SOFTWARE TOOLS AND METHODS FOR BUILDINGS FAULT DETECTION AND DIAGNOSTICS

Claudio Giovanni Mattera

Center for Energy Informatics
Maersk Mc-Kinney Moller Institute
Technical Faculty
University of Southern Denmark

SDU

CENTER FOR ENERGY INFORMATICS

February 2019

SUPERVISION

Principal supervisor: Hamid Reza Shaker, University of Southern Denmark
Co-supervisor: Muhyiddine Jradi, University of Southern Denmark

EVALUATION COMMITTEE

Jan Corfixen Sørensen (chair), University of Southern Denmark
Mohsen Soltani, Aalborg University
Robin Trulssen Bye, Norwegian University of Science and Technology

COLOPHON

claudio@mattera.it claudiomattera
0000-0002-3801-5617 P-3517-2018
Rw6VmLsAAAAJ

Typeset by the author in TeX Gyre Pagella 12pt, CABIN, Hasklig and *TeX Gyre Chorus*, using LuaLaTeX.
Figures created with TikZ and PGFPLOTS.
Printed by Print&Sign, University of Southern Denmark.

ABSTRACT

Energy production is one of the main activities causing environmental harm and climate change. For this reason, in the recent past, most government bodies and public and private institutions set ambitious goals for reduction in energy consumption and increase in energy efficiency. The buildings sector is one of the main actors in the energy usage scheme. It is estimated that buildings account for over one third of global energy consumption, and that they are responsible for a large amount of greenhouse gas emissions, which have major effect on climate change. Therefore, many policies and regulations for buildings energy efficiency have been issued, and many institutions have defined classes of highly energy efficient buildings.

However, a large difference between design goals and actual energy performance has been observed in many buildings, especially long time after construction. One of the main causes for this gap is due to building faults. Modern buildings have complex engineering designs, made up of several subsystems, which, in turns, are composed of many components. Several kinds of faults can affect all of these components, such as sensors faults, mechanical components failures, time wear or misconfiguration. Faults impact both occupants comfort and energy consumption, causing often significant energy waste.

Fault detection and diagnostics techniques have been successfully developed and used in many fields, such as avionics and process engineering, for many decades, however, their application on buildings is relatively recent. Many proposed methods were only tested on isolated or simulated components, since real data is scarce and not publicly available. The few available commercial solutions are still simplistic, and can only detect a small subset of possible faults.

Moreover, while many individual techniques have been proposed, there are no common and widespread approaches for overall fault detection and diagnostics.

In this thesis, a top-down approach for fault detection and diagnostics in buildings is proposed, where building systems are stacked in a hierarchy. In the first step, building data is validated. All building applications, including fault detection and diagnostics, require sane and validated data to operate correctly. Afterwards, using the building's energy distribution tree, the overall performance of the building is evaluated, and potential underperforming subsystems are identified. Once the scope is restricted, specialized methods are used for fault detection and diagnostics on the specific subsystems.

Ventilation systems are one of the most critical systems in buildings, and are responsible for a large share of energy consumption. Therefore, special focus was devoted to them, and they were considered as case study for three specialized fault detection and diagnostics methods. In the first, virtual redundancy was introduced inside a ventilation unit by exploiting physical relations between different measurement. In the second, consensus among multiple similar components was used to identify outliers in the air distribution system. Finally, a technical report of faults impact was prepared through simulating healthy and faulty conditions using a dynamic energy model of a building.

All techniques developed in this thesis were deployed and tested on a real building at the University of Southern Denmark. The building, built in 2015 and used for teaching and office work, is fully equipped with sensors and meters, and acts as a living lab for the university. When deployed, the techniques helped identifying faults and anomalous conditions in the building, such as uncalibrated CO_2 sensors which lead to reduced indoor air quality, oscillating temperature readings inside ventilation units, and rooms with anomalous air distribution patterns.

Finally, in the proposed techniques, all main classes of fault detection and diagnostics methods present in literature have been explored, i.e. rule-based methods, model-based methods and data-driven methods.

RESUMÉ

Energiproduktionen er en af hovedårsagerne til miljø skader og klimaforandringer. Af denne grund har de fleste statslige organer og offentlige og private institutioner, for nyligt, sat nogle ambitiøse mål for at reducere energiforbruget og forøge energieffektiviteten. Bygningssektoren er en af hovedforbrugerne. Det er estimeret at bygninger står for over en tredjedel af det globale energiforbrug, og at de er ansvarlige for en stor del af CO_2 udledningen, som har en stor indvirkning på klimaforandringerne. Derfor er der blevet indført mange regler og reguleringer for bygningers energieffektivitet, og flere institutioner har defineret klasser af energieffektive bygninger.

Der er observeret stor difference imellem design målet og den aktuelle energieffektivitet i mange bygninger, specielt lang tid efter konstruktionen af bygningen. En af de mange årsager til denne problematik er på grund af bygningsfejl. Moderne bygninger er designet komplekst, og de er lavet af mange forskellige subsystemer, som hver især er opbygget af mange komponenter. Mange forskellige typer af fejl kan påvirke alle disse komponenter, såsom sensor fejl, mekanisk komponentfejl, slid pga. driftstimer eller fejlkonfiguration. Fejl påvirker både beboernes komfort og deres energiforbrug, hvilket ofte forårsager betydeligt højere energiforbrug.

Tekniker til fejldetektion og diagnosticering er allerede blevet udviklet til mange forskellige områder, såsom flyelektronik og procesteknologi, men deres anvendelse på bygninger er relativ ny. Mange af de foreslående metoder er kun blevet testet på isolerede eller simulerede komponenter, da reelle data er mangelfulde og ikke offentligt tilgængelig. De få tilgængelige kommercielle løsninger er stadig simple og kan kun opdage en delmængde af de mulige opståede fejl. Selvom der er blevet foreslået mange individuelle metoder, er der ikke en almindelig og udbredt tilgang til overordnet fejldetektion og diagnostik.

I denne afhandling foreslås en top-down-metode til fejldetektion og diagnostik i bygninger, hvor bygningssystemer er hierarkisk opdelt. Som det første trin valideres bygningsdata. Alle bygningsapplikationer, herunder fejlregistrering og diagnostik, kræver gangbare og valideret data for at fungere korrekt. Derefter evalueres bygningens samlede ydeevne ud fra alle delelementer, og potentielle dårlige resultater af delsystemer identificeres. Når omfanget er blevet begrænset, anvendes specialiserede metoder til fejldetektion og diagnostik på de specifikke delsystemer.

Ventilationssystemer er et af de mest kritiske systemer i bygninger og er ansvarlige for en stor del af energiforbruget. Derfor blev der specielt fokuseret på dem, og de blev betragtet som casestudie for tre specialiserede fejldetekterings- og diagnostik-metoder. I den første metode blev virtuel redundans indført i en ventilationsenhed ved at udnytte de fysiske forhold mellem forskellige målinger. I den anden metode blev der brugt consensus imellem flere lignende komponenter til at identificere afvigende resultater i luftdistributionssystemet. Til sidst blev der lavet en teknisk rapport om fejlindvirkningen ved at simulere funktionsdygtige og defekte forhold, ved brug af en dynamisk energimodel af en bygning.

Alle teknikker udviklet i denne afhandling blev implementeret og testet på en virkelig bygning ved Syddansk Universitet. Bygningen er fra 2015 og bruges til undervisning og kontorarbejde, den er fuldt udstyret med sensorer og målere, og fungerer som et levende laboratorium for universitetet. Ved implementering hjalp teknikkerne med at identificere fejl og uregelmæssige forhold i bygningen, såsom ikke kalibrerede CO_2-sensorer, som medfører reduceret indendørs luftkvalitet, oscillerende temperaturaflæsninger inde i ventilationsenheder og rum med uregelmæssige luftfordelingsmønstre.

Til sidst, i de foreslåede teknikker, er der blevet undersøgt alle hovedklasser af fejldetektions- og diagnostik-metoder, som er til stede i litteraturen, dvs. rule-based, model-based og data-driven metoder.

PUBLICATIONS

MAIN AUTHOR PUBLICATIONS

The following are the publications, written as first author, included in this thesis.

[1] **Claudio Giovanni Mattera**, Sanja Lazarova-Molnar, Hamid Reza Shaker and Bo Nørregaard Jørgensen. 'A Practical Approach to Validation of Buildings' Sensor Data: a Commissioning Experience Report'. In: *Third International Conference on Big Data Computing Service and Applications* (*BigDataService*) (San Francisco, CA, USA, 6th–9th Apr. 2017). IEEE. 12th June 2017, pp. 287–292. DOI: 10.1109/BigDataService.2017.48.

[2] **Claudio Giovanni Mattera**, Muhyiddine Jradi and Hamid Reza Shaker. 'Online Energy Simulator for Building Fault Detection and Diagnostics Using Dynamic Energy Performance Model'. In: *International Journal of Low-Carbon Technologies* 13.3 (17th May 2018), pp. 231–239. ISSN: 1748-1325. DOI: 10.1093/ijlct/cty019.

[3] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'Fault Detection and Diagnostics in Ventilation Units Using Linear Regression Virtual Sensors'. In: *International Symposium on Advanced Electrical and Communication Technologies* (*ISAECT*) (Kenitra, Morocco, 21st–23rd Nov. 2018). IEEE. 24th Jan. 2019. DOI: 10.1109/ISAECT.2018.8618755.

[4] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'A Method for Fault Detection and Diagnostics in Ventilation Units Using Virtual Sensors'. In: *Sensors* 18.11 (14th Nov. 2018). ISSN: 1424-8220. DOI: 10.3390/s18113931.

[5] **Claudio Giovanni Mattera**, Hamid Reza Shaker and Muhyiddine Jradi. 'Consensus-based Method for Anomaly Detection in VAV Units'. In: *Energies* 12.3 (1st Feb. 2019). ISSN: 1996-1073. DOI: 10.3390/en12030468.

[6] **Claudio Giovanni Mattera**, Hamid Reza Shaker, Muhyiddine Jradi, Mathis Riber Skydt and Sebastian Skals Engelsgaard. 'Fault Detection in Ventilation Units using Dynamic Energy Performance Models'. In: *Sustainable Cities and Society* (2019). ISSN: 2210-6707. **Submitted**.

CO-AUTHOR PUBLICATIONS

The following are secondary publications, contributed as co-author during the Ph. D. project. They are not included in this thesis.

[7] Muhyiddine Jradi, Fisayo Caleb Sangogboye, **Claudio Giovanni Mattera**, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'A World Class Energy Efficient University Building by Danish 2020 Standards'. In: *Energy Procedia* 132 (Oct. 2017): *11th Nordic Symposium on Building Physics*, pp. 21–26. ISSN: 1876-6102. DOI: 10.1016/j.egypro.2017.09.625.

[8] Muhyiddine Jradi, Krzysztof Arendt, Fisayo Caleb Sangogboye, **Claudio Giovanni Mattera**, Elena Markoska, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'ObepME: An Online Building Energy Performance Monitoring and Evaluation Tool to Reduce Energy Performance Gaps'. In: *Energy and Buildings* 166 (1st May 2018), pp. 196–209. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2018.02.005.

[9] Krzysztof Arendt, Aslak Johansen, Bo Nørregaard Jørgensen, Mikkel Baun Kjærgaard, **Claudio Giovanni Mattera**, Fisayo Caleb Sangogboye, Jens Hjort Schwee and Christian T. Veje. 'Room-level Occupant Counts, Airflow and CO2 Data from an Office Building'. In: *The 16th ACM Conference on Embedded Networked Sensor Systems*. Proceedings of the First Workshop on Data Acquisition To Analysis (Shenzhen, China, 4th–7th Nov. 2018). ACM. New York, NY, USA, 2018, pp. 13–14. DOI: 10.1145/3277868.3277875.

[10] Muhyiddine Jradi, Na Liu, Aslak Johansen, Krzysztof Arendt, **Claudio Giovanni Mattera**, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'Dynamic Energy Model-Based Automatic Building Performance Testing for Continuous Commissioning'. In: *Proceedings of the 16th IBPSA International Conference Building Simulation 2019* (Rome, Italy, 2nd–4th Sept. 2019). International Building Performance Simulation Association. URL: http://buildingsimulation2019.org/. **Submitted**.

[11] Krzysztof Arendt, Anders Clausen, **Claudio Giovanni Mattera**, Muhyiddine Jradi, Aslak Johansen, Christian T. Veje, Mikkel Baun Kjægaard and Bo Nørregaard Jørgensen. 'Controleum: Multi-Objective Model Predictive Control Framework for Buildings'. In: *Proceedings of the 16th IBPSA International Conference Building Simulation 2019* (Rome, Italy, 2nd–4th Sept. 2019). International Building Performance Simulation Association. URL: http://buildingsimulation2019.org/. **Submitted**.

The pursuit of knowledge is hopeless, and eternal.

— Prof. Hubert J. Farnsworth

ACKNOWLEDGMENTS

It seems inevitable that Ph. D. students are unsatisfied with their thesis. The earliest contributions, made at the beginning of their career, when they are still callow, seem trivial and frivolous. Later contributions, made when talent and endeavour finally take over inexperience, are rushed to meet the final deadline, and appear cobbled together and unfinished. Eventually, they learn that research is a continuous flow: it is never finished, and it is never ending.

Nevertheless, a thesis is a substantial amount of work, and its completion is an important milestone and a chance for celebration. Ultimately, I take all the credit for this masterpiece. Beneath it, however, lies a massive team effort which must be properly acknowledged.

I will start by thanking my parents, my brother, and our cat, together with the rest of my family, who have been supporting and encouraging me since I have memory. They have been there for me when I was in school, when I attended university, when I studied abroad, when I started my professional life, and when I pursued my career abroad. They went through efforts and sacrifices but their support was never lacking. I do not show it often, but I do miss them, and I feel the burden of the distance between us.

Next in line are my principal supervisor and my co-supervisor, Hamid and Muhyiddine. Their guidance played an important role in my research and my thesis, which I would hardly have completed without them. I spare a word of gratitude also for my former principal supervisor and head of our center, Bo, and my former co-supervisor Sanja, who both guided me at the beginning of my Ph. D. when I started moving my first steps in this unfamiliar world.

I would like to thank all the people at the Center for Energy Informatics, both those with whom I had the pleasure to work directly, and also those whose I just enjoyed the company and the encouragement. Two of them deserve an explicit mention, if only because of all the time I took from their own tasks. Aslak, for having countless times lightened the mood during our battles against time-series madness and data quality, and Krzysztof, for the invaluable help in taming EnergyPlus and the FMI standard.

I reserve a special thank for Joseba, Teresa, and all the people I met at the Center for Supervision, Security and Automatic Control in Terrassa. I spent three months at a brilliant research group, where I learnt how to adapt my work to different approaches and techniques, a useful skill in my career. Under another perspective, I embarked on an epic road trip over several countries, which had an equal impact on my personal growth. I am truly grateful for this experience.

Most of my former teachers deserve a small recognition for putting me on the right track. I will mention one of them, prof. Amaldi, for giving me the final push to embark upon a Ph. D.

When it comes to thanking friends, I shall call myself lucky not to have enough room to name each and every one of them. Some I met in my childhood, and they are still by my side after many years. Some others I met later in life, and few of them even shared my same decision of wandering the world. Others again I met while living in Denmark, they welcomed me and made me feel at home. All of them contributed to shaping me as the person I am today. In particular, Daniela and Aisha take credit for going through this thesis and finding copious grammatical mistakes before the evaluation committee, and Martin and Jakob for translating the abstract to Danish.

Lastly—aside for support, encouragement, and guidance—financial aid must also be acknowledged as vital for any great endeavour. I would like to thank the Innovation Fund Denmark for seeing the potential in project COORDICY, of which my Ph. D. was a significant part, and allocating toward it generous funding.

CONTENTS

Abstract i

Resumé iii

Publications v

Acknowledgments ix

Contents xv

List of Figures xix

List of Tables xxi

List of Listings xxiii

List of Rules xxv

Acronyms xxvii

Glossary xxix

Structure of the Thesis xxxi

I Background 1

1 Introduction 5
1.1 Energy Efficiency and Environmental Impact 5

1.2 Buildings Energy Consumption 6
1.3 Energy Efficiency Strategies and Policies 7
1.3.1 European Union 10
1.3.2 Denmark 12
1.3.3 United States of America 13
1.4 Building Faults and Energy Waste 14
1.4.1 Impacts of Faults 16
1.5 Fault Detection and Diagnostics 19
1.6 Research Questions and Objectives 20
1.7 Energy Informatics and the COORDICY Project . . 20
1.7.1 Objective and Hypothesis 21
1.7.2 Project Organization 22
1.8 Methodology 25
1.9 Structure of the Thesis 27

2 State of the Art 29
2.1 Fault Detection and Diagnostics Methods 29
2.1.1 Process-History-Based Methods 31
2.1.2 Quantitative-Model-Based Methods 34
2.1.3 Qualitative-Model-Based Methods 36
2.1.4 Mixed Methods 37
2.1.5 Popularity of Fault Detection and Diagnostics Methods 38
2.2 Virtual Redundancy 38
2.3 Building Performance Simulation 40
2.4 Contributions 42
2.4.1 Holistic Contribution 42
2.4.2 Contributions of Individual Publications . . 45

3 A Hierarchical Framework for FDD 51
3.1 Sensors Validation: a Precondition to any Building Application 51
3.2 A Hierarchical Framework for FDD 53
3.3 Artificial Redundancy with Virtual Sensors 56
3.4 Consensus-based Anomaly Detection 58
3.5 Faults Simulation and Impact Assessment 60

4 Case Study: Building OU44 61
4.1 Data Collection 63
4.2 Ventilation System 64
4.3 Energy Model of the Building 68

II Publications 71

5 Sensors Validation 75
5.1 Introduction 76
5.2 Related Work 77
5.2.1 Data Validation: Overview and Classifications 77
5.2.2 State of the Art 79
5.3 Validation of Sensors and Meters Data 81
5.3.1 Validation Tests 82
5.4 Case Study 85
5.4.1 Results 86
5.4.2 Implications and Discussion 90
5.5 Conclusions 91

6 Online Energy Simulator 99
6.1 Introduction 100
6.2 State of the Art 101
6.2.1 Fault detection and diagnostics in Buildings 101
6.2.2 Building Simulation 102
6.3 Methodology for Fault Detection and Diagnostics in Buildings 105
6.4 Online Energy Simulator 108
6.4.1 Configuration 109
6.4.2 EnergyPlus Simulation Engine 111
6.5 Case Study: Building OU44 112
6.5.1 Monitoring Building Performance with Online Energy Simulator 113
6.5.2 Results 115
6.6 Conclusions 119
6.6.1 Future work 121

7 Linear Regression Virtual Sensors 127
7.1 Introduction 128
7.1.1 Problem statement 129
7.2 State of the Art 129
7.2.1 Fault detection and diagnostics 129
7.2.2 Virtual Sensors 130
7.3 Virtual Sensors in Ventilation Units 132
7.3.1 Fault Diagnostics 132
7.3.2 Measuring Deviations from Actual Sensors . 133
7.4 Case Study 134
7.4.1 Building OU44 134
7.4.2 Results 136
7.5 Conclusions 139

8 Non-linear and Statistical Virtual Sensors 145
8.1 Introduction 146
8.1.1 Problem Statement 147
8.2 State of the Art 148
8.2.1 Fault Detection and Diagnostics 148
8.2.2 Virtual Sensors 151
8.3 Material and Methods 154
8.3.1 Fault Diagnostics 155
8.3.2 Measuring Deviations from Physical Sensors 157
8.4 Results and Discussion 158
8.4.1 Building OU44 160
8.4.2 Results Using Linear Regression Models . . 161
8.4.3 Results Using Other Models 168
8.5 Conclusions and Future Directions 173
8.5.1 Conclusions 173
8.5.2 Future Directions 175

