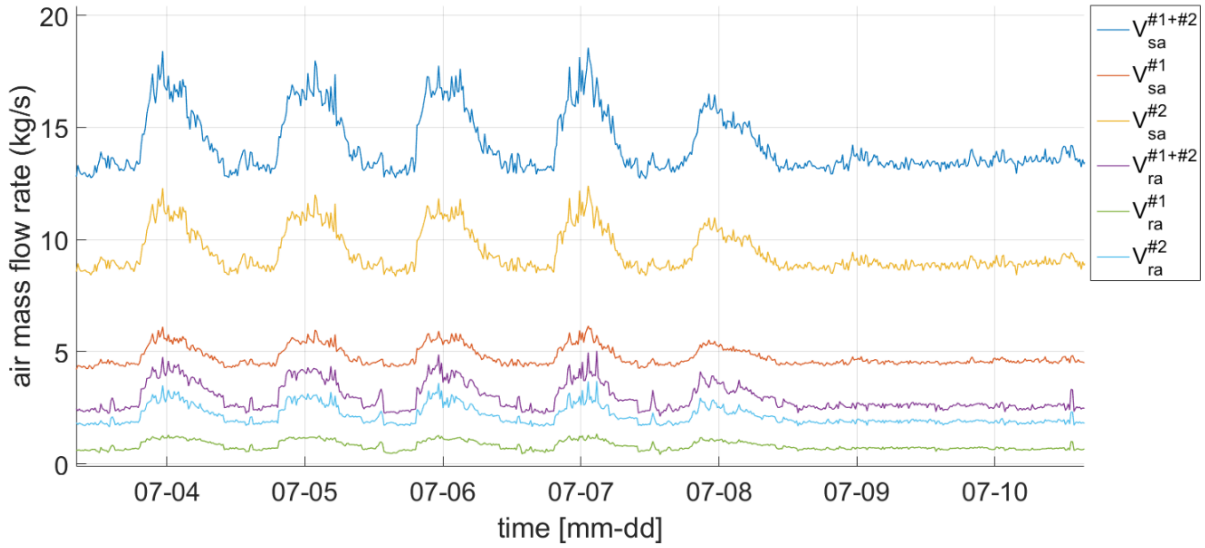


Figure 4.7 – Supply and return air mass flow rates during one week of July in 2016.

4.2.1. The mixing dampers

The mixing dampers are modulated between the closed position (at the dampers signal of 100%, the supply air flow rate is entirely taken from outdoor) and fully open (at the dampers signal of 0%, the maximum portion of return air flow is recirculated) when the outdoor air temperature varies between 10°C and 20°C (Figure 4.8).

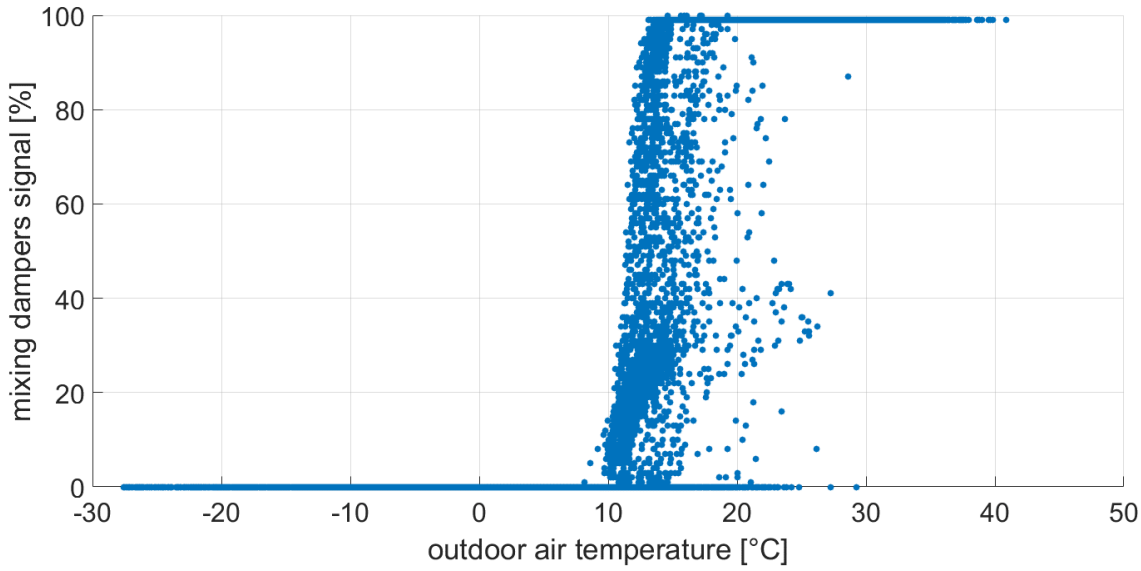


Figure 4.8 – Mixing dampers control signal against outdoor air temperature in 2016.

4.2.2. The heat recovery system

A recovery loop is used to recover heat from the rejected air stream coming from laboratories and washrooms, which is not recirculated. The recovered heat is then used to pre-heat the outdoor air stream when the outdoor air temperature drops below 8°C. Heat is recovered through one recovery coil (SR1-3), and it is transferred to the outdoor air stream through two coils (SC2-1 and SC2-2) (Figure 4.9). A constant speed pump (P03) drives a glycol based liquid (glycol/ethylene 50/50) through the recovery loop pipes. The constant speed pump P03 is rated at 11.8 L/s by the manufacturer. From spot measurements from 2014, the glycol constant flow rate resulted to be 10.0 L/s ± 0.5 L/s (Zibin 2014). In order to avoid frost issues at coil SR1-3, a three way valve (3V) is controlled to deviate part of the glycol, maintaining the glycol temperature at SR1-3 inlet (T_{hra}) higher than 4°C (Figure 4.10). When the recovery loop works, the mixing dampers are always fully open.

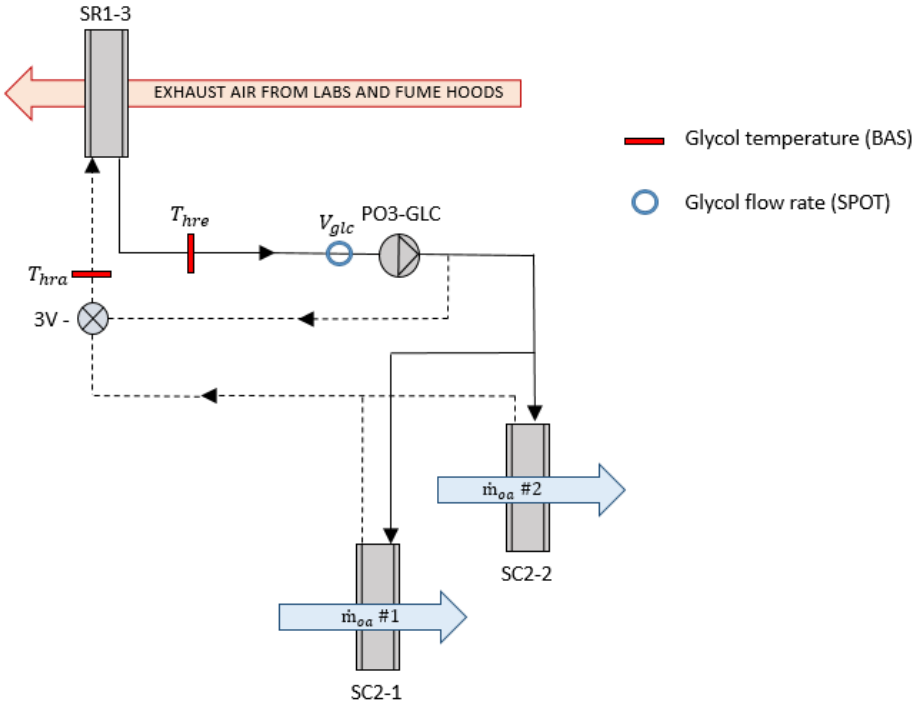


Figure 4.9 – Schematic of the case study recovery loop.

The rate at which the heat is recovered varies with the outdoor air temperature. It reaches its maximum (340 kW) at around T_{oa} = -6°C, and decreases to 250 kW for lower temperatures. Taking the glycol flow rate as constant (10.0 L/s as measured on 2014), the heat recovered through coil SR1-3 is calculated from the measurements of the glycol at the inlet and outlet of coil SR1-3 (Figure 4.11).

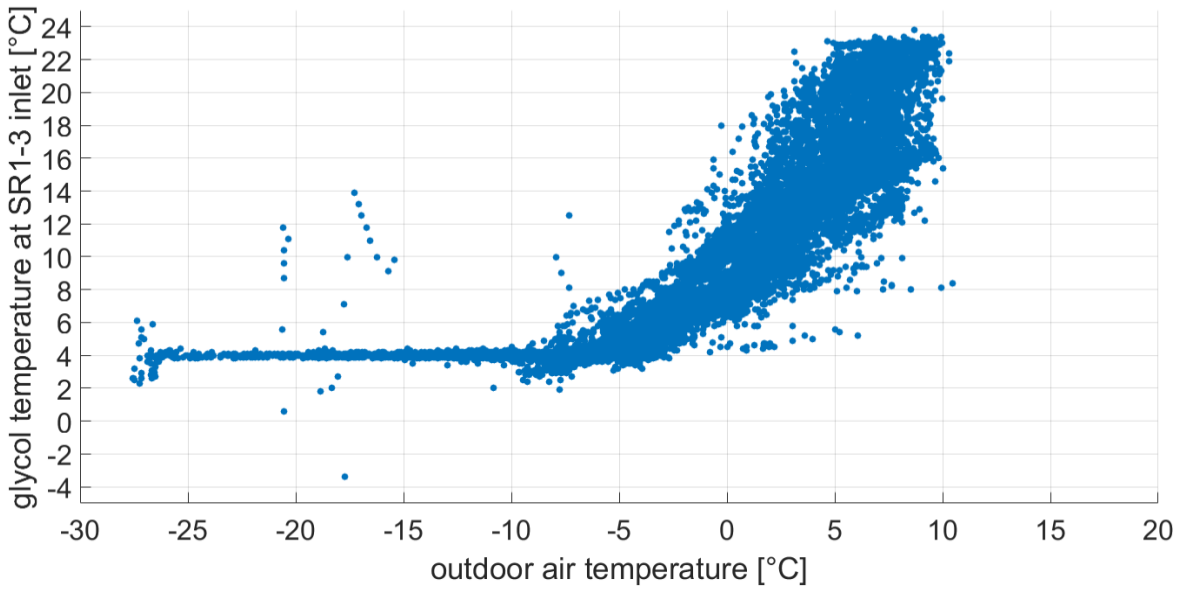


Figure 4.10 – Glycol temperature at SR1-3 inlet (T_{hra}) against the outdoor temperature in 2009.

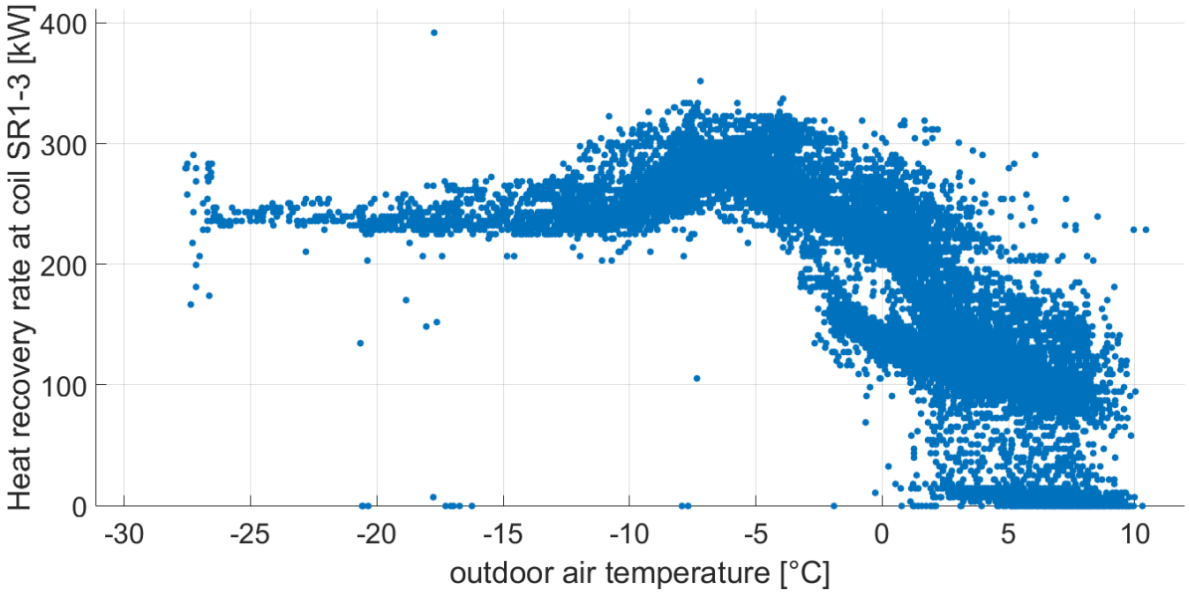


Figure 4.11 – Heat recovery rate at coil SR1-3 against the outdoor air temperature.

4.3. Measurements

The BAS collects measurements of variables of interest with a 15 minutes time step for control purpose. Over 220 variables are measured, including power inputs to equipment, water temperature and flow rate, and air temperature, relative humidity and flow rate, valves and dampers

modulation. Table 4.2 lists the variables available from BAS used in this study. The location the correspondent sensors is reported in figures 4.2, 4.6 and 4.9.

Table 4.2 - List of points measured by the BAS used in this research.

Description	Units	Variable names
Outdoor air		
Outdoor air temperature	°C	T_{oa}
Outdoor air relative humidity	%	RH_{oa}
Air Handling Units (#1 and #2)		
Supply air flow rate by fans	L/s	VA
Return air flow rate by fan	L/s	VR
Supply air temperature	°C	T_{sa}
Supply air relative humidity	%	RH_{sa}
Return air temperature	°C	T_{ra}
Return air relative humidity	%	RH_{ra}
Mixed air temperature	°C	T_{ma}
Cooling coil valve modulation	%	$CC_{\%}$
Heating coil valve modulation	%	$HC_{\%}$
Mixed air damper modulation	%	$MD_{\%}$
AHU Heat Recovery System		
Glycol temperature entering glycol HX	°C	T_{hra}
Glycol temperature leaving glycol HX	°C	T_{hre}
Pump operation status	ON/OFF	P03
Cooling Plant (#1 and #2)		
Chiller pumps operation status	ON/OFF	P_1 ; P_2
Condenser pumps operation status	ON/OFF	P_3 ; P_4
Supply chilled water temperature	°C	T_{CHWS}
Return chilled water temperature	°C	T_{CHWR}
Supply condenser water temperature	°C	T_{CNDS}
Return condenser water temperature	°C	T_{CNDR}
Electric power input to the chiller	kW	E_{CH}
Cooling towers supply water temperature	°C	$T_{CT\text{-}S}$
Cooling towers fan operation level	%	VFD_{CT}

4.3.1. Short-Term Measurements (STM) and Spot measurements

For the purpose of measurements validation, Short-Term Measurements (STM) and Spot measurements have been collected for several variables of interest at the AHU using calibrated

portable sensors. Short-term and spot measurements have been used to provide values for those variables which were not collected from the BAS: the air temperature at the mixing box inlets (after the recovery coil and behind the mixing dampers: T_{ac} and T_{rec}), the air relative humidity at mixing condition (RH_{ma}), and the glycol volume flow rate through the recovery loop pipes (V_{glc}).

The portable sensors used to measure the air temperature and relative humidity in AHU ducts are SmartReader data loggers by ACR Systems Inc. (ACR System Inc. 2012). Table 4.3 reports product specifications as given by the manufacturer. Those portable sensors have been recalibrated by the manufacturer in August 2013.

Table 4.3 – SmartReader product specifications (ACR System Inc. 2012).

Temperature	
Type	NTC Thermistor
Range	-40°C to 70°C
Accuracy	±0.2°C over the range of 0°C to 70°C
Relative Humidity	
Type	Capacitive thin polymer film
Range	0 to 95% RH (non-condensing)
Accuracy	±4% RH from10 to 90% RH

The glycol volume flow rate through the recovery loop pipes has been measured with a portable ultrasonic flow meter, Portaflow PT400, by Greyline instruments Inc. The air velocity meters, used to derive air flow rate before each of the four supply and two return fans have an operation range, as given by the manufacturer, between 0.1 m/s and 20 m/s, corresponding to a flow rate between 0.5 L/s and 70 L/s. The accuracy is stated to be between 0.2% and 2% of speed reading for velocidy higher than 0.2 m/s (Greyline instruments inc 2013). Table 4.4 reports the random and bias (fixed) errors, as well as the overall estimated uncertainty of the STM and BAS measurements from the AHU.

Table 4.4 – Sensors Uncertainty.

Sensors	Unit	Fixed (bias)	Random	Uncertainty
BAS - Air temperature at 30°C	°C	0.45	0.190	0.49
BAS - Air relative humidity	%	4.00	0.150	4.00
BAS - Glycol temperature at 30°C	°C	0.45	0.075	0.46
STM - Air temperature	°C	0.20	0.001	0.20
STM - Air relative humidity	%	0.80	0.230	0.83
Solar radiation	W/m^2	175.0	-	175.0

5. PRINCIPAL COMPONENT ANALYSIS (PCA) BASED FAULT DETECTION AND IDENTIFICATION

In this chapter the results from the development of a new PCA-based elliptical threshold model for the FD&I of HVAC equipment are presented. First the method is presented along is development using measurements of a chiller operation from summer 2009. Further the developed method is validated using measurements from seven following summer seasons, from 2010 to 2016.

5.1. PCA based method for FD&I

In order to describe chiller operation, several inter-correlated parameters have to be considered at the same time. The ongoing commissioning of chillers is a multivariate analysis problem. PCA has been used in order to reduce the number of initial variables, and remove the inter-correlation among them. Through the data transformation capabilities of the PCA, the initial set of j-variables from the training data set considered representative of the chiller operation, extracted from the BAS, is transformed into a reduced set ($k < j$) of variables called Principal Components (PCs), which are linear combinations of the initial j-variables and PCA coefficients. In other words, the initial j-variables are projected into a k-dimensional PC-based space. The transformed measurements in the PC-based space are called scores. As consequence of data normalization and PCA features, the projection of measurements into the PC-based space tends to be centered on the origin of the axes, and are here assumed to have a Gaussian distribution along each principal direction, around the origin of the axes.

5.1.1. Summary of the method

The PCA method for FD&I consists of three main steps: i) threshold model training; ii) outliers detection; and iii) variables identification. The threshold model, developed from the training data set through the projection in the PC-based space, defines the confidence region for

Most of the content of this chapter was published as Cotrufo and Zmeureanu. 2016. PCA-based method of soft fault detection and identification for the ongoing commissioning of chillers. *Energy and Buildings, 130*, 443-453.

normal operation conditions. In this study, the threshold model takes the shape of an ellipsoid into the k-dimensional PC-based space, which is centered at the origin of the axes. The scores that fall outside the ellipsoid are detected as outliers, which could be caused by sensor error, components degradation or change in the system operation. For each PC-based detected outlier, the initial variable which is responsible of the abnormal score is selected as the variable with the highest distance of the detected outlier from the axes origin along the corresponding variable's axis. The definition of variable's axis in the PCs-based space will be given in Section 5.1.3.4.

The proposed method uses an ellipsoidal threshold, a concept similar to Jackson (1991), to distinguish between normal and abnormal value of measurements in a building HVAC system. The projection into the PC-based space of observations of different years of operation (2010–2016) were normalized by the mean and standard deviation of observations calculated from the training dataset of the reference summer of 2009. The graphical representation of outliers by using the first two PCs in the PC-based space has a practical advantage for the building operation team, and should be implemented in current Building Automation Systems.

Most studies used the Q-statistic (or SPE) for the fault detection, and the Q contribution plot and sensor sensitivity index (SVI) to identify the sensor responsible of the abnormal measure. The identification procedure presented in this study selects, for each outlier detected in the PCs-based space, the variable corresponding to the highest distance of the outlier from the axes origin along the corresponding variable's axis. Some of previous studies used synthetic data without noise from simulation programs to test the proposed methods. This study uses real measurements that contain noise and errors or changes in operation, which are inherent to measurements from the Building Automation System of a university campus.

5.1.2. Operation data sets

Results from this chapter refers to the case when only chiller #2 (CH-2) is working. A dedicate PCA-based threshold model should be developed for each operation mode. According to Wang and Xiao (2004a), a single PCA-based model, which accounts for all the system variables, would not reach desirable levels of performance in HVAC FDD application. The application of the proposed PCA-based method to the operation of chiller CH-2 includes measurements from a dedicate list of variables inherent to the considered operation mode and available from the BAS (Table 5.1). The quality of available data was verified through inspection and profiles visualization. Although the considered building is quite new (the operation started in 2008), some

abnormal values were observed. The system operation, thus, is not really fully fault free, but representative of a real system operation, in which, after initial commissioning, some soft faults might occur and equipment performance might decrease over the time (Roth et al. 2008).

Table 5.1 – List of measured variables from the BAS trend data recorded every 15 min.

	Variables	Units	Symbols
1.	Outdoor air temperature	°C	T_{oa}
2.	Electrical power input to CH-2	kW	E_{CH}
3.	Supply chilled water temperature, CH-2	°C	T_{CHWS}
4.	Return chilled water temperature, CH-2	°C	T_{CHWR}
5.	Supply condenser water temperature, CH-2	°C	T_{CNDS}
6.	Cooling tower fan VFD signal, CT-2	%	VFD_{CT}
7.	Supply cooling tower water temperature	°C	$T_{CT\text{-}S}$

Available measurements of seven variables (Table 5.1) were split in two clusters: the *training* and *application* data sets. The *training* data set, which includes 672 observations (corresponding to one week measurements at 15 min time step) collected between May 25 and June19, 2009, is used as a reference set for the normal operation or "almost fault free" operation. The application data set contains 804 observations from June 29 to August 30, 2009. In addition to the original application data set from the BAS, a modified application data set was generated, in which the measurements of supply chilled water temperature from chiller CH-2 (T_{CHWS}), and the supply cooling tower water temperature from cooling tower CT-2 ($T_{CT\text{-}S}$) were replaced by a random Gaussian distribution that was generated using the corresponding mean and standard deviation values from the training data set of normal operation. For T_{CHWS} the mean value is equal to the set-point temperature of 6.7◦C, and the standard deviation is 0.07◦C. For $T_{CT\text{-}S}$, the mean value is 28.9◦C and the standard deviation is 0.31◦C. This modified data set is used for validation purposes.

5.1.3. Threshold model training

5.1.3.1. Transformation of original data

The original dataset of j-variables was transformed into a new data set made of a reduced k number (k < j) of variables named Principal Components (PCs). The PCA transformation and

data reduction is applied to measurements from the training period. The j-variables included in Table 5.1 have different units of measure and range of variation. According to Reddy (2011), in order to avoid that any variable overloads the projection of the observations into the PCs-based space, hiding the effect of the others, some sort of data normalization is needed. Given a training data set $X_{tr}(i;j)$, where i are the observations, and j are the variables, the normalization is defined as follows (eq. 5.1):

$$zX_{tr,j} = \frac{X_{tr,j} - \mu_{tr,j}}{\sigma_{tr,j}} \qquad \text{eq. 5.1}$$

where $zX_{tr,j}$ is the j-column of the normalized training data set; $X_{tr,j}$ is the j-column of the original training data set X_{tr}; $\mu_{tr,j}$ is the mean value of the j-column of the original training data set; $\sigma_{tr,j}$ is the standard deviation of the j-column of the original training data set.

The PCA transformation is applied to the normalized training data set $zX_{tr,j}$, and results consist of a j × j matrix Q_{tr} (eq. 5.2).The first column of Q_{tr} corresponds to the first principal component (PC#1), with the coefficients $q_{:,1}$ used to project the i-normalized observations along the first principal direction. The second column contains the coefficients of the second principal component (PC#2) and so on. The first raw corresponds to the first variable (T_{oa}) as listed in Table 5.1, the second raw corresponds to the second variable (E_{CH}), and so on. The order of variables always follows the list of Table 5.1.

$$Q_{tr} = \begin{pmatrix} +0.421 & +0.011 & +0.068 & +0.874 & +0.227 & +0.047 & -0.001 \\ +0.463 & +0.020 & +0.052 & -0.122 & -0.313 & -0.336 & +0.746 \\ -0.064 & +0.675 & +0.727 & -0.047 & +0.029 & +0.091 & +0.016 \\ +0.459 & +0.043 & +0.076 & -0.107 & -0.397 & -0.416 & -0.663 \\ +0.450 & -0.108 & -0.034 & -0.201 & -0.237 & +0.829 & -0.037 \\ +0.435 & +0.050 & -0.052 & -0.407 & +0.792 & -0.102 & -0.047 \\ +0.050 & +0.727 & -0.673 & +0.038 & -0.088 & +0.080 & +0.001 \end{pmatrix} \qquad \text{eq. 5.2}$$

The projection of measurements into the PC-based space consists of a linear combination of i-normalized observations with Q_{tr} matrix columns, resulting in a new matrix F_{tr}, whose elements are named scores (eq. 5.3). The scores are the new coordinates of observations into the PC-based space. As an example, the score f_{i1} of the i-observation along the first principal direction is calculated as follows (eqs. 5.4 and 5.5):

$$F_{tr} = zX_{tr} \cdot Q_{tr} \qquad \text{eq. 5.3}$$

$$f_{i1} = zx_{i,1} \cdot q_{1,1} + zx_{i,2} \cdot q_{2,1} + \ldots + zx_{i,j-1} \cdot q_{j-1,1} + zx_{i,j} \cdot q_{j,1} \qquad \text{eq. 5.4}$$

$$f_{i1} = zx_{i,1} \cdot 0.421 + zx_{i,2} \cdot 0.463 + \ldots - zx_{i,j-1} \cdot 0.435 + zx_{i,j} \cdot 0.050 \qquad \text{eq. 5.5}$$

where F_{tr} is the matrix of scores; zX_{tr} is the normalized training data set X_{tr}; Q_{tr} is the matrix of coefficients; f_{i1} is the score of the i-observation along the first principal direction; q_{jj} is the coefficient from Q_{tr} matrix corresponding to the j-variable and j-principal direction; and zx_{ij} is the normalized value of the j-variable at the i-observation.

The selection of PCs is a major issue in PCA variable reduction. Most of the variance of the initial data set is retained by first few PCs. According to Morrison and Donald (1976), PCs should be selected in such a way to explain the minimum cumulative variance of 75% in the initial data set. For Jolliffe (1986), it should be at least 70–80%, while for Ladd and Driscoll (1980) it should be equal to 80%. In this research work the minimum cumulative variance to be retained was 90% of the total variance explained by the whole initial training dataset. For the considered training data set, this condition corresponds to the use of the first three PCs, k = 3 (Figure 5.1).

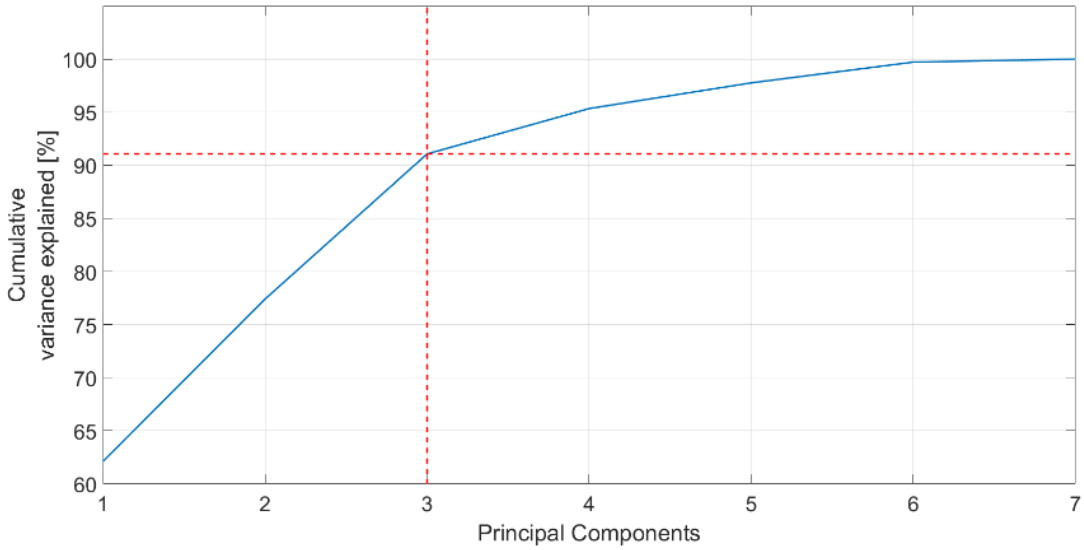


Figure 5.1– Cumulative variance explained by the principal components.

Hence the reduced Q_{tr} matrix contains only the first three columns of Q_{tr} from eq. 5.2. The remaining PCs were neglected because they explain only a very small variance from the original data set, which corresponds to random noise and some useless information from data set (Hu et

al., 2012).Scores evaluated through Eqs. 5.3–5.5, are distributed along the principal directions, around the origin of the axes. The plot of the scores into three two-dimensional PC-based spaces defined by PC#1, PC#2and PC#3 helps to visualize the different distributions (Figures 5.2-5.4).

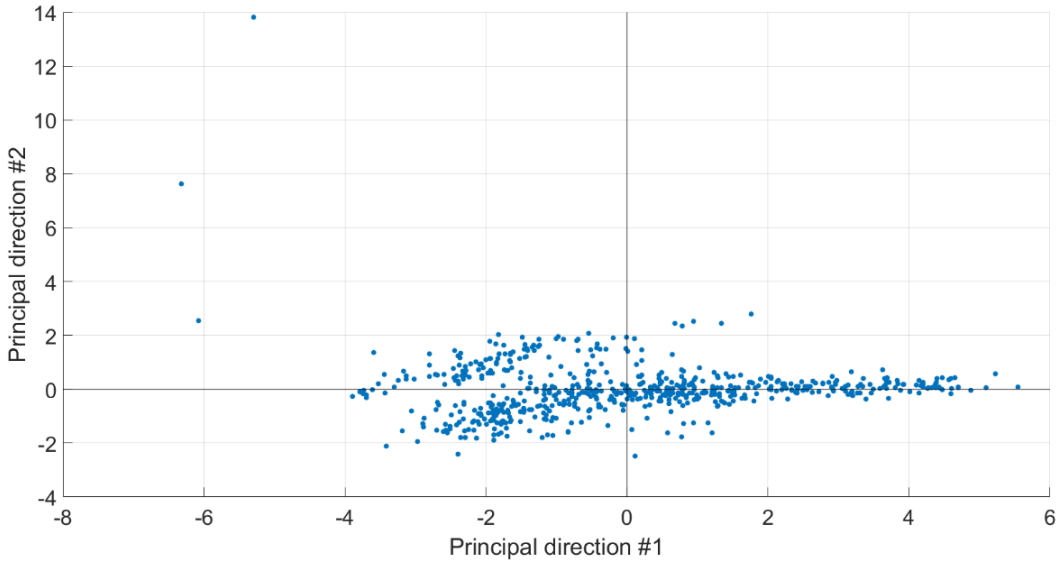


Figure 5.2 – Scores distribution into a PCs-based space defined by PC#1 and PC#2.

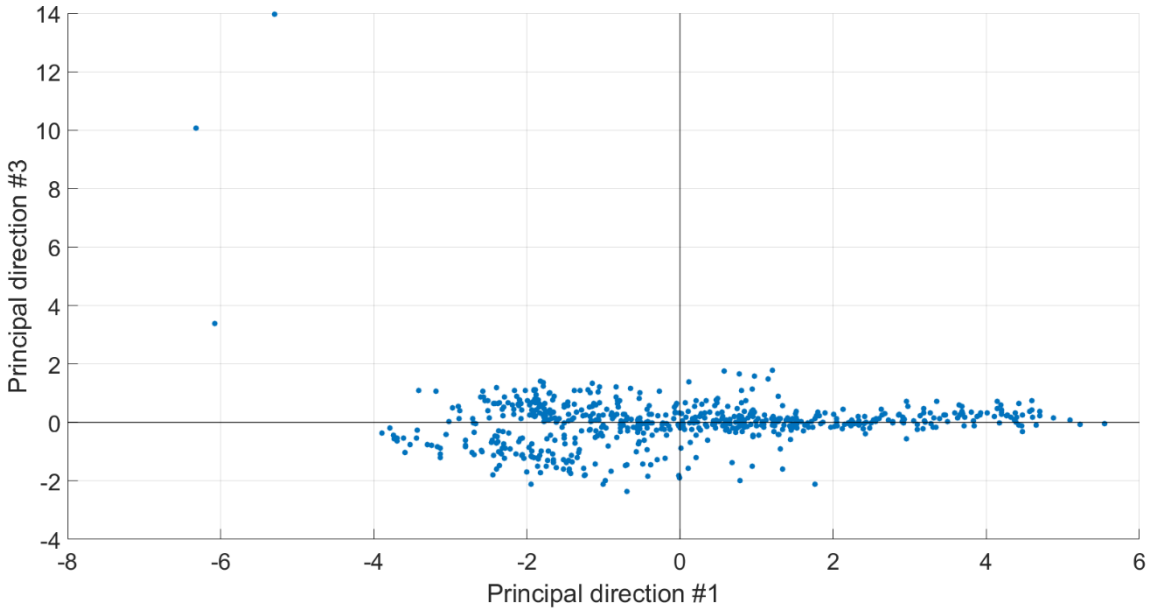


Figure 5.3 – Scores distribution into a PCs-based space defined by PC#1 and PC#3.

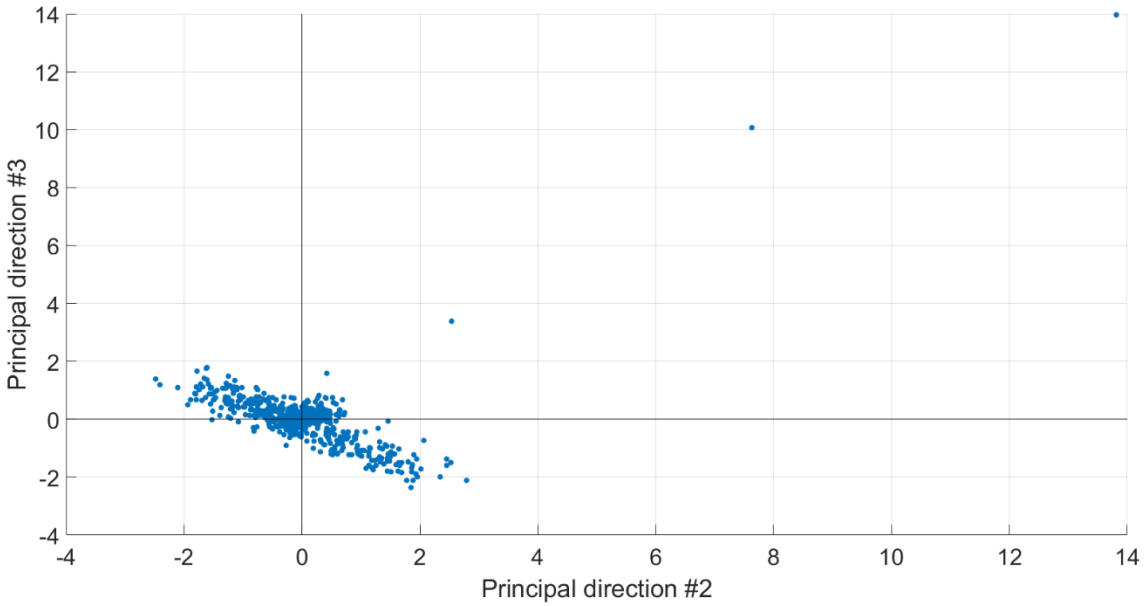


Figure 5.4 – Scores distribution into a PCs-based space defined by PC#2 and PC#3.

5.1.3.2. Threshold model for the training data set

The scores distribution along the principal direction is assumed to be a Gaussian distribution, with a different standard deviation for each direction. The threshold model draws the border of a region, within the k-dimensional PC-based space (in this case k = 3), within which scores correspond to normal operation conditions. Observations projected outside that border are labeled as abnormal events. The threshold model is formulated in terms of standard deviation of scores distribution along each principal direction. The scores and principal directions are calculated applying the PCA transformation to the training data set. An ellipsoidal formulation was selected for the threshold model having each k-semi-axis as function of the standard deviation of the scores distribution along the k-principal direction (eq. 5.6). A circular shape threshold was not used since it does not consider the difference between the standard deviation of scores along different principal directions. A parallelepiped or triangular shape thresholds were not considered because they do not account for the combined probability distribution due to the overlapping of several Gaussian distributions, with different standard deviations, along the k perpendicular axes (Figure 5.5). The *i*-observations projected into the PC-based corresponds to normal operation if eq. 5.7 is satisfied.

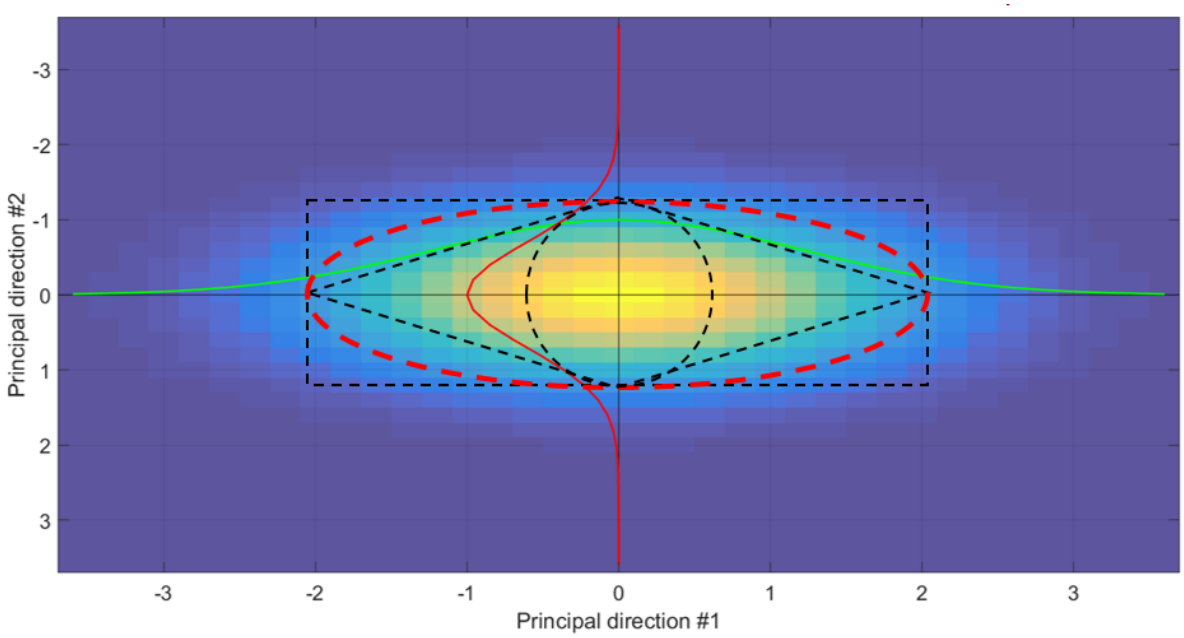


Figure 5.5 – Candidate threshold shapes (dash lines) against the color map of the combined probability distribution from the overlapping of two perpendicular Gaussian curves.

$$\sum_{j=1}^{k} \frac{f_{ij}^2}{(s\sigma_j)^2} = 1 \qquad \text{eq. 5.6}$$

$$\sum_{j=1}^{k} \frac{f_{ij}^2}{(s\sigma_j)^2} < 1 \qquad \text{eq. 5.7}$$

Where: $s\sigma_j = 1.96 \cdot \sigma_{tr,j}$, $s\sigma_j$ is the ellipsoid semi-axis along the j-principal direction; $\sigma_{tr,j}$ is the standard deviation of the scores distribution along the j-principal direction; and f_{ij} is the score of the *i*-observation along the *j*-principal direction of the training data set.

5.1.3.3. Detection of outliers in the PCs-based space

The detection of outliers consists of applying the threshold condition, given by eq. 5.7 and developed using the training data set, to the observations from the application data set, normalized and projected into the k-dimensional PCs-based space (k = 3). First the measurements from the application data set were normalized by using the mean values $\mu_{tr,j}$ and standard deviations $\sigma_{tr,j}$,

from the training data set (eq. 5.8) (Abdi and Williams 2010). For clarification, the subscript *tr* refers to the training data set, while *ap* refers to the application data set. Second, the normalized application data were projected into the PC-based space resulting in a new matrix F_{ap} of scores, by using the Q_{tr} matrix already calculated from the training data set (eq. 5.9).

$$zX_{j,ap} = \frac{X_{ap,j} - \mu_{tr,j}}{\sigma_{tr,j}} \qquad \text{eq. 5.8}$$

$$F_{ap} = zX_{ap} \cdot Q_{tr} \qquad \text{eq. 5.9}$$

where $zX_{ap,j}$ is the j-column of the normalized application data set; $X_{ap,j}$ is the j-column of the initial application data set.

The normalized measurements from the application data set were projected into the k-dimensional (k = 3) PC-based space using the first 3 columns of F_{ap}. Points located outside the ellipsoidal border, which was identified by the trained threshold model (eq. 5.6), were detected and marked as outliers. As an example, Figure 5.6 shows the normalized measurements from the application data set that were projected into a two-dimensional PC-based space defined by PC#1and PC#2. The scores corresponding to normal (blue) observations are within the ellipsoidal threshold border, while the scores corresponding to abnormal (red) observations are located outside the border.

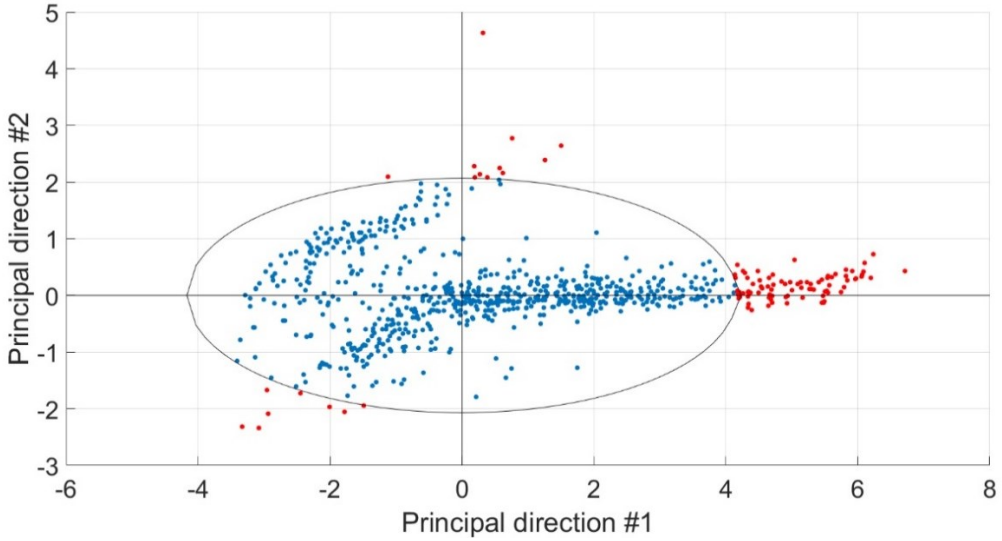


Figure 5.6 – Ellipsoidal threshold border separates the projections of normal observations (blue) from the projections of abnormal observations (red) from the normalized data set.

5.1.3.4. Identification of variables

This section presents the variable identification phase, which consists of identifying, for each detected outlier, which variables show abnormal values. The interpretation of the principal components, in order to extract information in terms of initial variables, is known in literature as the principal components interpretation.

There are different ways to interpret PCs and their relationship with initial variables. Several authors used the coefficients from Q matrix along with their linear combination with normalized data to identify strong correlations between a PC and specific variables (Hadley and Tomich 1986). According to Cadima and Jolliffe (1995) the coefficients are not appropriate to highlight this kind of correlation. In this section a new interpretation is proposed. As the outliers are scores, evaluated from linear combinations of variable's normalized measurements and PCA coefficients (eqs. 5.3-5.5), for each outlier one or several variables (Table 5.1) are expected to have an abnormal value and, thus, to be responsible for the score abnormal value and outlier detection.

The identification of variables (one or several) responsible for the outlier detection is based on the definition of axis of each *j* original variable in the k-dimensional PC-based space. We define a *j*-variable axis into the k-dimensional PC-based space as the straight line passing by the origin of the axes and by the point P, where the coordinates of point P are the first k values of the *j*-row of Q_{tr} matrix (eq. 5.2). For instance the axis of the third variable from Table 5.1 (T_{CHWS}), in a 2D PC-based space defined byPC#1 and PC#2, is the straight line passing through the origin of the axes and the point P($q_{3,1}$= −0.064; $q_{3,2}$= 0.675) (Figure 5.7).For each outlier in the PC-based space, the Euclidean distance between the outlier and the axis zero-value of each variable is calculated (Figure 5.8). The initial variable corresponding to the highest distance for that outlier is considered to be responsible for that outlier. Figure 5.8 shows, as an example in a 2D PC-based space defined by PC #1 and PC#2, the distance of a detected outlier (point S) from the zero-value of the axis corresponding to the supply chilled water temperature (T_{CHWS}). Since some variables have axes that are very close to each other (e.g., E_{CH} and T_{oa} in Figure 5.7), the zero-value of those axes are almost equal. Thus, for a given outlier, the Euclidean distance from the zero-value of those axes could be almost equal, and as a consequence errors in variables identification may occur. For this reason, the identification phase of this proposed method uses not only the first highest Euclidean distance, but also the second, and eventually the third highest distances.

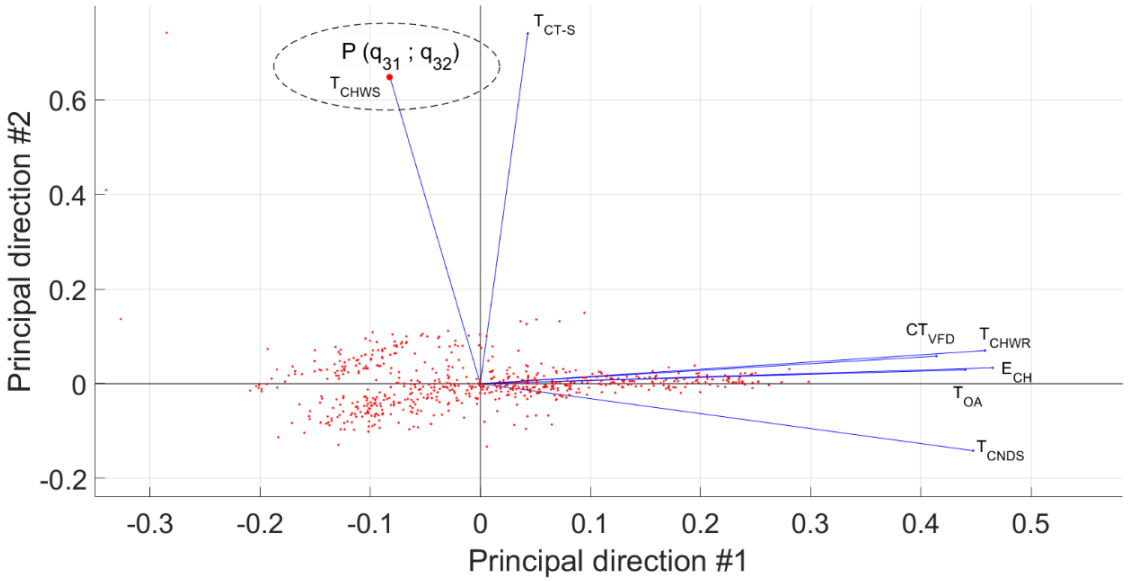


Figure 5.7 - Axes of the seven variables plotted into a two-dimensional PC-based space defined by PC#1 and PC#2 along with the scores scaled according to the maximum coefficient (Mathworks 2017).

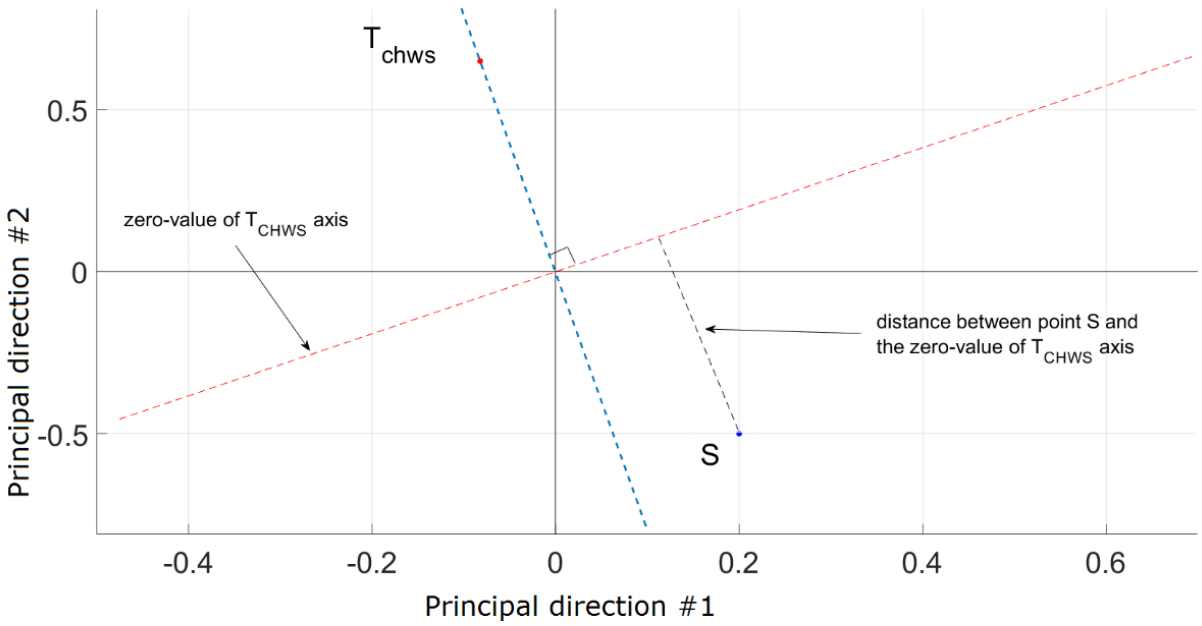


Figure 5.8 - Euclidean distance of point S from the zero-value of T_{CHWS} axis into a 2D PC-based space defined by PC#1 and PC#2.

5.1.4. Results from summer 2009

The threshold model, trained on a one week training period (672 observations from May 25 to June 19) was applied to an application data set (804 observations from June 29 to August 30, in 2009), to detect eventual outliers and test its effectiveness. The threshold model detected 139 outliers (Table 5.2).

Table 5.2 – Number and percentage of cases when each variable was responsible for outliers.

Variables	1st highest Euclidean distance		2nd highest Euclidean distance	
	no	[%]	no	[%]
T_{oa}	0	0.0	10	7.2
E_{CH}	28	20.1	46	33.1
T_{CHWS}	2	1.5	3	2.2
T_{CHWR}	4	2.9	37	26.6
T_{CNDS}	0	0.0	17	12.2
VFD_{CT}	52	37.4	26	18.7
T_{CT-S}	53	38.1	0	0.0
TOTAL	139	100	139	100

If the variables that influence the outliers are selected based on the highest Euclidean distance, $T_{CT\text{-}S}$ and VFD_{CT} together are the cause of 75.5% of outliers, followed by E_{CH} with 20.1%. If the second highest distance is used for the selection of variables, E_{CH} and T_{CHWR} together are the cause for about 59.7% of outliers. The proposed method detected possible errors in the application data set of summer 2009, related to the control or measurements of five variables from Table 5.1: E_{CH}, T_{CHWS}, T_{CHWR}, VFD_{CT}, and $T_{CT\text{-}S}$. Therefore the operation team should verify those five possible sources of errors.

In order to test the proposed method, the values of two variables in the application data set (T_{CHWS} and $T_{CT\text{-}S}$) were replaced with synthetic values representative of normal operation (modified application data set, Section 5.1.2). The model was applied to the modified application data set and, as expected, the two modified variables (T_{CHWS} and $T_{CT\text{-}S}$) do not appear anymore as the cause of outliers (Table 5.3). These results proved that PCA based method can effectively detect abnormal operation conditions from chiller operation data, and properly identify the variables which are responsible for the detected outliers.

Table 5.3 – Number of cases when each variable was responsible for outliers when using the modified application data set.

Variables	1st highest Euclidean distance		2nd highest Euclidean distance	
	no	[%]	no	[%]
T_{oa}	0	0.0	5	5.8
E_{CH}	29	33.7	44	51.2
T_{CHWS}	**0**	**0.0**	**0**	**0.0**
T_{CHWR}	4	4.7	21	24.4
T_{CNDS}	0	0.0	3	3.5
VFD_{CT}	53	61.6	13	15.1
T_{CT-S}	**0**	**0.0**	**0**	**0.0**
TOTAL	86	100	86	100

If the first three PCs are used, which explain 91.5% of the total variance in the initial training data set, the threshold model detects 139 outliers in the application data set. The detection has a lower performance when it uses only two PCs; it detects 101 outliers with the first two PCs (PC#1 and PC#2) that explain 77.4% of variance, 86outliers with PC#1 and PC#3 that explains 75.8% of variance, and45 outliers with PC#2 and PC#3 that explains 28.9% of variance.

5.1.5. Detection of outliers from 2009 to 2015

The proposed PCA-based method was used to assess the chiller performance over seven summer seasons from 2009 to 2015, from June 1st to August 30th. Only observations when CH-2 works alone were considered (Table 5.4). The entire data set of observations from the summer season of 2009, the first season of operation, were used as the reference data set, and used to train the threshold model on the PC-based space.

Table 5.4 – Number of observations included in each data sets.

Year	2009	2010	2011	2012	2013	2014	2015
Number of observations	2206	3236	3289	4241	2577	4131	4919

Figure 5.9 shows that the threshold model detected outliers (in red) in the 2009 training dataset; those are the outliers in the two-dimensional (2D) PCs-based space defined by PC#1 and PC#2, and not the outliers in the original measurement data set. The 2D threshold models contains about 75% of the total variance explained by the whole initial training data set (Figure 5.1), while

the threshold model composed of the first three PCs (3D model), PC#1, PC#2 and PC#3 (k = 3), contains 92%of the total variance. However, the 2D threshold models have the practical advantage of two-dimensional graphical representation. About 18% of the observation projections of 2009 on the PC-based space exceeded the threshold limit, e.g., they are outliers (Figure 5.11). The observations from the following summer seasons (2010–2015) were projected into the PCs-based space through eqs. 5.8 and 5.9, by using the j-mean values $\mu_{tr,j}$ and standard deviations $\sigma_{tr,j}$ calculated from the training data set of the entire summer 2009. For instance, Figure 5.10 shows that the same threshold model (developed along the 2009 data set) detected outliers (in red) in the application data set of summer 2015.

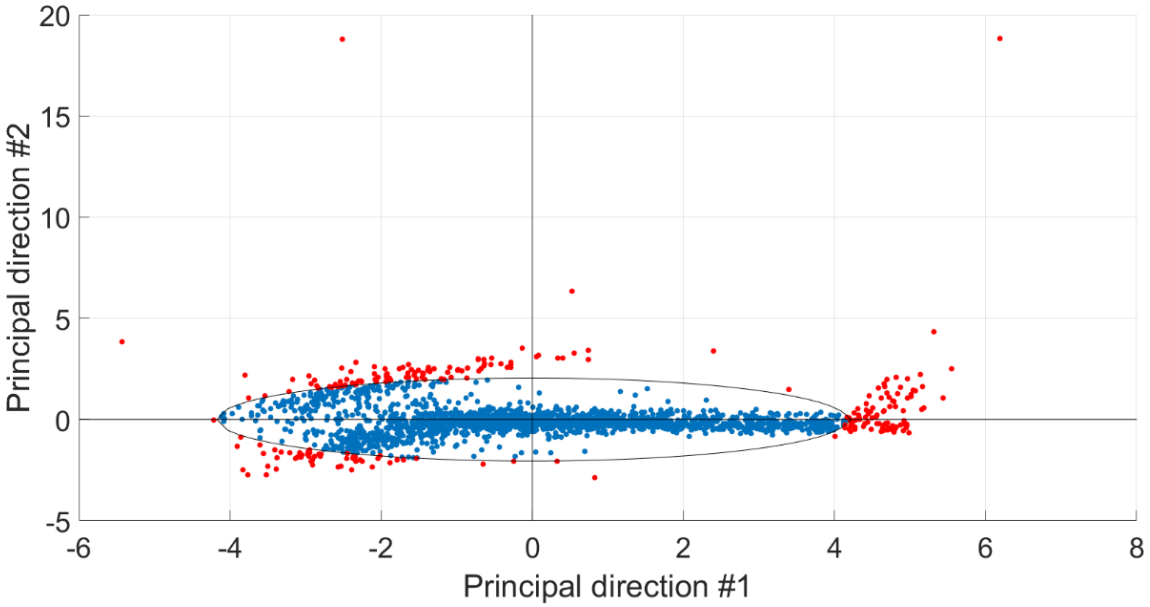


Figure 5.9 - Ellipsoidal threshold boundary separates the projections of normal observations (blue) from the projections of abnormal observations (red) along the 2009 data set.

From 2011 to 2015, the percentage of detected outliers over the entire seasonal data sets was lower than 10% (Figure 5.11), compared with 18% for the summer of 2009. Three combinations of PCs were used: (1) the first three PCs (3D model), PC#1, PC#2 andPC#3, (2) the first two PCs, PC#1 and #2, and (3) two PCs, PC#1 andPC#3. Since the 3D model contains 92% of the total variance explained by the whole initial training data set (Figure 5.1), one can expect that PC-based model is a good representation of the original measurement data set. The other 2D

models can also give close estimates of the number of outliers in the PC-based space, of about 11% for 2D model #2, and about 10% for 2D model #3.

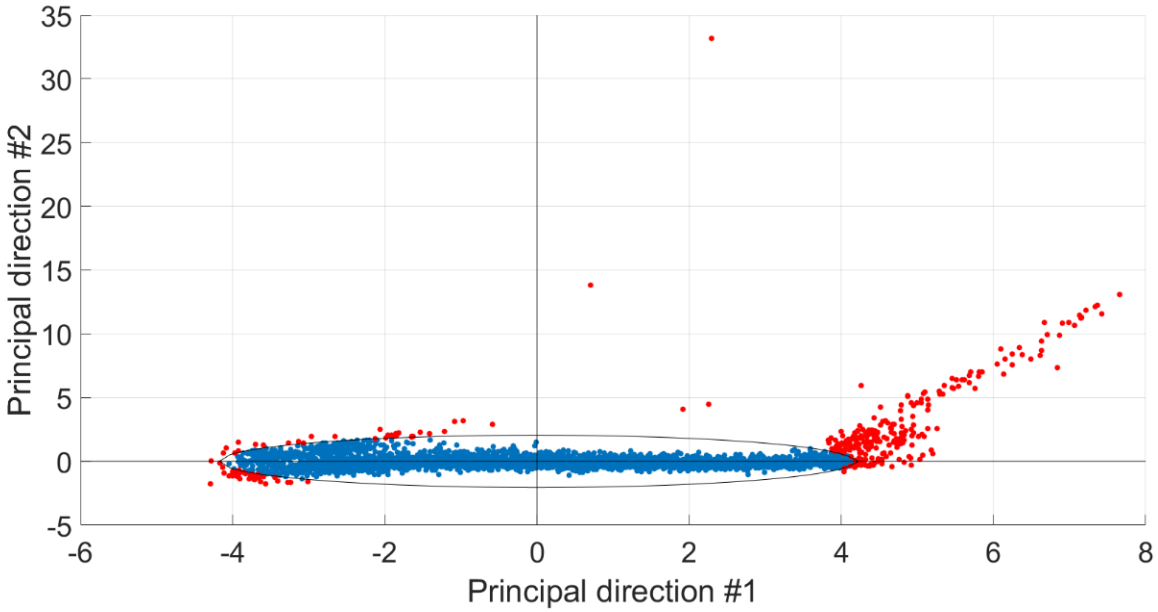


Figure 5.10 – Ellipsoidal threshold border separates the projections of normal observations (blue) from the projections of abnormal observations (red) along the 2015 data set.

As presented in Section 5.1.4, the proposed method detected possible errors related to the control or measurements, in the summer 2009, of five measurements: E_{CH}, T_{CHWS}, T_{CHWR}, VFD_{CT}, and $T_{CT\text{-}S}$. A higher percentage of faults along the 2009 data set could be explained by considering that in 2009 the system was in the first year of operation, and partially under commissioning. Corrections and adjustments in the system operation are responsible for variation of variable's range of variation and changing of correlations between the variables within the same training data set. The reduction of number of detected outliers during the following years (2010–2015) can be the result of the corrections undertaken during the first year of operation. However, there are still a few detected outliers due to sensors and component degradation, or changes in the operation conditions that normally occur in HVAC systems after the initial commissioning. The cooling tower fan VFD signal (VFD_{CT}) was responsible for 25–39% of outliers (Table 5.5), as noticed also by the operation team.

Table 5.5 – Variables identification for detected outliers from 2010 to 2015.