9 Consensus-based Anomaly Detection 183
9.1 Introduction 184
9.2 State of the Art 186
9.2.1 Problem statement 189

9.3 Consensus-Based Method for Anomaly Detection . 189
9.3.1 Data Preprocessing and Preparation 193
9.3.2 Episodes Involving Multiple Time-Series . . 196
9.3.3 Monitoring Multiple Time-Series 196
9.4 Case Study 198
9.4.1 Monitoring VAV Units 200
9.5 Conclusions 210

10 Faults Simulation 217
10.1 Introduction 218
10.2 State of the Art 219
10.2.1 Faults Simulation 222
10.2.2 Contributions 224
10.3 Method 225
10.3.1 Evaluation 225
10.4 Case Study: Building OU44 226
10.5 Experiments 227
10.5.1 Abrupt faults in air temperature loop 228
10.5.2 Gradual faults in air supply systems 228
10.5.3 Rules 229
10.5.4 Experiments Parameters 230
10.6 Results 235
10.6.1 Impact of Faults 238
10.6.2 Implications for Faults Prediction 242
10.7 Conclusion 242

III Conclusions 249

11 Future Research 253
11.1 Individual Publications 253
11.2 General Summary 254

12 Conclusion 257

Bibliography 261

LIST OF FIGURES

1.1 Historical atmospheric CO_2 level 6
1.2 Shares of electrical consumption in the EU and USA 8
1.3 CO_2 emissions in the EU and USA 9
1.4 Impact of common faults in American buildings in 2005 17
1.5 Impact of common faults in American buildings in 2017 18
1.6 Estimated energy savings for different buildings . . 19
1.7 Relations between work packages in project COORDICY 24

2.1 Classification of FDD methods based on the approach 30
2.2 Classification of FDD methods based on the equipment 32
2.3 Shares of FDD publications 39

3.1 An example of an energy distribution tree in a building 54
3.2 Hierarchy for buildings fault detection and diagnostics 55
3.3 Effect of faulty systems in consensus-based methods 59

4.1 Building OU44 62
4.2 Electrical energy distribution tree in building OU44 65
4.3 District heating distribution tree in building OU44 . 66
4.4 Ventilation unit in buildingOU44 67
4.5 Hot-water heating loop in building OU44 67
4.6 Ventilation shafts in building OU44 68
4.7 Building OU44 and its energy model 69

5.1 Threshold-based sensors 82
5.2 Data validation flowchart 83
5.3 Spikes test violations for CO_2 concentration level . . 87
5.4 Latency test violations for CO_2 concentration level . 87
5.5 Range test violations for CO_2 concentration level . . 88

5.6 Latency test violations for temperature 89
5.7 Range test violations for temperature 89
5.8 Range test violations for CO_2 concentration level after sensors were replaced 90

6.1 Distribution tree in a building for electrical energy . 106
6.2 Top-down approach in fault detection and diagnostics 107
6.3 Architecture of online energy simulator 108
6.4 Building OU44 at University of Southern Denmark . 113
6.5 Data from energy meters and simulation results . . 117
6.5 Data from energy meters and simulation results (cont.) 118
6.6 Comparison between room temperatures 119

7.1 Diagram of a ventilation unit 135
7.2 Comparison between post heat exchanger temperatures 138
7.3 Comparison between post heat exchanger airflows . 140

8.1 Categorization of FDD methods for buildings 148
8.2 Classification of FDD techniques for ventilation units 150
8.3 Categorization of virtual sensing techniques 152
8.4 Correlation between ventilation unit measurements 156
8.5 Confidence interval from virtual sensors 158
8.6 Diagram of a ventilation unit in building OU44 . . . 159
8.7 Diagram of a heating loop in building OU44 159
8.8 Physical sensors and linear regression virtual sensors 163
8.9 Acceptable ranges from linear virtual sensors for post heat exchanger temperature 166
8.10 Trends of variables in models A and B 167
8.11 Acceptable ranges from virtual sensors for post heat exchanger airflow 169
8.12 Physical sensors and ARMAX virtual sensors 171
8.13 Acceptable ranges from ARMAX virtual sensors for post heat exchanger temperature 172
8.14 Physical sensors and non-linear virtual sensors . . . 173
8.15 Acceptable ranges from non-linear regression virtual sensors for post heat exchanger temperature 174

9.1 Overview of consensus-based FDD method 191
9.2 Effect of faulty systems in consensus-based methods 192
9.3 Extracting events from time-series 194
9.4 Episodes involving multiple time-series 197
9.5 Building OU44 198
9.6 Moving average of deviations for ventilation unit 1 . 204
9.7 Moving average of deviations for ventilation unit 2 . 205
9.8 Moving average of deviations for ventilation unit 3 . 206
9.9 Moving average of deviations for ventilation unit 4 . 207
9.10 Moving average of deviations for classrooms 208
9.11 Correct behaviour of room ventilation 209
9.12 Anomalous behaviour of room ventilation 210

10.1 Overview 225
10.2 Ventilation unit 227
10.3 Results for experiments A and B during cold months 236
10.4 Post-HX temperature in experiments A and B 237
10.5 Daily average supply temperature in experiment C . 239
10.6 Fan energy consumption in experiments D and E . . 240
10.7 Fan efficiency in experiments D and E 241

LIST OF TABLES

2.1 Stack of building hierarchy covered by each publication 44
2.2 Methods used in each publication 46

4.1 Characteristics of building OU44 62

5.1 Related work in the different groups 79
5.2 Issues detected by each test 83
5.3 Implemented tests parameters 86
5.4 Number of streams and average frequency 86

6.1 Elapsed time for different simulation periods 120

7.1 Virtual Sensors Definitions 136
7.2 Prediction R^2 score for virtual sensors 137

8.1 Virtual sensors definitions for linear regression models 162
8.2 Coefficients for linear regression models 164
8.3 Prediction R^2 score for virtual sensors 165
8.4 Virtual sensors definitions for other models 170

9.1 Advantages of FDD methods 190
9.2 Room types in building OU44 199
9.3 Most frequent episodes for ventilation unit 1 202
9.4 Anomalous rooms in both experiments conducted . 202

10.1 Parameters for experiment A and B 232
10.2 Parameters for experiment C 233
10.3 Values for fan efficiency in experiment D 233
10.4 Values for pressure rise in experiment E 234

LIST OF LISTINGS

6.1 Example Mapping Configuration File 110

LIST OF RULES

10.1 Heater energy rule 229
10.2 Post-HX air temperature rule 230
10.3 Supply air temperature rule 230
10.4 Fan energy rule 230

ACRONYMS

AHU air handling unit.
ANN artificial neural network.
ARMAX autoregressive moving average with exogenous variables.

BIM building information model.
BMS building management system.

CSV comma separated value.

EPW Energy Plus weather.
EU European Union.

FDD fault detection and diagnostics.
FMI functional mock-up interface.
FMU functional mock-up unit.

GDP gross domestic product.

HVAC heating, ventilation and air conditioning.
HX heat exchanger.

ICT information and communication technology.
IDF Energy Plus input.

OU44 Odense undervisning 44.

PCA principal component analysis.

sMAP simple measurement and actuation profile.
SVM support vector machine.

USA United States of America.
UUID universally unique identifier.

VAV variable air volume.

GLOSSARY

CO_2 Carbon dioxide, gas produced by breathing and by several industrial processes.

Atmospheric CO_2 Atmospheric CO_2 concentration level is linked to global warming and climate change. 5, 7

Indoor CO_2 In buildings, CO_2 is used to measure indoor air quality. A CO_2 concentration level above few hundred ppm indicates stale air. 61, 66, 76, 90, 189

data-driven method *see* history-based method

district heating 'The distribution of thermal energy in the form of steam or hot water from a central source of production through a network to multiple buildings or sites, for the use of space or process heating' [12]. 61, 63, 64

dynamic energy performance model A model of the building that, given external variables such as the weather forecast and the occupancy profile as input, can be used to compute its energy consumption through a simulation. It is a useful tool for verifying whether a building performs as expected, or for testing different control strategies. 22, 26, 40, 43, 105, 225

energy distribution tree The hierarchical energy flow in a building, from the main meter to the individual components. 53, 63, 105

fault 'A fault is defined as an unpermitted deviation of at least one characteristic property of a variable from an acceptable behaviour. Therefore, the fault is a state that may lead to a malfunction or failure of the system' [13]. 15, 16, 19

HISTORY-BASED METHOD A method where historical data is used to train a black-box model of the system under test, which is used to predict or validate the behaviour of the system itself. 31, 78, 101, 130, 148, 186, 219

MODEL-BASED METHOD A method where a model of the system under test is created from first principles or other detailed knowledge about the system. The model is used to predict or validate the behaviour of the system itself. 34, 78, 102, 130, 149, 186, 220

NEARLY ZERO-ENERGY BUILDING 'A building that has a very high energy performance. The nearly zero or very low amount of energy required should be covered to a very significant extent by energy from renewable sources, including energy from renewable sources produced on-site or nearby' [12]. 10, 11, 61

PRIMARY ENERGY Energy available before any transformation. 6

RULE-BASED METHOD A method where the behaviour of the system under test is encoded through a set of rules obtained from experts knowledge, specifications or documentation, which are used to validate the behaviour of the system itself. 36, 78, 102, 130, 149, 186, 222

VIRTUAL SENSOR A sensor which uses a model to compute its readings, instead of measuring a physical quantity. 40, 57, 128, 145

STRUCTURE OF THE THESIS

This thesis follows the format of a collection of publications, i.e. it includes the publications produced during the Ph. D. project. It is divided into three separate parts.

Part I serves as a general introduction to the work carried out during the Ph. D. project. Chapter 1 introduces the context of this thesis and presents the motivation, the research questions and the methodology. The current state of the art of the field is reviewed in Chapter 2, where its gap is highlighted, and the contribution of the thesis is presented. Chapter 3 contains a short summary of the publications included in the thesis. In Chapter 4, the main case-study building OU44 is presented.

Part II contains the individual publications, both published and under review, included in this thesis. In Chapter 5, the topic of data validation is presented, together with a practical experience of sensors faults identified in the building. In Chapter 6, an online energy simulator is presented as a tool for hierarchical fault detection and diagnostics (FDD) on buildings. In Chapter 7, linear regression virtual sensors are used to introduce redundancy in ventilation units, and a FDD method is proposed to exploit such redundancy. The same method is improved in Chapter 8, where non-linear and statistical models are used to increase accuracy. In Chapter 9, consensus among multiple peers is exploited to remove the requirement of fault-free training data in data-driven FDD methods. Finally, the impact of faults on the building's energy consumption is assessed in Chapter 10, where a FDD method for ventilation units is proposed.

Part III concludes the thesis. In Chapter 11, future research directions are suggested in the context of the thesis. The findings of the thesis are summarized and elaborated in Chapter 12.

Part I

Background

This part serves as a general introduction to the work carried out during the Ph. D. project. Chapter 1 introduces the context of this thesis and presents the motivation, the research questions and the methodology. The current state of the art of the field is reviewed in Chapter 2, where its gap is highlighted, and the contribution of the thesis is presented. Chapter 3 contains a short summary of the publications included in the thesis. In Chapter 4, the main case-study building OU44 is presented.

CHAPTER 1
INTRODUCTION

In this chapter, the field of fault detection and diagnostics (FDD) in building systems is introduced in the context of energy efficiency plans and regulations. Afterwards, the objectives and the research questions of this thesis are defined, and the methodology followed during the Ph. D. project is described.

1.1 ENERGY EFFICIENCY AND ENVIRONMENTAL IMPACT

In the famous *The Stern Review*, Stern estimates that the effects of climate change on human society will be massive [14]. Without taking any measure to contrast the current trends in global warming, carbon emissions, deforestation and other climate change indicators, the global gross domestic product (GDP) will suffer a yearly loss of at least 5 %, up to a yearly loss of 20 % when wider risks and impacts are taken into consideration. Besides human economy, climate change will also significantly impact the global environment. The prevalence of extreme weather has been increasing over the whole world, causing damage and irreversible effects.

One of the key factors to climate change are carbon emissions. CO_2 level in the atmosphere has been relatively steady until the industrial revolution, when carbon emissions from human machinery made it rise at an increased pace, as it is shown in Figure 1.1. CO_2 in the atmosphere is the main cause for the greenhouse effect, i.e. the reduced cooling effect of the planet. Nowadays, CO_2 level in the atmosphere passed over 400 ppm, and it is increasing by 2 ppm every year. In order

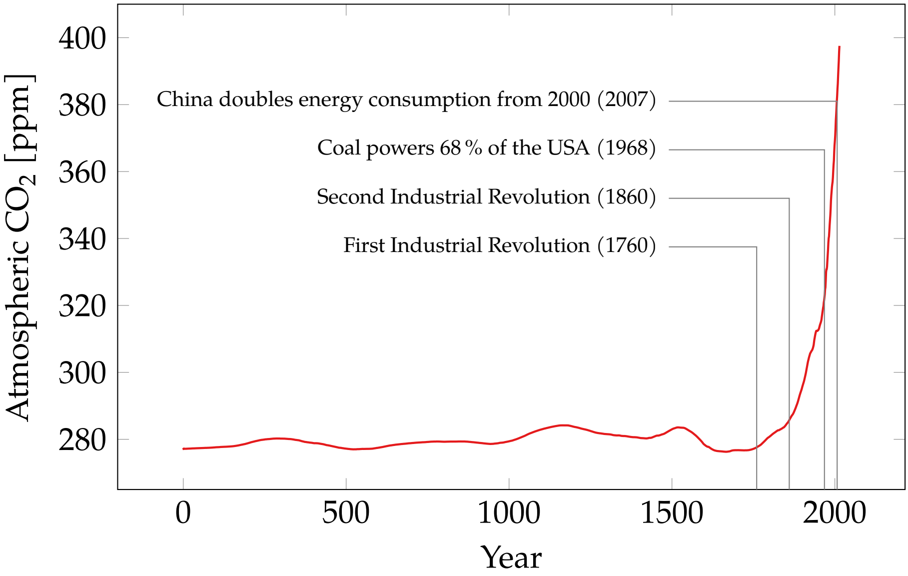

Figure 1.1: Historical atmospheric CO_2 level [15] with a few milestones in the worldwide energy consumption. The two industrial revolutions initiated a steep rise of CO_2 in the atmosphere, which increased together with technological development.

to avert a significant effect on global temperature, humankind should strive to maintain CO_2 below 550 ppm [14]. Since energy production is one of the main drives of carbon emissions and other climate change factors, most organizations and governments all over the world have been drawing long-term plans to reduce energy consumption.

1.2 BUILDINGS ENERGY CONSUMPTION

Buildings are responsible for a large share of global primary energy consumption and environmental impact. In the European Union (EU), they account for 40 % of the total energy used and a significant share of the total CO_2 emissions [12]. In the United States of America (USA), buildings accounted for about 41 % of primary energy consumption in 2010, 44 % more than the transportation sector and 36 % more than the industrial sector [16, § 1.1.3]. In the USA buildings are responsible for 40 % of CO_2 emissions.

Figure 1.2 shows the trend in electrical energy consumption for the residential and commercial, industrial and transports sectors in the EU and in the USA over the past decade. European data is available from the Eurostat database [17, Tab. ten00094, 18], and American data is available from *Monthly Energy Review - November 2018* [19, Tab. 7.6]). Residential and commercial sectors, taken together, represent more than two-thirds of the total energy consumption, which is double the amount of industry. Electrical consumption in transports sector is significantly lower, since petrol and its derivatives are used directly in vehicles.

A similar message can be read from Figure 1.3, which shows the impact of CO_2 emissions of different sectors in the EU and the USA. European data is available from *EU Energy in Figures*, [20, § 4.1.2], and American data is available from *Monthly Energy Review - November 2018* [19, Tabs. 12.2–12.5]). The shares of CO_2 emissions due to residential and commercial sector are different between the EU and the USA due to a different classification and measuring approaches. However, the two sectors are responsible for a significant share of the total emissions.

It appears evident how buildings are an important actor in global energy efficiency. In the last decades, governments, institutions, industries and the general public have become aware of the weight of buildings with respect to energy efficiency and environmental impact. The field of energy efficiency in buildings has attracted significant research focus, and buildings have become an important part of short- and long-term energy strategies and policies.

1.3 ENERGY EFFICIENCY STRATEGIES AND POLICIES

Several countries in the world defined short- and long-term strategies and policies for energy efficiency.

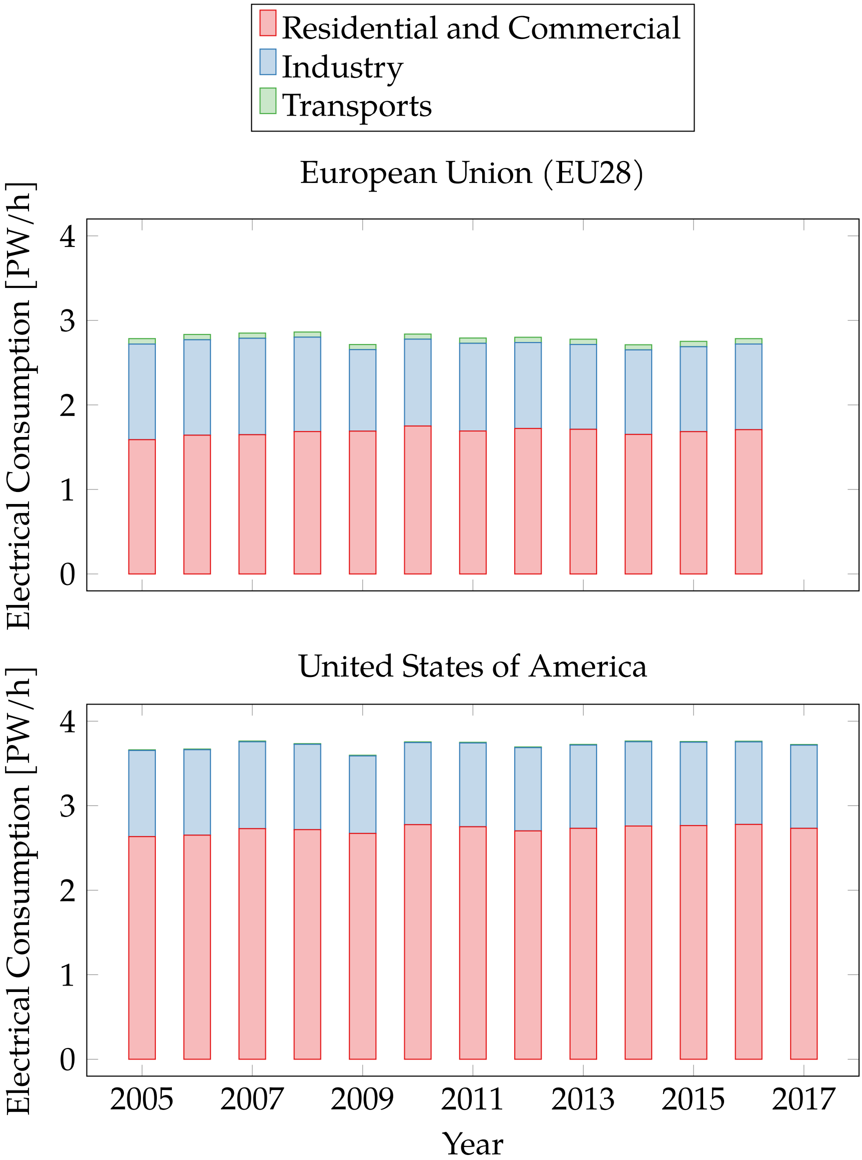

Figure 1.2: Shares of electrical energy consumption in the EU and USA. Residential and commercial sectors, taken together, are responsible for almost double of the industrial electrical consumption [17, Tab. ten00094, 19, Tab. 7.6].