Variables	2010		2011		2012		2013		2014		2015	
	no	[%]	no	[%]	no	[%]	no	[%]	no	[%]	no	[%]
T_{oa}	1	0.3	3	1.6	18	4.7	8	4.1	28	7.0	21	4.6
E_{CH}	1	0.3	0	0.0	9	2.3	2	1.0	31	7.7	3	0.6
T_{CHWS}	42	12.1	30	15.8	46	11.9	46	23.2	66	16.0	120	26.1
T_{CHWR}	11	3.2	2	1.0	22	5.7	1	0.5	14	3.5	25	5.4
T_{CNDS}	44	12.7	24	12.6	51	13.2	18	9.1	23	5.8	47	10.2
CT_{VFD}	105	30.4	37	19.5	141	36.4	77	38.9	100	25.0	164	35.7
T_{CT-S}	142	41.0	94	49.5	100	25.8	46	23.2	138	35.0	80	17.4

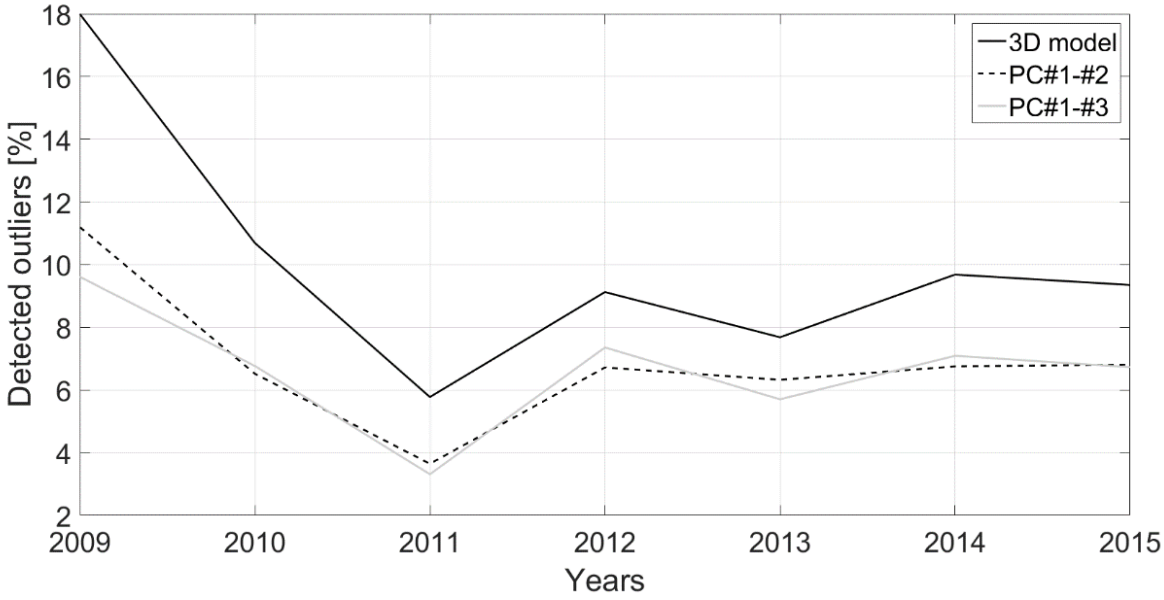


Figure 5.11 – Total percentage of detected outliers over the entire application data set size.

5.1.6. Detection of outliers in 2016

The same threshold model trained in the previous section, with the observations from the entire summer 2009, has been used to detect outliers among measurements from summer 2016. The data set from summer 2016 includes 1,496 observations for which chiller CH-2 was the only one to work. Figure 5.12 shows, for the three considered ellipsoidal threshold models based on PCA, the percentage of detected outliers over the entire sample of observation under the considered operation mode, from 2009 (the training data set) to 2016. A pick of detections is revealed by all

the three models in 2016, with a percentage of detection close to 100% for the 3D model. Figure 5.13 represents, in a 2D PCs-based space defined by PC#1 and #2, the projections of the observation from summer 2016 along with the threshold model trained with measurements from summer 2009. The cloud of projections lies just outside the ellipsoidal threshold, suggesting the occurrence of a change in the system operation level.

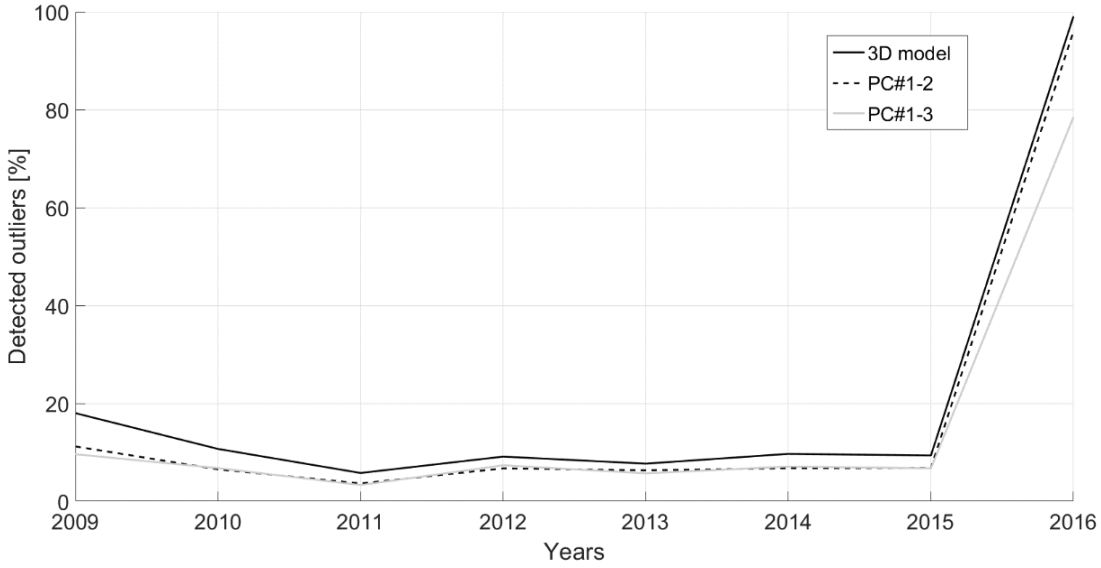


Figure 5.12 - Total percentage of detected outliers over the entire application data set size.

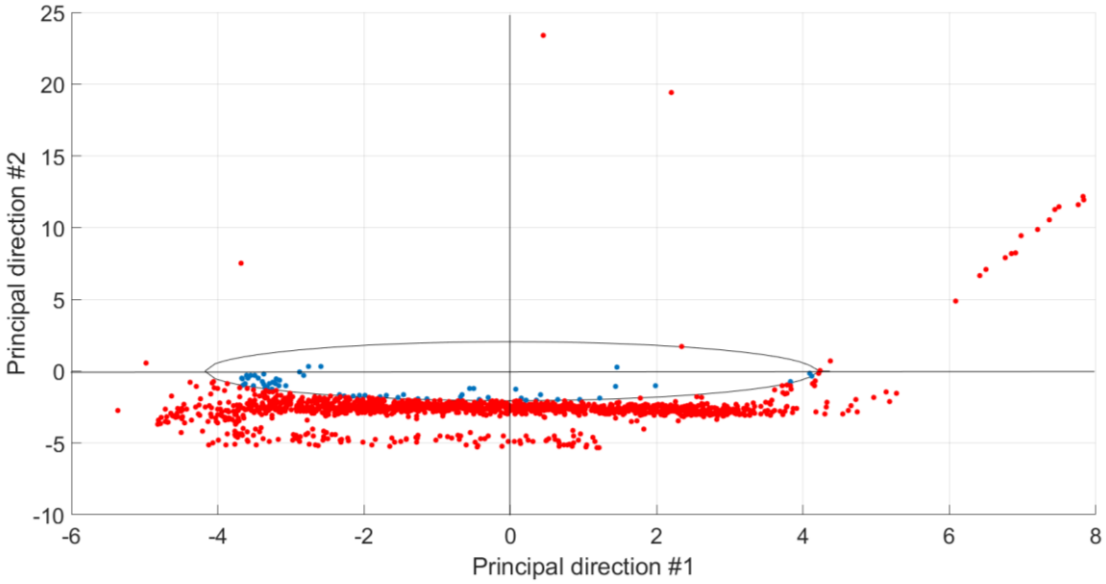


Figure 5.13 – Ellipsoidal threshold border separates the projections of normal observations (blue) from the projections of abnormal observations (red) along the 2016 data set.

In order to investigate the origin of the pick of detections in 2016 (Figure 5.12), the outcomes from the identification phase have been inquired: the first, second and third highest distances from the original variables zero-value axis have been considered for each outlier (Table 5.6). From considering the first highest distance, the supply chilled water temperature (T_{CHWS}) resulted to be responsible for 60.3% of the detected outliers, while the cooling tower supply water temperature ($T_{CT\text{-}S}$) was identified as the first responsible for the outliers in 19.1% of the cases. Through the second highest distance, $T_{CT\text{-}S}$ and T_{CHWS} were accounted for 59.4 and 20.3% of the outliers respectively. Finally, the cooling tower VFD signal (VFD_{CT}) was found to be responsible for the third highest distance for 59.7% of the outliers.

Table 5.6 – Number and percentage of cases when each variable was responsible for outliers.

Variables	1st highest distance		2nd highest distance		3rd highest distance	
	no	[%]	no	[%]	no	[%]
T_{oa}	0	0.0	18	5.4	0	0.0
E_{CH}	0	0.0	0	0.0	52	15.5
T_{CHWS}	202	60.3	68	20.3	12	3.6
T_{CHWR}	1	0.3	46	13.7	4	1.2
T_{CNDS}	18	5.4	0	0.0	63	18.8
CT_{VFD}	50	14.9	4	1.2	200	59.7
T_{CT-S}	64	19.1	199	59.4	4	1.2
TOTAL	335	100.0	335	100.0	335	100.0

The first variable to be considered for further investigation, thus, is the supply chilled water temperature. From data visualization, several abnormal values of T_{CHWS} were highlighted, as well as a change in the set-point from 6.7°C to 6.0°C (Figure 5.14). Although the highlights from data visualization can be accounted for some outliers, they are not enough to justify the detection of almost 100% of the observations from 2016 (Figures 5.12 and 5.13, and Table 5.6). Thus, the investigation effort was addressed to the second variable identified through the highest distance criterion (table 5.6): the cooling tower supply water temperature (T_{CT-S}). The $T_{CT\text{-}S}$ set-point was found to have changed between summer 2015 and 2016. Along the previous period (from 2009 to 2015) the $T_{CT\text{-}S}$ set-point was 29.0°C, while in 2016 this value dropped to 27.0°C.

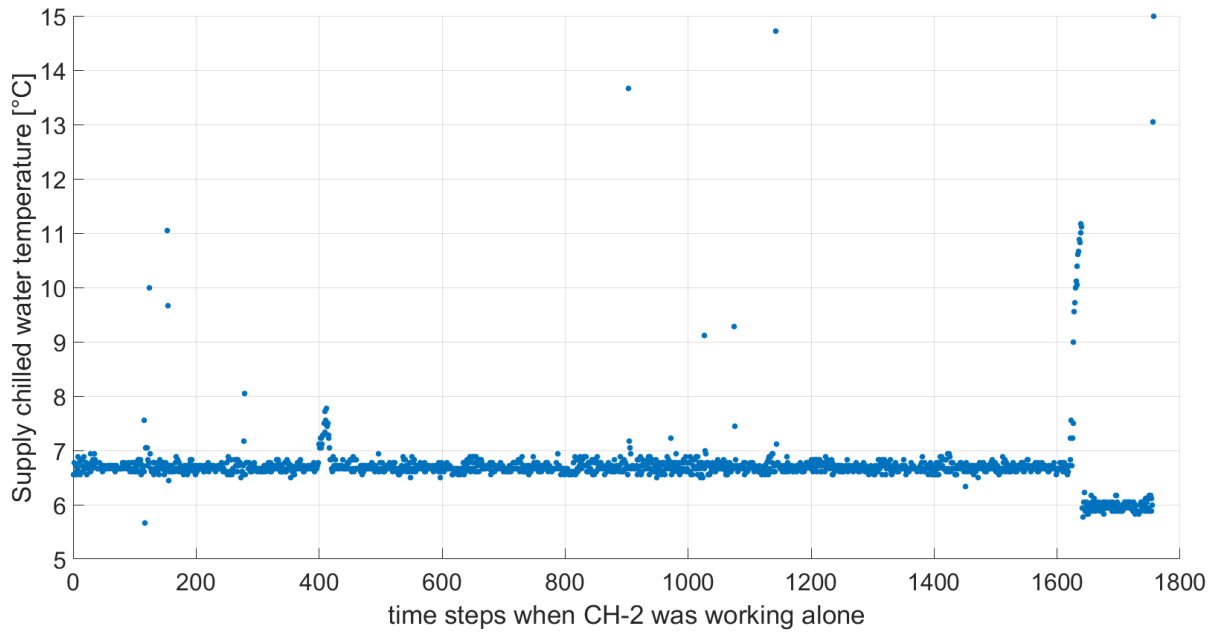


Figure 5.14 – Supply chilled water temperature measured at CH-2 when CH-2 was the only chiller to work, along summer 2016.

The faults detected along the summer season of 2016, thus, were not actual faults but changes of the set-points of the supply chilled water temperature (T_{CHWS}) and the cooling tower supply water temperature ($T_{CT\text{-}S}$). The proposed PCs-based threshold model effectively detected and identified any abnormal observations with respect to measurements from 2009 (the training data set), regardless if those abnormal values are due to faults or are inconsistent with the training period. As the system operation was modified (set-points changes), the model should be updated (re-trained using measurements from the summer season of 2016) before using it for further FD&I.

5.1.7. Discussion

A new PCA-based method for fault detection and identification in the chilled water system was presented in this paper, which used BAS trend data from an existing HVAC system in an institutional building. The proposed method proved to be effective for detecting abnormal measurements and for identifying the variables which are responsible for those outliers. The first case used one week of data (672 measurements at15 min time step) for training the threshold model, and then applied the model to an application data set (804 measurements between June 29 and August 30), both data sets being collected during the summer of 2009. The proposed method was effective in detecting outliers and identify faulty variables, by using data from a real system

operation. The threshold model detected 139 outliers of normal operation, and identified the variables that caused those abnormal operation outcomes. Synthetic data, representative of fault-free operation, have been used to replace values from two faulty variables in the application data set. Results from applying the PCA-method to this modified application data set validated the results from fault detection and identification from the original application data set.

In the second case, the measurements over the full summer of 2009 were used for training the threshold model, which was then applied to measurements of summer seasons of 2010–2016. The reduction of number of detected outliers between summer 2010 and 2015 can be due to the corrections in the first year of operation. There are still a few detected outliers due to sensors degradation which usually occurs in HVAC systems after initial commissioning. From summer 2016 almost 100% of the observations were detected as outliers (Figures 5.12 and 5.13). The proposed PCA-method identified two variables to be responsible for those outliers: the supply chilled water temperature and the cooling tower supply water. Those variables were found to undergo changes in the set-point values by the building operators. The PCA-method, thus, promptly detected the change in the chiller operation with respect to the training period (summer 2009), and effectively identified the variables characterized by new set-point values. If changes to the chiller operation are made, the PCA-method should be re-trained with measurements representative of the new operation profile.

The proposed method would give another tool to building operators to continuously verify and detect soft faults in sensors or degradation of equipment performance, which would help the scheduling for maintenance. Alarm signals would be sent to building operators when outliers are detected, and the identified responsible variable would be displayed. The method could be easily implemented in the control strategies using BAS trend data. Results from the proposed PCA-based method are easy to be interpreted, and advanced technical skills are not needed.

6. VIRTUAL FLOW METER FOR AHUs

Models for factor α have been developed for two AHU operation modes, to be used for the virtual measurement of the outdoor air flow rate (eq. 3.2):

i) **Operation mode #2: known mixing box inlet conditions**
The outdoor air temperature is taken as the temperature of the air at the mixing box inlet. The effect of the return fan on the return air temperature has been considered: a constant ΔT_{fan} = 1.8°C has been added to the return air temperature values from the BAS (Zibin 2014);

ii) **Operation mode #3: unknown mixing box inlet conditions**
The outdoor air stream is pre-heated before it reaches the mixing box, and the difference between measured outdoor air temperature and actual air temperature at the mixing box inlet cannot be ignored.

6.1. Factor α with known mixing box inlets conditions

In this section the three factor α first principle based models are first presented. The propagation of the uncertainty from the measurements to the predictions is evaluated for each model. In order to reduce the number of required inputs, three data driven models are developed, which predict the factor α using a reduced number of inputs. Once trained and validated, the data driven models can be used instead of the first principle based models. Finally the best first principle based and data driven models are selected based on models accuracy, uncertainty and required information.

6.1.1. First principle based models

Simplified energy balance (α_0)

The common formulation for the factor α based on air temperatures (eq. 3.5) is here reported (eq. 6.1). The factor α is given as function of the air temperatures at outdoor, return and mixed conditions.

$$\alpha_0 = \frac{T_{ma} - T_{ra}}{T_{oa} - T_{ra}} \qquad \text{eq. 6.1}$$

Part of the content of this chapter has been presented in a conference paper: Cotrufo et al. 2016. Virtual outdoor air flow meter for the ongoing commissioning of HVAC systems: lessons from a case study. *ASHRAE 2016 Winter Conference, Orlando, FL.*

Three balance equations system (α_s)

The factor α_s model is based on a system of three equations with three unknowns. The three equations are: i) the energy balance equation (eq. 6.2), ii) the water mass balance equation (eq. 6.3), and iii) the air mass balance equation (eq. 6.4). The three unknown variables from the considered case study are: the outdoor and recirculated air mass flow rates ($\dot{m}_{oa}$ and $\dot{m}_{rec}$ respectively), and the mixed air humidity ratio (x_{ma}).

$$\dot{m}_{sa} \cdot h_{ma} = \dot{m}_{oa} \cdot h_{oa} + \dot{m}_{rec} \cdot h_{ra} \qquad \text{eq. 6.2}$$

$$\dot{m}_{sa} \cdot x_{ma} = \dot{m}_{oa} \cdot x_{oa} + \dot{m}_{rec} \cdot x_{ra} \qquad \text{eq. 6.3}$$

$$\dot{m}_{sa} = \dot{m}_{oa} + \dot{m}_{rec} \qquad \text{eq. 6.4}$$

where $\dot{m}_{sa}$, $\dot{m}_{oa}$ and $\dot{m}_{rec}$ are the air mass flow rates at supply, outdoor and recirculated conditions respectively; h_{ma}, h_{oa} and h_{ra} are the air specific enthalpies at mixed, outdoor and return conditions respectively; and x_{ma} , x_{oa} and x_{ra} are the air humidity ratios at mixed, outdoor and return conditions respectively.

From combining the three above listed equations, the factor α is given by (eq. 6.5):

$$\alpha_s = \frac{C_{p,a} \cdot T_{ma} + x_{ra} \cdot (h_{fg} + C_{p,v} \cdot T_{ma}) - h_{ra}}{(h_{fg} + C_{p,v} \cdot T_{ma}) \cdot (x_{ra} - x_{oa}) + h_{oa} - h_{ra}} \qquad \text{eq. 6.5}$$

where: $C_{p,a}$ is the dry air specific heat at constant pressure; $C_{p,v}$ is the water vapor specific heat at constant pressure; h_{fg} is the water heat of vaporization.

Energy balance at the mixing box (α_e)

The factor α_e is given from the energy balance across the mixing box (eq. 3.4), combining eq. 3.1 and eq. 3.3. Eq. 3.4 is here reported (eq. 6.6):

$$\alpha_e = \frac{h_{ma} - h_{ra}}{h_{oa} - h_{ra}} \qquad \text{eq. 6.6}$$

The air humidity ratio at mixed condition (x_{ma}) is unknown, as an air humidity sensor after the mixing box is not installed (see Figure 4.6 from Case Study). Thus the mixed air humidity ratio is assumed to be equal to the supply air humidity ratio (x_{sa}), which can be derived from the available

measurements of the supply air temperature and relative humidity. This assumption ($x_{ma} = x_{sa}$) is valid under the following two conditions:

i. No humidification, as indicated by the humidifier valve modulation equal to 0%;
ii. No condensation occurs between mixing and supply conditions.

The second condition is imposed by considering the saturation degree of air at the cooling coil (ASHRAE 2009). Measurements for which the saturation degree is close to 100% are excluded from the dataset as well as measurements taken when the air was humidified. The outdoor and return air enthalpies are evaluated by using the measured air temperature and relative humidity.

6.1.2. Uncertainty analysis

The uncertainty propagated from the initial measurements, through the above presented models to the predictions, is evaluated according to section 3.5 (Uncertainty analysis). The measurements mean value (μ) and random error (R) values have been derived from the *training period* (from April 7 to May 5, in 2015), and the bias (fixed) error (B) as stated by the sensors manufacturer (Tables 6.1 and 6.2).

Table 6.1 – Measurements mean value and random error as observed during the training period (from April 7 to May 5, in 2015), and bias error as stated by the sensors manufacturer.

	T_{ma} [°C]	T_{sa} [°C]	T_{oa} [°C]	T_{ra} [°C]	RH_{sa} [%]	RH_{oa} [%]	RH_{ra} [%]
μ	11.5	15.0	8.6	21.5	60.3	38.3	32.2
R	0.08	0.08	0.10	0.08	8.3	1.4	0.5
B	0.45	0.45	0.45	0.45	4.0	4.0	4.0

Table 6.2 – Statistical indices and uncertainty of the factor α first principle based formulations.

Formulation	Mean [-]	Std [-]	N. of input [-]	Bias error [-]	Random error [-]	Uncertainty [-]
α_0	0.71	0.18	3	0.08	0.03	0.08
α_s	0.71	0.18	5	0.13	0.02	0.13
α_e	0.76	0.09	6	0.12	0.12	0.17

The three factor α models have different uncertainty values (Table 6.2), as the bias and random errors propagate through different mathematical formulations (eqs. 6.1, 6.5 and 6.6). Specifically, the more complex is the mathematical formulation and the greater is the number of input measurements, higher the propagated uncertainty is expected. For instance the first model gives $\alpha_0 = 0.71 \pm 0.08$, compared to the third model $\alpha_e = 0.76 \pm 0.17$.

6.1.3. Data driven models

The three first principle based models from eqs. 6.1, 6.5 and 6.6 (α_0, α_s and α_e) require for 3, 5 and 6 input measurements respectively. With the goal of reducing the number of those required measurements, three data driven models have been developed which predict the factor α, as predicted by the first principle based models, using a reduced number of predictors. From a preliminary correlation analysis (Figure 6.1), the best regressor among others has been found to be the difference between mixed and outdoor air temperatures. Quadratic regression models have been developed which use the term ($T_{ma} - T_{oa}$) to predict factors α_0 (model *a*) and α_s (model *b*). For the prediction of factor α_e, a quadratic regression model of the term ($T_{ma} - T_{oa}$) showed limited prediction capability. Thus, a multivariate regression model has been developed using four regressors (outdoor air relative humidity, RH_{oa}; mixed air temperature, T_{ma}; outdoor air enthalpy, h_{oa}; and return air enthalpy, h_{ra}), which resulted in the highest Pearson's coefficients of correlation to the factor α_e (model *c*).

The three models haven been trained along the *training period* (April 7 to May 5, 2015) and validated on the *validation period* (May 5-12, 2015) (Table 6.3). Figures 6.2 shows results from the factor α data driven models along the training period. Statistical indices from comparison between first principle based and data driven models are given in (Table 6.4). Figure 6.2.c presents data from only two consecutive days because, as required by the formulation of the factor α_e conditions, several measurements are excluded from the data set to satisfy two conditions: i) no saturation, and ii) no condensation occurring in the AHU between mixed and supply conditions.

Table 6.3 –Prediction models trained along the training period (April 7 to May 5, 2015).

Prediction models for α	Trained model's formulation
model *a*	$\alpha_0 = 1.004 - 0.071 \cdot (T_{ma} - T_{oa}) - 0.002 \cdot (T_{ma} - T_{oa})^2$
model *b*	$\alpha_s = 1.002 - 0.079 \cdot (T_{ma} - T_{oa}) - 0.002 \cdot (T_{ma} - T_{oa})^2$
model *c*	$\alpha_e = 1.598 - 0.001 \cdot RH_{oa} - 0.055 \cdot T_{ma} + 0.0155 \cdot h_{oa} - 0.009 \cdot h_{ra}$

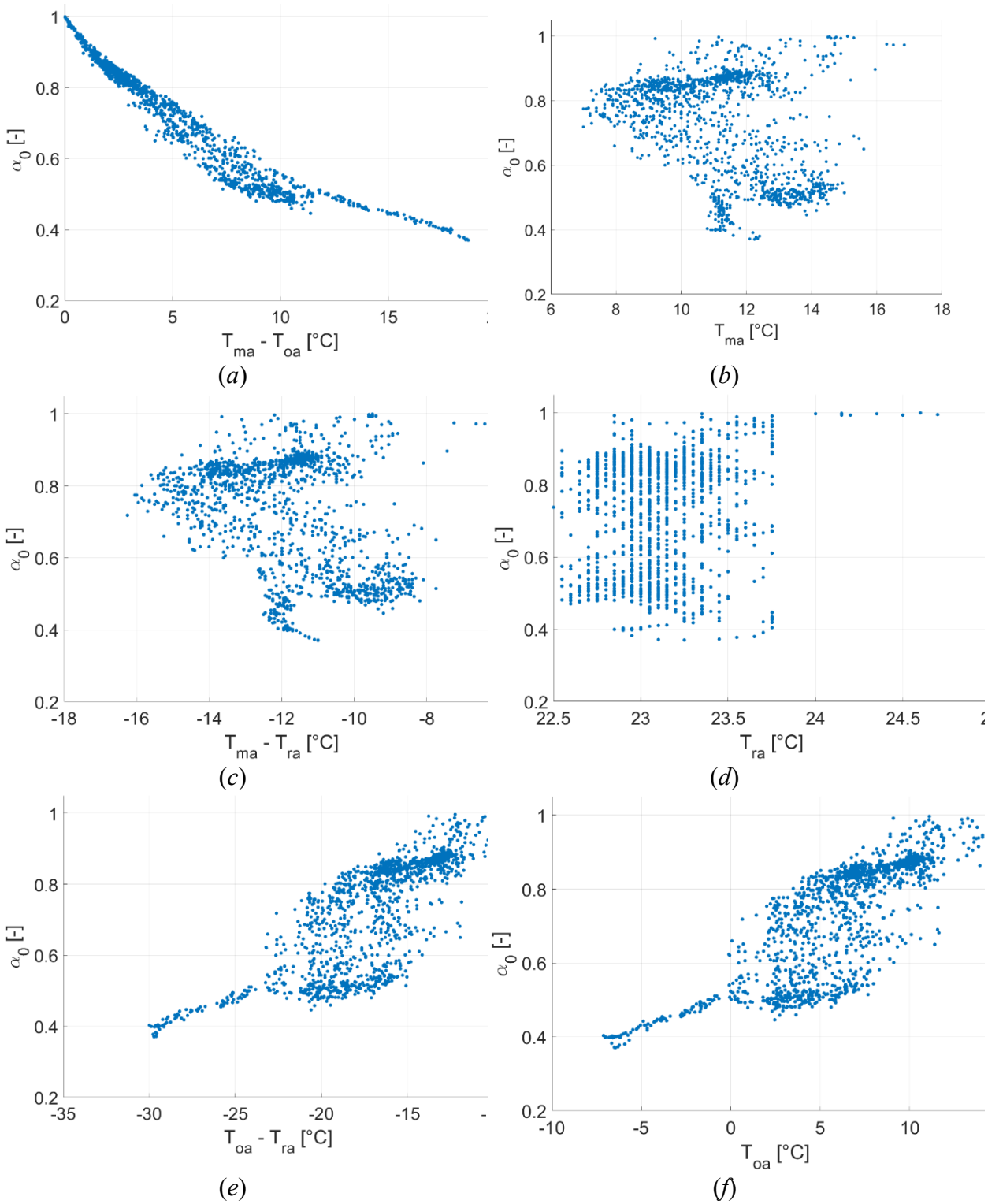


Figure 6.1 – Sensitivity analysis for the best predictor of the factor α, *training period* (April 7 to May 5, 2015).

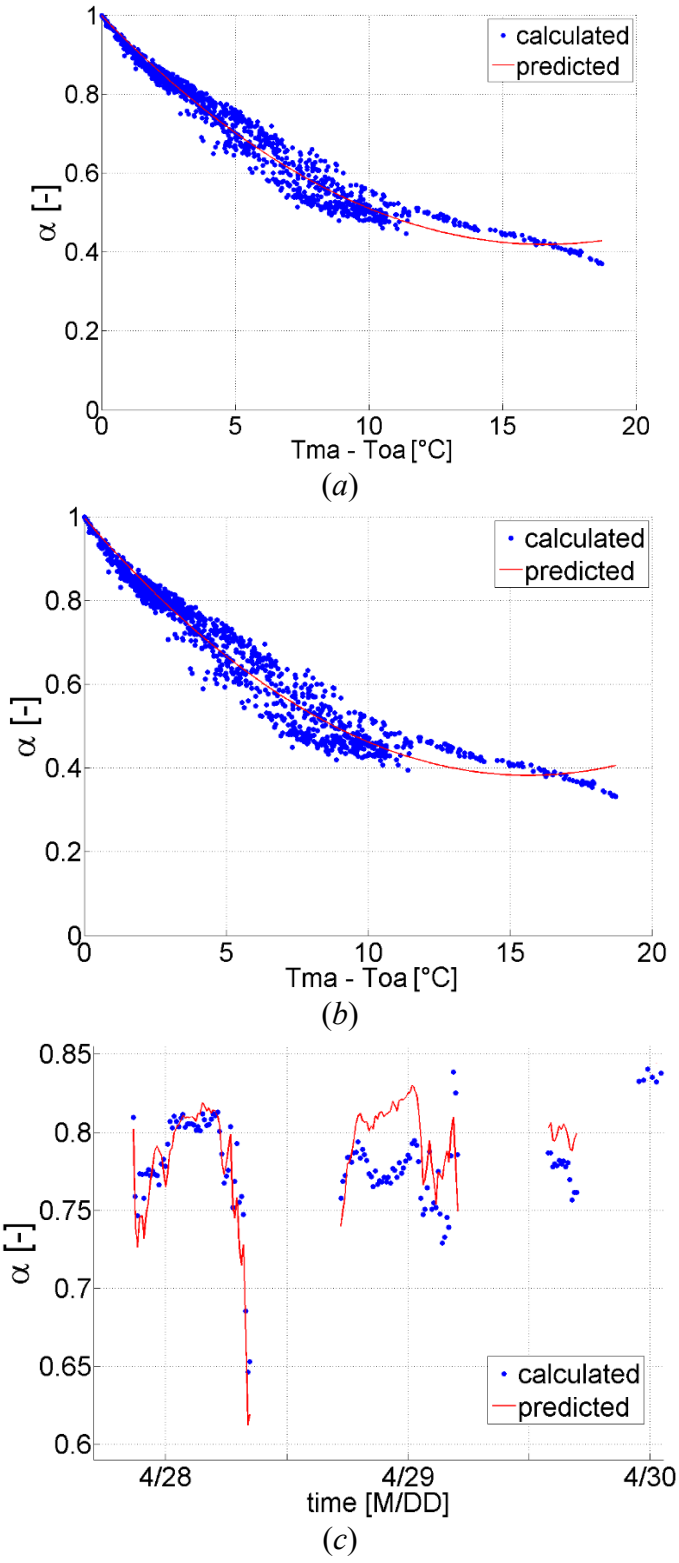


Figure 6.2 – Factor α predictions from the regression models (red line) and from the first principle based formulations (blue points) along the *training period*.