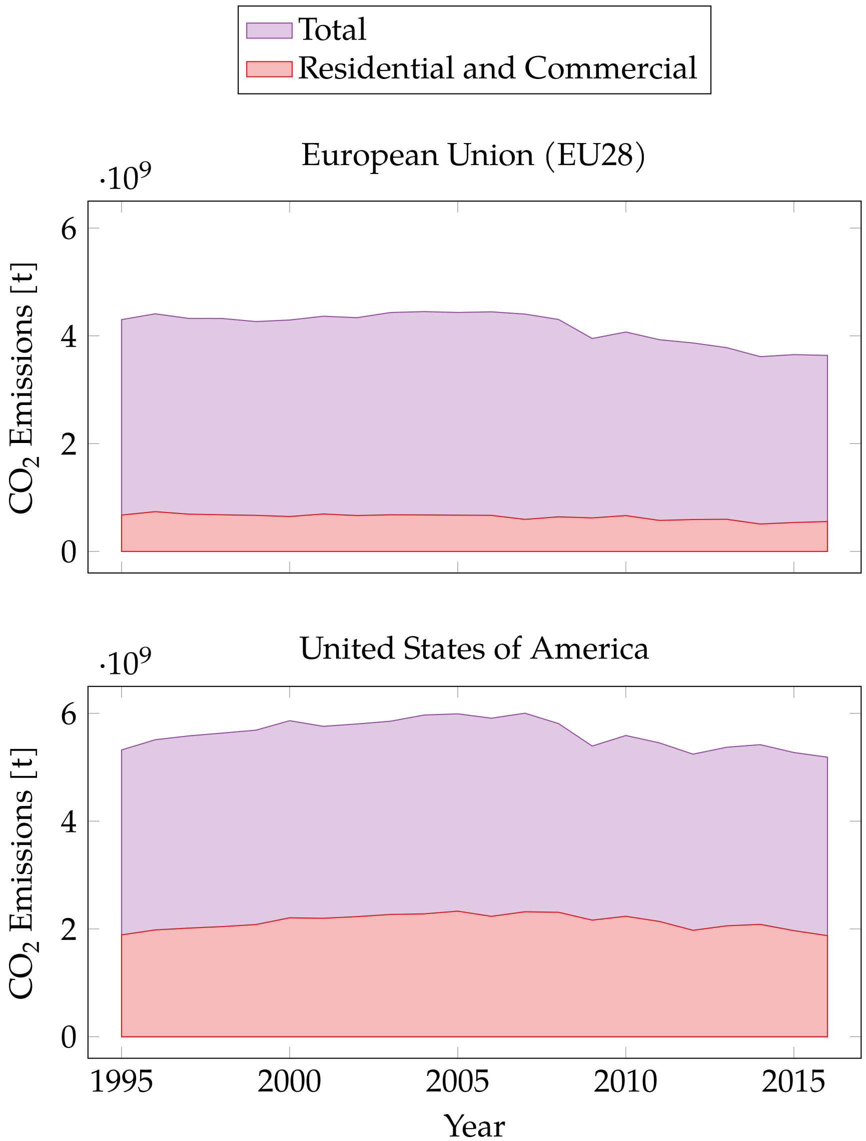

Figure 1.3: CO_2 emissions due to buildings and total in the EU and USA [20, § 4.1.2, 19, Tabs. 12.2–12.5]. The large difference between the two geopolitical blocks is due to a different classification and measuring approach. However, residential and commercial sectors account for a large part of CO_2 emissions in both of them.

1.3.1 European Union

In 2007, the EU proposed and defined its famous '20 20 by 2020' energy efficiency plan. [21]. The title came from the two major goals that the union set in its resolution by 2020:

- greenhouse gases emission will be reduced by at least 20 % compared to the levels of 1990, and by a more ambitious 30 % in case of an international agreement;
- renewable energies will reach a share of at least 20 % of total energy consumption.

The union sees also opportunities aside from the environmental impact. Less dependency on oil and gas imports will result in significant savings and reduced dependency on other international actors. Moreover, the long term strategy is expected to create a significant number of jobs in the renewable energies and other eco-industry sectors.

On the longer term, the EU has a major goal of halving the CO_2 emissions by the year 2050 compared to the levels in the year 1990. The 20 20 energy efficiency plan is indeed one of the steps toward this more ambitious goal.

Other plans to follow the 20 20 plan have been discussed and drafted in the past years. In 2016, a proposal for a new directive has been published, advocating for an extension of energy efficiency targets of 30 % by the year 2030 [22].

European buildings

In all European directives and resolutions about energy efficiency, buildings are explicitly mentioned as an important sector. In 2010, the union published the 'Energy Performance of Buildings Directive' [12]. The directive defines a general framework for assessing energy efficiency in buildings, minimum requirements for newly constructed and renovated buildings and building equipment, and promoted certifications for highly energy efficient buildings. The directive puts significant focus on *nearly zero-energy buildings*, i.e. buildings with very high energy efficiency, which should be covered by renewable energy sources. By the year 2018, all newly constructed public buildings in the EU are required to be nearly zero-energy buildings and, by the year 2020, the requirement is extended to every newly constructed building.

Different agencies within the union have been promoting and financing research and development of nearly zero-energy buildings, such as the Executive Agency for Small and Medium-sized Enterprises [23]. The EU has also promoted several agencies and offices to keep track of the building sector, such as the Buildings Performance Institute Europe, the Building Stock Observatory or Build Up, the European portal for energy efficiency in buildings.

Reports tracking the progress with respect to the directive for energy performance in buildings stated that progress was visible in many member states [24], however, only about half of the estimated energy savings was achieved [25]. A series of recommendations have been produced to increase compliance with the directive, such as

- increasing transparency about future regulations;
- promoting further education to improve skills in the buildings workforce;
- distribution of examples of good and bad practices;
- increasing financial support to member states;
- clarifying compliance requirements;
- promoting the value of energy performance certifications to final users;
- standardizing certification methodologies.

In 2018, a new directive for energy performance in buildings was published [26]. Member states shall define a long term renovation strategy for their national stock of buildings, both public and private. The strategy should aim to a complete renovation by the year 2050 so that all buildings are transformed in highly energy efficient buildings.

1.3.2 Denmark

Aside from goals imposed by the EU, member states have also specific national goals and strategies. Denmark long-term objective for energy efficiency is to remove any dependency on fossil fuels. In order to pursue its long-term objective of going fossil-free in 2050, the country defined shorter-term, intermediate goals in its Energy Strategy 2050 [27].

The year 2020 will be an important milestone in the energy strategy. The plan states that fossil fuels usage will be reduced by 33 % compared to 2009, and the share of renewable energy sources will increase by 33 % compared to 2009. Moreover, the focus on energy efficiency will result in reducing primary energy consumption by 6 % compared to 2006.

Danish buildings

The Danish government recognizes the importance of buildings in global energy consumption, and buildings design and construction are regulated in order to achieve safety, occupants comfort and energy efficiency. Energy efficiency goals have been encoded as a framework for energy efficient buildings, which has been updated over the past decade to reflect the current objectives and technological advancements.

In 2010, the Danish buildings regulations BR10 stated that, for newly constructed building, the annual energy consumption for heating, ventilation, cooling, domestic hot water and lighting must not exceed 71.3 $\mathrm{kW\,h/m^2}$, plus 1650 kW h divided by the heated floor area [28, § 7.2.3]. For buildings constructed from 2015 onward, the annual energy consumption must not exceed 41 $\mathrm{kW\,h/m^2}$, plus 1100 kW h divided by the heated floor area [28, § 7.2.4.2].

In 2015, a new version of the buildings regulations was released under the name of BR15, which introduced the new class for highly energy efficient buildings constructed after 2020 with an ambitious efficiency level. For buildings in class 2020, the annual energy consumption must not exceed 25 $\mathrm{kW\,h/m^2}$ [29, § 7.2.4.2]. The most recent regulations, BR18, released in 2018, included class 2020 as a voluntary requirement, i.e. its efficiency is no longer mandatory for buildings constructed after the year 2020 [30, chap. 25].

The introduction of class 2020 at a national level in 2015, while at European level its efficiency will be required only from 2020, has been praised by reports from the EU [25]. This allowed the Danish sector to adapt their products to the future standards in advance, and already resulted in highly efficient products to reach the market.

1.3.3 UNITED STATES OF AMERICA

The USA have also objectives with respect to increasing energy efficiency and reducing environmental impact.

In the year 2013, the Alliance Commission on National Energy Efficiency Policy adopted the ambitious goal of doubling energy productivity by the year 2030 compared to levels in 2011 [31]. Estimated benefits of reaching such goal are significant. Individual households would achieve over 1000 $ yearly savings in energy consumption, over one million new jobs would be created in the related industry, and both carbon emissions and oil imports would decrease by one-third.

The commission proposed a series of recommendations in order to reach the objective set, such as

- increasing finance opportunities for research and development in energy efficiency;
- increasing federal funding;
- strengthening regulations for vehicles, buildings and other equipment;
- promoting education of energy users.

AMERICAN BUILDINGS

Government bodies in the USA have been regulating the buildings sector and proposed plans to improve energy efficiency. Until 2012, the Department of Energy published a yearly report about energy consumption in buildings [16]. Since 2012, monthly reports are available for the general public, which contain fewer aggregate data and can be used for less coarse analysis of energy consumption over time [19].

In 2015, the Department of Energy made a common definition of *zero energy buildings* [32]. The advantage of a shared definition at national level is to make the evaluation and certification procedures uniform and unambiguous across the country, both for private users and companies.

Along the same direction, the Building Energy Codes Program was promoted by the Department of Energy as a way to improve energy efficiency in buildings and inform end users [33]. These actions have been taken with the goal of a long-term renovation of the buildings stock. It is estimated that by 2030 18 % of buildings will be more recent than the current year, and many more will be renovated. Reports suggest that the program resulted in 11.5 % energy savings before 2015 [34].

1.4 BUILDING FAULTS AND ENERGY WASTE

Buildings and building equipment are complex and massive systems. Not only they are large in the sense of physical size, e.g. a commercial building such a university lecture hall can reach thousands of square meters spanning several floors, they are also large in the sense of the amount of subsystems and their interactions. Moreover, modern buildings contain a significant amount of intelligence and control.

While, in the past, buildings only had manual lighting switches in each room and a centralized heating system operating with fixed schedules and manually controlled radiators, in the latest decades their complexity had increased enormously. Modern buildings are automatically and centrally managed by a building management system (BMS), which controls all building subsystems such as lighting, heating, ventilation and air conditioning (HVAC). The current state of the building is recorded through a complex network of sensors and meters, which measure indoor conditions such as temperature and air quality, and operating quantities such as ventilation rate, light intensity level, heating and cooling signals.

Every single component of every subsystem, as for every man-made product, is subject to wearing, misconfiguration and, in general, faults. A fault is any instance of a component which does not perform its task as expected. The definition of faults is somewhat vague and their nature can be extremely heterogeneous. Isermann define a fault as ' an unpermitted deviation of at least one characteristic property of a variable from an acceptable behaviour. Therefore, the fault is a state that may lead to a malfunction or failure of the system' [13]. Examples of faults are

- a stuck sensor which returns constant readings regardless of the actual measurement;
- a noisy sensor which returns inaccurate readings;
- a worn-out fan which consumes more energy to maintain the same airflow compared to its design value;
- a misplaced energy meter that monitors the wrong subsystem;
- a wrong schedule where Sundays are registered as working days;
- a wrong configuration where an office is classified as a storage unit.

Faults impact building operations in two different, but not necessarily distinct, ways. They may cause occupants discomfort or energy waste.

Occupancy discomfort happens when the building's operation is degraded such that the indoor conditions are no longer within the acceptable range for humans comfortable conditions. E.g. a broken heating system during winter could cause the indoor temperature to fall below 15 °C. Another example is a broken light bulb, which makes a room dark.

Energy waste, on the other hand, happens when the system consumes more energy than it should according to its design. E.g. insufficient insulation cause an increase in heating. Another example is simultaneous heating and cooling of a space.

The two aspects often oppose each other, i.e. a fault causes either occupants discomfort or energy waste, but sometimes a fault results in both. E.g. broken lights, while obviously causing discomfort, actually result in energy savings. Similarly, insufficient insulation does not affect occupants, if the increased heating results in appropriate indoor temperature. On the other hand, a setpoint wrongly set to 10 °C during the summer has the double effect of making a room extremely uncomfortable, and of unnecessarily increasing cooling.

Faults causing occupants discomfort are usually easily detectable, at least because said occupants would complain to the building management. Faults causing energy waste, on the other hand, are more subtle and difficult to detect. If no system is set up to monitor a building and no maintenance operation is scheduled, faults can go unnoticed for a very long time. For this reason, most research about faults is done in the context of energy waste.

1.4.1 Impacts of faults

Estimating the impact of faults with respect to energy waste is a difficult task.

In 2005, the 13 most common faults in buildings were estimated to be responsible for over 99.6 TW h and 3.3 billion $ energy waste in the USA [35, Tab. 2.1, 36, Tab. 1]. Figure 1.4, from the same report, shows the energy and financial impact of the most common faults in buildings [35, Tab. 2.1, 36, Tab. 1]. Figure 1.5 shows a more recent report from 2017, focused on small commercial building [37, Appx. C].

While the numbers are on different scales and the specific faults do not exactly coincide between the two reports, the overall trend is the same. Duct and building envelope leakage is the largest culprit for energy waste, followed by situations when ventilation and lighting are turned on when no occupants are in the building.

The yearly financial impact of these faults ranges from 3.3 billion $ to 17.3 billion $, with 3.3 billion $ being the most conservative estimate, according to the former report. The latter report estimates over 6 billion $ yearly waste due to faults.

Figure 1.6 shows the estimated energy savings for different categories of buildings when performing recommissioning operations, which includes addressing existing faults [38].

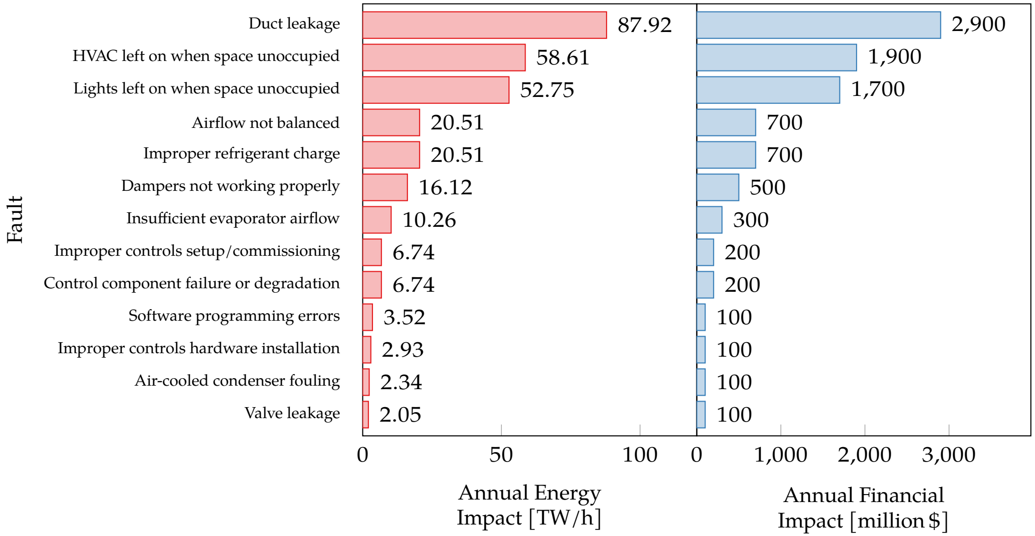

Figure 1.4: Impact of common faults in American buildings in 2005 [35, Tab. 2.1, 36, Tab. 1].

Figure 1.5: Impact of common faults in small commercial American buildings in 2017 [37, Appx. C].

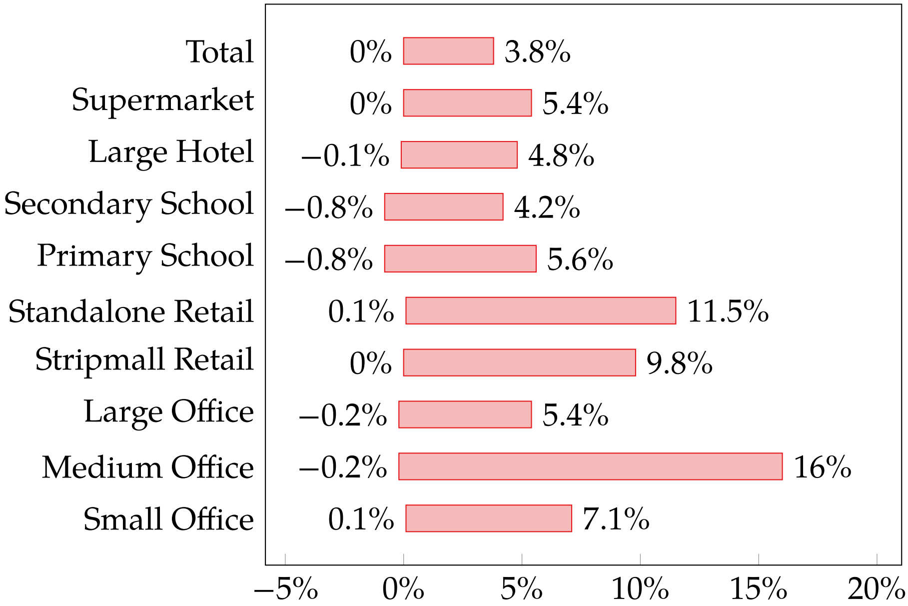

Figure 1.6: Estimated energy savings for different categories of buildings when performing recommissioning operations [38].

1.5 FAULT DETECTION AND DIAGNOSTICS

The traditional way to detect and fix faults is to perform periodic recommissioning, i.e. going through the building from the grounds up, taking note of everything that does not work, and fix it. The main disadvantage of this approach is that such operations are costly and, therefore, are only scheduled once in long periods. For this reason, faults have effects for a long time before they are discovered and addressed.

With technological improvements, modern buildings are able to collect data, and such data can be used to detect faults presence without physical interventions. In the past few decades, therefore, FDD for building systems has emerged as a new field. Many methods have been proposed for detecting faults and issues with building equipment [39, 40, 41].

1.6 RESEARCH QUESTIONS AND OBJECTIVES

The overall research question for this thesis is the following.

> How can faults in buildings energy facilities be automatically identified and explained?

Sub-questions are the following.

- What are the characteristics of faulty behaviour?
- What kind of data and infrastructure support would be necessary for effective FDD?
- Which techniques are effective in FDD?

The objectives of this thesis, therefore, are to develop a framework for FDD in building systems and to investigate which methods are effective.

1.7 ENERGY INFORMATICS AND THE COORDICY PROJECT

The work done in this thesis was part of the COORDICY project. COORDICY is a joint Danish–American interdisciplinary research project for advancing ICT approaches in the context of energy efficiency in buildings. The project aimed to contribute to Denmark's objective of reducing energy consumption by 50 % and 75 % in existing and newly constructed buildings by 2050 [27], and the USA's objective of doubling energy productivity by 2030 [31].

One of the underlying assumptions of the COORDICY project is that in order to achieve high energy efficiency in buildings or in other sectors, it is essential to make use of techniques from information and communication technology (ICT), energy engineering and computer science. This recent discipline goes under the name of *energy informatics*. Energy informatics studies solutions for increasing the intelligence of energy systems by developing solutions for intelligent control, system diagnostics, performance monitoring and usage analysis.

The main goal of the project is, therefore, to develop a holistic and comprehensive ICT framework for increasing energy efficiency in buildings.

1.7.1 Objective and Hypothesis

Buildings represent a significant part of energy consumption. In the recent years, several countries set their own goal to increase building energy efficiency, such as EU 20 20 [21], the US Better Buildings Initiative [42] or the International Energy Agency EBC programme [43]. The main objective of the project is to use ICT techniques to effectively improve buildings energy efficiency without compromising occupants' comfort.

The project has three main hypotheses, summarized in the following.

Closing the Energy Gap in Energy-Efficient Public and Commercial Buildings

Nowadays, there exist several standard and references to certify high energy efficient buildings, such as ENERGY Star [44], LEED [45] and Green Globes. However, it often happens that their actual performance is significantly worse than the one predicted during design [46, 47, 48, 49]. The cause for this gap is due to the unpredictable effect of numerous factors, such as occupants behaviour, weather conditions, thermal dynamics, construction materials, building systems and control strategies. Comparing the actual performance with the predicted one is a good estimate of the gap.

The project hypothesis is that ICT methods can be used to estimate the performance gap and that data-driven diagnostics methods can be used to precisely identify its causes.

Advancing energy performance by increasing the building's intelligence quotient

It is an accepted notion that increasing building intelligence has positive effects on the energy efficiency of buildings. However, the trade-off is that high intelligence adds complexity during commissioning and operation. In the latest years, a new notion of building intelligence quotient suggests that buildings where systems are coordinated together have higher intelligence than buildings where systems are managed independently.

The project hypothesis is that energy efficiency can be improved through increasing building intelligence by combining multi-objective coordination of decentralized systems with predicted energy performance. This can allow buildings to meet their design performance and even to advance their energy class.

Balancing deep energy-retrofits and building intelligence quotient

Existing buildings, constructed in the past decades when laxer regulations were in force, can undergo energy-oriented retrofit operations to improve their energy efficiency. However, extensive retrofit operations might not be the most cost-effective solution to improve energy efficiency. Since energy performance strongly depends on occupancy behaviour, significant improvements can be obtained by increasing building intelligence. Therefore, the optimal approach is a balance between the two strategies.

The project hypothesis is that using an energy model of the building to simulate several trade-offs between retrofit operations and an increase of building intelligence can estimate the most cost-effective strategy for performance improvement.

1.7.2 Project organization

The project is organized into seven work packages.

Work package 1: Sensing and Modeling of Occupant Behavior

The objective of this work package is to develop methods for identifying, estimating and predicting the behaviour of building's occupants.

Work package 2: Building Modeling and Simulation

The objective of this work package is to develop a model for simulating the building's operations, with a particular focus on its energy consumption. The model makes use of the geometrical, structural and thermal properties, together with occupancy models from work package 1 and weather forecast.

Work package 3: Building Operating System Services Platform

The objective of this work package is to develop a platform with support for construction typologies, thermal properties, building systems layout, and occupant detection.

Work package 4: Multi-objective Coordination Framework

The objective of this work package is to develop a multi-objective coordination framework for controlling the building operations to achieve optimal energy performance. The framework is able to follow multiple objectives over decentralized and decoupled building systems.

Work package 5: Building Diagnostics Framework

The objective of this work package is to develop a framework for detecting and diagnosing the causes for the difference in actual and predicted energy performance.

Work package 6: Tool Suite Construction

The objective of this work package is to develop a suite of tools integrating the results of the other work packages.

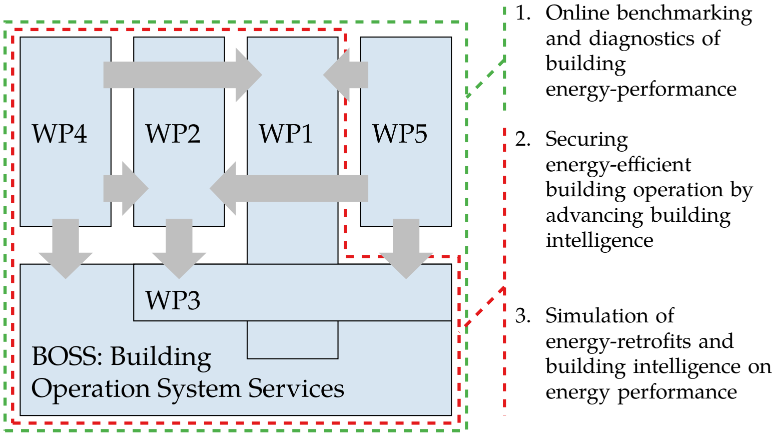

Figure 1.7: Relations between work packages in the project COORDICY.

WORK PACKAGE 7: CASE STUDIES

The objective of this work package is to investigate the three research questions of the project on real buildings.

The buildings considered for this work package are:

- GreenTech Center, at GreenTech;
- Building Odense undervisning 44 (OU44), at University of Southern Denmark;
- Sustainability Base, at NASA Ames Research Center;
- Sutardja Dai Hall, at University of California, Berkeley.

The relations between the work packages in the project COORDICY are summarized in Figure 1.7.

The work performed during this Ph. D. is mainly focused on work package 5. However, some of the tools and methods developed have impacted on other work packages as well. The online energy simulator, presented in Chapter 6, has been a significant part of work package 2, and it has been used for testing the control framework in work package 4. Software developed internally to the center, such as building drivers to access weather forecast, and software libraries to access real-time data storage, are related to work packages 6 and 3. Finally, all the methods presented in this thesis have been deployed or tested on building OU44, which is one of the case studies of work package 7.

1.8 METHODOLOGY

The methodology used in this thesis is loosely inspired by the 'case-study' methodology, which was thoroughly discussed by Runeson et al. in the field of software engineering [50]. The methodology is akin to the experimental-setup methodology, where the researcher designs a controlled experiment to investigate a phenomenon, eliminating most of the disturbance and external influences. In the case-study methodology, on the other hand, the phenomenon is studied in its natural context.

The case-study methodology consists of the following five major steps.

1. Design of the case study;
2. preparation for data collection;
3. collection of the evidence;
4. analysis of the collected data;
5. reporting of the results.

At the beginning of the project, an extensive literature review has been performed. The major databases collecting publications and research papers in the field of FDD for building systems were queried, and a list of publications was obtained. Those publications were read and categorized according to the methods used and the specific building equipment they focused on, and some of the most interesting ones were selected. This initial step was useful to gather an initial overview of the field and the popular techniques and methods, and to understand the current state of the art.

The second step was to analyse the available case studies. Buildings from a few institutions were available within the project, in particular, building OU44 which is physically located on campus. The following characteristics of building OU44 has been reviewed, by reading its technical documentation, interviewing the maintenance crew, and performing field investigation.

- Physical envelope;
- technical subsystems and equipment;
- building usage;
- data sensing and collection infrastructure;
- data distribution infrastructure;
- control strategies.

Building OU44 offered ambitious features, such as real-time collection of a large amount of data and a detailed dynamic energy performance model of the building which can be used for accurate simulations. However, the sensing and distribution infrastructure was lacking in many aspects and the simulation workflow was cumbersome. An action plan was defined to address these inadequacies by developing necessary tools and libraries, in what could be defined as the preparation for data collection.

Libraries for accessing the data storage system were developed for two popular programming languages, Python and Java. Those were robust and featureful libraries supporting both batch and real-time data processing, which have been improved over the Ph. D. project and were also used by other projects at the center[1]. Additional software tools were developed, such as building drivers for collecting additional data, and middleware to ease the simulations of the building's behaviour.

Requirements for the software libraries and tools were not exhaustively defined at the beginning of this step. Instead, a more agile and flexible development approach has been used, and development has been alternating with requirements definition during the project.

Once these initial steps were completed, several iterations of evidence collection, data analysis and reporting were performed. In each of these iterations, a specific niche in the field was identified in the context of the project and of the research questions.

[1]Some of the libraries were released to the public, and they are available at `https://sdu-cfei.github.io/cfei-smap/library.html` for Python and `https://sdu-cfei.github.io/java-libraries/` for Java.

The case-study methodology was then repeated at a lower scale. Objectives were defined, and a case study was designed for the specific iteration. A specialized data analysis tool was developed to perform FDD on the case study. Finally, a report of the findings was compiled, often in form of a scientific paper submitted to a peer-reviewed journal or conference.

1.9 STRUCTURE OF THE THESIS

The rest of the thesis is structured as it follows. The current state of the art of the field is reviewed in Chapter 2, where its gap is highlighted, and the contribution of the thesis is presented. Chapter 3 contains a short summary of the publications included in the thesis. In Chapter 4, the main case-study building OU44 is presented.

Part II contains the individual publications, both published and under review, included in this thesis. In Chapter 5, the topic of data validation is presented, together with a practical experience of sensors faults identified in the building. In Chapter 6, an online energy simulator is presented as a tool for hierarchical FDD on buildings. In Chapter 7, linear regression virtual sensors are used to introduce redundancy in ventilation units, and a FDD method is proposed to exploit such redundancy. The same method is improved in Chapter 8, where non-linear and statistical models are used to increase accuracy. In Chapter 9, consensus among multiple peers is exploited to remove the requirement of fault-free training data in data-driven FDD methods. Finally, the impact of faults on the building's energy consumption is assessed in Chapter 10, where a FDD method for ventilation units is proposed.

Part III concludes the thesis. In Chapter 11, future research directions are suggested in the context of the thesis. The findings of the thesis are summarized and elaborated in Chapter 12.

CHAPTER 2

STATE OF THE ART

In this chapter, the state of the art for fault detection and diagnostics (FDD) in building systems is reviewed. In addition to that, a short review of the state of the art is also presented for virtual sensing techniques and building simulations, which are fields covered in some of the publications included in this thesis. Finally, the gap in the state of the art is identified, and the contributions of this thesis—and of the individual publications included—are summarized.

2.1 FAULT DETECTION AND DIAGNOSTICS METHODS

The panorama of FDD for building systems in the past two decades has been reviewed by Katipamula et al. in 2005 [39, 40], and their work was updated by Kim et al. in 2018 [41]. The authors present a comprehensive review of almost 200 publications.

The authors propose two different schemes for classifying FDD studies, one based on the approach used, and one based on the specific equipment under test. In the former, studies are divided in process-history-based, quantitative-model-based and qualitative-model-based methods, and are further divided according to the specific technique, as shown in Figure 2.1. Each of these families of methods has different advantages, disadvantages and trade-offs, as well as implementation constraints and caveats.

In the latter, studies are divided depending on whether the method was applied to systems or equipment available in small buildings, large buildings, or both, as shown in Figure 2.2. Studies are further divided depending on the specific system or equipment: air conditioners and heat pumps, chillers and cooling towers, air handling units (AHUs)

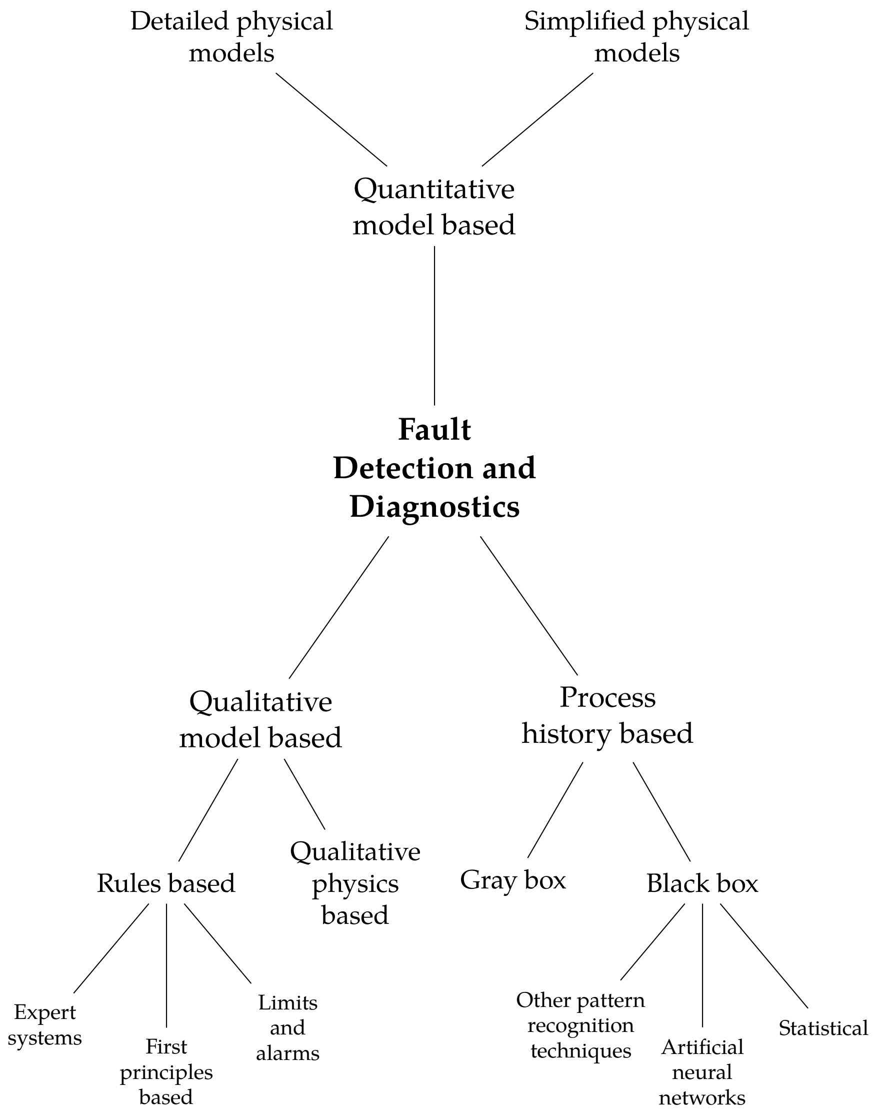

Figure 2.1: Classification of FDD methods based on the approach [41]. In quantitative-model-based methods, a detailed model of the system is created from first principles. In qualitative-model-based methods, qualitative relationships between the systems components are used to develop a model of the system. In process-history-based methods, historical data is used to train a model of the system .

and variable air volume (VAV) units, fan coil units, commercial refrigerators, lighting equipment, water heaters, or the entire building. Each building system presents unique challenges and has a different impact on both building operations and energy consumption.

2.1.1 PROCESS-HISTORY-BASED METHODS

History-based methods, often called *data-driven* methods in the literature, exploit historical data for FDD. Historical data is used to train a black-box model of the system under test, which is used to predict and validate the system itself.

Most methods define two separate phases: an offline 'training' phase and an online 'testing' phase. In the former, historical data is fed to the black-box model and used to estimate its hidden parameters. The output of the training phase is a black-box model which predicts the state of the system under test. The training phase occurs only once, or once in a while, when a new model must be generated, and does not have significant execution time constraints. Therefore, complex techniques which require significant training time and computational power are still suitable for this approach.

In the testing phase, the black-box model is used to predict and validate the state of the real system. When the system is operating normally, its behaviour matches the one learned by the model. Faults, on the other hand, impact the behaviour of the system such that it no longer matches the predictions from the model. Therefore, when the system deviates from the model a fault is detected.

Black-box models represent mappings between input and output variables, without any information about the physics of the system. For this reason, data-driven methods can be applied when the internals of the system are unknown, as long as historical data is available. This has another advantage: a data-driven method can be seamlessly applied to different systems without significant changes. After the method undergoes a new training phase, which does not require any customization or often even manual operation, a new model is trained with historical data from the new system and it can be deployed.

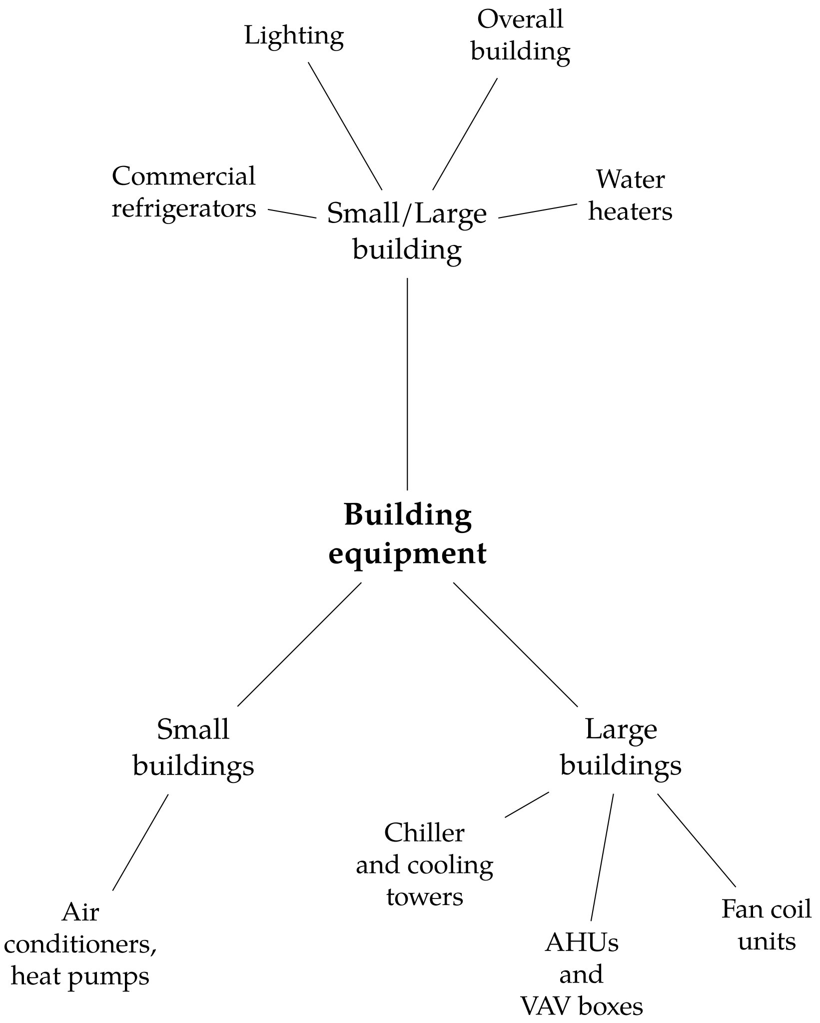

Figure 2.2: Classification of FDD methods based on the building equipment [41]. Air conditioners and heat pumps are usually only found in small buildings. Large buildings, on the other hand, have often larger equipment such as cooling towers, AHUs and fan coil units. Some components, such as water heaters and lighting infrastructure, are present in both small and large buildings .

A single model can only learn the behaviour of a healthy system. When its predictions stop matching the real system, the only possible conclusion is that the system is no longer healthy. A popular approach to extend data-driven methods, so that they can be used to precisely diagnose the specific fault affecting the system, is to train additional models of the faulty systems. In this case, the state of the system is compared with predictions from several models, each representing the system under a specific fault, including the special case of no faults. The model matching the current state of the system is the one corresponding to the current fault. If no models match the current state, an unknown fault is happening.

The main disadvantage of data-driven methods is the requirement for historical data. Training data must be available for all operational profiles of the system, e.g. during both summer and winter and during working and non-working hours. This requirement makes deployment on newly constructed buildings impractical, since historical data is not available during the first operational period. For the same reason, these methods cannot accurately model transients and other fast dynamics phenomena, or infrequent operations which do not appear often in historical data.

Another issue is that historical data must be *fault-free*, i.e. it must be collected from a healthy system. If instead it was collected from a system affected by a fault, the resulting model would represent a faulty system and would hence classify faulty behaviour as healthy. This issue is particularly significant for faults present since construction: it is impossible to recognise an anomalous behaviour by only looking at the process history if it was present since construction. If used for fault diagnostics, data-driven methods require *labelled faulty* historical data, i.e. data generated by systems affected by known faults, which are rarely available.

Another disadvantage of data-driven methods is their inaccuracy. Simple and understandable models, e.g. linear regression models, cannot capture complex interactions between inputs and outputs. On the other hand, complex models such as artificial neural network (ANN) or statistical machine learning techniques are able to learn complex interactions but suffer from increasing complexity and overfitting.

Therefore, without using knowledge about the physics of the system it is difficult to accurately predict its behaviour.