Table 6.4 – Statistical indices of goodness of fit for the regression models.

Factor α prediction model	Training		Validation	
	R^2 [-]	CV-RMSE [%]	R^2 [-]	CV-RMSE [%]
model *a* (α_0)	0.97	2.8	0.87	4.3
model *b* (α_s)	0.96	3.8	0.80	4.6
model *c* (α_e)	0.85	1.4	0.78	2.2

6.1.4. Validation through the mixed air temperature

Direct measurements of the outdoor air flow rate were not available from this case study. The reference values of the factor α needed for validation, thus, could not be derived from eq. 3.1. An alternative, indirect strategy has been used to validate the three presented factor α formulations (eqs. 6.1, 6.5 and 6.6) and the corresponding data driven models. The predictions of the factor α from each of the three data driven models are used, along with measurements of the outdoor and return air conditions, to derive the mixed air humidity ratio $x_{ma,\alpha}$ (eq. 6.7), the mixed air specific enthalpy $h_{ma,\alpha}$ (eq. 6.8), and the mixed air temperature $T_{ma,\alpha}$ (eq. 6.9). The derived mixed air temperature ($T_{ma,d}$) is thus compared to the direct measurements of the same variable (T_{ma}) (table 6.5 and figures 6.3).

$$x_{ma,\alpha} = \alpha \cdot (x_{oa} - x_{ra}) + x_{ra} \qquad \text{eq. 6.7}$$

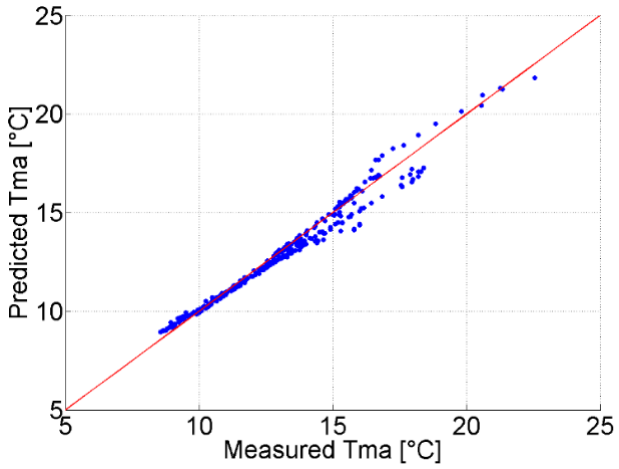
$$h_{ma,\alpha} = \alpha \cdot (h_{oa} - h_{ra}) + h_{ra} \qquad \text{eq. 6.8}$$

$$T_{ma,\alpha} = \frac{h_{ma,d} - x_{ma,d} \cdot h_{fg}}{c_{p,a} + x_{ma,d} \cdot c_{p,v}} \qquad \text{eq. 6.9}$$

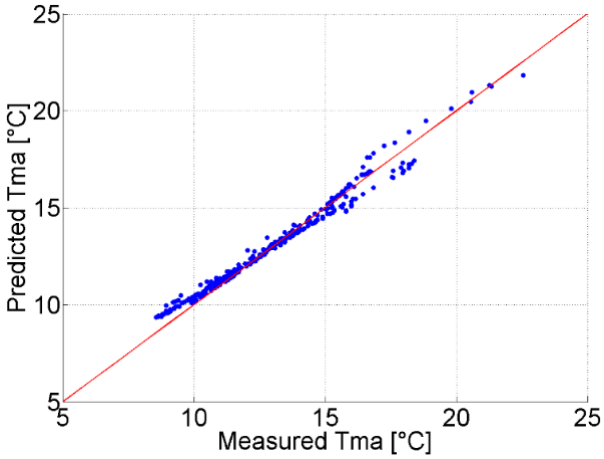
where $T_{ma,\alpha}$ is the derived mixed air temperature, $x_{ma,\alpha}$ is the derived mixed air humidity ration, and $h_{ma,d}$ is the derived mixed air enthalpy calculated from the derived $x_{ma,\alpha}$ and the measured T_{ma}.

Table 6.5 – Comparison between measured Tma and derived Tma,α mixed air temperatures.

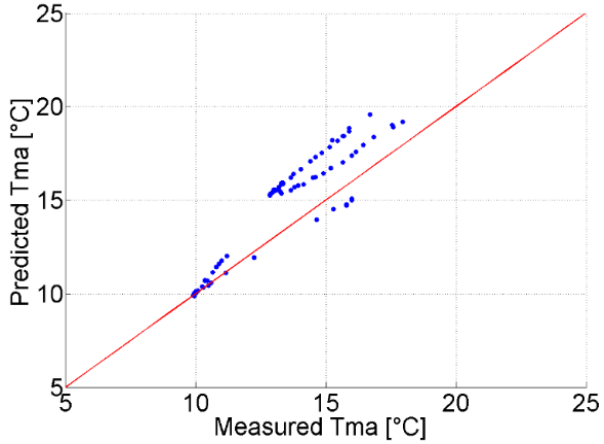
Factor α prediction model	R^2 [-]	MAE [°C]
model *a* (α_0)	0.97	0.27
model *b* (α_s)	0.97	0.26
model *c* (α_e)	0.67	1.56



(*a*)



(*b*)



(*c*)

Figure 6.3 – Regression models validation through the mixed air temperature.

6.1.5. Discussion

The three first principle based models of the factor α lead to different predictions uncertainty (Table 6.2). The uncertainty analysis allows to evaluate the models feasibility: the uncertainty from factor (eq. 6.1) α_0 is the smallest one ($U_0 = 0.08$ and α varies from 0 to 1), which makes the a_0 model the best candidate for the development of a virtual air flow meter. The uncertainty from models α_s and α_e (eqs. 6.5 and 6.6) is 0.13 and 0.17 respectively. Due to the error propagation, more complex formulations lead to higher uncertainty values. Thus, simple mathematical formulations, with a reduced number of required inputs, should be preferred. The uncertainty from α_s and α_e is more than 10% of the entire range of variation of the factor α. This circumstance may prevent from the effective implementation of those factor α formulations in an air VFM (eq. 3.2).

The regression models based on the regressor *Tma - Toa* allow for the physical interpretation of system operation (Figures 6.2.a and 6.2.b). Models *a* and *b* showed good prediction accuracy: $R^2 = 0.97$ and 0.96, respectively along the training period, $R^2 = 0.87 = 0.80$ respectively along the validation period. Model *c* shows the poorest accuracy, with $R^2 = 0.85$ and 0.78 for training and validation periods respectively (Table 6.4). The CV-RMSE values are acceptable for the first two models.

From models validation, models *a* and *b* provide good statistical indices (Table 6.5): $R^2 =$ 0.97 for both and MAE equal to 0.27°C and 0.26°C, respectively. However, model *b* is trained on the factor α_s from eq. 6.5, which requires more sensors (5) than the common formulation of α_0 from eq. 6.1. The three air temperature sensors needed for α_0 are commonly already installed in AHU for control. Furthermore, the factor α_0 formulation has the smallest uncertainty value (Table 6.2). It is important to highlight that the models validation is only based on T_{ma} because the outdoor air flow rate was not available from the BAS, and direct verification of α was not possible.

The temperature-based model of the factor α (a_0) should be preferred when compared to the other two models. Also, the data driven model *a* uses the term ($T_{ma} - T_{oa}$) as regressor, thus requires for a reduced number of sensors if compared to model *c*.

6.2. Heat recovery operation mode

The factor α formulation from eq. 3.6, which accounts for the actual outdoor air temperature at the mixing box inlets (T_{ac} and T_{rec}), is here used as a reference (eq. 6.10).

$$\alpha_{ref} = \frac{T_{ma} - T_{rec}}{T_{ac} - T_{rec}} \qquad \text{eq. 6.10}$$

The lack of measurements of T_{ac} is overcome coupling eq. 6.10 with the equation of the energy balance at the heat recovery coils (SC2-1 and SC2-2) (eq. 6.11). The left part from eq. 6.11 is the heat recovered by the recovery system at coil SC1-3, the right part is the heat gained by the outdoor air stream passing through coils (SC2-1 and SC2-2). The factor α, thus, is derived from a system of two equations (eq. 6.10 and eq. 6.12) and two unknown (the factor α and T_{ac}) (eq. 6.12). The third unknown variable, the recirculated air temperature T_{rec}, is here replaced by the return air temperature T_{ra}, which is available from the BAS.

$$\rho_{glc} \cdot V_{glc} \cdot C_{p,glc} \cdot (T_{hre} - T_{hra}) = \alpha \cdot (\rho_{air} \cdot V_{sa}) \cdot C_{p,air} \cdot (T_{ac} - T_{oa}) \qquad \text{eq. 6.11}$$

$$\alpha_a = \frac{Q_{hr} - \rho_{air} \cdot V_{sa} \cdot C_{p,air} \cdot (T_{ma} - T_{ra})}{\rho_{air} \cdot V_{sa} \cdot C_{p,air} \cdot (T_{ra} - T_{oa})} \qquad \text{eq. 6.12}$$

where $C_{p,air}$ is the dry air specific heat at constant pressure, $C_{p,air} = 1.006$ kJ/(kg K); ρ_{glc} is the glycol density, kg/m^3; V_{glc} is the glycol flow rate, L/s·10^{-3}; $C_{p,glc}$ is the glycol specific heat, $C_{p,glc} = 3.35$ kJ/(kg K); and T_{hre} and T_{hra} are the glycol temperatures respectively before and after the pre-heating coil, °C.

The air temperature difference between return and recirculated conditions depends on features which are unique of each AHU system (e.g. return fan temperature rise, ducts size and insulation level, etc.) and it may not be always negligible. A white box model of the thermodynamic processes of the air flow between return and recirculated conditions would require additional information, modelling skills and time. Design information may be difficult to find, and the overall approach is likely to result time consuming. Thus, in order to account for the air temperature variation between the return condition (T_{ra}) and the recirculation mixing box inlet (T_{rec}), a grey box modelling approach is preferred. Using short-term measurements of the actual T_{rec}, an inverse model is developed which predicts the recirculated air temperature ($T_{rec,p}$) from

measurements available from the BAS. A liner regression model has been successfully used (eq. 6.13).

$$T_{rec,p} = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 \quad \text{eq. 6.13}$$

where $T_{rec,p}$ is the predicted recirculated air temperature, °C; a_0 , a_1 and a_2 are the regression coefficients; and x_1 and x_2 are the measurements of the identified most correlated variables.
The predicted recirculated air temperature is then used to replace the return air temperature in eq. 6.12 (eq. 6.14).

$$\alpha_b = \frac{Q_{hr} - \rho_{air} \cdot V_{sa} \cdot C_{p,air} \cdot (T_{ma} - T_{rec,p})}{\rho_{air} \cdot V_{sa} \cdot C_{p,air} \cdot (T_{rec,p} - T_{oa})} \quad \text{eq. 6.14}$$

The factors α from eq. 6.12 (α_a) and eq. 6.14 (α_b) are compared to α reference values from eq. 6.10 (α_{ref}), where T_{ra} is replaced by T_{rec}, and both T_{ac} and T_{rec} have been measured with portable calibrated sensors.

6.2.1. The equivalent AHU single system

The proposed factor α formulations (eqs. 6.12 and 6.14) can be applied at each of the two sub-systems from the considered case study (figures 4.6 and 4.9) if the heat transfer rate ($Q_{hr,1}$ and $Q_{hr,2}$) at each of the two coils (SC1-2 and SC2-2) is known. This is not the case, as only the total heat transfer across coil SR1-3 ($Q_{hr} = Q_{hr,1} + Q_{hr,2}$) can me estimated. Thus, an *equivalent* single AHU system has been derived, using the measurements of the supply air volume flow rate at each sub-system. Taking the air density as constant, the parameters of the new equivalent system have been derived as weighted average values of the factor α (eq. 6.15), and the mixed air temperature (T_{ma}) (eq. 6.16).

$$\alpha_{ref} = \frac{\alpha_1 \cdot V_{sa}^{\#1} + \alpha_2 \cdot V_{sa}^{\#2}}{V_{sa}} \quad \text{eq. 6.15}$$

$$T_{ma} = \frac{T_{ma}^{\#1} \cdot V_{sa}^{\#1} + T_{ma}^{\#2} \cdot V_{sa}^{\#2}}{V_{sa}} \quad \text{eq. 6.16}$$

Where α_1 and α_2 are the factor α for systems #1 and #2 respectively, [-]; $V_{sa}^{\#1}$ and $V_{sa}^{\#2}$ are the air volume flow rates supplied by sub-systems #1 and #2 respectively, L/s; and $T_{ma}^{\#1}$ and $T_{ma}^{\#2}$ are the mixed air temperatures at sub-systems #1 and #2 respectively, °C.

6.2.2. Measurements of the outdoor air temperature at the mixing box inlet (T_{ac})

The temperature of the air in the AHU, just after the heat recovery coils SC2-1 and SC2-2 (T_{ac}), is a variable of interest for this research work. The air temperature T_{ac} is not currently measured by the BAS. The compact structure of the mixing box makes difficult to measure the effective air temperature just after the heat recovery coils, before the outdoor air stream starts to mix up with the recirculated air stream (Figure 6.4). In order to identify the optimal location where T_{ac} can be properly measured, a set of six air temperature portable sensors (from #5 to #10) has been placed, in system #1, just after the heat recovery coil SC2-1, on the opposite side from the mixing dampers (Figure 6.5). Measurements of the air temperature ($T_{ac}^{\#1}$) have been collected when the heat recovery coil was not working and the mixing dampers were open, along a ten days period between October and November in 2016. Collected values have been compared to measurements of the outdoor air temperature before the heat recovery coil (T_{oa}). When the heat recovery system is off, the air temperature after the recovery coil ($T_{ac}^{\#1}$) is expected to be the same at outdoor condition ($T_{ac}^{\#1} = T_{oa}$). Figure 6.6 shows, for each of the six sensors, the Mean Absolute Error (MAE) and the maximum Absolute Error (AE_{max}) between $T_{ac}^{\#1}$ and T_{oa}. A black dash line reports the uncertainty of the difference between those two measurements (U = 0.53°C). Air temperature differences smaller than the correspondent uncertainty have no engineering meaning, and are not taken in consideration. Except for location #5, measurements from all the other locations gave have MAE lower than the uncertainty. Location #6 has been selected for the measurements of the air temperature after the recovery coil.

The occurrence of air stratification at the outdoor air mixing box inlet has been verified collecting measurements of the air temperature at point #5 at different heights (Figure 6.7) along the same ten days period between October and November in 2016. The air temperature difference among the collected points was always smaller than the correspondent uncertainty. Thus, the air temperature distribution at the mixing box inlet has been taken homogenous.

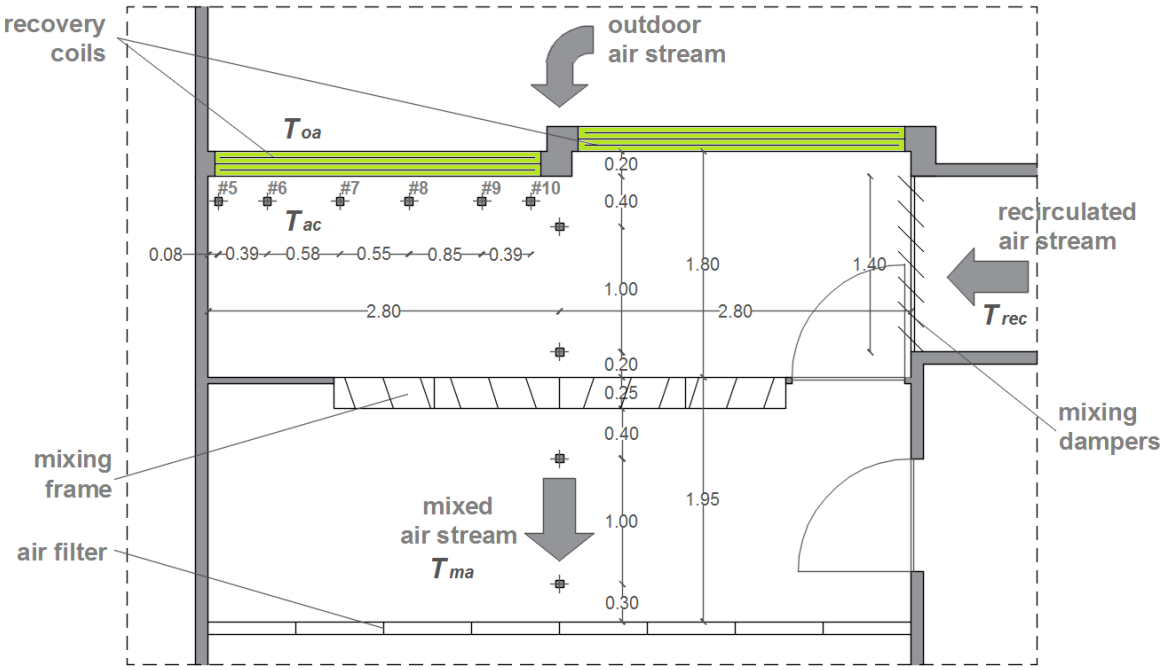


Figure 6.4 – Schematic of the mixing box from one AHU sub-system (dimension in meters).

Figure 6.5 – Picture: air temperature portable sensors placed just after the heat recovery coil.

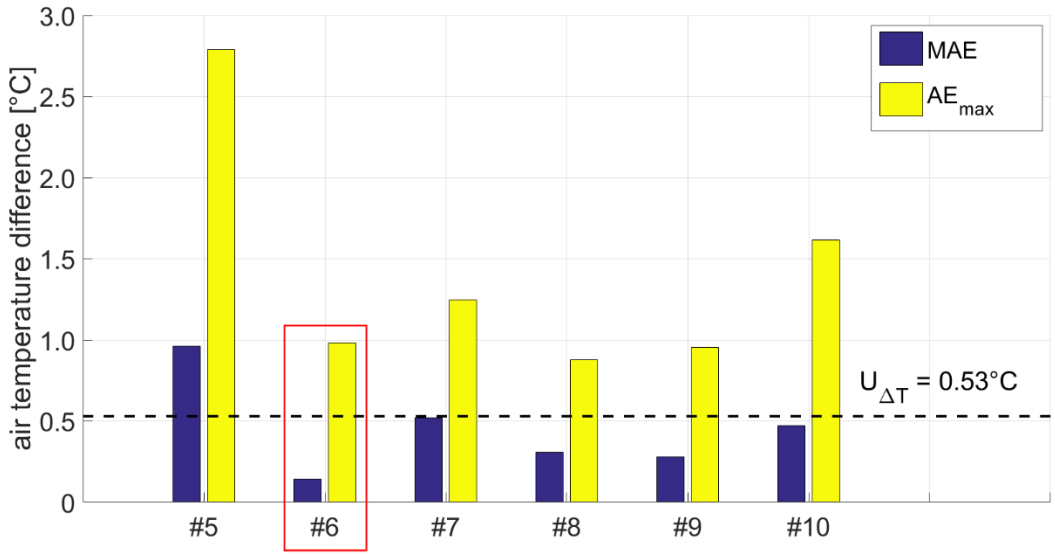


Figure 6.6 – MAE and AE_{max} between air temperatures before and after the heat recovery coil at different locations, while the heat recovery loop is not working, along ten days between October and November 2016.

Figure 6.7 - Picture: air temperature portable sensors at different heights after the heat recovery coil while the heat recovery loop is not working.

6.2.3. Trend data validation

Data validation has been performed on the trend data from the BAS collected between November 20, in 2016, and January 12, in 2017. Physical and theoretical limits have been verified according to ASHRAE (2005b), and measurements behind those limits have been discharged.

Figure 6.8 shows the glycol temperature difference (ΔT_{glc}) across the recovery coil SR1-3. Two time intervals were identified (red dash lines in figure 6.8) which present abnormally fast variations of the parameter ΔT_{glc} (the glycol temperature difference across coil SR1-3). Interval 1 was found to be due to the return fans from laboratories and washrooms which were off during few hours, even though the outdoor air temperature was lower than 8.0°C and the recovery loop was working. Thus, the recovery coil did not recover any heat along these measurements. This may be due to maintenance operation by the building operators, a fault at the fans or a bug in the control code. Interval 2 includes fast oscillations of the glycol temperature difference, up to 5.0-6.0°C, at each 15-30 minutes, which were found to be correlated to the outdoor air temperature. Figure 6.9 shows those consecutive variations of the glycol temperature, $|\Delta T_{glc}(t+1) - \Delta T_{glc}(t)|$, plotted against the outdoor air temperature. A marked change in variations magnitude occurs at around 0.0°C. Although the recovery loop is started when the outdoor air temperature drops below 8.0°C, above 0.0°C the outdoor air stream is not cold enough to always establish a quasi-steady state thermal exchange with the heat recovery coils SR2-1 and SC2-2 (blue points in figure 6.9). The ΔT_{glc} fast variations are thus due to transient heat transfer. When $T_{oa} < 0.0$°C, the observation are considered to be representative of a quasi-steady state heat transfer at SR1-3. Only the smallest 95% of ΔT_{glc} variations has been retained (as the remaining 5% includes outliers and abnormal values) which corresponds to a maximum ΔT_{glc} variation equal to 0.5°C (orange points in figure 6.9). The maximum ΔT_{glc} variation has been used as criteria to detect transient operation, and applied to the entire data set, including measurements when 0.0°C < T_{oa} < 8.0°C (dash line in figure 6.9).

From data validation, two windows of measurements where identified: window #1 (December 21-25, 2016), here named *training period,* and window #2 (from December 27, 2016 to January 6, 2017), named *validation period*.

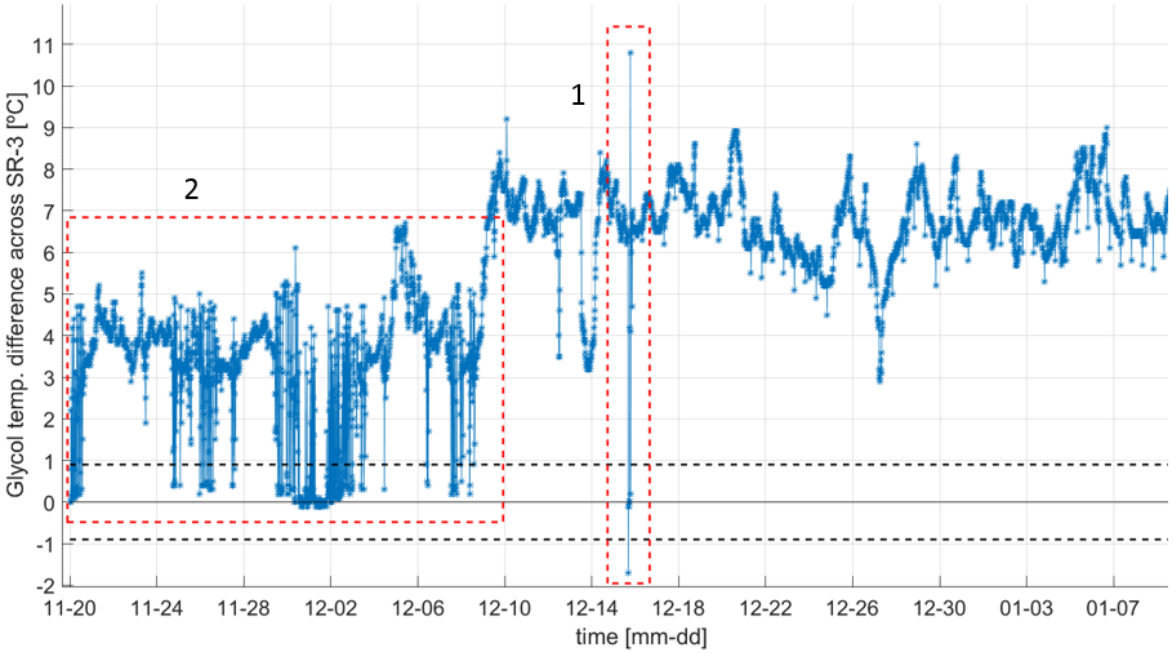


Figure 6.8 – Glycol temperature difference across the recovery coil SR1-3.

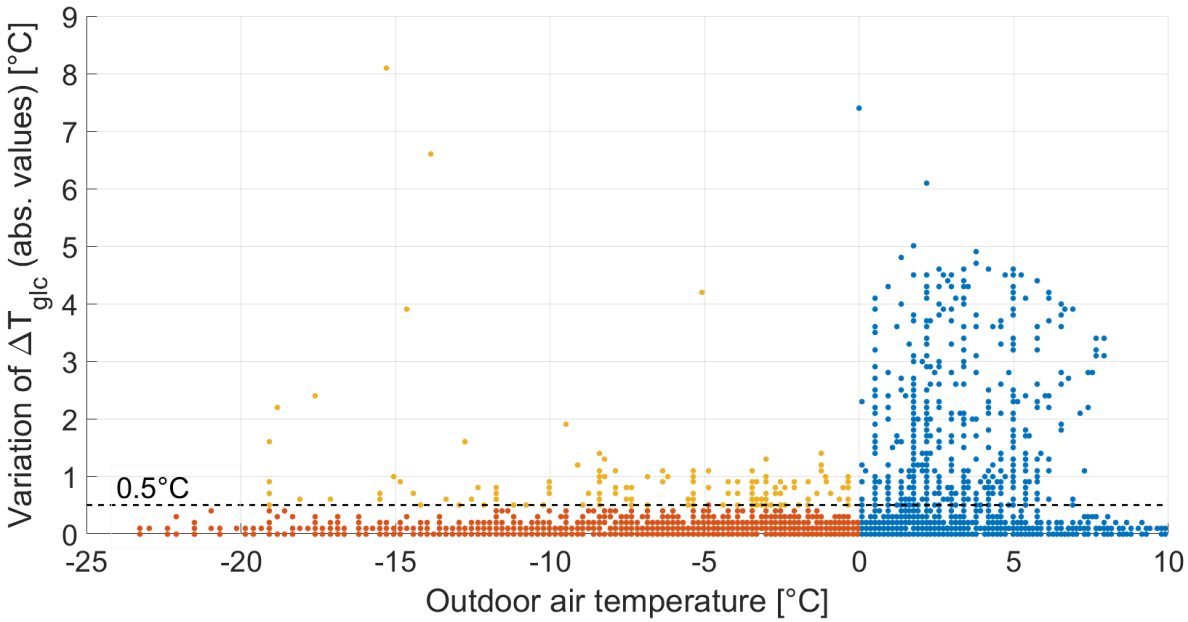


Figure 6.9 – Variations of the glycol temperature difference (absolute values) versus the outdoor air.

6.2.4. Models training

The *training dataset* (December 21-25, 2016) is here used to: i) train the regression model from eq. 6.13, which will predict values of the recirculation air temperature; and ii) derive the actual value of the glycol flow rate through the heat recovery system (V_{glc}).

6.2.4.1 The prediction model for the recirculation air temperature ($T_{rec,p}$)

Short-term measurements (STM) of the recirculated air temperature (T_{rec}) have been collected along the training period using portable calibrated sensors. The difference between the return (T_{ra}) and recirculated (T_{rec}) air temperatures has been found to be correlated to the outdoor air temperature: lower was the outdoor air temperature, larger was the difference between T_{ra} and T_{rec} (Figure 6.10).

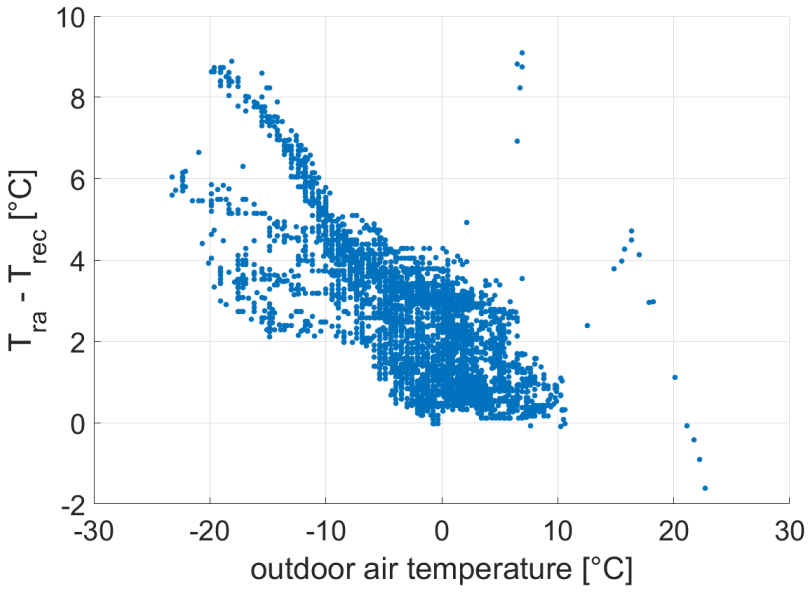


Figure 6.10 – Air temperature difference between return and recirculated conditions plotted against the outdoor air temperature.

The air handling system is installed in a non-conditioned space, and the long, non-insulated recirculation duct is exposed to the ambient temperature which is strictly correlated to the outdoor condition. A regression model, thus, is developed which predicts the recirculated air temperature from measurements of the outdoor and return air temperatures (eq. 6.17). The goodness of fit of the predictions from Eq. 11 to the short-term measurements of the recirculated air temperature is evaluated in terms of statistical indices: MAE = 0.45°C, CV-RMSE = 3.3%, and R^2 = 0.43.

$$T_{rec,p} = 1.24 + 0.17 \cdot T_{oa} + 0.81 \cdot T_{ra} \quad \text{eq. 6.17}$$

where $T_{rec,p}$ is the predicted value of the recirculated air temperature, °C.

6.2.4.2 The glycol flow rate (V_{glc})

The recovery loop pump (P03-GLC) is rated by the manufacturer at 11.8 L/s constant flow rate. This value was verified with spot measurements from a portable ultrasonic flow rate. Two times, in April 2014 and February 2017, the glycol flow rate was measured and found to be 10.0 and 9.7 L/s respectively, with 0.5 L/s uncertainty. The discrepancy between measurements from 2014 and 2017 can be explained by the uncertainty due to the flow meter. The only available location to install the portable flow meter, in both 2014 and 2017, was just before the pump (Figure 6.11). It was not possible to comply with the minimum distance between the flow meter and any obstruction or liquid flow distortion, as required by the flow meter user's guide (Greyline instruments inc. n.d.).

Figure 6.11 – Pictures of the portable flow meter installed just before pump P03-GLC.

Thus, measurements from the *training period* were used to derive the actual glycol constant flow rate from the energy balance at the recovery coils (eq. 6.18). The derived average glycol constant flow rate was found to be $V_{glc,d}$ = 8.2 L/s (Figure 6.12), with a standard deviation σ = 0.25 L/s, and U_{Vglc} = ±0.9 L/s. Figure 6.13 reports the glycol constant flow rate as given by manufacturer specification, measured in 2014 and 2017, and derived from eq. 6.18, along with the correspondent uncertainty intervals in red.

$$V_{glc,d} = \frac{\alpha \cdot (\rho_{air} \cdot V_{sa}) \cdot C_{p,air} \cdot (T_{ac} - T_{oa})}{\rho_{glc} \cdot C_{p,glc} \cdot (T_{hre} - T_{hra})} \qquad \text{eq. 6.18}$$

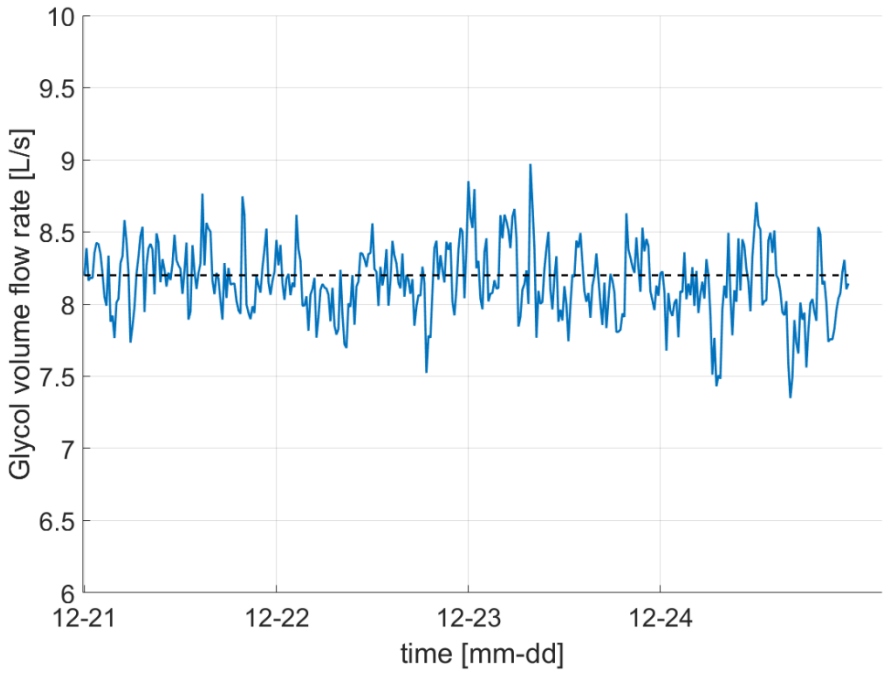


Figure 6.12 – Glycol volume flow rate derived from measurements and eq. 6.18.

The discrepancy between spot measurements and the value derived from eq. 6.18 is larger than the involved uncertainties. This circumstance, together with the non-optimal location of the portable flow meter, makes the spot measurements from 2014 and 2017 unusable. Thus, for the implementation of the presented VFMs, the glycol constant flow rate as derived from eq. 6.18 has been used (V_{glc} = 8.2 L/s).

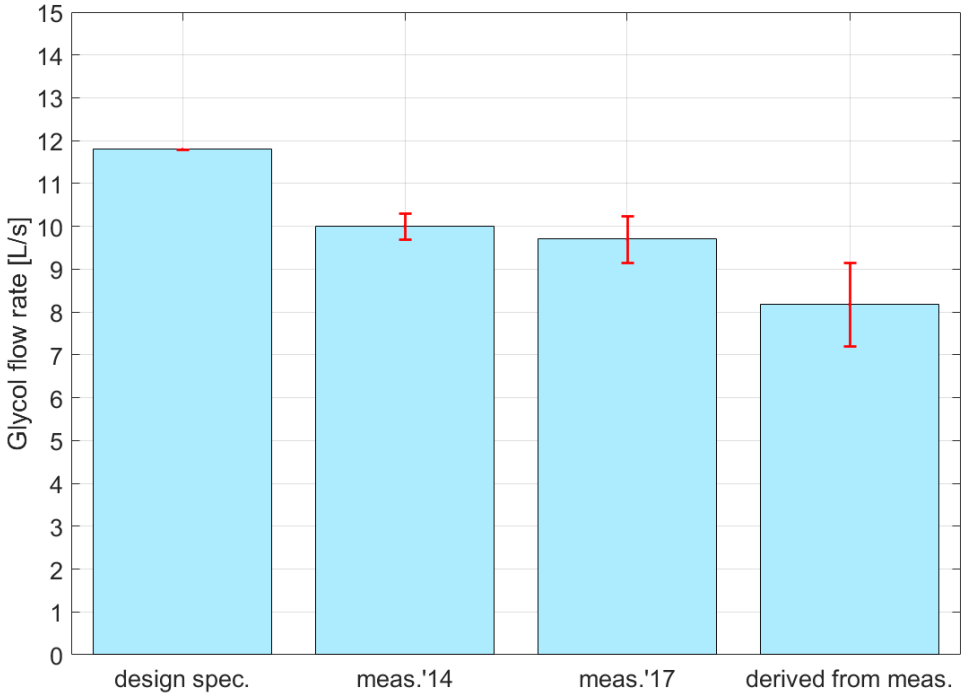


Figure 6.13 – Glycol constant flow rate values (bars) and uncertainties (red)

6.2.5. Validation of the factor α models

The two presented models of the factor α (α_a from eq. 6.12, and α_b from eq. 6.14) have been applied to measurements from the *validation* data set (from December 27 in 2016 to January 6 in 2017). Reference values of the factor α from eq. 6.10 (α_{ref}) have been used for comparison. Values of the factor α from the common formulation from eq. 6.1 (α_0) are used to show the improvement of predictions due to the use of the actual (measured or derived) mixing box inlets air temperatures.

In figure 6.14 the values of the four factors α (α_a, α_b, α_0 and α_{ref}) are plotted during few hours from the *validation* period. Table 6.6 reports the statistical indices from comparing the factor α predictions from eqs. 6.1, 6.12 and 6.14 (α_0, α_a, α_b) against the reference values from eq. 6.10 (α_{ref}).

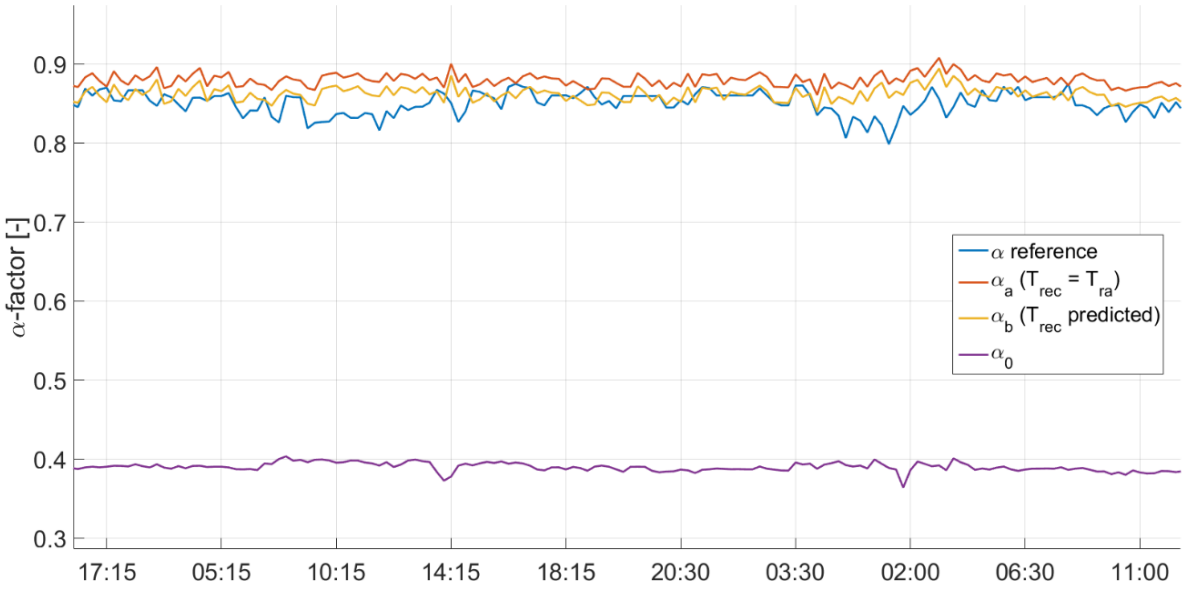


Figure 6.14 – Factor α predictions from the reference model (α_{ref}), the two proposed models (α_a and α_b), and the base case (α_0) during 24 hours along the *training* period.

Table 6.6 – Statistical indices from comparison against factor α reference values (α_{ref}).

	MBE [-]	MAE [-]	AE_{max} [-]	CV-RMSE [%]
Training period (window #1) December 21-25, 2016				
α_0	-0.44	0.44	0.48	53.2
α_a	0.045	0.045	0.09	5.7
α_b	0.026	0.026	0.07	3.7
Validation period (window #2) December 27, 2016 - January 6, 2017				
α_0	-0.46	0.46	0.50	54.0
α_a	0.03	0.03	0.08	4.2
α_b	0.01	0.02	0.08	2.7

6.2.6. Uncertainty analysis and Bias (fixed) error estimation and removal

The propagated uncertainty was estimated for each considered factor α model: eq. 6.10 (α_{ref}), eq. 6.12 (α_a) and eq. 6.14 (α_b) along the *training data* set. Table 6.7 reports the measurements mean value (μ) and random error (R) as observed along the considered period, along with the bias (fixed) error B as stated by the sensors manufacturer.

Table 6.7 – Measurements mean value and random error as observed during the training period (December 21-25, in 2016), and bias error as stated by the sensors manufacturer.

	V_{glc} [L/s]	V_{sa} [L/s]	T_{hra} [°C]	T_{hre} [°C]	T_{ma} [°C]	T_{oa} [°C]	T_{ra} [°C]	T_{ac} [°C]	T_{rec} [°C]
μ	8.2	14,618	8.9	15.0	13.7	1.0	22.3	12.9	19.9
R (±)	0.3	240.0	0.17	0.17	0.35	0.55	0.10	0.01	0.01
B (±)	0.85	222.3	0.45	0.45	0.45	0.45	0.45	0.2	0.2

The propagated uncertainty has been estimated to be ±0.05, ±0.42 and ±0.56 for α_{ref}, α_a and α_b, respectively. According to section 3.5.1 (Bias (fixed) error estimation and removal) an estimation of the contribution of the bias (fixed) error to the residual from reference values was evaluated for both models α_a and α_b. Assuming the random errors to be symmetrical around the mean value, the estimated contribution of the bias (fixed) error to the overall uncertainty of α_a and α_b is given by the MBE along the *training period* (Table 6.6): MBE_a = 0.045 and MBE_b = 0.026 for α_a and α_b, respectively (Table 6.6). Figures 6.15 and 6.16 show for both α_a and α_b the distribution of frequency of the residuals from the reference values α_{ref} along the *training period*. The correspondence between the MBE values from table 6.6 and the highest frequency residual values (Figures 6.15 and 6.16) validate the assumption that the random error around the mean was symmetrical.

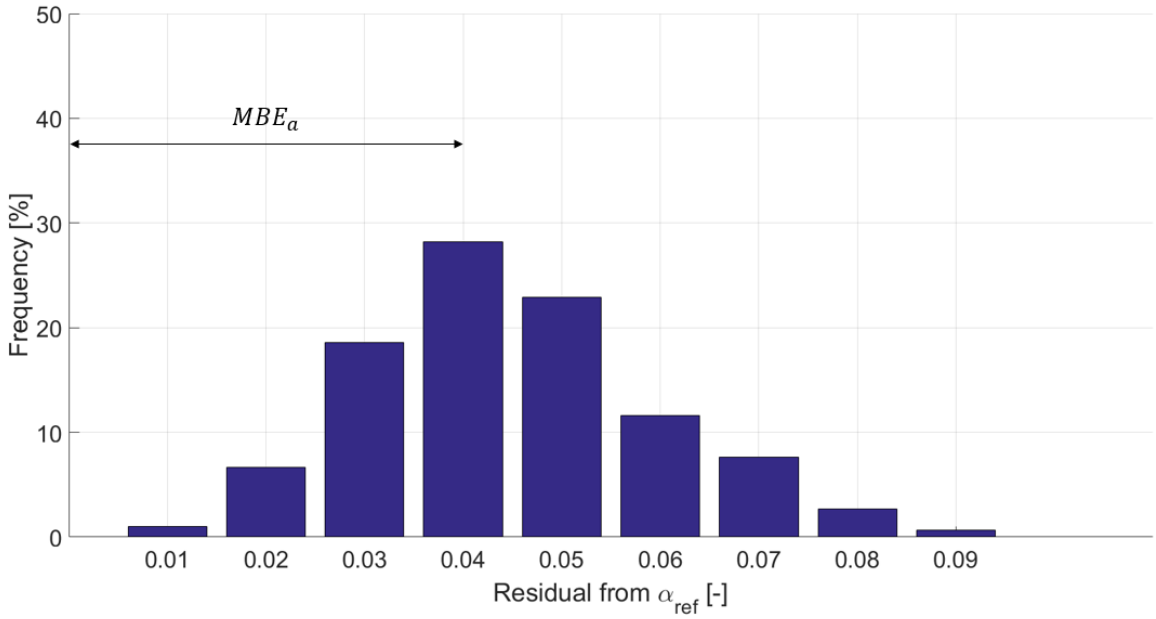


Figure 6.15 - α_a distribution of frequency of the residuals from α_{ref} along the *training period*.

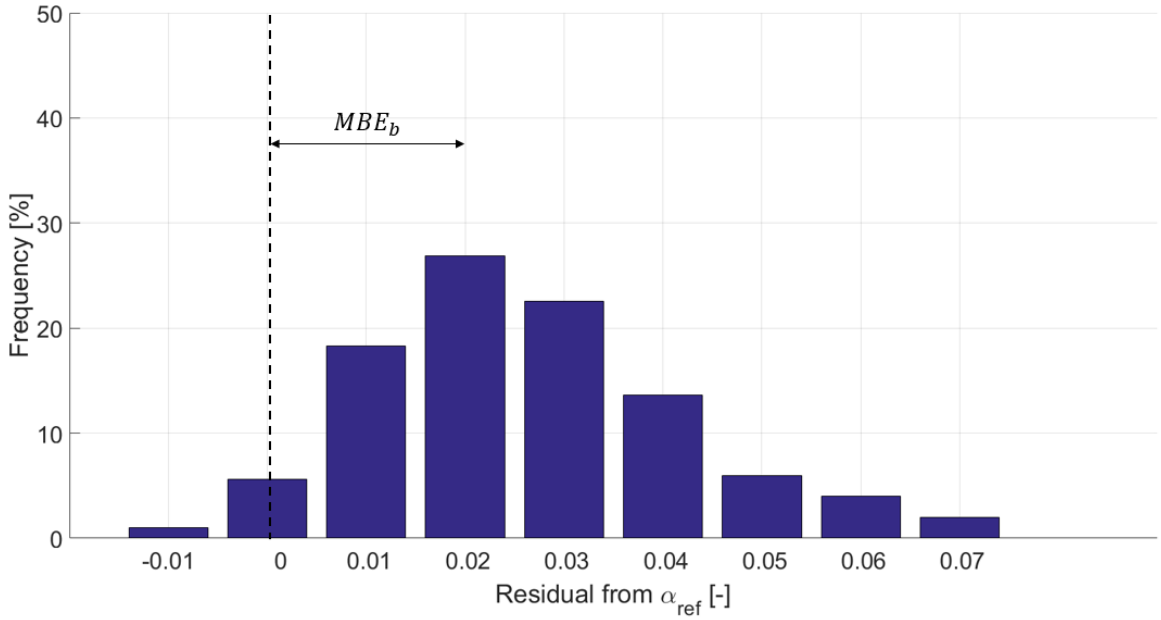


Figure 6.16 – α_b distribution of frequency of the residuals from α_{ref} along the *training period*.

The contribution of the bias (fixed) error estimated along the *training period*, is subtracted by further predictions along the *validation period* (eqs. 6.19 and 6.20) (section 3.5.1). Those new predictions are said to be *unbiased*, and the models from eqs. 6.19 and 6.20 ($\alpha_{a,c}$ and $\alpha_{b,c}$) are said to be *calibrated*. Predictions from calibrated models along the *validation period*, have been compared to α_{ref}, and the statistical indices from comparison are given in Table 6.8.

$$\alpha_{a,c} = \frac{Q_{hr} - \rho_{air} \cdot V_{sa} \cdot C_{p,air} \cdot (T_{ma} - T_{ra})}{\rho_{air} \cdot V_{sa} \cdot C_{p,air} \cdot (T_{ra} - T_{oa})} - MBE_a \qquad \text{eq. 6.19}$$

$$\alpha_{b,c} = \frac{Q_{hr} - \rho_{air} \cdot V_{sa} \cdot C_{p,air} \cdot (T_{ma} - T_{rec,p})}{\rho_{air} \cdot V_{sa} \cdot C_{p,air} \cdot (T_{rec,p} - T_{oa})} - MBE_b \qquad \text{eq. 6.20}$$

where MBE_a and MBE_b are the mean bias errors for factors α_a and α_b respectively along the training period, $\alpha_{a,c}$ and $\alpha_{b,c}$ are the initial factors α_a and α_b calibrated with MBE_a and MBE_b respectively.

Table 6.8 – Statistical indices from comparison between calibrated models and reference values (α_{ref}) along the validation period (from December 27 in 2016, to January 6 in 2017).

	MBE [-]	MAE [-]	AE_{max} [-]	CV-RMSE [%]
Validation period (window #2) December 27, 2016 - January 6, 2017				
$\boldsymbol{\alpha_a}$	-0.009	0.015	0.09	2.3
$\boldsymbol{\alpha_b}$	-0.008	0.016	0.10	2.4

6.2.7. Discussion

When the AHU is working in heat recovery mode and the heat loss through the recirculation duct cannot be neglected, the common formulation for the outdoor air ratio (factor α_0 from eq. 6.1) cannot be used (Figure 6.14 and Table 6.6). Reference values of the factor α (α_{ref}) were obtained with short-term measurements of the actual air temperature at the mixing box inlets (eq. 6.10). Two new formulations for the factor α (eqs. 6.12 and 6.14) were proposed which do not need for additional permanent sensors.

Along BAS trend data of the heat recovery system from the considered period (from November 20 in 2016 to January 6 in 2017), a transient thermodynamic equilibrium occurred several times. From data visualization, a criterion was derived to detect the occurrence of transient operation. When transient operation was detected, the correspondent measurements were discharged from the data set. From a four days dataset (*training dataset,* December 21-25, 2016), which included measurements from the BAS and from portable sensors, a T_{rec} prediction model (eq. 6.17) and the actual glycol flow rate (eq. 6.18) were derived.

The factor α was predicted through the two proposed formulations (eqs. 6.12 and 6.14), and predictions were compared against the reference values from eq. 6.10 (Figure 6.14 and Table 6.6). The statistical indices from table 6.6 prove that both the proposed models (eqs. 6.12 and 6.14) are highly effective in predicting the outdoor air ration. Along the *validation period* the MBE and MAE are lower than 0.03, which correspond to the 3% of the factor α full range of variation. Also, the AE_{max} is lower than 0.1, and CV-RMSE does not exceed 5.0%.

The MBE from the training period has been used to estimate and remove the effect of the bias error from the predictions of the two presented formulations over the validation period (Table 6.8). As a result, the MBE dropped under 0.01 (1.0% of the factor α full range of variation): MBE = -0.08 and MBE = -0.09 for α_a and α_b respectively.

Table 6.9 lists the long-term measurements required by each of the two proposed formulations (α_a and α_b) and the ones required by the reference α formulation (α_{ref}). Although the two proposed formulations require more measurements than the model used as reference (six permanent sensors for both α_a and α_b against three for α_{ref}), those sensors are commonly already installed and measurements are available from BASs as they are used for AHU control. The implementation of the presented models, thus, does not need any additional physical sensor, which would come at additional cost. On the other hands, although the α_{ref} formulation (eq. 6.10) only

needs for three air temperature measures to be implemented, two over three are usually not collected from BAS. In order to implement eq. 6.10, two new dedicated physical sensors should be purchased and installed.

Table 6.9 – List of permanent (x) sensors required by each VFM model.

Factor α	T_{hre} [°C]	T_{hra} [°C]	T_{oa} [°C]	T_{ac} [°C]	T_{ma} [°C]	T_{ra} [°C]	T_{rec} [°C]	V_{sa} [L/s]
α_{ref}				**x**	x		**x**	x
α_a	x	x	x		x	x		x
α_b	x	x	x		x	x		x

The uncertainty of each considered formulation was found to be $U_{ref} = \pm 0.05$, $U_a = \pm 0.42$ and $U_b = \pm 0.56$ for the reference model and models *a* and *b* respectively.

As the bias (fixed) error was derived in terms of MBE, and removed from the predictions of α along the validating period, the propagated random error for each α-prediction model is the only source of error affecting the *unbiased* α predictions. Because of its simple formulation and small random errors affecting the input variables, α_{ref} (eq. 6.10) is affected by an extremely small random error, $R_{ref} = \pm 0.001$. The random error for models *a* and *b* was found to be $R_a = \pm 0.3$ and $R_b = \pm 0.4$ respectively, which means that the predictions of α from the *calibrated* formulations (eqs. 6.19 and 6.20) are expected to fall within the intervals ±0.3 and ±0.4 around the true value, with a normal distribution and a 95% confidence level.

Although maximum uncertainty of ±0.3 or ±0.4 would make the predictions of α unreliable for every time step, the statistical indices from Table 6.8 indicate that the average residual between predictions and α_{ref} is small (MAE = 0.015 for α1 and MAE = 0.016 for α2). Thus, the two new models should be used for the estimation of the daily average value of the factor α, and hence of the daily average outdoor air flow rate.

7. VIRTUAL SENSORS FOR THE AIR PROPERTIES AT THE AHUs MIXING BOX

The methods for the virtual measurements or virtual calibration of air properties sensors in AHUs introduced in section 3.4 are here implemented and validated.

7.1 Method A

Method A is here applied to re-calibrate the outdoor air temperature (T_{oa}) sensor, using measurements from October 19, 2016 (Figure 7.1). The corrected values are compared with the STM. Table 7.1 reports statistical indices of the comparison between the STM and the faulty BAS measurements, and between the STM and the corrected BAS measurements. The overall uncertainty affecting the difference between STM and BAS measurements is ±0.53ºC for air temperature and ±4.1% for air relative humidity.

Table 7.1 – Statistical indices of the outdoor air temperature STM comparison to the BAS faulty measurements, and to the BAS corrected measurements, with method A, on October 19, 2016.

Comparison	AE_{max}	MAE	CV-RMSE
STM vs BAS	4.0ºC	1.2ºC	9.7%
STM vs BAS corrected	0.7ºC	0.2ºC	1.7%

Similarly, method A was used to re-calibrate the sensors of the remaining five air properties from Figure 3.5, by considering one faulty sensor at a time. For this purpose a dedicate data set has been generated for each variable to be corrected. The dataset includes faulty measurements of the considered variable along with fault-free values of the remaining five variables. Figures 7.2 and 7.3 show the results from correction of the mixed air temperature and outdoor air relative humidity respectively. Statistical indices of the difference between the STM and faulty BAS measurements, and between the STM and BAS corrected measurements are given in Tables 7.2 to 7.6.

Part of the content of this chapter has been presented in a conference paper: Cotrufo and Zmeureanu. 2016. A New Algorithm for Sensors Verification and Correction in Air Handling Units. *eSim 2016 Conference, Hamilton, ON.*

Table 7.2 – Statistical indices of the mixed air temperature STM comparison to the BAS faulty measurements, and to the BAS corrected measurements, with method A, on October 19, 2016.

Comparison	AE_{max}	MAE	CV-RMSE
STM vs BAS	5.0°C	3.1°C	22.9%
STM vs BAS corrected	0.43°C	0.2°C	1.4%

Table 7.3 – Statistical indices of the outdoor air humidity STM comparison to the BAS faulty measurements, and to the BAS corrected measurements, with method A, on October 19, 2016.

Comparison	AE_{max}	MAE	CV-RMSE
STM vs BAS	38.7%	32.5%	38.4%
STM vs BAS corrected	5.1%	2.1%	2.9%

Table 7.4 – Statistical indices of the mixed air humidity STM comparison to the BAS faulty measurements, and to the BAS corrected measurements, with method A, on October 19, 2016.

Comparison	AE_{max}	MAE	CV-RMSE
STM vs BAS	30.0%	30.0%	43.0%
STM vs BAS corrected	3.5%	0.8%	1.6%

Table 7.5 – Statistical indices of the return air temperature STM comparison to the BAS faulty measurements, and to the BAS corrected measurements, with method A, on October 19, 2016.

Comparison	AE_{max}	MAE	CV-RMSE
STM vs BAS	5.0°C	3.1°C	17.1%
STM vs BAS corrected	3.38°C	1.1°C	6.8%

Table 7.6 – Statistical indices of the return air humidity STM comparison to the BAS faulty measurements, and to the BAS corrected measurements, with method A, on October 19, 2016.

Comparison	AE_{max}	MAE	CV-RMSE
STM vs BAS	25.0%	25.0%	53.3%
STM vs BAS corrected	7.5%	1.9%	5.6%

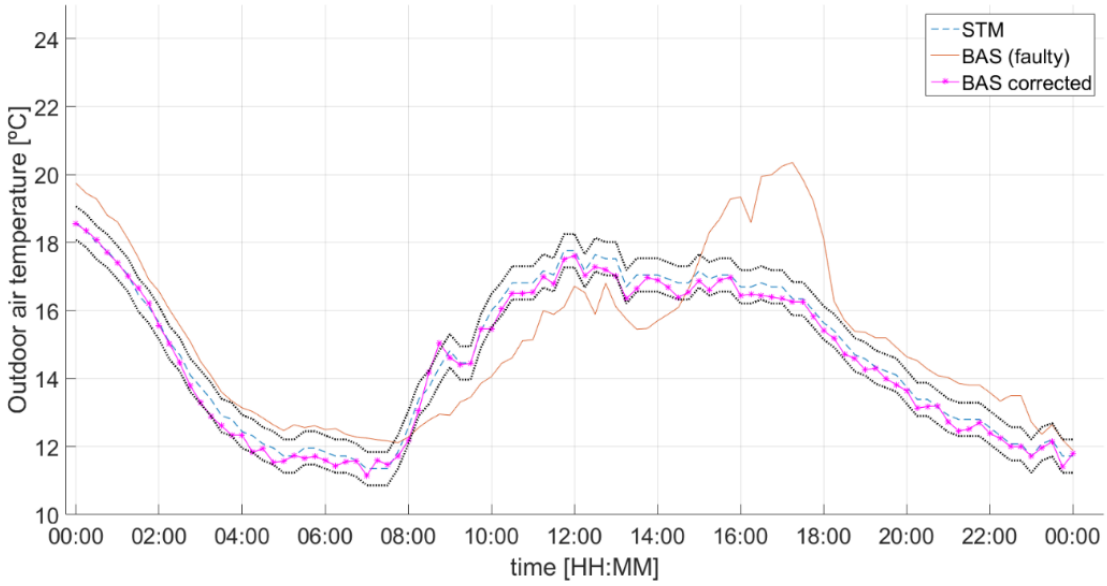


Figure 7.1 – Outdoor air temperature: STM (dash line), BAS (solid line) and corrected (line-asterix) measurements, and uncertainty (dot black lines) on October 19, in 2016.

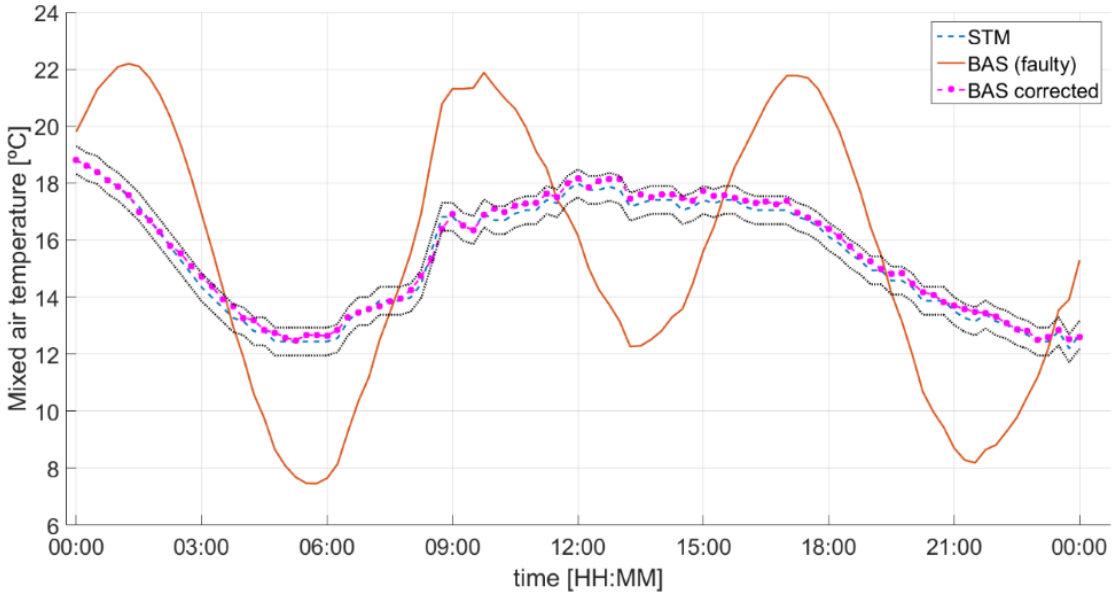


Figure 7.2 – Mixed air temperature: STM (dash line), BAS (solid line) and corrected (line-asterix) measurements, and uncertainty (dot black lines) on October 19, in 2016.

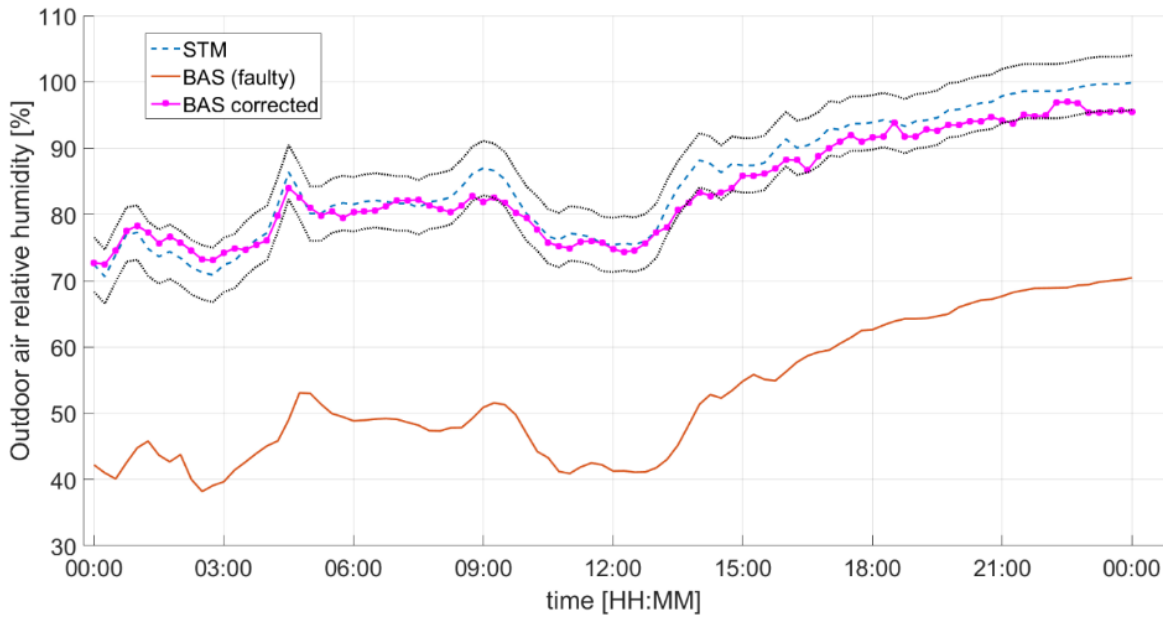


Figure 7.3 – Outdoor air relative humidity: STM (dash line), BAS (solid line) and corrected (line-asterix) measurements, and uncertainty (dot black lines) on October 19, in 2016.