2.1.2 Quantitative-model-based methods

Quantitative-model-based methods, often called simply *model-based methods* in the literature, make use of detailed knowledge of the system. A model of the system under test is created from first principles, representing the mathematical dependencies and interactions between its components. Similarly to process history based methods, the model is used to predict and validate the state of the real system.

In quantitative-model-based methods, the 'training' phase is performed manually by scientists and technicians, who create the model by studying in detail the system. Depending on the complexity and accuracy of the model, this step can involve the estimation of a few parameters, which can be done by using historical data. The final output is again a model that predicts the behaviour of the system under test.

Quantitative models are often much more accurate than black-box models, due to explicit modelling of physical interactions. For instance, a black-box model can only learn the relation between indoor and outdoor temperatures from data alone. A physical model, on the other hand, could estimate the heat loss, using the heat transfer equation and the precise coefficients for the wall materials and the heat gain, using information on the weather forecast, radiators layout, and building schedules.

As mentioned before, data-driven methods require a significant amount of historical data covering all the operational profiles. Quantitative models, on the other hand, can predict the behaviour of the system during transients and fast dynamics phenomena, as well as during infrequent operations. The accuracy of quantitative models is, in general, limited by the accuracy of the model, which is a trade-off with its complexity.

In order to properly diagnose the specific fault affecting the system, different models are used to predict the behaviour of faulty systems, such as in data-driven methods. For quantitative model based methods, however, labelled faulty historical data is not necessary. The model can instead be directly modified to represent a faulty system. E.g. the efficiency of a fan can be lowered to simulate wearing, or readings from a sensor can be set to a constant value to simulate a stuck sensor.

The main disadvantages of quantitative-model-based methods are the complexity of the models and the time and effort required to create them. For data-driven methods, the training phase can last for a few days when using the most complex black-box model over a large amount of data. On the other hand, creating a physical model of the system can take several weeks and may require to analyse in details the documentation of the system, or even on-site investigation, e.g. to confirm construction materials or rooms layout.

The higher complexity not only affects the time necessary to create the model, but also the time necessary to execute it. While most of the data-driven methods can be executed immediately on real-time data, quantitative models, depending on their complexity, may take several minutes or even hours. For this reason, it is often infeasible to use quantitative-model-based methods for fast, real-time FDD.

The strong dependency of these models on the specific system under test is one of the reasons for their high accuracy. However, this is also a reason for low extensibility. Not only parameters, but also the structure of the model itself is strongly tied to the system. A different system might have different geometry, or might be ruled by different physical equations. Therefore, a model suitable for one system cannot be used for a different one, and it must be redesigned from scratch.

Finally, complex models often contain parameters that cannot be obtained from the documentation of the system alone. In this case, historical data can be used for parameter estimation. While this is an effective way to estimate parameters, it carries along the same drawbacks from data-driven models, i.e. the necessity of fault-free historical data or the impossibility of deploying the method on newly constructed systems. While the impact of faulty historical data may be smaller on quantitative models, since it only affects a part of it, it could decrease the accuracy.

2.1.3 Qualitative-model-based methods

Qualitative-model-based methods, often conflated in rule-based methods, make use of a priori knowledge of the system. Such knowledge, usually provided by system documentation or by experts in the field, is used to create sets of rules or qualitative physics systems. These systems represent the behaviour of the system at a higher level and they are easily understandable.

While simple in principle, qualitative-model-based methods have several advantages. Their implementation is straightforward and, when formal documentation is available, they can be derived automatically without human intervention. Field experts can reason about the models and compare them to their understanding of the system. Therefore, models can be validated without the use of historical data.

These methods often represent the state of the system at a coarse level. This makes the methods robust to numerical uncertainties and noise, which do not affect the quality of the results. Qualitative-model-based methods, therefore, are often not significantly impacted by noisy sensors, inaccurate parameters or human error in schedules.

Many commercial building management systems (BMSs) offer rudimentary FDD through rule-based alarms. Alarms can be used to detect faults with respect to the building operations, e.g. when the temperature in the room rises above a threshold, an alarm is triggered. Another usage is to detect unnecessary energy consumption, e.g. an alarm can be triggered when lights are on in an empty room, or when both heating and cooling are on in the same space.

While the simplicity of qualitative-model-based methods carries many advantages, it also comes with trade-offs and drawbacks. Complex dynamics require a huge number of narrow and specific rules, which would both make the set of rules unwieldy and difficult to maintain, and would remove the ability to reason about them and validate them. Qualitative-model-based methods, therefore, are only effective to represent simple, high-level behaviour.

As with quantitative model based methods, qualitative-model-based methods are strongly dependent on the specific system. A set of rules designed for a specific system would not necessarily work for a different one. For the same reason, qualitative physics may differ between the two systems, making a common model impossible or inaccurate. Compared with quantitative models, however, this drawback is slightly mitigated by the simplicity of creating a new model.

2.1.4 MIXED METHODS

Methods from each category have different trade-offs and are suitable for different kinds of systems. Hybrid approaches that make use of multiple methods are also common, in order to exploit advantages and reduce the disadvantages of individual methods. Using multiple methods also increases robustness and reliability: if multiple independent models detect a fault, the chance for a false alarm is lower.

Quantitative models contain sometimes parameters that are difficult or impossible to measure on the real system. E.g. an energy model of the building may require parameters representing the plug load energy consumption due to room level equipment. When available, historical data can be used to estimate such parameters and to increase the accuracy of the model.

Data-driven and quantitative models are often sensitive to noise in measurement and numerical uncertainty, contrary to qualitative models. Fuzzy logic, often used in rule-based methods, can be used to increase the coarseness of data which in turns increases the robustness of the method at the expense of its accuracy.

Accuracy in data-driven methods can be increased by structuring their model with information about the physics of the system. E.g. known relations between input and output variables, such as linear or quadratic, can be enforced by selecting an appropriate regression model. Sensitivity analysis can also be used to filter out variables with a lower impact, reducing the computational power required for training the model.

2.1.5 POPULARITY OF FAULT DETECTION AND DIAGNOSTICS METHODS

Kim et al. report the shares of publications in each category, which are shown in Figure 2.3 [41]. The vast majority of FDD publications make use of process-history-based methods, while only a relatively small number use quantitative-model-based methods.

The authors speculate that the popularity of data-driven methods is due to the reduced modelling complexity. Training a data-driven model requires large computational power, which is becoming cheaper over time. On the other hand, creating a set of rules requires significant human effort, and even more does creating a quantitative model.

Another reason for this skew between the methods could be the increasing availability of data in the field. While historically buildings were not equipped with many sensors and meters, buildings constructed in the past decade, perhaps due to decreasing sensors price and complexity, have extended infrastructure for data collection.

With respect to building systems, AHUs and VAV units attract the most research. AHUs, VAV units, and other heating, ventilation and air conditioning (HVAC) equipment are often the most critical systems in buildings and they account for the largest share of energy consumption. Therefore, addressing faults in these systems has the potential of large energy saving compared to other smaller equipment such as lighting systems or fan coils.

2.2 VIRTUAL REDUNDANCY

Redundancy in a system is an effective tool for FDD. When the same quantity is measured twice, either both measurements match within a certain threshold, or they deviate from each other. In the that case, at least one of the measurements is incorrect. When multiple independent measurements are available, it is possible to precisely isolate the faulty one.

Physical redundancy, however, is seldom available in building systems. Duplicating a sensor, a meter or an actuator increases the cost

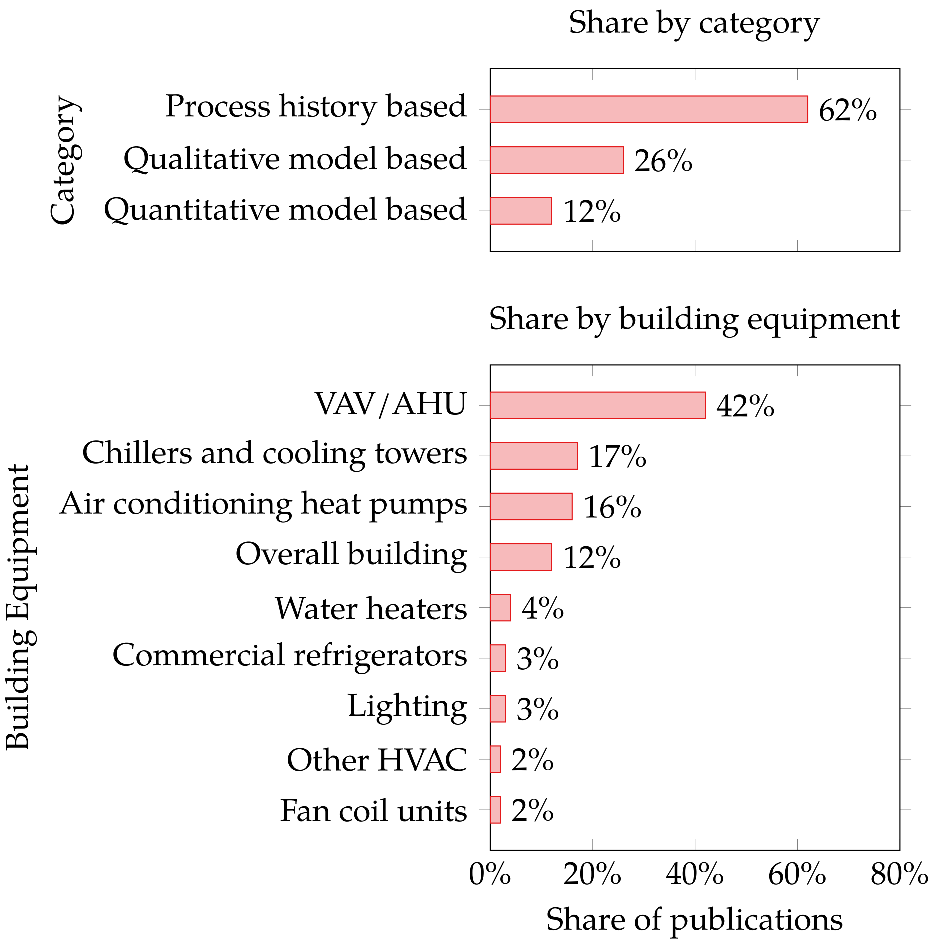

Figure 2.3: Shares of reviewed FDD publications by category and by building equipment. The majority of publications propose data-driven methods, while quantitative-model-based methods are seldom used. AHUs and VAV units are the equipment most covered in publications by a large margin, while for others, such as lighting, only few methods are presented. The review covered 197 publications [41].

and the complexity of the system. In some cases, it is even infeasible to duplicate a component due to space limitations. Virtual redundancy, on the other hand, does not have these disadvantages. Interactions and relationships between quantities inside the system can be exploited to obtain the same readings from different sources.

Li et al. present a comprehensive review of virtual sensing techniques for building systems [51]. The authors propose three categorizations, depending on the characteristic of the measurement, the application purpose, and the modelling method.

In the first categorization, virtual sensing techniques can either be used to measure a quantity during a transient or at steady state. The former case is useful when the system behaviour reacts quickly to changes in input variables, and it is useful for automatic control. When the behaviour has instead slower dynamics, steady-state modelling is usually enough. FDD and performance monitoring applications often fall into the steady-state case.

In the second categorization, techniques are either used for observing hidden state or for monitoring and diagnostics. Some quantities cannot be measured directly, either because of complexity constraints or because it is technically impossible to construct the specific sensor. For instance, no physical sensor could measure the efficiency of a motor, but a virtual one could compute it by dividing the generated workload by the consumed energy. In the second case, virtual sensors are used to introduce redundancy in the system in order to monitor performances and perform FDD.

The final categorization depends on the technical methods used to develop the virtual sensors. The authors follow an approach similar to the one presented for generic FDD methods, i.e. virtual sensors methods are divided between history-based methods, quantitative-model-based methods and qualitative-model-based methods.

2.3 BUILDING PERFORMANCE SIMULATION

Simulating the energy performance of a building is a useful technique for FDD. It makes it possible to assess the building performance at higher design stages but also after construction and during its normal operational life. Moreover, it allows testing control strategies and recommissioning processes without disrupting the operation in the real building.

Many software tools are available for simulation. Two of the most popular ones in the current panorama are EnergyPlus and Modelica. EnergyPlus is a specific engine developed by the Lawrence Berkeley National Laboratory for simulating energy consumption in buildings [52]. Modelica is a more generic software for simulating technical systems [53]. It can simulate an entire building, however, it is more often used to simulate independent subsystems such as AHUs or individual rooms. Other general-purpose tools can be used as well, such as Matlab, Simulink, or even general-purpose programming languages such as Python, Java or C++.

Clarke et al. discuss the overall topic of building performance simulations [54]. Three main aspects are considered as objectives for building performance simulation solutions, i.e.

- high integrity representation, in order to accurately model the real systems;
- coupling of different domain models which may require different modelling strategies;
- design process integration with building constructors.

The current and future requirements for building performance simulation are discussed, accounting for micro-grid infrastructure, integrating urban energy management, and internet energy systems. In particular, the authors stress the lack of collective effort to standardize the design and implementation of solutions for building performance simulation, which results in duplicated work in the field.

The potential of simulating the behaviour of buildings and comparing their expected and real performance for FDD and continuous commissioning is discussed by Costa et al. [55] in the context of the Building EQ program [56]. Visualization techniques are described as a fundamental tool to facilitate manual FDD operated by the building management, and the authors propose a layered visualization strategy, where detailed information is hidden until a fault is suspected at a higher level. Several visualization techniques are shown and suggested for specific FDD instances. Finally, building performance simulation are a useful tool for testing different control strategies.

2.4 CONTRIBUTIONS

The contribution of this thesis is two-fold. On one hand, the topics covered by the publications included in this thesis and their sequence were carefully laid out to produce a holistic contribution to the field at a larger scope. On the other hand, each publication covers a well-defined niche in the field and advances the current state of the art within its scope.

2.4.1 HOLISTIC CONTRIBUTION

FDD for building systems is a relatively young field and, while it spun the interest of many researchers who presented interesting, effective and advanced methods, it is still far from maturity. In particular, no comprehensive frameworks for FDD that cover the entirety of building systems have been proposed, and most of the publications only focus on ad hoc techniques for individual components or subsystems. Therefore, while many techniques have been proposed, no integrated solution that can be easily deployed on a real building is available. Commercial solutions are also at their infancy and, while some BMS offer FDD features, they are usually rudimentary, such as threshold-based alarms or simple rules, and advanced techniques are not yet implemented.

In this thesis, an integrated framework for FDD in building systems is presented. The framework, whose main foundations are presented in Chapter 6 [2], considers the following system hierarchy.

1. *Whole building*.
2. *System*. One of the main building systems, such as HVAC, lighting, or air distribution system.

3. *Subsystem*. A smaller section of a system. E.g. for HVAC, it could be a specific ventilation unit, for the lighting system, it could be a lighting zone and, for air distribution systems, it could be an individual room.
4. *Component*. An individual component inside a subsystem. E.g. a sensor in a ventilation unit or in a specific room.

At first, the infrastructure for sensing and collecting building data is validated. Correct building data is necessary for any application, including FDD. This aspect is described in Chapter 5 and in the conference paper 'A Practical Approach to Validation of Buildings' Sensor Data: a Commissioning Experience Report' [1].

Once the infrastructure for reading the state of the building can be trusted, faults are detected using a top-down approach by comparing the building's energy performance with the expected one, obtained from simulation using its dynamic energy performance model. If the overall performances are lower than expected, they are compared recursively with less aggregate meters, by traversing the energy distribution tree, until the specific under-performing subsystems are identified. This aspect is described in Chapter 6 and in the journal paper 'Online Energy Simulator for Building Fault Detection and Diagnostics Using Dynamic Energy Performance Model' [2].

At this point, specialized FDD methods for the specific subsystems can be used to detect and precisely diagnose the component responsible for the reduced performances. Due to their large contribution to energy consumption in buildings, in this thesis, the focus has been on ventilation units and air distribution systems. Chapters 7 to 10, and their relative publications [3, 4, 5, 6], present some specialized methods for FDD on these systems.

In conclusion, this framework allows to systematically link reduced performances of a building to the faulty subsystem and, afterwards, to the faulty component.

Table 2.1 shows the stack of the building hierarchy covered by the publications included in this thesis. The publications cover the entire stack of building systems, from the whole building layer to the individual components.

Table 2.1: Stack of the building hierarchy covered by each publication. The publications included in this thesis cover the entire stack of building systems, from the whole building layer to the individual components. .

	Chapter 5	**Chapter 6**	**Chapters 7, 8**	**Chapter 9**	**Chapter 10**
	Sensors validation [1]	Online simulator [2]	Virtual sensors [3, 4]	Consensus [5]	Faults simulation [6]
Whole Building					
System					
Subsystem					
Component					

2.4.2 CONTRIBUTIONS OF INDIVIDUAL PUBLICATIONS

Each of the individual publications included in this thesis, available in Chapters 5 to 10, has a contribution in its own specific niche.

Moreover, the methods used in these publications cover all the main groups in the categorization of FDD methods presented by Kim et al. and summarized in Section 2.1, i.e. data-driven, model-based and rule-based methods. Table 2.2 summarizes the methods used by each publication.

SENSORS VALIDATION

Many publications, especially those that use simulations, assume ideal infrastructure conditions. Data is assumed to be available continuously with no gaps and to be collected with a fixed, common frequency. This assumption, however, does not necessarily hold for real-world systems. In Chapter 5, a framework for validating input data is presented [1]. A particularly novel contribution was a test for detecting missing data when threshold-based sensors are used. Those sensors report readings with time-variant frequency, for which common methods in the literature would produce a high number of false alarms.

Due to the complexity and cost of building systems, it is often impractical to build a custom 'testing' unit, and, instead, researchers can only study real 'production' ones. This makes it difficult to collect data from genuinely faulty systems, since causing faults on actually used buildings might deteriorate or disrupt their operation. Therefore, most research is done using simulated data, real data with simulated faults or, rarely, real data with artificially induced faults. In Chapter 5, instead, an actual, unplanned fault on an existing building is detected and reported to the facility management. Later, it could be observed from data that the fault was fixed.

ONLINE ENERGY SIMULATOR

As mentioned in Section 2.3, using simulations is a popular approach for generating data and testing FDD methods in building systems.

Table 2.2: Methods used in each publication with respect to the common classification of FDD methods, i.e. data-driven, model-based and rule-based. Each of the main methods is covered by one or more publications.

	Chapter 5 Sensors validation [1]	**Chapter 6** Online simulator [2]	**Chapters 7, 8** Virtual sensors [3, 4]	**Chapter 9** Consensus [5]	**Chapter 10** Faults simulation [6]
Data-driven					
Model-based					
Rule-based					

Moreover, simulations are useful for other applications such as evaluating recommissioning options, control strategies, and occupancy analysis. The current state-of-the-art building simulation engines, however, require significant manual operations and are difficult to automate.

Users need to locate and fetch input data for the simulation. Data must often be preprocessed, multiple time-series might have to be merged, resampled, or converted to a different unit. Finally, they need to be formatted according to the specific simulation engine. Users must manually run a simulation, sometimes modifying the model itself to point to the current data files, and specifying the desired simulation period. Sharing the results with other researchers and making them available to other applications is also problematic. Simulation results are usually available as unstructured comma separated value (CSV) files, which lack metadata, and are so large that are difficult to share by email or any other traditional way. Moreover, these procedures are specific for each simulation engine, which only adds complexity and possibility for human errors.

In Chapter 6, a tool for automating these procedures is presented [2]. The tool can locate and fetch data from a common data storage system, run the simulation on the desired period, and make the results available on the same data storage system. The simulation is run through a common interface implemented by many engines, which makes it easier to replace the model. Similarly, data is exchanged through a common protocol that can be implemented by applications and data storage systems. Human intervention is only necessary once, to generate an initial configuration, and the tool can regularly run simulations on real-time data by itself.