7.2 Method B

In the absence of a physical sensor, the variable of interest is virtually measured (modelled) in terms of measurements from other sensors related to the thermodynamic process within the AHU. The paper focuses on the prediction of the temperature and relative humidity of outdoor air that enters the AHU. The same approach can also be used to virtually measure the air properties at mixed or return conditions.

There are three models B for the three different operation modes of an AHU.

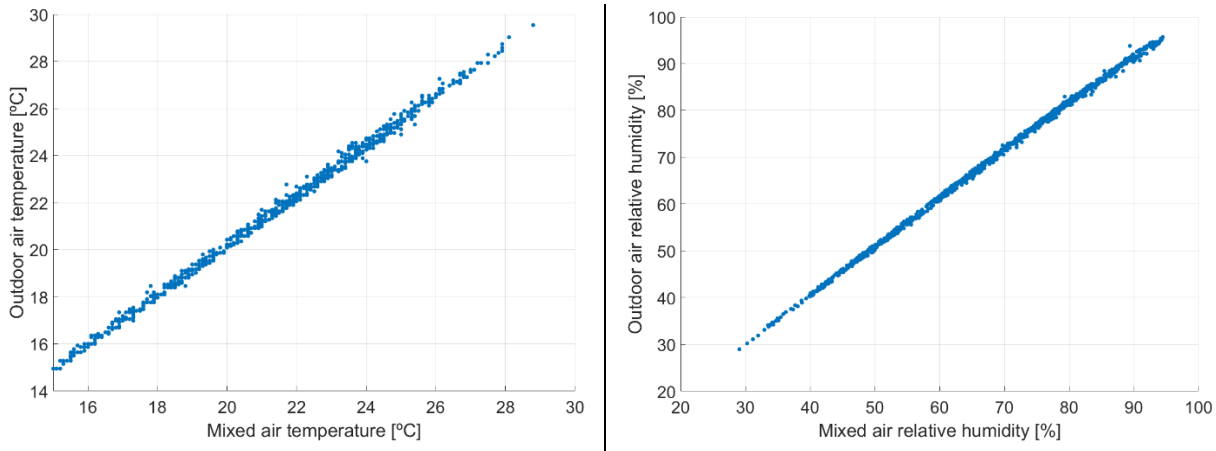
7.2.1 Models B1

Under operation mode #1, the AHU economizer works with 100% outdoor air. The outdoor air (*oa*) flow reaches the mixing box without changes of the thermodynamic properties. Hence, the variables of interest are derived from the measurements of the mixed air (*ma*) (eqs. 7.1 and 7.2). Those relations are well known and commonly used for proactive tests to verify the sensors calibration status (e.g. Fernandez et al. 2009).

$$T_{oa} = T_{ma} \qquad \text{eq. 7.1}$$

$$RH_{oa} = RH_{ma} \qquad \text{eq. 7.2}$$

Measurements from September 2 to 18, in 2015, are used for the operation mode #1. The outdoor air properties are predicted with eqs. 7.1 and 7.2 (Figures 7.4 and Table 7.7).



(*a*) (*b*)

Figure 7.4– Models B1: *a*) comparison between the outdoor and mixed air temperature (eq. B1), and *b*) comparison between the outdoor and mixed air relative humidity.

Table 7.7 – Statistical indices from models B1.

Eq.	AE_{max}	MAE	CV-RMSE
7.1	0.85ºC	0.2ºC	1.4%
7.2	4.4%	1.4%	2.1%

The uncertainty affecting the difference between STM outdoor and BAS mixed measurements, was estimated as ±0.49ºC for air temperature and ±4.1% for relative humidity. For both air temperature and relative humidity, the average absolute difference between outdoor and mixed conditions (MAE) is lower than the corresponding overall uncertainty (Table 7.7).

7.2.2 Models B2

Under the operation mode #2, the mixing dampers are modulated to control the recirculated air flow rate. The STM are used to train the prediction models (eqs. 7.3 to 7.6) that predict the outdoor and mixed air temperature and relative humidity from the measurements of the most correlated variables:

$$T_{oa} = a_0 + a_1 \cdot T_{ma} + a_2 \cdot MD_{\%} \qquad \text{eq. 7.3}$$

$$RH_{oa} = b_0 + b_1 \cdot RH_{ma} + b_2 \cdot RH_{ra} \qquad \text{eq. 7.4}$$

$$T_{ma} = c_0 + c_1 \cdot T_{oa} + c_2 \cdot MD_{\%} \quad \text{eq. 7.5}$$

$$RH_{ma} = d_0 + d_1 \cdot RH_{oa} + d_2 \cdot RH_{ra} \quad \text{eq. 7.6}$$

where a_i, b_i, c_i and d_i are the coefficients of the prediction models, $MD_{\%}$ is the mixing dampers control signal, and RH is the relative humidity.

Measurements from October 16 and November 11, 2015, are used for the operation mode 2. Models B2 (eqs. 7.3 to 7.6) have been trained with the STM for one-week period between October 16 and 23 (672 observation at 15 minutes time step). The uncertainty from models predictions has been evaluated as well (Table 7.8). For validation purpose, the predictions from models B2 have been compared to STM in terms of statistical indices. The difference between models predictions and STM has an uncertainty itself, which is given in Table 7.8.

The results over the validation period of 17 days (October 24 to November 11) are given in Figures 7.5 and Tables 7.9 and 7.10 for the outdoor air properties (eqs. 7.3 and 7.4), and in Figures 7.6 and Tables 7.11 and 7.12 for the mixed air properties (eqs. 7.5 and 7.6).

Table 7.8 – Model B2 trained with the STM for one-week (October 16 to 23).

Eq.	Trained model	Models uncertainty	Difference uncertainty
7.3	$T_{oa} = -2.133 + 1.043 \cdot T_{ma} + 0.021 \cdot MD_{\%}$	±0.47°C	±0.55°C
7.4	$RH_{oa} = -0.281 + 1.842 \cdot RH_{ma} - 1.042 \cdot RH_{ra}$	±8.47%	±8.51%
7.5	$T_{ma} = 3.003 + 0.873 \cdot T_{oa} - 0.016 \cdot MD_{\%}$	±0.43°C	±0.47°C
7.6	$RH_{ma} = -0.62 + 0.395 \cdot RH_{oa} + 0.862 \cdot RH_{ra}$	±3.79%	±3.88%

Table 7.9 – Comparison between the outdoor air temperature STM and model B2 (eq. 7.3) predictions.

Time interval	AE_{max}	MAE	CV-RMSE
Training	1.19ºC	0.3ºC	3.4%
Validation	3.13ºC	0.4ºC	4.8%

Table 7.10 – Comparison between the outdoor air relative humidity STM and model B2 (eq. 7.4) predictions.

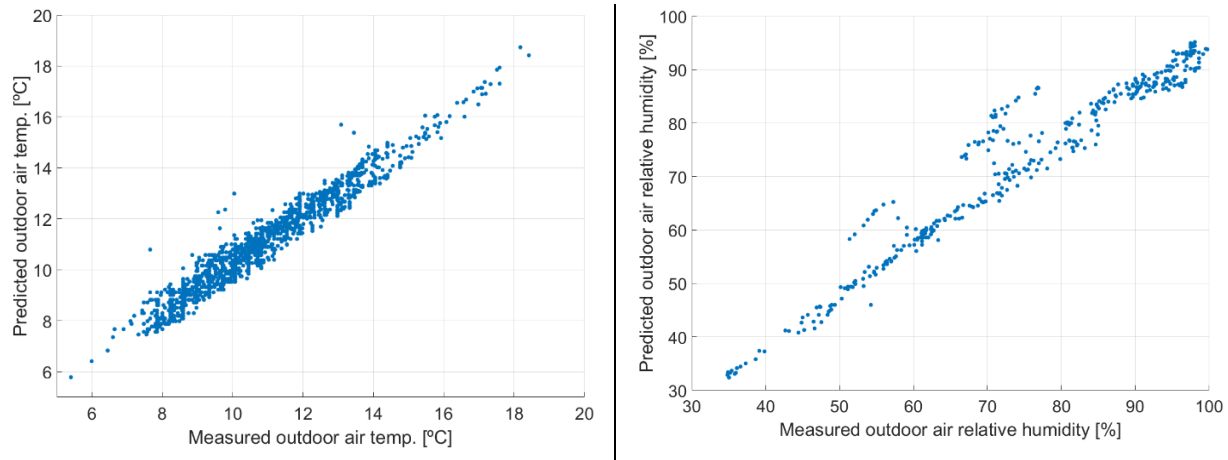
Time interval	AE_{max}	MAE	CV-RMSE
Training	14.6%	3.2%	4.7%
Validation	11.4%	4.1%	6.3%

Table 7.11 – Comparison between the mixed air temperature STM and model B2 (eq. 7.5) predictions.

Time interval	AE_{max}	MAE	CV-RMSE
Training	1.15ºC	0.3ºC	2.8%
Validation	2.73ºC	0.4ºC	3.7%

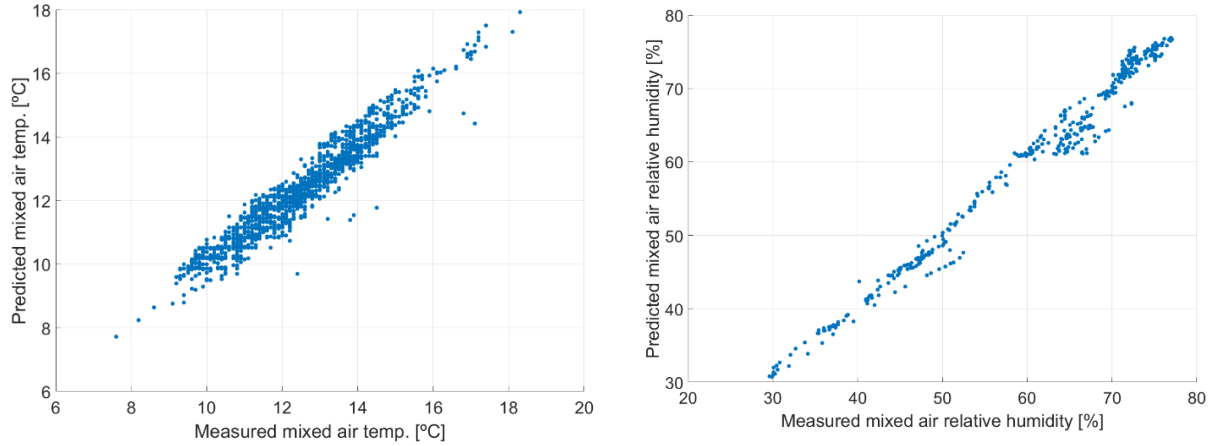
Table 7.12 – Comparison between the mixed air relative humidity STM and model B2 (eq. 7.6) predictions.

Time interval	AE_{max}	MAE	CV-RMSE
Training	7.8%	1.5%	2.6%
Validation	5.9%	1.3%	2.9%



(a) *(b)*

Figure 7.5 – Models B2 – validation period: comparison between STM and predicted values of: *a*) outdoor air temperature (eq. 7.3); and *b*) outdoor air relative humidity (eq. 7.4).



(a) *(b)*

Figure 7.6 – Models B2 – validation period: comparison between STM and predicted values of: *a*) mixed air temperature (eq. 7.5); and *b*) mixed air relative humidity (eq. 7.6).

7.2.3 Models B3

Under operation mode #3, the mixing dampers are fully open, and the outdoor air stream is preheated by the heat recovery coils (SC2-1 and SC2-2) before it reaches the mixing box. A correlation analysis identified the best regressors to predict the outdoor and mixed air properties. The STM are used to train the prediction models (eqs. 7.7 to 7.10) used to predict the outdoor and mixed air temperature and relative humidity.

$$T_{oa} = m_0 + m_1 \cdot T_{ma} + m_2 \cdot T_{hre} \quad \text{eq. 7.7}$$

$$RH_{oa} = n_0 + n_1 \cdot RH_{ma} + n_2 \cdot RH_{ra} \quad \text{eq. 7.8}$$

$$T_{ma} = p_0 + p_1 \cdot T_{oa} + p_2 \cdot T_{hre} \quad \text{eq. 7.9}$$

$$RH_{ma} = q_0 + q_1 \cdot RH_{oa} + q_2 \cdot RH_{ra} \quad \text{eq. 7.10}$$

where m_i, n_i, p_i and q_i are the coefficients of the prediction models, and T_{hre} is the glycol temperature at the heat recovery coil outlet.

Measurements from January 1 to 30, 2016, are selected for the operation mode #3, when the recovery loop was on. Models B3 (eqs. 7.7 to 7.10) are trained with the STM for one-week period between January 1 and 7, in 2016 (Table 7.13), and validated with the STM measurements from the remaining 23 days. The uncertainty from models predictions, and the uncertainty from

the difference between models predictions and STM are given in table 7.13. The results over the validation period of 23 days are given in Tables 7.14 and 7.15 and Figures 7.7 for the outdoor air properties, and in Tables 7.16 and 7.17 and Figures 7.8 for the mixed air properties.

Table 7.13 – Models B3 trained with the STM for one-week (January 1 to 7).

Eq.	Trained model	Models uncertainty	Difference uncertainty
7.7	$T_{oa} = -29.362 + 0.888 \cdot T_{ma} - 0.251 \cdot T_{hre}$	±0.42°C	±0.49°C
7.8	$RH_{oa} = 22.919 + 1.009 \cdot RH_{ma} + 0.359 \cdot RH_{ra}$	±4.29%	±4.36%
7.9	$T_{ma} = 28.16 + 0.977 \cdot T_{oa} - 1.081 \cdot T_{hre}$	±0.69°C	±0.71°C
7.10	$RH_{ma} = -0.344 + 0.045 \cdot RH_{oa} + 1.525 \cdot RH_{ra}$	±6.10%	±6.16%

Table 7.14 – Comparison between the outdoor air temperature STM and model B3 (eq. 7.7) predictions.

Time interval	AE_{max}	MAE	CV-RMSE
Training	2.32ºC	0.8ºC	-15.0
Validation	5.02ºC	0.8ºC	-20.0

Table 7.15 – Comparison between the outdoor air relative humidity STM and model B3 (eq. 7.8) predictions.

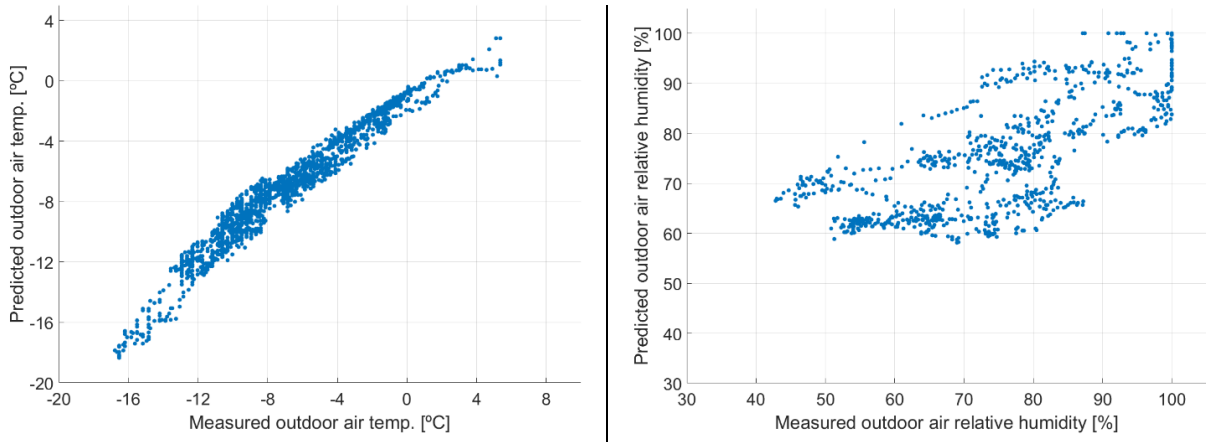
Time interval	AE_{max}	MAE	CV-RMSE
Training	23.8%	8.1%	14.0%
Validation	24.9%	8.1%	13.7%

Table 7.16 – Comparison between the mixed air temperature STM and model B3 (eq. 7.9) predictions.

Time interval	AE_{max}	MAE	CV-RMSE
Training	2.59ºC	0.6ºC	9.6%
Validation	5.05ºC	0.5ºC	8.0%

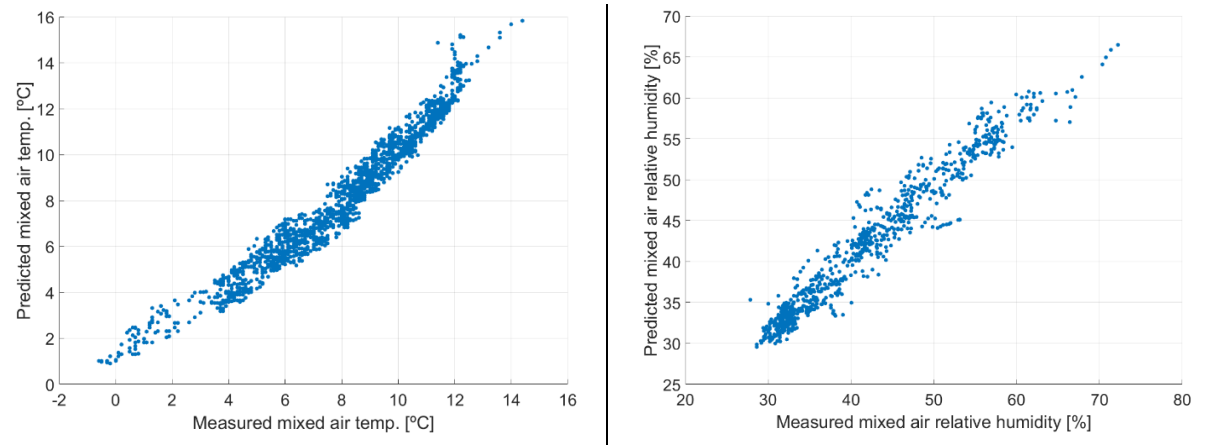
Table 7.17 – Comparison between the mixed air relative humidity STM and model B3 (eq. 7.10) predictions.

Time interval	AE_{max}	MAE	CV-RMSE
Training	14.0%	1.6%	5.4%
Validation	9.4%	1.6%	5.3%



(a) *(b)*

Figure 7.7 – Models B3 – validation period: comparison between STM and predicted values of: *a*) outdoor air temperature (eq. 7.7); and *b*) outdoor air relative humidity (eq. 7.8).



(a) *(b)*

Figure 7.8 – Models B3 – validation period: comparison between STM and predicted values of: *a*) mixed air temperature (eq. 7.9); and *b*) mixed air relative humidity (eq. 7.10).

7.3 Method C

The outdoor air sensor installed at the Loyola has been found to be faulty. Figure 7.9 highlights the outdoor air sensor faulty measurements as recorded by the BAS, when compared with the STM. The plot of hourly error (Figure 7.10) revealed that during two time intervals, one between 8:00 and 10:00 (window #1) and one between 13:00 and 20:00 (window #2), the average hourly error eh exceeded the uncertainty limits. Statistical indices from comparison between faulty BAS measurements and STM along the training period are listed in Table 7.18. From the on-site inspection, it was found that the temperature sensor location has been the cause of the faulty measurements in the BAS: the thermistor was in contact with a metal deck, which is directly exposed to the external conditions on the North-West building facade. The thermal inertia of the metal deck affects the sensor records in the morning, between 8:00 and 10:00, when the solar radiation does not reach the North-West façade and the outdoor air temperature rises faster than the metal deck temperature; thus the air temperature values recorded by the BAS are lower than the reference STM values. Between 13:00 and 20:00, the metal deck is directly exposed to the solar radiation and overheats, affecting the sensor records; thus the air temperature values recorded by the BAS are higher than that values from STM measurements. Hence, two different regression models, one per each window, are proposed (eqs. 7.11 and 7.12) that predict the hourly error eh.

$$e_h = r_h \cdot RH_{oa} \qquad \text{eq. 7.11}$$

$$e_h = s_h \cdot I_s \qquad \text{eq. 7.12}$$

where r_h and s_h are the coefficients of prediction models; and I_s is the total solar radiation on a horizontal surface.

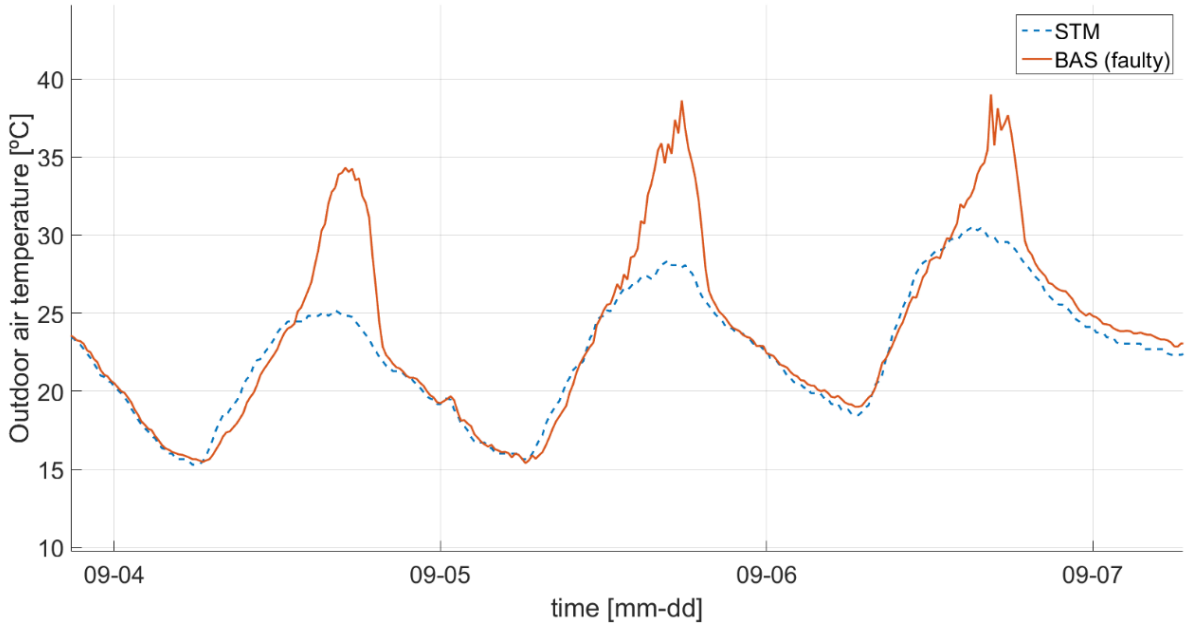


Figure 7.9 - Comparison of outdoor air temperature measurements from BAS (orange solid line) and the STM (blue dash line): September 4 to 7, 2015.

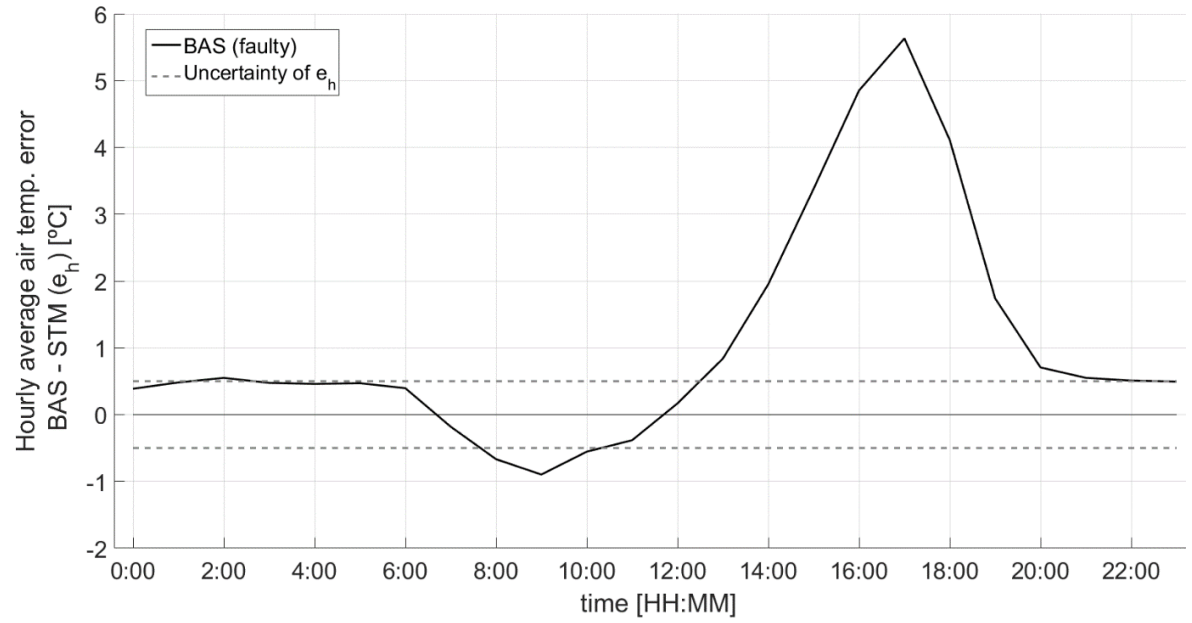


Figure 7.10 – Hourly average temperature error between BAS and STM over the time interval September 3 to 10, 2015.

Table 7.18 – Comparison between faulty BAS measurements and STM values for both windows from both training and validation data sets.

Training period data set (September 3-10, 2015)			
Window	AE_{max}	MAE	CV-RMSE
#1	1.82°C	0.9°C	4.4%
#2	10.69°C	2.9°C	13.3%
Validation period data set (September 11-19, 2015)			
Window	AE_{max}	MAE	CV-RMSE
#1	3.05°C	1.5°C	9.1%
#2	6.55°C	1.8°C	9.7%

The STM from seven days period (September 3-10, 2015) are used for the training of the correction models aimed to correct outdoor air temperature faulty measurements (Table 7.19). The outdoor air temperature corrected with Method C is displayed in green in Figure 7.11.

Table 7.19 – Model C trained with the STM for one-week (September 3-10, 2015).

Equation	Hourly bin (h)	Trained models	Model prediction uncertainty
7.11	*8:00 – 9:00*	$e_h = -0.0092 \cdot RH_{oa}$	±0.04°C
	9:00 – 10:00	$e_h = -0.0141 \cdot RH_{oa}$	±0.06°C
7.12	*13:00 – 14:00*	$e_h = 3.148 \cdot 10^{-4} \cdot I_s$	±0.05°C
	14:00 – 15:00	$e_h = 8.352 \cdot 10^{-4} \cdot I_s$	±0.15°C
	15:00 – 16:00	$e_h = 17.324 \cdot 10^{-4} \cdot I_s$	±0.30°C
	16:00 – 17:00	$e_h = 31.771 \cdot 10^{-4} \cdot I_s$	±0.55°C
	17:00 – 18:00	$e_h = 51.123 \cdot 10^{-4} \cdot I_s$	±0.89°C
	18:00 – 19:00	$e_h = 81.121 \cdot 10^{-4} \cdot I_s$	±1.42°C
	19:00 – 20:00	$e_h = 110.57 \cdot 10^{-4} \cdot I_s$	±1.93°C

The uncertainty of the measurements of the outdoor air relative humidity (RH_{oa}) and solar radiation (I_s) propagate through the trained models (table 7.19) and affect the predicted values of e_h. According to section 3.5, the propagated uncertainty has been added to the initial uncertainty which affect the estimation of the term e_h: $U_{\Delta T}$ = 0.49°C (and reported in figures 7.12 and 7.13 as red dot lines). Figure 7.12 shows, for each hourly bin along the training period, the hourly average error e_h value between the faulty BAS measurements and STM (black solid line), and between BAS corrected measurements and STM (red dash line). The uncertainty of error e_h between BAS faulty measurements and STM (grey dash lines), and between BAS corrected values and STM (red dot lines) are showed as well. Figure 7.13 show the results from Method C along the validation period. The statistical indices of the error e_h are given in Table 7.20.

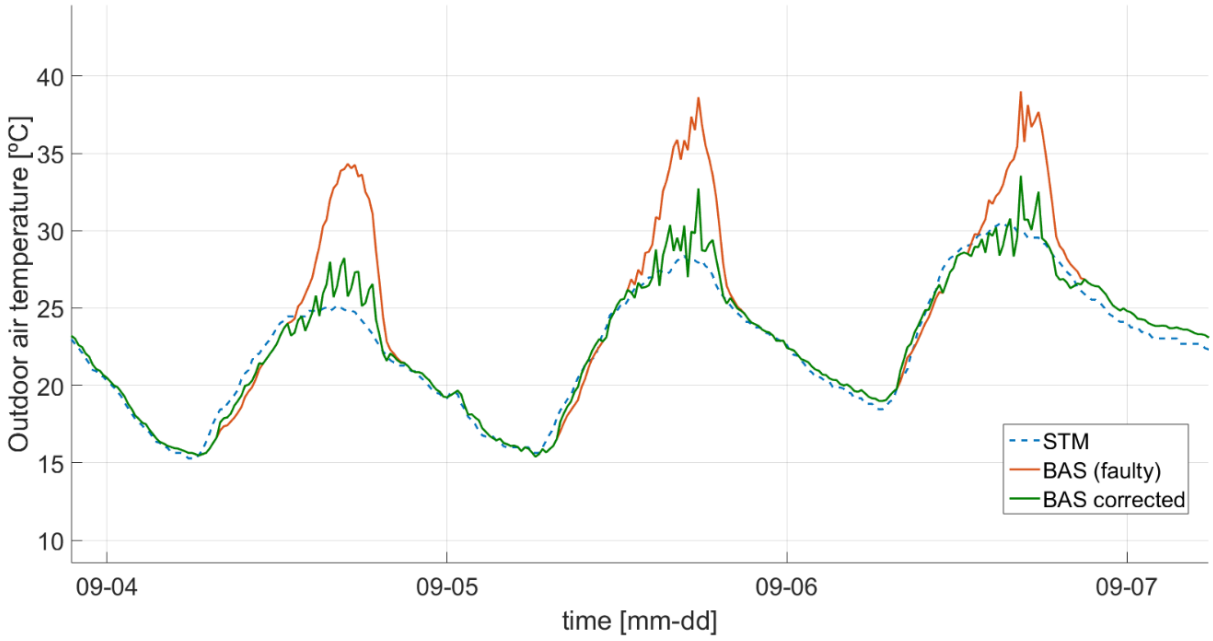


Figure 7.11 - Comparison of outdoor air temperature measurements from BAS (orange solid line), from BAS corrected values (green solid line), and from STM (blue dash line): September 4 to 7, 2015.

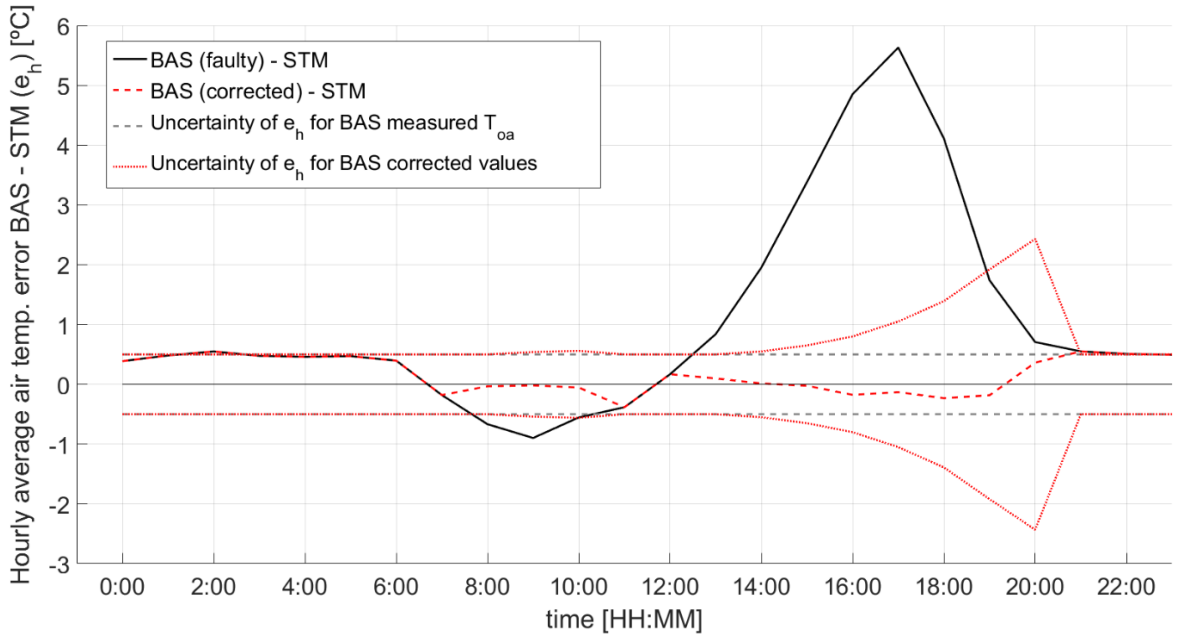


Figure 7.12 – Hourly average error between BAS measurements and STM, before (black solid line) and after (red dash line) the correction through Method C (Sept 3 – Sept 10, in 2015).

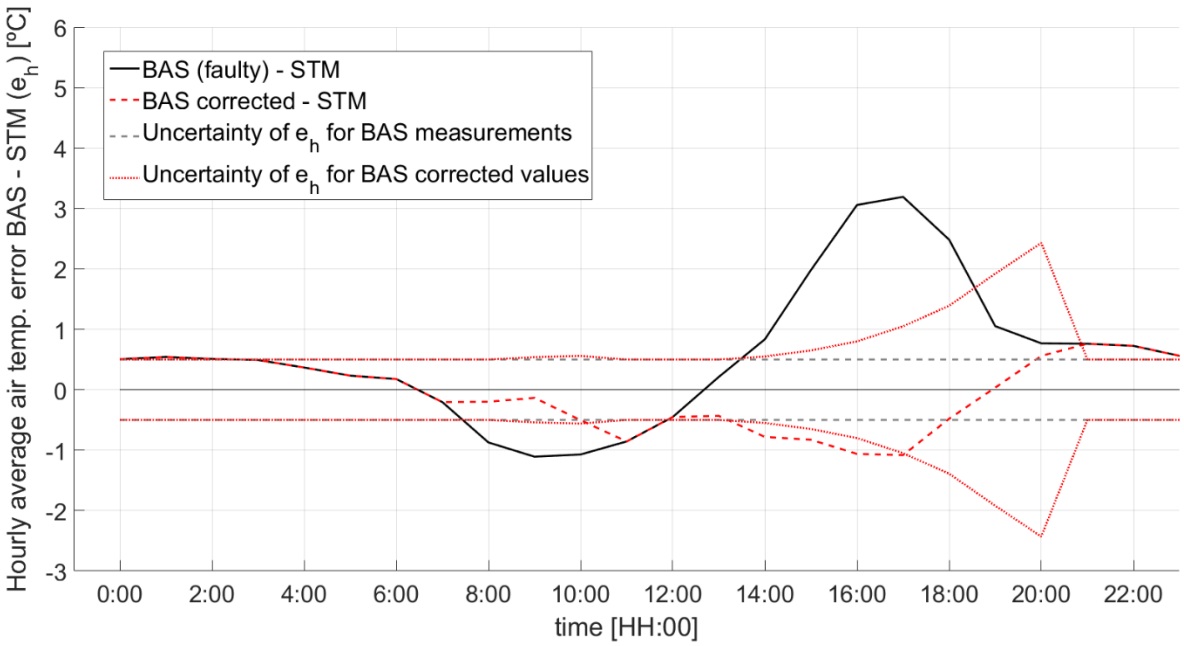


Figure 7.13 – Hourly average error between BAS measurements and STM, before (black solid line) and after (red dash line) the correction through Method C (Sept 11 – Sept 19, in 2015).

Table 7.20 – Statistical indices from comparison between corrected BAS measurements and STM along both the training and the validation periods.

Training period			
Window	AE_{max}	MAE	CV-RMSE
#1	1.41ºC	0.6ºC	2.9%
#2	4.81ºC	1.2ºC	5.6%
Validation period			
Window	AE_{max}	MAE	CV-RMSE
#1	2.63ºC	1.3ºC	7.6%
#2	3.30ºC	1.0ºC	5.9%

7.4 Discussion

Method A uses a first principle based model to self-correct the measurements from a faulty air properties sensor at the AHU economizer. It does not require the model training. If a recovery loop pre-heat the outdoor air stream (operation mode #3), the air properties after the pre-heating coil should be measured. When Method A was applied to correct faulty values of the outdoor air temperature, all the statistical indices from the comparison against the reference STM remarkably improved (Table 7.1). The MAE was lower than the residual overall uncertainty (0.2°C against 0.49°C). A similar effectiveness has been observed when Method A was used to correct faulty values from other four air properties (Tables 7.2, 7.3, 7.4 and 7.6). When Method A was applied to correct faulty values of the return air temperature (Table 7.5), the average observed absolute residual (MAE) was twice the corresponding uncertainty, meaning that the observed residuals along the 24 hours data set cannot be entirely justified by measurements uncertainty, but are partially due to the method accuracy.

Method B consists of developing linear regression models to predict the air properties, for different AHU economizer operation modes. Model B1 (eqs. 7.1 and 7.2) and Model B2 (eqs. 7.3,

7.4, 7.5 and 7.6) proved high prediction effectiveness (Tables 7.7, 7.9, 7.10, 7.11 and 7.12). The MAE was lower than the corresponding difference uncertainty for both air temperature and relative humidity. Under operation mode #3, from comparison between predictions and STM the maximum and average absolute error is larger than the difference uncertainty: AE_{max} = 5.02°C and MAE = 0.8°C against ±0.42°C uncertainty for the outdoor air temperature, and AE_{max} = 24.9% and MAE = 8.1% against ±4.29% uncertainty for the outdoor air relative humidity. Similar indices were obtained for the virtual prediction of the mixed air properties (eqs. 7.9 and 7.10). Simple linear regression models with two regressors were not able to properly model the heat transfer between coils SC2-1 and SC2-2 and the outdoor air stream. Non-linear models should be considered for further development of Method B under operation mode #3. Method B requires only one or two regressors, compared to Method A (five or six regressors).

The proposed methods are intended to replace physical sensors for the measurement of the outdoor air properties. The uncertainty of typical physical sensors available on the market are used as reference against the prediction uncertainty from the proposed models. In this case study the outdoor air physical sensors have an uncertainty of ±0.49°C for the air temperature, and ±4.1% for the air relative humidity. The prediction uncertainties of models from Method A and Method B are of the same magnitude of their reference values, thus are considered to be acceptable. Method A has a higher computational cost, as it uses an iterative process at each time step. On the other hand, Method B require better thermodynamic and modelling skills for the development of the prediction models.

Method C is applied to correct measurements of an outdoor air temperature sensor, found to be faulty. The sensor error was modelled using the variables which cause the fault itself (e.g. the solar radiation). Technical and analytical skills are required. STM of the faulty variable allowed

for properly modelling the sensor faulty behavior. Linear regression models were developed which predict a correction term for each hourly interval within the two identified windows. The overall improvement of the measurements accuracy is remarkable along the training period (Figure 7.12). Along the validation period (September 11–19, 2015), the improvement of the measurements accuracy is less marked (Figure 7.13). The lower effectiveness along the validation period can be explained by the different hourly error profiles between training (Figure 7.12) and validation period (Figure 7.13). To reduce the impact of weather variation, different weather day types should be considered. Dedicated models would be developed for each considered weather day type, and used when the correspondent weather day type occurs.

8. A NEW DEFINITION OF VIRTUAL SENSOR

In this section a new formulation of the definition of *virtual sensor* is proposed, limited to the case of building applications. The need of making the subject clear is introduced in the literature review, section 2.4.1 ("The need for a new definition of *virtual sensor*"). The new definition intends to highlight key characteristics that every model should comply with in order to be named *virtual sensor* (or *virtual meter*). In process control the term *virtual sensor* is used for models which allow the measurement of process parameters "*where direct measurement is too expensive or even not possible"* (Dementjev et al. 2010). Similarly, within the field of HVAC monitoring, the additional cost due to purchasing a new physical sensor, and practical issues related to physical sensors installation, maintenance and accuracy, appear to be the two most recurrent reasons for which new virtual sensors are proposed:

"A virtual sensor uses low-cost measurements and mathematical models to estimate a difficult to measure or expensive quantity." (Li et al. 2011)

"…** measuring and monitoring SCFM* [supply air flow] *in rooftop air-conditioning units (RTUs) by using the conventional SCFM metering devices are very costly and more than often problematic.**" (Yu et al. 2011)

"Direct refrigerant pressure measuring practices are expensive and more than often problematic." (Li and Braun 2009a)

"Virtual sensors are usually used when the targeted monitoring or control value is not directly or only expensively measureable (e.g. hostile environments), or only measureable with large delays (e.g. dead-time processes)." (Ploennigs et al. 2011)

"*A model correcting a single-point mixed-air temperature measurement for inaccuracy caused by thermal stratification and non-uniform velocity distributions* […]" (Hjortland and Braun 2016)

"[…] *the non-intrusive virtual flow meter introduced in this paper provides a solution to one of the measurement barriers and challenges: a low cost, reliable energy metering system at the AHU level."* (Swamy et al. 2012)

Twenty-eight scientific papers have been reviewed (Table 8.1) that explicitly use the term *virtual sensors* (or *virtual meter*) to refer to models used for indirectly measuring variables of interest in HVAC. Twenty papers out of twenty-eight explicitly state that a virtual sensor allows to overcome practical issues, as sensor installation, maintenance, recalibration, and uncertainty. Common practical issues are, for instance, the air stratification in AHU ducts, which makes difficult to measure the air temperature with a single-point sensor (Lee and Dexter 2005 ; Wichman and Braun 2009), or the compact structure of HVAC installation which may not guarantee optimal locations for sensors installation (Mcdonald et al. 2014). Moreover, nineteen of the twenty-eight reviewed papers identified the additional cost of a new physical sensor as one of the main reasons to implement virtual sensors. Only four out of the twenty-eight reviewed publications do not mention the need to overcome the increase of costs and the rise of practical issues behind the use of virtual sensors. Those two identified key aspects (practical issues and additional costs) are here used to formulate a new definition of *virtual sensor*.

In HVAC monitoring, thus, virtual sensor applications are considered to be a valid alternative to physical sensors, which can be expensive and/or difficult to be implemented. It is essential here to remark that those physical sensors *exist*, and the reasons why a virtual sensor should be preferred to them, are strictly correlated to the features of the specific products currently present on the market: i) the cost, and ii) the installation and maintenance requirements. Thus, the new *virtual sensor* definition should include this aspect.

The proposed new definition of virtual sensor is the following:

A virtual sensor uses a mathematical model along with measurements from other installed sensors to derive the value of a variable of interest, which could be directly measured with a physical sensor available on the market if the cost and practical issues do not prevent the use of such a physical sensor.

Table 8.1 – List of reviewed papers which use the term virtual sensor (or virtual meter).

Reference	Additional cost	Practical issues	description of the practical issue
Alhashme and Ashgriz 2016			
Andiroglu et al. 2016		X	lack of optimal location for sensors installation
Fernandez et al. 2009			
Hjortland and Braun 2016		X	air stratification
Kusiak et al. 2010		X	sensor failure, improper installation, poor maintenance
Li and Braun 2007a	X		
Li and Braun 2009a	X	X	improper practice
Li and Braun 2009b	X	X	
Li et al. 2011	X	X	
McDonald and Zmeureanu 2014	X		
McDonald et al. 2014	X	X	lack of optimal location for sensors installation
Mishukov and Horyna, 2015	X	X	
Padilla et al. 2015			
Ploennings et al. 2011	X	X	
Reppa et al. 2014	X	X	invasive action required for sensor installation
Song et al. 2012	X	X	installation, accuracy

Song et al. 2013		X	uncertainty
Swamy et al. 2012	X	X	
Tahmasebi and Mahdavi 2013			
Wang et al. 2014	X	X	
Wichman and Braun 2009		X	Compact structure, air stratification
Yu et al. 2011a		X	
Yu et al. 2011b	X	X	
Yu et al. 2011c	X	X	
Yu et al. 2011d	X	X	
Zhao et al. 2012	X		
Zhao et al. 2012b	X	X	maintenance and calibration
Zmeureanu and Vanderbrook 2015	X		

According to the above proposed definition, a model which estimate the mixed air temperature using measurements of other variables would be said to be a *virtual sensor*, as physical sensors for the measurements of air temperature in AHU ducts are currently available on the market. The chiller COP formulation, as function of the thermal load at the evaporator and the electrical power input to the chiller (eq. 8.1), should not be named *virtual sensor*, as a physical sensor capable to *directly* measure the chiller COP does not exist.

$$\text{COP} = \frac{Q_{EV}}{E} \qquad \text{eq. 8.1}$$

where Q_{EV} is the thermal load at the chiller evaporator, and E is the chiller power input.

Zhao et al. (2012) proposed a method to diagnose fouling in chiller condensers using measurements of temperature and pressure of the refrigerant, and named it *virtual condenser fouling sensor*. If the new definition of virtual sensor given above is accepted, a method which use temperature and pressure measurements to diagnostic condenser fouling should not be named virtual sensor. On the other hands, because physical sensors for fluids temperature within pipes exist, the non-invasive model proposed by Gorman et al. (2013) to derive the fluid temperature using measurements of the pipe wall surface temperate, should be referred to as *virtual fluid temperature sensor*.

9. CONCLUSIONS

This research work proposed several new methods and models to enhance the practical feasibility of the HVAC systems ongoing commissioning. All the proposed methods and models only need for measurements commonly available from the BAS, while STM have been used for models development and validation. The goal was to enlarge the families of available methods to support energy managers and building operators in the non-stop commissioning of HVAC.

The central plant and an AHU installed at the Concordia University Loyola Campus in Montreal were used as case study in this research. Measurements from the BAS collected at 15 minutes time step, along with STM and spot measurements from portable calibrated sensors have been used for models development and validation.

The proposed methods and models include a FD&I method for chillers to benchmark the chiller fault-free operation and detect abnormal values along further observations. The chiller's benchmark consists of an ellipsoidal threshold model which identify a region, within a PCs based space, where the fault-free observations are likely to fall. The abnormal observations from seven summer seasons (from 2009 to 2015) have been effectively detected and identified. Moreover, when in 2016 the set-points of the chilled water supply temperature (T_{CHWS}) and the cooling tower supply temperature ($T_{CT\text{-}S}$) were modified, the proposed FD&I method detected the change in the chiller operation and properly identified the two variables whose set-point was changed.

One of the main objectives of this thesis was to enlarge the families of available virtual sensor for building application. Several virtual air flow meters have been presented and investigated which virtually measure the outdoor air flow rate into the AHU. When the mixing box inlets air conditions are known (heat recovery system is off), three models were proposed, along with the assessment of their prediction capability. The simplest formulation of the outdoor air ratio (factor α) was preferred as it leads to the smallest prediction uncertainty (±0.08). Two models were proposed for the case when the air conditions at the mixing box inlets are not measured (heat recovery system is on): model *a* and *b*, with prediction uncertainty ±0.42 and ±0.56 respectively. The bias error affecting the models prediction was estimated and removed, reducing the models overall uncertainty to ±0.3 and ±0.4. Although those uncertainty values would make the predictions of the factor α at each time step unreliable, the results show good agreement between models predictions and reference values (MAE = 0.015 and MAE = 0.016 for models *a* and *b*

respectively). Thus, the two proposed models should be used for the estimation of the daily average value of the factor α, and hence of the daily average outdoor air flow rate.

Three methods for the virtual measurement or the virtual calibration of faulty air properties sensors have been presented and validated. Methods A and B showed good prediction and self-correction capabilities when the heat recovery system was off: for both the methods the MAE was lower than the correspondent uncertainty along the validation period. When the heat recovery system was turned on (operation mode 3), the liner regression models from Method B (eqs. 7.7 to 7.10) were not capable to properly model the heat transfer across the heat recovery coils (SC2-1 and SC2-2): e.g. MAE = 0.8ºC against ±0.49ºC difference uncertainty, and MAE = 8.1% against ±4.36% difference uncertainty, for the outdoor air temperature and relative humidity respectively. Method C effectively corrected the outdoor air temperature faulty measurements along the training period. During the validation period its correction capability decreased as consequence of the changed average weather condition. Classification based on weather day types should be considered to enhance Method C effectiveness. Method A is based on an energy balance, and thus it does not require technical skills. Methods B and C need for a reduced number of measurements, however technical and modelling skills are required. For each case a method should be selected based on available information, technical knowledge, modelling skills and time.

A new definition of *virtual sensors* for building applications was proposed with the intent of contributing to the recent research interest in virtual sensors for building systems. The proposed new definition focuses on the variable of interest predicted or corrected by a virtual sensor model. A physical sensor should already be available on the market to measure such a variable of interest.

9.1 Contributions

The present research work leads to the following contributions:

1. Development of a new fault detection and identification method for chillers which uses measurements commonly available from BAS;

2. Development of new virtual flow meter models for the virtual measurement of the outdoor air flow rate into AHUs, under two different operation modes:
    a. The heat recovery system is off: the air temperatures at the mixing box inlets are known;
    b. The heat recovery system is on: the temperature of the air after the heat recovery coil and at recirculate conditions are not measured;

3. Development of new methods for the virtual measurement or virtual calibration of faulty sensors of the air temperature and relative humidity in AHUs under different operation modes:
    a. Method A: a first principle based method for virtual measurement or virtual calibration;
    b. Method B: data driven models for the virtual measurement of variables of interest;
    c. Method C: data driven models for virtual calibration of faulty sensors;

4. Formulation of a new definition of *virtual sensors* for building applications.

9.2 Recommendation for future work

The new methods and models proposed in this research work support building operators and energy managers in the ongoing commissioning of HVAC systems, providing new strategies to extract information from BAS measurements. In order to improve the proposed methods and models, and enlarge their potentiality, the following items are recommended for future work:

- Assess the FD&I method capability for HVAC equipment other than chillers (e.g. heat exchangers, fans, etc.);
- Assess the FD&I method capability for HVAC whole systems and sub-systems (e.g. mixing boxes, AHUs, cooling plants, etc.);
- Test the FD&I method on other chiller case studies;
- Investigate the implementation of density-based clustering techniques within the ellipsoidal threshold to enhance the PCA-based method effectiveness to detect abnormal observations;
- Evaluate the effectiveness of the proposed VFM models *a* and *b* for AHU under pre-cooling mode;
- Method B3: test other data-driven models to properly model the non-linear correlation between the heat transfer at the recovery coil and the outdoor air temperature;
- Method C: include weather day types based clustering to enhance the method's virtual calibration effectiveness;
- Integrate the proposed PCA-based FD&I method and the virtual sensor models and methods in order to provide a complete fault detection, identification and correction method for HVAC equipment.

References

Abdelalim, A., W. O'Brien, and Z. Shi. 2017. "Development of Sankey Diagrams to Visualize Real HVAC Performance." *Energy and Buildings*.

Abdi, H. and L. J. Williams. 2010. "Principal Component Analysis." Pp. 433–59 in *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2.

ACR System Inc. 2012. URL: www.acrsystems.com/product/smartreader-plus-2. Last access: 11/15/2017.

Alhashme, M. and N. Ashgriz. 2016. "A Virtual Thermostat for Local Temperature Control." *Energy and Buildings* 126:323–39. (http://dx.doi.org/10.1016/j.enbuild.2016.05.045).

Andiroglu, E., G. Wang, L. Song, and K. Kiamehr. 2016. "Development of a Virtual Pump Water Flow Meter Using Power Derived from Comprehensive Energy Loss Analysis." *Science and Technology for the Built Environment* 22(2):214–26.

ASHRAE. 2005a. *ASHARE Guideline 0-2005 - The Commissioning Process*. ASHRAE.

ASHRAE. 2005b. *Guideline 2-2005 - Engineering Analysis of Experimental Data*. edited by I. American Society of Heating, Refrigerating and Air-Conditioning Engineers. Atlanta.

ASHRAE. 2009. *Handbook Foundamental 2009*.

Brambley, M., R. Pratt, D. Chassin, S. Katipamula, and D. Hatley. 1998. "Diagnostics for Outdoor Air Ventilation and Economizers." *ASHRAE Journal* 40(10):49–55.

Brambley, M., N. Fernandez, W. Wang, K. Cort, H. Cho, H. Ngo and J. K. Goddard. 2011. *Final Project Report : Self-Correcting Controls for VAV System Faults Filter / Fan / Coil and VAV Box Sections*.PNNL-20452.

Bynum, J. D., D. E. Claridge, and J. M. Curtin. 2012. "Development and Testing of an Automated Building Commissioning Analysis Tool (ABCAT)." *Energy and Buildings* 55:607–17. (http://dx.doi.org/10.1016/j.enbuild.2012.08.038).

Cadima, J. and I. T. Jolliffe. 1995. "Loading and Correlations in the Interpretation of Principle Compenents." *Journal of Applied Statistics* 22(2):203–14.

Castro, N. S. and H. Vaezi-Nejad. 2005. "CITE-AHU, an Automated Commissioning Tool for Air-Handling Units." Pp. 1–12 in *National Conference on Building Commissioning*. (http://www.bcxa.org/ncbc/2005/proceedings/TM_04_Castro_NCBC2005.pdf).

Choinière, D. 2008. "DABO TM : A BEMS ASSISTED ON-GOING COMMISSIONING

TOOL." in *National Conference on Building Commissioning*.

Chung, W., Y. V. Hui, and Y. Miu Lam. 2006. "Benchmarking the Energy Efficiency of Commercial Buildings." *Applied Energy* 83(1):1–14.

Claridge, D. E., C. H. Culp, M. Liu, S. Deng, W. D. Turner, and J. S. Haberl. 2000. "Campus-Wide Continuous Conunissioning of University Buildings." *Buildings. ACEEE 2000 Summer Study on Energy Efficiency in Buildings* 101–12.

Claridge, D. E., W. D. Turner, M. Liu, S. Deng, G. Wei, C. H. Culp, H. Chen and S. Cho. 2004. "Is Commissioning Once Enough?" *Energy Engineering* 101(4):7–19.

Comstock, M. C., J. E. Braun, and E. A. Groll. 2001. "The Sensitivity of Chiller Performance to Common Faults." *HVAC&R Research* 7(3):263–79.

Cortes, C. and V. Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 20(3):273–97.

Cotrufo, N. and R. Zmeureanu. 2015. A principal component analysis-based approach for the ongoing commissioning of centrifugal chillers. *CISBAT 2015, Lausanne, CH.*

Cotrufo, N. and R. Zmeureanu. 2016. PCA-based method od soft fault detection and identification for the ongoing commissioning of chillers. *Energy and Buildings, 130*, 443-453.

Cotrufo, N., L. Natale and R. Zmeureanu. 2016. Virtual outdoor air flow meter for the ongoing commissioning of HVAC systems: lessons from a case study. *ASHRAE 2016 Winter Conference, Orlando, FL.*

Cotrufo, N. and R. Zmeureanu. 2016. A New Algorithm for Sensors Verification and Correction in Air Handling Units. *eSim 2016 Conference, Hamilton, ON.*

Crawford, R. R., R. G. Dykowsky, and S. E. Czajkowski. 1991. "A Segmented Linear Least Squares Modeling Procedure for Nonlinear HVAC Components." 11–18 in *ASHRAE Transaction*.

Cui, J. and S. Wang. 2005. "A Model-Based Online Fault Detection and Diagnosis Strategy for Centrifugal Chiller Systems." *International Journal of Thermal Science* 44:986–99.

Deltacontrols, n.d. URL: www.deltacontrols.com/products/facilities-management/supervisory-software/enteliweb. Last access: 11/15/2017.

Dementjev, A., H. D. Ribbecke, and K. Kabitzsch. 2010. "Improvement of Prediction Performance for Data-Driven Virtual Sensors." in *15th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 2010*.

Dong, B., C. Cao, and S. E. Lee. 2005. “Applying Support Vector Machines to Predict Building Energy Consumption in Tropical Region.” *Energy and Buildings* 37(5):545–53.

Draper, N. R. and H. Smith. 1981. *Applied Regression Analysis*. 2nd ed. edited by J. & S. Wiley. New York: Inc.

Du, Z. and X. Jin. 2007. “Detection and Diagnosis for Multiple Faults in VAV Systems.” *Energy and Buildings* 39(8):923–34.

Dunia, R., S. J. Qin, T. F. Edgar, and T. J. Mcavoy. 1996. “Identification of Faulty Sensors Using Principal Component Analysis.” *Process Systems Engineering* 42(10):2707–2812.

Efficiency Valuation Organization. 2007. *International Performance Measurement and Verification Protocol Concepts and Options for Energy and Water Savings*.(www.evo-world.org.).

FacilityDynamics Engineering, n.d.. URL: www.facilitydynamics.com/software/pacrat/. Last access: 11/15/2017.

Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth. 1996. “From Data Mining to Knowledge Discovery in Databases.” *American Association for Artificial Intelligence* 17(3):37–54. (http://www.aaai.org/aitopics/assets/PDF/AIMag17-03-2-article.pdf).

Fels, M. F. 1986. “PRISM: An Introduction.” *Energy and Buildings* 9(1-2):5–18.

Fernandez, N., M. Brambley, and S. Katipamula. 2009. *Self-Correcting HVAC Controls: Algorithms for Sensors and Dampers in Air-Handling Units*. PNNL-19104.

Gorman, J. M., E. M. Sparrow, and J. P. Abraham. 2013. “Differences between Measured Pipe Wall Surface Temperatures and Internal Fluid Temperatures.” *Case Studies in Thermal Engineering* 1(1):13–16.

Granderson, J. and G. Lin. 2016. “Building Energy Information Systems: Synthesis of Costs, Savings, and Best-Practice Uses.” *Energy Efficiency* 9(6):1369–84.

Greyline instruments inc. n.d. URL: www.greyline.com/index.php/component/virtuemart/irrigation-instruments/portaflow-pt400-portable-transit-time-flow-meter-detail?Itemid=0. Last access: 11/15/2017.

Hadley, D. L. 1993. “Daily Variations in HVAC System Electrical Energy Consumption in Response to Different Weather Conditions.” *Energy and Buildings* 19(3):235–47. (http://www.sciencedirect.com/science/article/pii/037877889390031O).

Hadley, D. L. and S. D. Tomich. 1986. “Multivariate Statistical Assessment of Meteorological

Influences on Residential Space Heating." Pp. 9.132–39.145 in *Proceedings of the ACEEE*.

Haves, P., T. Salsbury, D. Claridge, and M. Liu. 2001. *Use of Whole Building Simulation in on-Line Performance Assessment: Modeling and Implementation Issues*. (http://escholarship.org/uc/item/4g7923xn.pdf).

Haves, P. 1999. "Overview of Diagnostic Methods." in *Proceedings of the workshop on diagnostics for commercial buildings: From research to practice*.

Hjortland, A. L. and J. E. Braun. 2016. "Virtual Sensors for RTU Air-Side Diagnostics." *Science and Technology for the Built Environment* 22(2):189–200.

Hong, W. C. 2009. "Hybrid Evolutionary Algorithms in a SVR-Based Electric Load Forecasting Model." *International Journal of Electrical Power & Energy Systems* 31(7):409–17.

IEA. 2010. *Annex 47 Report 3: Commissioning Cost-Benefit and Persistence of Savings*.

IEA. 2016. *Key World Energy Statistics*. (http://www.oecd-ilibrary.org/energy/key-world-energy-statistics-2009_9789264039537-en).

Jackson, J. E. 1991. *A User's Guide to Principal Components*. 1st ed. New York: John Wiley & Sons, Inc. (http://doi.wiley.com/10.1002/0471725331).

Jia, Y. and T. A. Reddy. 2003. "Characteristic Physical Parameter Approach to Modeling Chillers Suitable for Fault Detection, Diagnosis, and Evaluation." *Journal of Solar Energy Engineering* 125(3):258–65.

Jolliffe, I. T. 1986. *Principal Component Analysis*. Springer. edited by J. & S. Wiley. New York.

Katipamula, S. and M. Brambley. 2005. "Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems — A Review, Part I." *HVAC&R Research2* 11(1):3–25.

Katipamula, S. and M. Brambley. 2005. "Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems — A Review, Part II." *HVAC&R Research* 11(2):169–87.

Katipamula, S., T. A. Reddy, and D. E. Claridge. 1998. "Multivariate Regression Modeling." *Journal of Solar Energy Engineering* 120:177–84.

Kavousi-Fard, A. and F. Kavousi-Fard. 2013. "A New Hybrid Correction Method for Short-Term Load Forecasting Based on ARIMA, SVR and CSA." *Journal of Experimental & Theoretical Artificial Intelligence* 25(4):559–74.

Krarti, M., J. F. Kreider, D. Cohen, and P. Curtis. 1998. "Estimation of Energy Savings for Building Retrofits Using Neural Networks." *Journal of Solar Energy Engineering* 120: 211–16.