VIRTUAL SENSORS

Virtual sensors are an effective method for introducing redundancy in a system without the drawbacks of physical redundancy, such as increased cost and complexity. They are especially useful in ventilation units, where many sensors are available, though rarely duplicated. Some authors applied virtual sensors in building systems in the past, however, they are often used for observing unmeasured quantities rather than for diagnostics.

In Chapter 7, a method for creating virtual sensors in ventilation units using linear regression from historical data is presented [3]. While model-based virtual sensors in ventilation units are available in the literature, data-driven methods lack in this specific topic. Data-driven methods trade accuracy for simplicity, and have the advantage of not depending on a detailed description of the specific system. They are, therefore, easy to deploy to different ventilation units.

This method is improved in Chapter 8 [4], where linear regression models are extended in two orthogonal directions to account for interactions not accurately representable through linear relationships. First, by more accurately representing relationships between measurements using non-linear models. Secondly, by taking into account recent history using statistical models.

Consensus-based anomaly detection

Data-driven methods are popular for FDD in building systems, as shown in Figure 2.3 on Page 39. Their major drawback, however, is the requirement for fault-free historical data, which is rarely available. For this reason, they are difficult to deploy in practice.

Consensus-based methods, on the other hand, trade this requirement with one for multiple identical components. The aggregate behaviour of a group of components gobbles the small contributions from the faulty ones and can, therefore, be used to train data-driven models. In Chapter 9, a consensus-based method for FDD at room level is presented [5]. The interactions between VAV units and CO_2 level are converted to a set of qualitative episodes, which are weighted with their number of times they occur. VAV units which exhibit frequent uncommon episodes are identified as anomalous and are flagged for additional investigation.

Consensus-based methods are popular for control and decision support, however, they are rarely used for FDD, especially in building systems. Besides the specific anomaly detection method, therefore, the contribution of this chapter is to promote consensus-based methods.

FAULTS SIMULATION

Ventilation units are one of the most complex and critical systems in buildings, and often responsible for significant energy consumption. For this reason, many researchers work on FDD methods applied to ventilation units and their components. However, some of the components have received more attention than others. Many publications focus on VAV units, due to their immediate effect on indoor air distribution, on chillers and refrigeration systems, probably due to their widespread presence in warmer climates, and on heating coils, which are significant in colder areas. Some other components such as heat exchangers (HXs) have instead been overlooked or even ignored in the literature.

In Chapter 10, a rule-based method for FDD in ventilation units is presented and specifically applied to hot-water loops and heat exchangers [6]. Under-performing subsystems are first isolated using the method presented in Chapter 6, and a set of rules is then used to diagnose the specific fault. The dynamic energy model of the building is used to simulate faulty conditions and the method is tested on such generated data. The impact in terms of increased energy consumption is also reported for the selected faults.

Moreover, in the chapter, two experiments are defined where gradual faults are introduced in the building. Gradual faults can be detected before their impact becomes significant, allowing pre-emptive maintenance. This is a first step in the direction of *faults prediction*, which aims to detect faults before they even manifest, improving maintenance schedules and reducing operation disruption.

CHAPTER 3

A HIERARCHICAL FRAMEWORK FOR DATA VALIDATION AND FAULT DETECTION AND DIAGNOSTICS IN BUILDING SYSTEMS

In this chapter, the publications included in the thesis are summarized in the context of a hierarchical framework for data validation and fault detection in building systems.

3.1 SENSORS VALIDATION: A PRECONDITION TO ANY BUILDING APPLICATION

In buildings, a large number of heterogeneous devices are used to generate data. Physical sensors measure indoor and outdoor environmental conditions such as temperature, CO_2 concentration level, humidity, light intensity and weather conditions. Other sensors measure the status of building's equipment such as valve position of variable air volume (VAV) units and radiators, artificial light level, or thermostat's status. Energy meters measure the energy consumption of different subsystems such as lighting, rooms plug load, heating, ventilation and air conditioning (HVAC), elevators or other building's facilities. Where available, occupants counting camera report the number of people who entered the monitored areas.

All buildings applications use part of the recorded data to perform their tasks. The building management system (BMS) uses the current building status, sometimes augmented with weather and occupancy forecasts, to control the actuators and activate the HVAC systems. Forecasting applications use historical data in their model, either for parameter estimation or for statistical and black-box model construction. Fault detection and diagnostics (FDD) applications compare the current status of the building with the expected one to confirm that the operations are correct. Building's data is, therefore, a fundamental component of every building application.

Measuring devices, however, are physical devices and are subject to wearing, misconfiguration and failures. Therefore, the very first step in FDD must be the validation of building's data.

Typical physical sensors faults are the following.

Abrupt failures Sensors readings are stuck to a fixed value, which can be plausible or implausible (e.g. negative CO_2 concentration level).

Biased output Sensors readings have a fixed offset, e.g. reported temperature is always 5 °C higher than the actual value. This is often caused by improper, or even entirely missing, calibration.

Precision degradation The variance of sensors readings increases, which results in increased noise in the signal.

Data issues are not only caused by physical sensors faults, but also by issues in the sensing infrastructure.

Data communication failures A sensing network is necessary to propagate readings from the sensors' location to a central data storage. Such network is itself subject to faults and other issues. Network nodes and connection can fail and prevent recording of measurements from sensors.

Infrastructure misconfiguration Many components in the sensing infrastructure have configuration parameters that can be wrong due to mistakes in the initial configuration, or unrecorded changes over time. Sensors' and meters' unit can be different from the actual measurements, e.g. kW h when the measurement is in J. Synchronization issues between subsystems can disorder time-series, e.g. misconfigured timezone or drifting clocks.

Sensors wrong placement Some sensors are susceptible to placement. E.g. temperature sensors placed over a radiator would overestimate the room temperature when the radiator is on. In the same way, CO_2 concentration level sensors located next to a window would be sensitive to air drafts. Finally, large objects could shield people from occupancy sensors, which would incorrectly record a room as empty.

In Chapter 5, a method for sensors validation is presented. A set of rules defining the expected characteristics of time-series are used to define validation tests for their readings. Four tests are defined.

Range test Readings must fall within the physical range of their time-series. E.g. indoor temperature must be within 15 °C to 30 °C, CO_2 value must be strictly positive.

Latency test In threshold-based sensors, readings are recorded with varying frequency, which makes it difficult to detect missing data. A maximal delay is defined for each time-series, based on the dynamics of the system.

Spike test Due to physical constraints, the quantity measured by a time-series cannot change too sharply in a short time. A maximal absolute variation is defined for each time-series based on the physical properties of the measured quantity.

Monotonicity test Some time-series record cumulative values and can only increase over time.

3.2 A HIERARCHICAL FRAMEWORK FOR FAULT DETECTION AND DIAGNOSTICS

In building systems, energy enters the system at one location, and it flows to the rest of the components through an *energy distribution tree*. Separate energy meters are usually located on some of the nodes and measure the amount of energy consumed by all nodes below them. An example of an energy distribution tree is shown in Figure 3.1.

A dynamic energy performance model of the building is used to simulate its behaviour, in particular, to obtain the expected energy

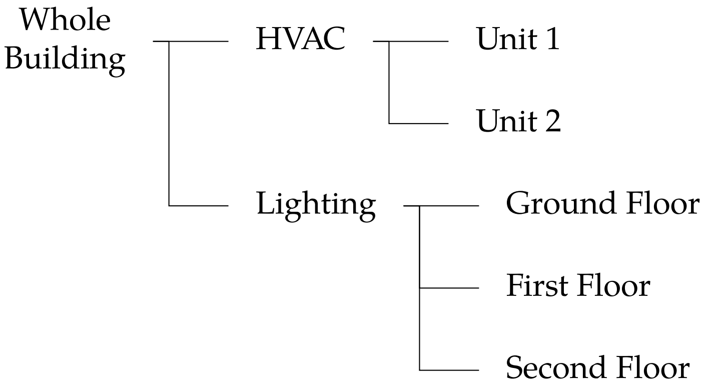

Figure 3.1: An example of an energy distribution tree in a building.

consumption at each node in the energy distribution tree. The model represents with high accuracy the physical envelope of the building, including zone shapes, specific materials, wall sizes, and windows location, but also the operation of all relevant subsystems, such as HVAC, down to individual components and ducts level, lighting and district heating. Given the weather forecast, schedule configuration and an occupancy profile, the model can generate the expected energy consumption, but also the expected indoor conditions.

When the building's energy performance is close to the expectation, the building is considered operating correctly. If, on the other hand, the building is consuming more energy than expected, lower nodes in the energy distribution tree are visited in order to isolate the underperforming subsystems, as shown in Figure 3.2. Once such subsystem is isolated, the faulty component is diagnosed using a specialized FDD method for the specific subsystem.

The framework considers the following hierarchy of four layers.

Level 1. Whole building energy performance assessment. At the top of the hierarchy, the overall energy performance is compared with the expected one. If the building performs as expected, no faults are significantly affecting it. Otherwise, further investigation must be performed.

Level 2. System level. At this level, the performance at the system level

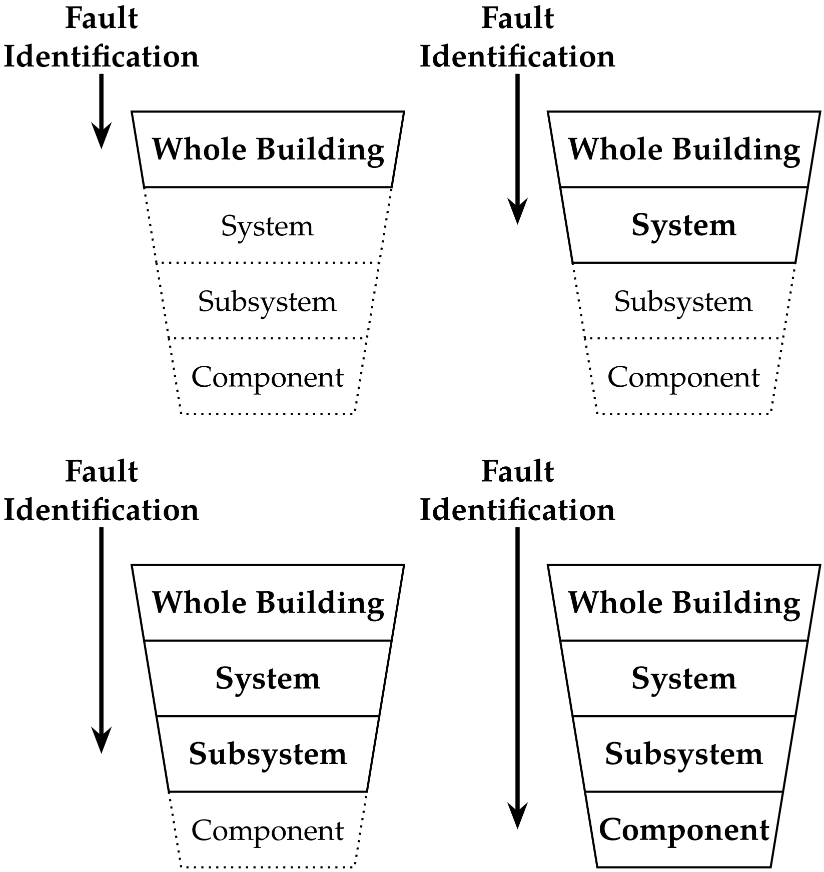

Figure 3.2: Hierarchy for buildings fault detection and diagnostics. Deviations from expected performance are first detected at the whole building level, then the scope is restricted to under-performing system and subsystem. The faulty component is finally diagnosed by using specialized FDD methods.

is compared with the expected one. The affected system, such as ventilation, lighting or heating is isolated, and the scope of the investigation is reduced.

Level 3. Subsystem level. At this level, the specific subsystem is isolated. E.g. when investigating the ventilation system, the specific unit is found or, when investigating the ventilation distribution system or the lighting system, the specific room is found.

Level 4. Component level. At the bottom of the hierarchy, the specific faulty component is isolated. E.g. the specific sensors or actuator inside a ventilation unit or inside a room.

Several simulation engines exist for building performance simulations, such as EnergyPlus [52] or Modelica [57]. Most of them, however, require considerable human intervention for setting up and running the simulation. Data needs to be collected and converted to a suitable format, the model should be configured with the specific inputs file paths, and output results must be manually shared.

In Chapter 6, a tool is developed for automating these procedures. The tool uses two standard interfaces: the simple measurement and actuation profile (sMAP) for data acquisition and publishing, and the functional mock-up interface (FMI) for abstracting the simulation execution. After preparing an initial configuration, the tool can run periodically, e.g. once per day, and generate the results in a common data storage.

3.3 ARTIFICIAL REDUNDANCY WITH VIRTUAL SENSORS

A system has redundancy when multiple components serve the same purpose, and the system could continue operating when one of them is not working. This concept can refer to actuators, sensors, and other equipment. E.g. in a system with two fans, if one is broken, the other is still able to achieve the same airflow, even if at a lower efficiency. In sensing equipment, redundancy makes it possible to validate readings among a group of sensors. E.g. if the temperature is measured by two sensors, they must agree, otherwise, one of them is faulty.

Redundancy is an effective resource for FDD, however, physically duplicating sensors and actuators is seldom feasible. Additional components increase the cost, and sometimes physical constraints such as size prevent to achieve physical redundancy.

Virtual sensors, on the other hand, have no additional cost and are not subject to physical constraint. A virtual sensor consists of a model which computes its readings from other inputs, i.e.

$$S' = f(S_0, S_1, \dots, S_n).$$

Virtual sensors have two main purposes, i.e. observation and diagnostics. In the former, a virtual sensor is designed to estimate a quantity that is not, or cannot, be measured. E.g. there are no sensors which can measure the efficiency of an engine, however, a virtual sensor could estimate it, by comparing the generated physical work with the consumed energy. In the latter, a virtual sensor is designed to estimate the same quantity of a physical sensor and, therefore, validate its readings.

In Chapter 7, a method for designing virtual sensors inside ventilation units is presented. Ventilation units contain several physical sensors and, therefore, are good candidates for deploying additional virtual sensors. The quantities measured from each physical sensor are not completely independent, instead, they are correlated with each other.

Virtual sensors are constructed to replicate the values of existing sensors. Virtual readings are computed from linear regression models using other sensors as input. Three physical sensors are considered, i.e. temperature, fan speed and airflow and, for each of them, two virtual sensors are constructed using different inputs.

Two different metrics are defined to determine whether two sensors deviate from each other, i.e. the coefficient of determination and acceptable intervals. The coefficient of determination is a measure of how much a linear model fits the real data. Higher values indicate that the virtual sensor closely follows the physical ones and, therefore their readings agree. Values close to zero, on the other hand, indicate that the two sensors disagree.

In acceptable intervals, virtual sensors are configured to return a band instead of a single reading. If the values from the physical sensor fall within such band, the two sensors agree, otherwise, they disagree. The width of the band is computed using the maximal error obtained during the training phase.

In Chapter 8, the same method is improved using non-linear regression models and statistical models. The two alternatives improve linear regression models along two different directions. In the former case, more complex dynamics between simultaneous readings from other sensors can be represented using non-linear relationships. In the latter case, linear regression models are augmented with the recent history of the duplicated sensor.

3.4 CONSENSUS-BASED ANOMALY DETECTION

Data-driven methods have many advantages, e.g. they do not require a detailed understanding of the system under test and they can be easily generalized and adapted to different systems. They have, however, one major drawback. Namely, they require fault-free historical data, which is seldom available in real systems. This makes it difficult to deploy those methods in practice.

When a system consists of multiple identical or similar components, however, a different approach is possible. Assuming the majority of the components is working correctly and only a small part of them are faulty, aggregate historical data can be used for training data-driven models. The contribution of the faulty components is diluted among the healthy ones and it will not affect the resulting model, as illustrated in Figure 3.3. Moreover, faulty components can be isolated because their behaviour differs from the aggregate consensus.

In Chapter 9, a consensus-based method for anomaly detection on air distribution systems is presented. The air distribution system is a good candidate for consensus-based methods, since VAV units and rooms are often similar in size and operations. Even when rooms are

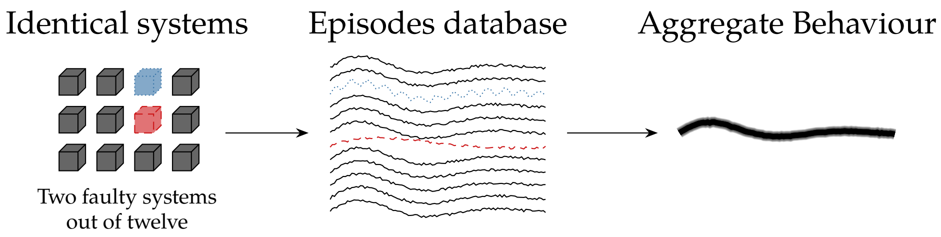

Figure 3.3: Effect of faulty systems in consensus-based methods. If a small number of systems used in training are faulty, their contributions will be diluted among the ones by correctly operating systems.

not completely identical, it is at least possible to find groups of similar rooms depending on some measure such as the following.

Grouping by room type Rooms of the same type such as offices, conference rooms, kitchens or corridors, have similar size and usage patterns.

Grouping by location Rooms in the same location may be subject to the same external interferences. E.g. rooms on the same side of the building have the same solar radiation patterns, and rooms on the same floor have similar insulation due to other floors.

Grouping by related subsystem Sometimes different groups of rooms are related to different subsystems, such as ventilation shafts or lighting floor structure.

Time-series data related to the air distribution system, i.e. room CO_2 concentration level and VAV unit position, are preprocessed to extract qualitative events and episodes. A database of common episodes is generated and updated over time. Rooms whose VAV units often exhibit uncommon episodes are flagged as anomalous.

In order to avoid classifying as anomalous rooms which are actually belonging to a different group, two instances of the method are executed. In one, rooms are grouped by ventilation unit, in the other, rooms are grouped by type. The same rooms are flagged as anomalous in both experiments.

3.5 FAULTS SIMULATION AND IMPACT ASSESSMENT

Buildings are large, complex and, in particular, expensive systems. Testing FDD methods on building systems is difficult, since it is rarely feasible to artificially cause faults without disrupting the normal activities and operations in the buildings. A popular approach, instead, is to use a model of the system to simulate its behaviour when affected by faults.

In Chapter 10, a method for FDD on ventilation units is presented and tested on a model of a real building. A dynamic energy performance model of the building is used for simulating the behaviour of the building. Five experiments are designed and, in each of them, a specific fault is introduced in the unit by modifying parameters in the model. A set of rules is used to validate the operation of each component.

Two different types of faults were considered, i.e. abrupt and gradual faults. In abrupt faults, there is no transition between the healthy and faulty conditions. In gradual faults, on the other hand, the fault intensity increases gradually over time.

In addition to that, the dynamic energy performance model is also used to assess the impact of faults on the building's energy consumption. This gives insights on the effect of faults, both in relative and absolute terms. Some of the faults considered do not visibly change the building's energy consumption, while others have a significant impact.

CHAPTER 4

CASE STUDY: BUILDING OU44

Odense undervisning 44 (OU44) is a recently constructed building at the University of Southern Denmark [7], and it has been the main case study for the project COORDICY, this thesis, and for all publications included in this thesis. The building, shown in Figure 4.1, is located at the main university campus, in Odense, Denmark. It was built in 2015, and it was designed to comply with European 20 20 goals for nearly zero-energy buildings [21]. The building, whose characteristics are shown in Table 4.1, consists of three floors plus a basement and it is mainly used for teaching and office work. It contains over 200 rooms between classrooms, study areas, auditoriums and offices, and around 1000 people visit it on weekdays to attend classes, join study groups or work at their offices. An 80 m^2 large system of photovoltaic panels with a power capacity of 12 kW is located on the roof of the building, and the generated electricity is used to reduce the amount absorbed from the distribution grid.

Ventilation is provided by 4 ventilation units, and a building management system (BMS) controls the indoor climate to maintain a good comfort level for occupants. Indoor air quality is measured by CO_2 concentration level and the air is circulated to maintain such level below national regulations. The indoor temperature is controlled through ventilation air temperature, which is heated through a district heating hot water loop. In addition to that, several rooms have radiators, which are also heated up using district heating. During warm months, ventilation is also used to provide natural cooling to the building by using outdoor air.

Figure 4.1: Building OU44 at University of Southern Denmark, campus Odense.

Table 4.1: Characteristics of building OU44.

Construction Year	2015
Total Area	8000 m^2
Floors	4
Rooms	135
Study Areas	8
Classrooms	19
Offices	48
Ventilation Units	4

4.1 DATA COLLECTION

From the data collection perspective, the building is divided into the following three main subsystems.

- Room-level data;
- ventilation unit data;
- energy meters data.

Measurements at room level are collected through a KNX bus [58] from several sensors such as indoor climate, i.e.

- temperature [°C],
- CO_2 concentration level [ppm],
- light intensity [lx],

and room usage sensors, i.e.

- occupancy [boolean];
- booking [boolean].

Additional sensors are available for ventilation and radiators signals and statuses, shader position and lighting configuration. Some rooms have additional meters and sensors such as separate meters for plug load or relative humidity sensors. Rooms on the ground floor have window opening sensors, however, those are only used for security monitoring, and not for operation control. A weather station, located on the external hull of the building, records temperature, wind speed, solar radiation, and rain intensity.

Data in the ventilation subsystem are available through the BMS. Two classes of data are collected, i.e. room-level data, which is replicated from the room-level subsystem, and data from the ventilation units themselves. The building contains several energy meters that record the energy consumption at different aggregation levels, both for electricity and district heating, represented as energy distribution trees in Figures 4.2 and 4.3. All energy meters are accessible through an EnergyKey system.

Furthermore, 17 3D stereo vision occupants counting cameras are installed at the entrances of the building, in corridors and in four selected test rooms. Those cameras track people crossing the entrances of rooms and floors and provide an estimate of the number people occupying those areas.

Those subsystems have heterogeneous interfaces, however, building drivers continuously gather data from them and publish them to a centralized data storage using the simple measurement and actuation profile (sMAP) protocol. Therefore, data is accessible through a standard interface and available to several applications such as occupancy prediction [59] and model development and calibration [60].

4.2 VENTILATION SYSTEM

Building OU44 contains four identical ventilation units, each of which serves one of its vertical corners: north-west, north-east, south-west, south-east. Figure 4.4 shows a schematic diagram of a ventilation unit. Air circulates from outside to the building, and outside again, through two main ducts, pushed by two fans located in each duct. In the figure, air enters the building from the lower-left corner, goes through a filter, and through a heat exchanger (HX), then it is heated up by a heater and, finally, it enters the supply shaft, from which it will circulate to the rooms. On the way back, air leaves the rooms and enters the extract shaft, it goes through a filter and through the heat exchanger and, finally, it is pushed out of the building. The heat exchanger is used to recover heat from exhaust air, in order to reduce the energy required to heat up inlet air through the heater. Sensors record several measurements inside the ventilation unit, such as air flow, pressure, temperature, humidity, fan speed, and electrical consumption, at five main locations: inlet, post-HX, supply, extract, and exhaust.

The heater, shown in detail in Figure 4.5, uses a hot-water heating loop coming from district heating to heat up incoming air. Sensors and energy meters record water flow, incoming and outgoing temperatures, and thermal energy consumption.

From the main supply shaft, air enters individual rooms through variable air volume (VAV) units, as shown in Figure 4.6. Each room independently opens its VAV unit according to the required ventilation,

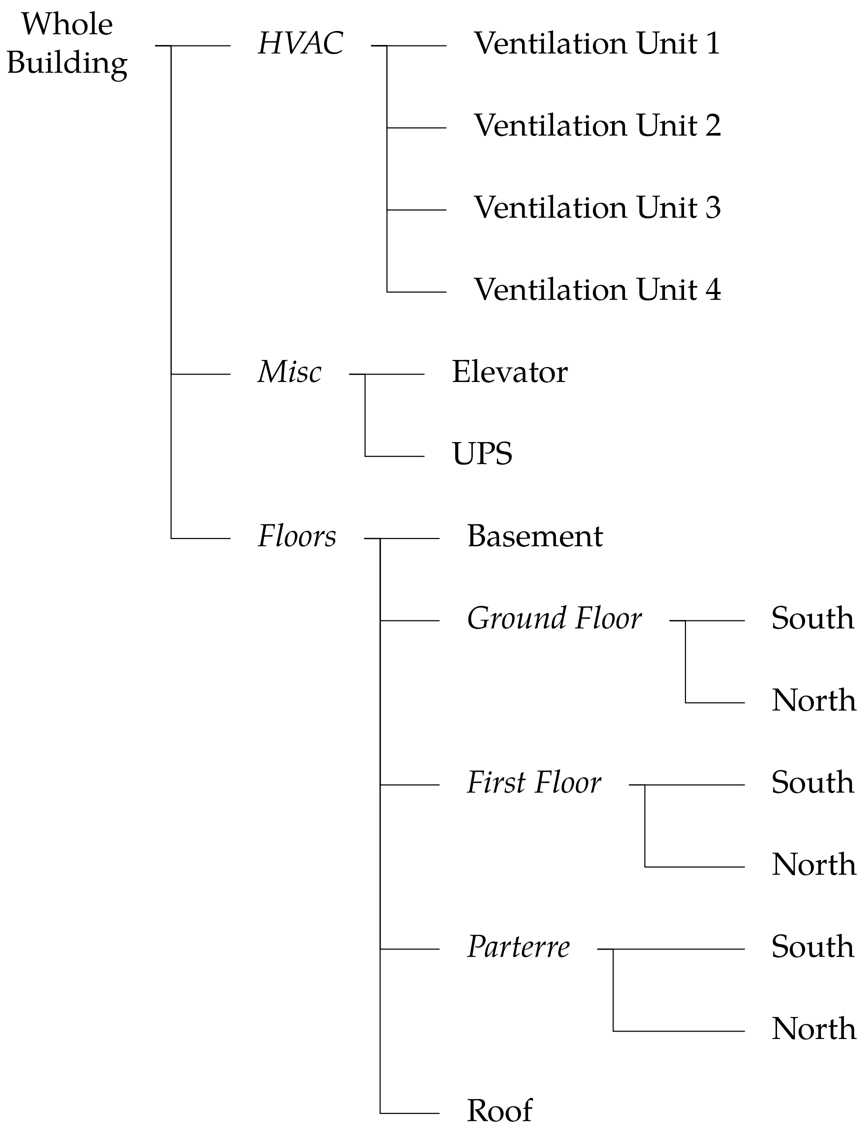

Figure 4.2: Electrical energy distribution tree in building OU44. Meters in *italic* do not exist as independent meters, but they can be obtained by aggregating their children.

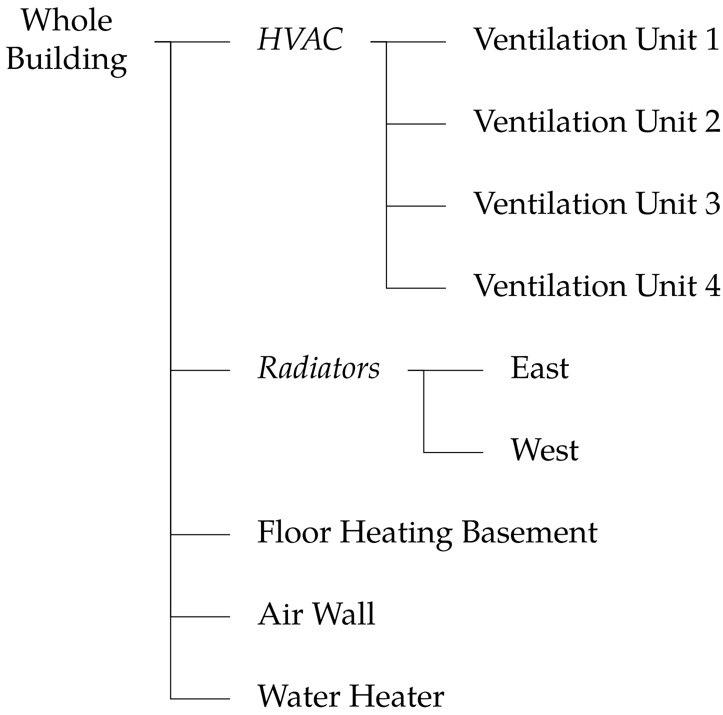

Figure 4.3: District heating distribution tree in building OU44. Meters in *italic* do not exist as independent meters, but they can be obtained by aggregating their children.

which depends on both air quality, measured by CO_2 concentration level, and temperature. When CO_2 level reaches the thresholds of 600 ppm, 750 ppm and 900 ppm, the VAV unit opens by 45 %, 70 % and 100 %. On the way down, a 100 ppm hysteresis is used to avoid frequent movement of VAV units. VAV units are also controlled by temperature. Their position is increased gradually when the temperature is above the threshold to increase the intake of outdoor air, which results in natural cooling.

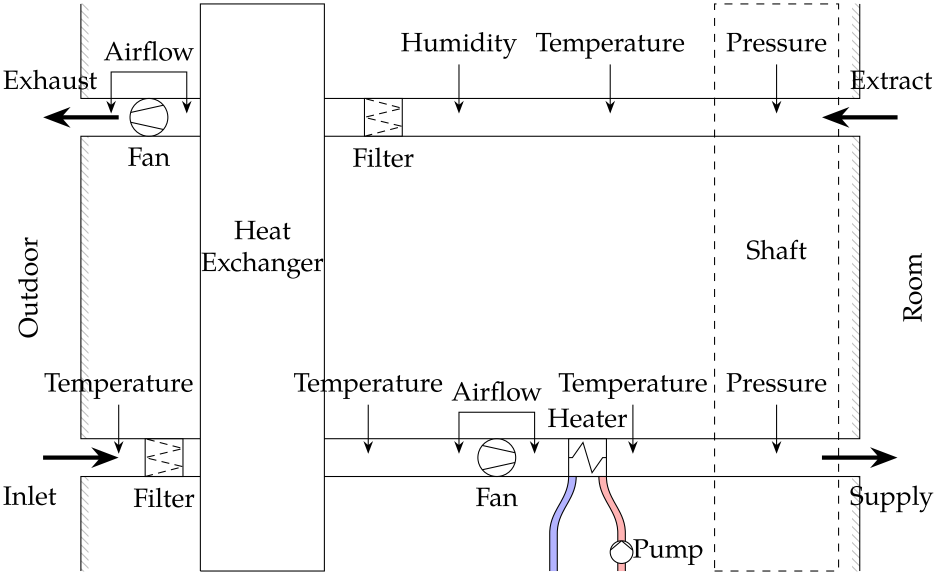

Figure 4.4: Ventilation unit in building OU44. Air enters the building, goes through a heat exchanger, it is heated up, and enters the main shaft from which it circulates the building. On the way out, heat is recovered from exhaust air to reduce the energy required by the hot-water heating loop.

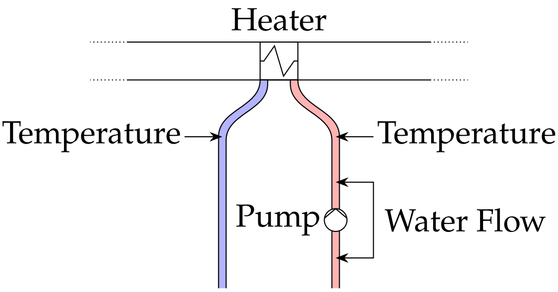

Figure 4.5: Hot-water heating loop in building OU44.

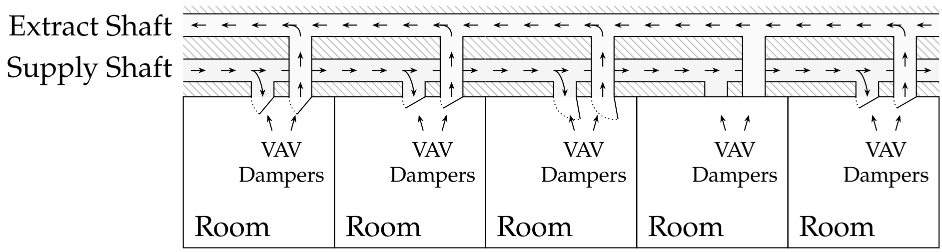

Figure 4.6: Ventilation shafts in building OU44. Air enters the room from the supply shaft and leaves to the extract shaft. Supply and extract VAV dampers are synchronized, i.e. they are always open at the same level. Dampers in different rooms are independent.

4.3 ENERGY MODEL OF THE BUILDING

A detailed dynamic energy model of the building was developed for building OU44 [7]. The model, whose 3D rendering is shown in Figure 4.7 against a picture of the building, can be used to simulate the behaviour of the building according to weather conditions and occupancy schedules, using the simulation engine EnergyPlus. EnergyPlus is an industry standard tool for simulation of building performance, developed at Lawrence Berkeley National Laboratory in the United States of America (USA) [52].

Building details were initially obtained from the building information model (BIM) provided by the constructor. At first, the geometry of the building was replicated in the Sketchup Pro software. Structural entities, such as the main envelope, rooms and corridors, were accurately represented in the model. Using OpenStudio plugin for Sketchup Pro, the model was augmented with information about the technical systems, such as ventilation components, system loads and operation schedules. Finally, the model was exported as an EnergyPlus model in the form of an Energy Plus input (IDF) file, including other characteristics, such as CO_2 and temperature sensors, operational setpoints and other parameters. Several iterations of this pipeline were performed, until a complete and accurate model was obtained.

Figure 4.7: Building OU44 and its energy model.

Part II

Publications

In this part, are reported, one per chapter, the following publications.

[1] **Claudio Giovanni Mattera**, Sanja Lazarova-Molnar, Hamid Reza Shaker and Bo Nørregaard Jørgensen. 'A Practical Approach to Validation of Buildings' Sensor Data: a Commissioning Experience Report'. In: *Third International Conference on Big Data Computing Service and Applications (BigDataService)* (San Francisco, CA, USA, 6th–9th Apr. 2017). IEEE. 12th June 2017, pp. 287–292. DOI: 10.1109/BigDataService.2017.48.

[2] **Claudio Giovanni Mattera**, Muhyiddine Jradi and Hamid Reza Shaker. 'Online Energy Simulator for Building Fault Detection and Diagnostics Using Dynamic Energy Performance Model'. In: *International Journal of Low-Carbon Technologies* 13.3 (17th May 2018), pp. 231–239. ISSN: 1748-1325. DOI: 10.1093/ijlct/cty019.

[3] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'Fault Detection and Diagnostics in Ventilation Units Using Linear Regression Virtual Sensors'. In: *International Symposium on Advanced Electrical and Communication Technologies (ISAECT)* (Kenitra, Morocco, 21st–23rd Nov. 2018). IEEE. 24th Jan. 2019. DOI: 10.1109/ISAECT.2018.8618755.

[4] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'A Method for Fault Detection and Diagnostics in Ventilation Units Using Virtual Sensors'. In: *Sensors* 18.11 (14th Nov. 2018). ISSN: 1424-8220. DOI: 10.3390/s18113931.

[5] **Claudio Giovanni Mattera**, Hamid Reza Shaker and Muhyiddine Jradi. 'Consensus-based Method for Anomaly Detection in VAV Units'. In: *Energies* 12.3 (1st Feb. 2019). ISSN: 1996-1073. DOI: 10.3390/en12030468.

[6] **Claudio Giovanni Mattera**, Hamid Reza Shaker, Muhyiddine Jradi, Mathis Riber Skydt and Sebastian Skals Engelsgaard. 'Fault Detection in Ventilation Units using Dynamic Energy

Performance Models'. In: *Sustainable Cities and Society* (2019). ISSN: 2210-6707. **Submitted**.

Figures and tables were minimally edited to fit the layout of this thesis, and they are otherwise identical to the ones in aforementioned publications. Text was copied verbatim, except for bibliography reference numbers and styles which are kept uniform for the entire thesis, and for minor spelling adaptations and corrections.

A local bibliography is reported at the end of each chapter, listing all references cited in the corresponding publication. All references are also repeated in the global bibliography at the end of this thesis, on page 261.

CHAPTER 5

SENSORS VALIDATION: A PRECONDITION TO ANY BUILDING APPLICATION

This chapter is a cosmetic adaptation of the following conference paper.

Claudio Giovanni Mattera, Sanja Lazarova-Molnar, Hamid Reza Shaker and Bo Nørregaard Jørgensen. 'A Practical Approach to Validation of Buildings' Sensor Data: a Commissioning Experience Report'. In: *Third International Conference on Big Data Computing Service and Applications* (*BigDataService*) (San Francisco, CA, USA, 6th–9th Apr. 2017). IEEE. 12th June 2017, pp. 287–292. DOI: 10.1109/BigDataService.2017.48

The paper was presented at the IEEE Third International Conference on Big Data Computing Service and Applications in San Francisco, USA, on Sunday 9 April 2017.

ABSTRACT

Often manually performed commissioning processes on building's sensors fail to systematically validate that all building's sensors operate correctly. This is so because manual processes are tedious and only inspect a limited number of sensors. As a result, sensors are often uncalibrated, biased or somehow faulty, impacting building's behaviour, comfort level and energy usage.

We present a practical approach to automatically validate data from all building's sensors. We designed and implemented four different

tests to detect out-of-range values, spikes, latency issues and non-monotonous values. Our tests are based on expert knowledge and do not need historical data.

We ran the validation tests on a newly constructed building at the campus of the University of Southern Denmark. As a result we identified two types of faulty behaviours in the building's sensors: CO_2 sensors reporting biased values and temperature sensors' readings exhibiting high latency.

We show how automatic data validation for building sensors enhances the processes of detecting issues which could severely impact building's operations, and were otherwise going unnoticed. Thus, we emphasize the importance of performing data validation as a necessity for a correct building operation.

5.1 INTRODUCTION

In the past several years buildings have become more and more intelligent [61]. Complex building management systems (BMSs), often, manage buildings in their entirety: ventilation, heating, lighting, and other relevant subsystems. To operate, a BMS requires access to the status of the corresponding building, both at room level and at coarser levels. For instance, when the CO_2 concentration levels in a zone are too high, the ventilation is turned on to improve the air quality. Similarly, when the temperature in a room is below its setpoint, the heating is turned on, and when a room has been empty for a certain period, the lights are turned off.

The status of a building is sampled by an increasingly large number and variety of sensors: CO_2, temperature, air flow, water flow, humidity, light intensity, occupancy, energy meters, etc. If data collected from a building is incorrect, it would impact the correct operation of the corresponding BMS. If a BMS is fed low quality data, it would produce low quality results. It is, therefore, important to perform data validation to ensure that the quality meets the requirements.

Nevertheless, in many buildings no systematic validation is performed before they are handed over to final users, even if basic tests do not require complex setups. Facility management personnel manually validate only a small subset of all sensors and might assume that the data is correct and that the building's systems are operating correctly, as well as mistake data issues with system issues. E.g. they might blame the BMS for not handling heating correctly when it is the temperature sensors that are producing lower values than real.

In this paper we propose a series of practical data validation tests that can be automatically performed on new buildings, regardless of how data is collected, and further customized. We report the application and testing of this approach on a real case study building. We, furthermore, discuss the identified building's issues and their implications on the building's operations and performance.

Most research papers on data validation make use of simulated data [62, 63, 64, 65, 66, 67], real data with simulated faults (e.g. by adding a fixed amount to model a biased sensor) [68, 69, 70, 71, 72] or real data with artificially induced faults (e.g. by manually blocking a valve) [73, 74]. In this paper we document the testing of our proposed method on real data from our case-study building, through which we identified significant faults. We informed the facility management team about the discovered faults and they, subsequently, attended to them. Furthermore, we adapted tests existing in literature to threshold-based sensors.