Kusiak, A. and M. Li. 2010. “Cooling Output Optimization of an Air Handling Unit.” *Applied Energy* 87(3):901–9. (http://dx.doi.org/10.1016/j.apenergy.2009.06.010).

Kusiak, A., M. Li, and H. Zheng. 2010. “Virtual Models of Indoor-Air-Quality Sensors.” *Applied Energy* 87(6):2087–94. (http://dx.doi.org/10.1016/j.apenergy.2009.12.008).

Kusiak, A. and G. Xu. 2012. “Modeling and Optimization of HVAC Systems Using a Dynamic Neural Network.” *Energy* 42(1):241–50.

Lam, J. C., K. K. W. Wan, K. L. Cheung, and L. Yang. 2008. “Principal Component Analysis of Electricity Use in Office Buildings.” *Energy and Buildings* 40(5):828–36.

Le Cam, M., A. Daoud, and R. Zmeureanu. 2016. “Forecasting Electric Demand of Supply Fan Using Data Mining Techniques.” *Energy* 101:541–57. (http://dx.doi.org/10.1016/j.energy.2016.02.061).

Le Cam, M., R. Zmeureanu, and A. Daoud. 2017. “Cascade-Based Short-Term Forecasting Method of the Electric Demand of HVAC System.” *Energy* 119:1098–1107. (http://dx.doi.org/10.1016/j.energy.2016.11.064).

Le Cam, M. 2016. “Short-Term Forecasting of the Electric Demand of HVAC Systems.” Ph.D. Thesis, BCEE, Concordia University, Montreal.

Lee, P. S. and A. L. Dexter. 2005. “A Fuzzy Sensor for Measuring the Mixed Air Temperature in Air-Handling Units.” *Measurement* 37(1):83–93.

Lee, S. E. and P. Rajagopalan. 2008. “Building Energy Efficiency Labeling Programme in Singapore.” *Energy Policy* 36(10):3982–92.

Lennox, J. and C. Rosen. 2002. “Adaptive Multiscale Principal Components Analysis for Online Monitoring of Wastewater Treatment.” *Water Science & Technology* 45(4-5):227–35. (http://www.ncbi.nlm.nih.gov/pubmed/11936638).

Li, H. and J. E. Braun. 2007. “Decoupling Features and Virtual Sensors for Diagnosis of Faults in Vapor Compression Air Conditioners.” *International Journal of Refrigeration* 30(3):546–64.

Li, H. and J. E. Braun. 2009. “Virtual Refrigerant Pressure Sensors for Use in Monitoring and Fault Diagnosis of Vapor-Compression Equipment.” *HVAC&R Research* 15(3):597–616.

Li, H. and J. E. Braun. 2009b. “Development , Evaluation , and Demonstration of a Virtual Refrigernt Charge Sensor.” *HVAC&R Research* 15(1):117–36.

Li, H., D. Yu, and J. E. Braun. 2011. “A Review of Virtual Sensing Technology and Application in Building Systems.” *HVAC&R Research* 17(5):619–45.

Li, Q., Q. Meng, J. Cai, H. Yoshino, and A. Mochida. 2009a. "Applying Support Vector Machine to Predict Hourly Cooling Load in the Building." *Applied Energy* 86(10):2249–56.

Li, Q., Q. Meng, J. Cai, H. Yoshino, and A. Mochida. 2009b. "Predicting Hourly Cooling Load in the Building: A Comparison of Support Vector Machine and Different Artificial Neural Networks." *Energy Conversion and Management* 50(1):90–96.

Li, S. and J. Wen. 2014. "A Model-Based Fault Detection and Diagnostic Methodology Based on PCA Method and Wavelet Transform." *Energy and Buildings* 68:63–71. (http://linkinghub.elsevier.com/retrieve/pii/S0378778813005410).

Liu, M. 1999. "Improving Building Energy System Performance by Continuous Commissioning." *Energy Engineering* 96(5):46–57. (http://www.tandfonline.com/doi/abs/10.1080/01998595.1999.10530472).

Luo, R., M. Misra, and D. M. Himmelblau. 1999. "Sensor Fault Detection via Multiscale Analysis and Dynamic PCA." *Industrial & Engineering Chemistry Research* 38(4):1489–95. (http://pubs.acs.org/doi/abs/10.1021/ie980557b).

Lutes, R. G, S. Katipamula, B. A. Akyol, N. D. Tenney, J. N. Haack, K. E. Monson and B. J. Carpenter. 2014. *VOLTTRON : User Guide*.

Mathworks. 2017. URL: www.mathworks.com/help/stats/biplot.html. Last access: 11/15/2017.

Mcdonald, E.. 2014. "Development and Testing of a Virtual Flow Meter for Use in Ongoing Commissioning of Commercial and Institutional Buildings." MASc Thesis, BCEE, Concordia University, Montreal.

Mcdonald, E. and R. Zmeureanu. 2014. "Virtual Flow Meter to Estimate the Water Flow Rates in Chillers." Pp. 200–209 in *ASHRAE Transaction*, vol. 120.

Mcdonald, E., R. Zmeureanu, and D. Giguère. 2014. "Virtual Flow Meter for Chilled Water Loops in Existing Buildings." in *In Conference eSim*.

Mihai, A. 2014. "Calibration of a Building Energy Model Using Measured Data for a Research Center." MASc Thesis, BCEE, Concordia University, Montreal, QC.

Monfet, D. and R. Zmeureanu. 2011. "Ongoing Commissioning Approach for a Central Cooling and Heating Plant." *ASHRAE Transactions* 117(1):908–24.

Monfet, D. and R. Zmeureanu. 2012. "Ongoing Commissioning of Water-Cooled Electric Chillers Using Benchmarking Models." *Applied Energy* 92:99–108. (http://dx.doi.org/10.1016/j.apenergy.2011.10.019).

Morrison, P. and F. Donald. 1976. *Multivariate Statistical Methods*. 2nd ed. edited by McGraw-Hill. New York.

Natural Resources Canada. 2016. *Energy Efficiency Trends in Canada 1990 to 2013*.

Nie, H., G. Liu, X. Liu, and Y. Wang. 2012. "Hybrid of ARIMA and SVMs for Short-Term Lod Forecasting." *Energy Procedia* 16(C):1455–60.

Padilla, M., D. Choinière, and J. A. Candanedo. 2015. "A Model-Based Strategy for Self-Correction of Sensor Faults in Variable Air Volume Air Handling Units." *Science and Technology for the Built Environment* 21(7):1018–32.

Ploennigs, J., A. Ahmed, B. Hensel, P. Stack, and K. Menzel. 2011. "Virtual Sensors for Estimation of Energy Consumption and Thermal Comfort in Buildings with Underfloor Heating." *Advanced Engineering Informatics* 25(4):688–98.

Pranatyasto, T. N. and S. J. Qin. 2001. "Sensor Validation and Process Fault Diagnosis for FCC Units under MPC Feedback." *Control Engineering Practice* 9(8):877–88.

Reddy, T. A.. 2011. *Applied Data Analysis and Modeling for Energy Engineers and Scientists*. Springer S. (http://link.springer.com/10.1007/978-1-4419-9613-8).

Reddy, T. A., D. Niebur, K. Andersen, P. Pericolo, and G. Cabrera. 2003. "Evaluation of the Suitability of Different Chiller Performance Models for On-Line Training Applied to Automated Fault Detection and Diagnosis (RP-1139)." *HVAC&R Research* 9(4):385–414.

Reddy, T. A. and D. Claridge. 1994. "Using Synthetic Data to Evaluate Multiple Regression and Principal Component Analyses for Statistical Modeling of Daily Building Energy Consumption." *Energy and Buildings* 21(1):35–44. (http://linkinghub.elsevier.com/retrieve/pii/0378778894900140).

Roth, K., D. Westphalen, and J. Brodrick. 2008. "Ongoing Commissioning." *ASHRAE Journal* 50(3):66–70.

Santos, J. and L. Brightbill. 2000. "Automated Diagnostics from DDC Data - PACRAT." in *8th National Conference on Building Commissioning*. Kansas City, MO.

Schein, J. and S. T. Bushby. 2005. "Fault Detection & Diagnostics for AHU and VAV Boxes." *ASHRAE Journal* 47(7):58–63.

Shadpour, F. and J. Kilcoyne. 2015. "Criteria for Building Automation Dashboards." *ASHRAE Journal* 57(5):28–36.

Song, L., I. Joo, and G. Wang. 2012. "Uncertainty Analysis of a Virtual Water Flow Measurement

in Building Energy Consumption Monitoring." *HVAC&R Research* 18(5):997–1010.

Swamy, A., L. Song, and G. Wang. 2012. "A Virtual Chilled-Water Flow Meter Development at Air Handling Unit Level." Pp. 1013–20 in *ASHRAE Transactions*, vol. 118.

Tan, H. and A. Dexter. 2006. "Estimating Airflow Rates in Air-Handling Units from Actuator Control Signals." *Building and Environment* 41(10):1291–98.

Teeter, J. and M. Y. Chow. 1998. "Application of Functional Link Neural Network to Hvac Thermal Dynamic System Identification." Pp. 170–76 in *IEEE Transactions on Industrial Electronics*, vol. 45.

Thompson, J. and B. Moore. 2011. "Re-/retro-Commissioning: The Best Kept Secret You Can't Afford Not to Know." *Strategic Planning for Energy and the Environment* 31(2):49–68.

U.S. Department of Energy. 2012. *Buildings Energy Databook 2011*. edited by J. D. Kelso and D. I. Ltd. Silver Spring, Mariland. (http://buildingsdatabook.eren.doe.gov/DataBooks.aspx).

Wang, G., Li Song, E. Andiroglu, and G. Shim. 2014. "Investigations on a Virtual Airflow Meter Using Projected Motor and Fan Efficiencies." *HVAC&R Research* 20(2):178–87.

Wang, S. and J. Cui. 2005. "Sensor-Fault Detection, Diagnosis and Estimation for Centrifugal Chiller Systems Using Principal-Component Analysis Method." *Applied Energy* 82(3):197–213.

Wang, S. and F. Xiao. 2004a. "AHU Sensor Fault Diagnosis Using Principal Component Analysis Method." *Energy and Buildings* 36(2):147–60.

Wang, S. and F. Xiao. 2004b. "Detection and Diagnosis of AHU Sensor Faults Using Principal Component Analysis Method." *Energy Conversion and Management* 45(17):2667–86. (http://linkinghub.elsevier.com/retrieve/pii/S019689040300373X).

Wichman, A. and J. E. Braun. 2009. "A Smart Mixed-Air Temperature Sensor A Smart Mixed-Air Temperature Sensor." *HVAC&R Research* 15(1):101–15.

Xu, X., F. Xiao, and S. Wang. 2008. "Enhanced Chiller Sensor Fault Detection, Diagnosis and Estimation Using Wavelet Analysis and Principal Component Analysis Methods." *Applied Thermal Engineering* 28(2-3):226–37.

Yan, B., X. Li, A. M. Malkawi, and G. Augenbroe. 2017. "Quantifying Uncertainty in Outdoor Air Flow Control and Its Impacts on Building Performance Simulation and Fault Detection." *Energy and Buildings* 134:115–28. (http://dx.doi.org/10.1016/j.enbuild.2016.10.028).

Yan, B., X. Li, W. Shi, X. Zhang, and A. Malkawi. 2017. "Forecasting Building Energy Demand

under Uncertainty Using Gaussian Process Regression : Feature Selection , Baseline Prediction , Parametric Analysis and a Web-Based Tool." in *International Building Performance Simulation Association*. San Francisco.

Yokoyama, R., T. Wakui, and R. Satake. 2009. "Prediction of Energy Demands Using Neural Network with Model Identification by Global Optimization." *Energy Conversion and Management* 50(2):319–27. (http://dx.doi.org/10.1016/j.enconman.2008.09.017).

Yu, D. H., Haorong Li, and Yuebin Yu. 2011. "A Gray-Box Based Virtual SCFM Meter in Rooftop Air-Conditioning Units." *Journal of Thermal Science and Engineering Applications* 3(1):011005.

Yu, D. H., H. R. Li, L. Ni, and Y. S. Yu. 2011a. "An Improved Virtual Calibration of a Supply Air Temperature Sensor in Rooftop Air Conditioning Units." *HVAC&R Research* 17(5):798–812.

Yu, D. H., H. R. Li, L. Ni, and Y. S. Yu. 2011b. "Virtual Calibration of a Supply Air Temperature Sensor in Rooftop Air Conditioning Units." *HVAC&R Research* 17(1):31–50.

Zhao, X., M. Yang, and H. Li. 2012. "A Virtual Condenser Fouling Sensor for Chillers." *Energy and Buildings* 52:68–76.

Zibin, N.. 2014. "A Bottom-up Method to Calibrate Building Energy Models Using Building Automation System (BAS) Trend Data." MASc Thesis, BCEE, Concordia University, Montreal.

Zibin, N., R. Zmeureanu, and J. Love. 2016. "Automatic Assisted Calibration Tool for Coupling Building Automation System Trend Data with Commissioning." *Automation in Construction* 61:124–33. (http://dx.doi.org/10.1016/j.autcon.2015.10.004).

Zmeureanu, R. 2002. "Prediction of the COP of Existing Rooftop Units Using Artificial Neural Network and Minimum Number of Sensors." *Energy* 27(9):889–904.

Appendix A

The set of steps and equations to be implemented at each iteration of the correction process from Method A are here given for the correction of (1) the outdoor air relative humidity, (2) the mixed air temperature, (3) the mixed air relative humidity, (4) the recirculated air temperature, and (5) the recirculated air relative humidity.

A1 Outdoor Air Relative Humidity

A vector of candidate correction terms for the outdoor air relative humidity is defined.

$$dRH = [dRH_1 , dRH_2 , \dots , dRH_{j-1} , dRH_j] \quad \text{eq. A1.1}$$

For each *j* term of vector *dRH*, the iterative process includes the following steps:

a. the *j* correction term from vector *dRH* is added (eq. A1.2) to the faulty outdoor air relative humidity, RH_{oa}, giving the *j* modified faulty variable (RH_{oa}^{j});

$$RH_{oa}^{j} = RH_{oa} + dRH_j \quad \text{eq. A1.2}$$

b. the *j* modified faulty variable (RH_{oa}^{j}) is used, along with the five remaining fault-free variables, to evaluated a *j* modified outdoor air humidity ratio x_{oa}^{j} (eq. A1.3) , a *j* modified outdoor air specific enthalpy value h_{oa}^{j} (eq. A1.4), and a *j* α-factor, α_j (eq. A1.5);

$$x_{oa}^{j} = 0.62198 \cdot \frac{\mathrm{Pv}_{oa}^{j}}{P - Ps_{oa}} \quad \text{eq. A1.3}$$

$$h_{oa}^{j} = C_p \cdot T_{oa} + x_{oa}^{j} \cdot \left(h_{fg} + C_v \cdot T_{oa}\right) \quad \text{eq. A1.4}$$

$$\alpha_j = \frac{h_{ma} - h_{rec}}{h_{oa}^{j} - h_{rec}} \quad \text{eq. A1.5}$$

where $\mathrm{Pv}_{oa}^{j} = f(RH_{oa}^{j})$, and $\mathrm{Ps}_{oa} = f(T_{oa})$ are the partial pressure and the saturation pressure of water vapor respectively.

c. a value of the *j* modified faulty air humidity ratio is derived from the correspondent α_j ($x_{oa,\alpha}^{j}$) (eq. A1.6) and used to estimate the *j* modified air relative humidity derived from α_j ($RH_{oa,\alpha}^{j}$) (eq. A1.7);

$$x_{oa,\alpha}^{j} = \frac{(x_{ma} - x_{rec})}{\alpha_j} + x_{rec} \quad \text{eq. A1.6}$$

$$RH_{oa,\alpha}^{j} = \frac{\mathrm{Pv}_{oa,\alpha}^{j}}{\mathrm{Ps}_{oa}} \quad \text{eq. A1.7}$$

where $\mathrm{Pv}_{oa,\alpha}^{j} = f(x_{oa,\alpha}^{j})$, and $\mathrm{Ps}_{oa} = f(T_{oa})$ are the partial pressure and the saturation pressure of water vapor respectively.

d. the *j* term $\Delta RH_{BAS,\alpha}^{j}$ is evaluate (eq. A1.8) as the absolute value of the difference between RH_{oa}^{j} and $RH_{oa,\alpha}^{j}$.

$$\Delta RH_{BAS,\alpha}^{j} = \left| RH_{oa}^{j} - RH_{oa,\alpha}^{j} \right| \qquad \text{eq. A1.8}$$

The optimal correction term dRH^{*} from the candidate correction terms vector *dRH* is identified as the one which corresponds to the minimum $\Delta RH_{BAS,\alpha}^{j}$ (eq. A1.9), and used to correct faulty measurements of the of the outdoor air relative humidity (eq. A1.10) before being used by the BAS for control purpose.

$$RH_{oa}^{*} = RH_{oa}^{j} \Longrightarrow \min(\Delta RH_{BAS,\alpha}^{j}) \qquad \text{eq. A1.9}$$

$$RH_{oa}^{*} = RH_{oa} + dRH^{*} \qquad \text{eq. A1.10}$$

A2 Mixed air Temperature

A vector of candidate correction terms for the mixed air temperature is defined.

$$dT = [dT_1\,,\, dT_2\,,\, \ldots\,,\, dT_{j-1}\,,\, dT_j] \qquad \text{eq. A2.1}$$

For each *j* term of vector *dT*, the iterative process includes the following steps:

a. the *j* correction term from vector *dT* is added (eq. A2.2) to the faulty mixed air temperature, T_{ma}, giving the *j* modified faulty variable (T_{ma}^{j});

$$T_{ma}^{j} = T_{ma} + dT_{\mathrm{j}} \qquad \text{eq. A2.2}$$

b. the *j* modified faulty variable (T_{ma}^{j}) is used, along with the other measured fault-free air property (in this case the mixed air relative humidity, RH_{ma}), to evaluated the *j* modified air humidity ratio, x_{ma}^{j} (eq. A2.3), and the *j* modified air enthalpy, h_{ma}^{j}, (eq. A2.4), which is then used to derive a *j* α-factor, α_{j} (eq. A2.5);

$$x_{ma}^{j} = 0.62198 \cdot \frac{Pv_{ma}}{P - \mathrm{Ps}_{ma}^{j}} \qquad \text{eq. A2.3}$$

$$h_{ma}^{j} = C_p \cdot T_{ma}^{j} + x_{ma}^{j} \cdot \left(h_{fg} + C_v \cdot T_{ma}^{j}\right) \qquad \text{eq. A2.4}$$

$$\alpha_{\mathrm{j}} = \frac{h_{ma}^{j} - h_{rec}}{h_{oa} - h_{rec}} \qquad \text{eq. A2.5}$$

where Pv_{ma} =*f*(RH_{ma}), and Ps_{ma}^{j} =*f*(T_{ma}^{j}) are the partial pressure and the saturation pressure of water vapor respectively.

c. a value of the *j* modified mixed air humidity ratio and specific enthalpy are derived from α_j (eq. A2.6 and eq. A2.7), which are used to derive the value of the j modified faulty variable ($T_{oa,\alpha}^{j}$) through eq. A2.8;

$$x_{ma,\alpha}^{j} = \alpha_j \cdot (x_{oa} - x_{rec}) + x_{rec} \quad \text{eq. A2.6}$$

$$h_{ma,\alpha}^{j} = \alpha_j \cdot (h_{oa} - h_{rec}) + h_{rec} \quad \text{eq. A2.7}$$

$$T_{ma,\alpha}^{j} = \frac{h_{ma,\alpha}^{j} - h_{fg} \cdot x_{ma,\alpha}^{j}}{C_p + C_v \cdot x_{ma,\alpha}^{j}} \quad \text{eq. A2.8}$$

d. the *j* term $\Delta T_{BAS,\alpha}^{j}$ is evaluate (eq. A2.9) as difference between T_{ma}^{j} and $T_{ma,\alpha}^{j}$.

$$\Delta T_{BAS,\alpha}^{j} = T_{ma}^{j} - T_{ma,\alpha}^{j} \quad \text{eq. A2.9}$$

Finally, the optimal correction term dT^*, from the candidate correction terms vector *dT,* is identified as the one which corresponds to the minimum $\Delta T_{BAS,\alpha}^{j}$ (eq. A2.10).

$$dT^* = dT_j \Longrightarrow \min(\Delta T_{BAS,\alpha}^{j}) \quad \text{eq. A2.10}$$

The identified optimal correction term is used to correct the faulty variable (eq. A2.11), before it is used by the BAS for control purpose.

$$T_{ma}^{*} = T_{ma} + dT^* \quad \text{eq. A2.11}$$

A3 Mixed air Relative Humidity

A vector of candidate correction terms for the mixed air relative humidity is defined.

$$dRH = [dRH_1, dRH_2, \ldots, dRH_{j-1}, dRH_j] \quad \text{eq. A3.1}$$

For each *j* term of vector *dRH*, the iterative process includes the following steps:

a. the *j* correction term from vector *dRH* is added (eq. A3.2) to the faulty mixed air relative humidity, RH_{ma}, giving the *j* modified faulty variable (RH_{ma}^{j});

$$RH_{ma}^{j} = RH_{ma} + dRH_j \quad \text{eq. A3.2}$$

b. the *j* modified faulty variable (RH_{ma}^{j}) is used to evaluated a *j* modified mixed air humidity ratio x_{ma}^{j} (eq. A3.3) , a *j* modified mixed air specific enthalpy value h_{ma}^{j} (eq. A3.4), and finally a *j* α-factor, α_j (eq. A3.5);

$$x_{ma}^{j} = 0.62198 \cdot \frac{\mathrm{Pv}_{ma}^{j}}{P - Ps_{ma}} \qquad \text{eq. A3.3}$$

$$h_{ma}^{j} = C_p \cdot T_{ma} + x_{ma}^{j} \cdot \left(h_{fg} + C_v \cdot T_{ma}\right) \qquad \text{eq. A3.4}$$

$$\alpha_j = \frac{h_{ma}^{j} - h_{rec}}{h_{oa} - h_{rec}} \qquad \text{eq. A3.5}$$

where $\mathrm{Pv}_{ma}^{j} = f(RH_{ma}^{j})$, and $\mathrm{Ps}_{ma} = f(T_{ma})$ are the partial pressure and the saturation pressure of water vapor respectively.

c. a value of the *j* modified faulty air humidity ratio is derived from the correspondent α_j ($x_{ma,\alpha}^{j}$) (eq. A3.6) and used to estimate the *j* modified air relative humidity derived from α_j ($RH_{ma,\alpha}^{j}$) (eq. A3.7);

$$x_{ma,\alpha}^{j} = \alpha_j \cdot (x_{oa} - x_{rec}) + x_{rec} \qquad \text{eq. A3.6}$$

$$RH_{ma,\alpha}^{j} = \frac{\mathrm{Pv}_{ma,\alpha}^{j}}{\mathrm{Ps}_{ma}} \qquad \text{eq. A3.7}$$

where $\mathrm{Pv}_{ma,\alpha}^{j} = f(x_{ma,\alpha}^{j})$, and $\mathrm{Ps}_{ma} = f(T_{ma})$ are the partial pressure and the saturation pressure of water vapor respectively.

d. the *j* term $\Delta RH_{BAS,\alpha}^{j}$ is evaluate (eq. A3.8) as difference between RH_{ma}^{j} and $RH_{ma,\alpha}^{j}$.

$$\Delta RH_{BAS,\alpha}^{j} = RH_{ma}^{j} - RH_{ma,\alpha}^{j} \qquad \text{eq. A3.8}$$

Finally, the optimal correction term dRH^{*} from the candidate correction terms vector *dRH* is identified as the one which corresponds to the minimum $\Delta RH_{BAS,\alpha}^{j}$ (eq. A3.9), and used to correct faulty measurements of the of the mixed air relative humidity (eq. A3.10) before being used by the BAS for control purpose.

$$RH_{ma}^{*} = RH_{ma}^{j} \Longrightarrow \min(\Delta RH_{BAS,\alpha}^{j}) \qquad \text{eq. A3.9}$$

$$RH_{ma}^{*} = RH_{ma} + dRH^{*} \qquad \text{eq. A3.10}$$

A4 Recirculated air Temperature

A vector of candidate correction terms for the recirculated air temperature is defined.

$$dT = [dT_1 , dT_2 , \dots , dT_{j-1} , dT_j] \quad \text{eq. A4.1}$$

For each *j* term of vector *dT*, the iterative process includes the following steps:

a. the *j* correction term from vector *dT* is added (eq. A4.2) to the faulty recirculated air temperature, T_{rec}, giving the *j* modified faulty variable (T_{rec}^{j});

$$T_{rec}^{j} = T_{rec} + dT_j \quad \text{eq. A4.2}$$

b. the *j* modified faulty variable (T_{rec}^{j}) is used, along with the other measured fault-free air property (in this case the recirculated air relative humidity, RH_{rec}), to evaluated the *j* modified air humidity ratio, x_{rec}^{j} (eq. A4.3), and the *j* modified air enthalpy, h_{rec}^{j}, (eq. A4.4), which is then used to derive a *j* α-factor, α_j (eq. A4.5);

$$x_{rec}^{j} = 0.62198 \cdot \frac{Pv_{rec}}{P - \mathrm{Ps}_{rec}^{j}} \quad \text{eq. A4.3}$$

$$h_{rec}^{j} = C_p \cdot T_{rec}^{j} + x_{rec}^{j} \cdot \left(h_{fg} + C_v \cdot T_{rec}^{j}\right) \quad \text{eq. A4.4}$$

$$\alpha_j = \frac{h_{ma} - h_{rec}^{j}}{h_{oa} - h_{rec}^{j}} \quad \text{eq. A4.5}$$

where $Pv_{rec} = f(RH_{rec})$, and $Ps_{rec}^{j} = f(T_{rec}^{j})$ are the partial pressure and the saturation pressure of water vapor respectively.

c. a value of the *j* modified recirculated air humidity ratio and specific enthalpy are derived from α_j (eq. A4.6 and eq. A4.7), which are used to derive the value of the j modified faulty variable ($T_{rec,\alpha}^{j}$) through eq. A4.8;

$$x_{rec,\alpha}^{j} = \frac{(\alpha_j \cdot x_{oa} - x_{ma})}{\alpha_j - 1} \quad \text{eq. A4.6}$$

$$h_{rec,\alpha}^{j} = \frac{(\alpha_j \cdot h_{oa} - h_{ma})}{\alpha_j - 1} \quad \text{eq. A4.7}$$

$$T_{rec,\alpha}^{j} = \frac{h_{rec,\alpha}^{j} - h_{fg} \cdot x_{rec,\alpha}^{j}}{C_p + C_v \cdot x_{rec,\alpha}^{j}} \quad \text{eq. A4.8}$$

d. the *j* term $\Delta T_{BAS,\alpha}^{j}$ is evaluate (eq. A4.9) as difference between T_{rec}^{j} and $T_{rec,\alpha}^{j}$.

$$\Delta T_{BAS,\alpha}^{j} = T_{rec}^{j} - T_{rec,\alpha}^{j} \quad \text{eq. A4.9}$$

Finally, the optimal correction term dT^* from the candidate correction terms vector dT is identified as the one which corresponds to the minimum $\Delta T^j_{BAS,\alpha}$ (eq. A4.10).

$$dT^* = dT_{(j)} \Longrightarrow \min(\Delta T^j_{BAS,\alpha}) \quad \text{eq. A4.10}$$

The identified optimal correction term is used to correct the faulty variable (eq. A4.11), before it is used by the BAS for control purpose.

$$T^*_{rec} = T_{rec} + dT^* \quad \text{eq. A4.11}$$

A5 Recirculated air Relative humidity

A vector of candidate correction terms for the recirculated air relative humidity is defined.

$$dRH = [dRH_1\ ,\ dRH_2\ ,\ \dots\ ,\ dRH_{j-1}\ ,\ dRH_j] \quad \text{eq. A5.1}$$

For each j term of vector dRH, the iterative process includes the following steps:

a. the j correction term from vector dRH is added (eq. A5.2) to the faulty recirculated air relative humidity, RH_{rec}, giving back the j modified faulty variable (RH^j_{rec});

$$RH^j_{rec} = RH_{rec} + dRH_j \quad \text{eq. A5.2}$$

b. the j modified faulty variable (RH^j_{rec}) is used to evaluated a j modified recirculated air humidity ratio x^j_{rec} (eq. A5.3) , a j modified recirculated air specific enthalpy value h^j_{rec} (eq. A5.4), and finally a j α-factor, α_j (eq. A5.5);

$$x^j_{rec} = 0.62198 \cdot \frac{\mathrm{Pv}^j_{rec}}{P - Ps_{rec}} \quad \text{eq. A5.3}$$

$$h^j_{rec} = C_p \cdot T_{rec} + x^j_{rec} \cdot \left(h_{fg} + C_v \cdot T_{rec}\right) \quad \text{eq. A5.4}$$

$$\alpha_j = \frac{h_{ma} - h^j_{rec}}{h_{oa} - h^j_{rec}} \quad \text{eq. A5.5}$$

where $\mathrm{Pv}^j_{rec} = f(RH^j_{rec})$, and $\mathrm{Ps}_{rec} = f(T_{rec})$ are the partial pressure and the saturation pressure of water vapor respectively.

c. a value of the j modified faulty air humidity ratio is derived from the correspondent α_j ($x^j_{rec,\alpha}$) (eq. A3.6) and used to estimate the j modified air relative humidity derived from α_j ($RH^j_{rec,\alpha}$) (eq. A3.7);

$$x_{rec,\alpha}^{j} = \frac{(\alpha_j \cdot x_{oa} - x_{ma})}{\alpha_j - 1} \quad \text{eq. A5.6}$$

$$RH_{rec,\alpha}^{j} = \frac{\mathrm{Pv}_{rec,\alpha}^{j}}{\mathrm{Ps}_{rec}} \quad \text{eq. A5.7}$$

where $\mathrm{Pv}_{rec,\alpha}^{j} = f(x_{rec,\alpha}^{j})$, and $\mathrm{Ps}_{ma} = f(T_{rec})$ are the partial pressure and the saturation pressure of water vapor respectively.

d. the *j* term $\Delta RH_{BAS,\alpha}^{j}$ is evaluate (eq. A5.8) as difference between RH_{rec}^{j} and $RH_{rec,\alpha}^{j}$.

$$\Delta RH_{BAS,\alpha}^{j} = RH_{rec}^{j} - RH_{rec,\alpha}^{j} \quad \text{eq. A5.8}$$

Finally, the optimal correction term dRH^{*} from the candidate correction terms vector *dRH* is identified as the one which corresponds to the minimum $\Delta RH_{BAS,\alpha}^{j}$ (eq. A5.9), and used to correct faulty measurements of the of the recirculated air relative humidity (eq. A5.10) before being used by the BAS for control purpose.

$$RH_{rec}^{*} = RH_{rec}^{j} \Longrightarrow \min(\Delta RH_{BAS,\alpha}^{j}) \quad \text{eq. A5.9}$$

$$RH_{rec}^{*} = RH_{rec} + dRH^{*} \quad \text{eq. A5.10}$$