The rest of the paper is organized as follows. The current state of the art is reviewed in Section 5.2. The data validation tests are introduced in Section 5.3. Section 5.4 presents the case study and discusses results and implications. Finally, conclusions are drawn in Section 5.5.

5.2 Related Work

5.2.1 Data Validation: Overview and Classifications

Data validation belongs to the more general field of fault detection and diagnostics (FDD), on which depends successful operation and performance of smart buildings [75]. FDD methods analyze data from buildings to detect and identify faults. The basis for any higher level

FDD approach is correct data, implying that data validation must be the very first step.

Erroneous data is caused by faults in the measuring processes, when collected data points do not adequately represent the measurements. On the other hand, data may appear faulty while correctly representing a faulty system, and it is therefore important to avoid misdetecting one for the other [70]. E.g. a too low temperature record can be caused by a biased sensor or by a broken heating unit.

Similarly to one classification of FDD methods, data validation methods can be divided in three groups: model-based methods, data-driven methods and rule-based methods [39]. Model-based methods use physical knowledge of the system to produce estimate data points and compare them with measured ones. Data-driven methods rely on fault-free historical data which is used to learn a black-box model of the system. Rule-based methods checks whether rules obtained through expert knowledge hold for sampled sensors data.

Model-based methods need to be designed and tuned for the system under test, therefore they are often too complex or too expensive to set up. Data-driven methods are more flexible and easier to adapt but require flawless historical data, which is not always available. Methods based on expert knowledge do not require neither detailed models nor historical data: they are therefore suitable to be deployed on buildings during the commissioning, when no data has been recorded yet or it has not been validated, i.e. methods in this group are the initial step to establish ground truths.

Another way to differentiate data validation methods is whether they consider single or multiple data streams. Methods in the first group are flexible and can be adapted to many kinds of streams. Methods in the second group exploit correlations between different streams and potentially allow validating complex interactions, but they require either redundancy—sensors in similar environments should report similar values—or historical data—from which patterns are extracted and compared with real time data.

Table 5.1 shows which of the reviewed methods belong to each group. The method proposed in this paper belongs to the upper left cell, it considers single data streams and assumes no historical data is available.

Table 5.1: Related work in the different groups.

	No historical data	Historical data
Single streams	[76, 77, 78, 62, 63, 79, 80, 72, 81, 73]	[76, 77, 78, 82, 73]
Multiple streams	[70]	[76, 77, 68, 69, 71, 64, 74, 65, 66, 67]

5.2.2 STATE OF THE ART

In the following we summarize the state of the art of relevant advances in data validation. We begin by a general overview, followed by a focus on data-driven methods, and finally we review the remaining significant approaches.

GENERAL OVERVIEW

Siao et al. present a literature review on data validation methods. Simple tests include physical range check based on sensor's range; local realistic range, based on sensor's location and condition, possibly obtained through statistical analysis; gaps detection; flat lines detection; gradient test; tolerance band methods; and physical redundancy checks. More complex tests include statistical analysis to detect outliers, drift detection using exponential weighted moving average methods, spatial consistency methods, analytical redundancy (to check quantities correlated in a physical model), gross error detection, multivariate statistical methods (e.g. principal component analysis (PCA)) and data mining techniques [83].

Pires et al. present a review and classification of data validation methods used for mobile health applications. The authors divide those methods in three groups: faulty data detection methods, data correction methods, and other assisting techniques or tools dealing with hardware errors [84].

Cugueró-Escofet et al. identify 6 levels of data validation tests—communications, physical range, trend, equipment state, spatial consistency and time series consistency—divided in low and high level tests. Low level tests concern a single sensor, while high level tests exploits correlation among different sensors. The tests, obtained from expert knowledge, are applied in sequence to incoming data which is scored accordingly. In case one or more tests fail, a data reconstruction method is used to produce valid data [78].

Branisavljević et al. consider an ordered sequence of validation methods to be applied to data: detection of zero values, detection of constant values, range check for physical limits, range check for historical limits, statistical univariate test, statistical multivariate tests, artificial neural network (ANN), non-linear models, SVM and physical model. Data is augmented with contextual information (part of day and weather conditions) and validation methods perform better when tuned separately on each class [76].

DATA-DRIVEN METHODS

Castello et al. present two applications to handle data validation and correction and data provenance for buildings. Provenance includes information about the transformations through which data undergo (unit conversion, resampling and filtering). The authors present three experimental buildings as case studies [77].

Hou et al. propose a combined rough sets and ANN method for detecting biases on HVAC chillers. Several rules are defined to split data in subsets corresponding to different operating conditions, and historical fault-free data is used to train ANNs for each subset. The ANNs compute then an estimate confidence interval for the bias [68].

Sharifi et al. propose a MPPCA model for non-linear sensor faults detection. The input space is divided in few locally linear regions and on each of them a PPCA model is trained. When a new measurement is available, it is first mapped to the correct region and then validated. A drawback of the proposed method is that in case of large error it is difficult to obtain the correct region and, therefore, to correctly isolate the fault [74].

Tsang et al. propose a method to validate sensor data using polynomial predictive filter and fuzzy logic. Three sets of fuzzy rules are considered: data is in range, data frequency is in range, and data variance satisfies the F-ratio test. Polynomial predictive filter is applied to historical data to obtain estimates for the rules' lower and upper bounds [73].

MODEL-BASED AND RULE-BASED METHODS

Tsang proposes a gray model method for sensor data validation. The authors consider three fault indicators: limit indicator, where a signal or its rate of change are out of prescribed bounds, jump indicator, when there is a sudden change in the signal corresponding to a spike, and noise indicator, when there is a change in the predicted signal's error [81].

Näsi et al. propose fuzzy limits centered on the signal's current mean to identify outliers within the range limits. The authors compare distribution-based and density-based limits, showing that the latter are less sensitive to non-evenly divided values. Adaptive fuzzy limits contain a damping term based on distance from the average, in order to prevent the outliers to affect the limits [79, 80].

5.3 VALIDATION OF SENSORS AND METERS DATA

The methodology presented in this paper is mostly inspired by the M1–M3 tests from [76], threshold-based tests from [77] and level 0–2 tests from [78]. Implementations from these papers were not available to deploy on our building. Moreover, the mentioned approaches assumed sensors can be sampled with constant frequency, which was not the case in our setup. Therefore, we designed and implemented our own methodology.

Sensors perform measurements and make them available to the BMS. We assume sensors to be threshold-based, i.e., to continuously sense values and produce a new reading only when the difference from the previous one is larger than a threshold (see Figure 5.1). Rate-of-change tests, while popular in literature [62, 63, 71, 79, 80, 72, 81, 73],

Figure 5.1: Threshold-based sensors produce a new reading when the difference from the previous value is larger than a threshold (1 °C).

are ineffective for threshold-based sensors, since their rate-of-change is constant ($\pm$ the threshold).

We define the following terms:

Value the physical value sensed by a sensor;
Reading the act of receiving a value from a sensor;
Record a pair (t, v), denoting a reading at instant t with reported value v;
Sensor threshold difference between the current value and the last record needed for a sensor to make a new reading.

Figure 5.2 shows an overview of the data path in our system. When a reading happens on a sensor, a record is stored on the BMS. Over time a sensor produces a time series of records $[(t_0, v_0), (t_1, v_1), (t_2, v_2), \ldots]$. A *driver* is a program running on the BMS which forwards readings to a centralized data storage. Validation tests are performed on all incoming records going to the data storage.

5.3.1 Validation Tests

We implemented four different tests to validate records from the building sensors: range, latency, spikes and monotonicity. Each test can detect different issues (Table 5.2).

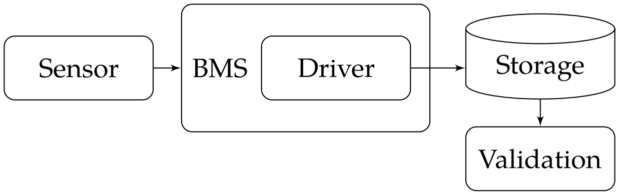

Figure 5.2: Data validation flowchart. Sensors push records to the BMS. A driver collects records from the BMS and forwards it to a centralized data storage, on which validation is performed.

RANGE TEST

Sensors from the building measure different physical quantities, e.g. CO_2 concentration level in the air, light intensity in rooms, temperature of air in rooms or in ventilation units or temperature of water in the heating system and humidity in rooms. Most of these quantities should have measure within a given range. E.g. in normal rooms CO_2 concentration level cannot be lower than in atmosphere and air temperature should be within comfort level range. Sensible ranges can be obtained for many of the measured quantities from expert knowledge, sensors data sheets and validated historical data. Given upper and lower bounds v_{min} and v_{max} the record (t, v) is labeled as erroneous if $v \lneq v_{\text{min}} \lor v \gneq v_{\text{max}}$.

Table 5.2: Issues detected by each test.

Issue \ Test	Range	Latency	Spikes	Monotonicity
Sensor bias	✓			
Misplaced sensor	✓	✓		
Driver fault		✓	✓	✓
Accuracy degradation		✓		
Communication problem		✓		✓

LATENCY TEST

For threshold-based sensors at any time the uncertainty of a measure is twice the sensor-threshold. Long periods without a reading can be caused by a decrease in sensor accuracy, a sensor hard failure, communication problems or BMS failure.

Given a maximal latency $\Delta t_{\max}$ and two consecutive records (t_0, v_0) and (t_1, v_1) *the former record* is labeled as erroneous if $t_1 - t_0 > \Delta t_{\max}$. For this test it is not the records themselves which are erroneous, but the interval between them. More readings were expected between t_0 and t_1, therefore the record (t_0, v_0) was labeled as erroneous, since it is not known until when its value can be trusted.

SPIKES TEST

A spike is a large variation in a very short time window. Occasionally a driver may fail to parse a record from the BMS, or the sensor itself can generate an erroneous value. Typical examples are zero, negative numbers, or random numbers. Sometimes spikes are invalid values (e.g. a negative CO_2 concentration level) which are easy to filter out, but they can also be valid (e.g. temperature can be negative).

A naïve way to check for spikes is to check when the difference between two consecutive records is above a given threshold. This is however susceptible to false positives: for instance, if the building's communication network was down for a short period of time the first record after it is back on line might be significantly different from the previous one, but this should not be considered a spike. Given the parameters δ_v and δ_t and two consecutive records (t_0, v_0) and (t_1, v_1), the latter record is labeled as a spike if

$$\frac{v_1 - v_0}{t_1 - t_0} > \frac{\delta_v}{\delta_t}. \tag{5.1}$$

MONOTONICITY TEST

Some meters, in particular energy meters that record the total energy consumption, record an incremental quantity. Such values are monotonically increasing, as it is not possible to recover previously consumed energy, and therefore every record value must be greater or equal to the previous one. Given two consecutive records (t_0, v_0) and (t_1, v_1) the latter record is labeled as erroneous if $v_1 \ngeq v_0$.

5.4 CASE STUDY

In this paper we present building Odense undervisning 44 (OU44) as a case study. The building is located on the campus of the University of Southern Denmark. It contains classrooms, offices and study rooms, it has been operating and collecting data since October 2015.

Every room has the following sensors:

- Temperature [°C], relevant for heating and ventilation;
- CO_2 [ppm], relevant for ventilation;
- PIR [boolean], room occupancy;
- Light [lx], relevant for automatic lighting control.

Some rooms have additional sensors or meters. For instance some have separate meters for plug load or sensors for humidity. Four rooms are equipped with occupancy counting cameras that provide an estimate of people in the room. In addition to that, the building has a weather station that records external temperature, wind speed, rain and solar radiation. There are also several energy meters: for heating, ventilation, hot water, lighting, plug load, usually aggregate by floor or area.

All sensors are accessible through a KNX bus [58] and broadcast records to the BMS according to their configuration. Custom drivers fetch data from the BMS and publish it to a centralized data base using simple measurement and actuation profile (sMAP) protocol, so that it is available to other applications, like occupancy prediction [59] and model development and calibration [60].

Validation tests were executed on all the available rooms for a number of selected sensors and meters. Table 5.3 shows a summary of the tests and the sensors along with the corresponding parameters.

Table 5.3: Implemented tests parameters.

Test \ Sensor	Temperature	Humidity	CO_2	Energy Meter
Range	10–40 °C	0–100 %	350–1200 ppm	
Latency	180 min	20 min	10 min	
Spikes	20 °C in 5 min	5 % in 10 min	300 ppm in 10 min	
Monotonicity				Yes

Table 5.4: Number of streams and average frequency.

	Temperature	Humidity	CO_2	Energy Meter
Number of streams	140	10	231	88
Average frequency	30 min	5 s	5 min	1 min

Details about number of streams and average frequency are shown in Table 5.4.

5.4.1 RESULTS

RESULTS FOR CO_2 SENSORS

Figure 5.3 shows spikes test violations for CO_2 concentration level for selected room and period. Some spikes have value zero, which is impossible for CO_2, while some other spikes have unusually large, but in principle plausible, values.

Figure 5.4 shows latency test violations for CO_2 concentration level for a selected room and a time period. There are two intervals with missing data due to a driver crash. The data was, however, still available on the BMS, although not yet forwarded to the storage.

Figure 5.5 shows range test violations for CO_2 concentration level for a selected room and a time period. Results for most of the rooms in the building are similar to these. For the given room, the CO_2 sensor was reporting CO_2 concentration levels lower than the current atmospheric level. The CO_2 sensors have an accuracy of ± 125 ppm, therefore, some violations were expected when the CO_2 concentration level was close to the minimum. However, records were consistently below range

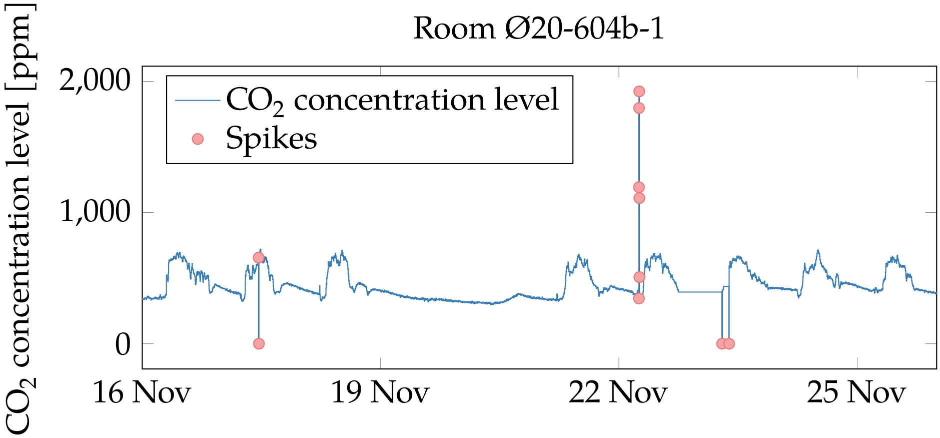

Figure 5.3: Spikes test violations for CO_2 concentration level.

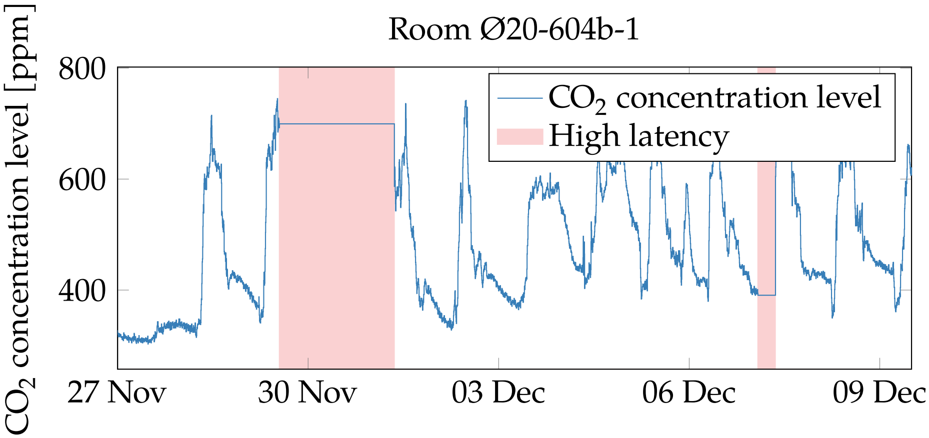

Figure 5.4: Latency test violations for CO_2 concentration level.

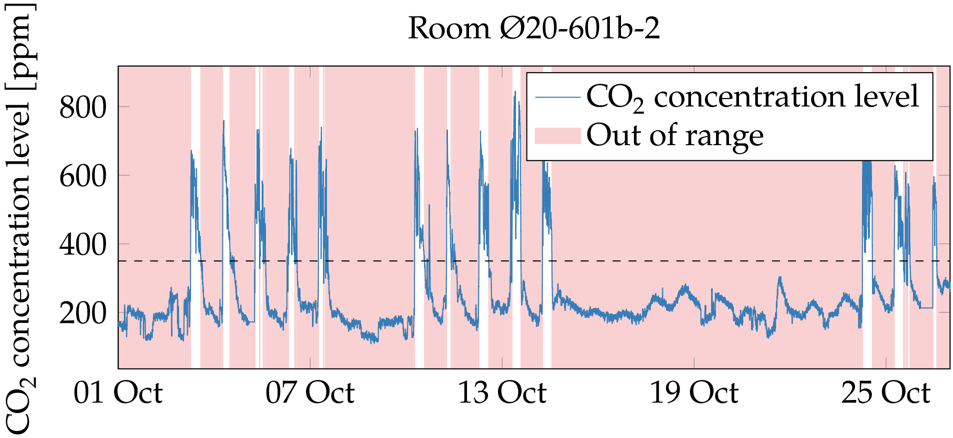

Figure 5.5: Range test violations for CO_2 concentration level.

nearly all the time, sometimes for several days, which suggests these sensors are faulty.

RESULTS FOR TEMPERATURE SENSORS

Figure 5.6 shows latency test violations for temperature measurements for a selected room and time period. Even if the maximal latency was rather large (180 min), there were many erroneous intervals. The sensor-threshold in the sensor was set to 0.1 °C, so more frequent readings were expected. However, most of the faults occurred during the night, when the room was empty and temperature should stay close to the setpoint. It is also possible that boxes enclosing sensors isolate them too much from the environment and they, therefore, shield them from high-frequency variations.

Figure 5.7 shows range test violations for temperature for selected room and period. For a short time during the night the temperature dropped below the lower bound. Since this was an isolated instance, it occurred during the night and the temperature went back to the setpoint in the morning, it might suggest that a window was left open during the night.

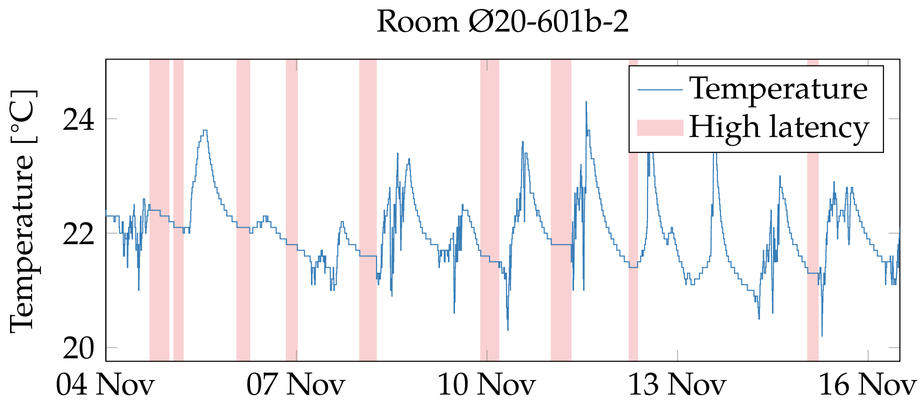

Figure 5.6: Latency test violations for temperature.

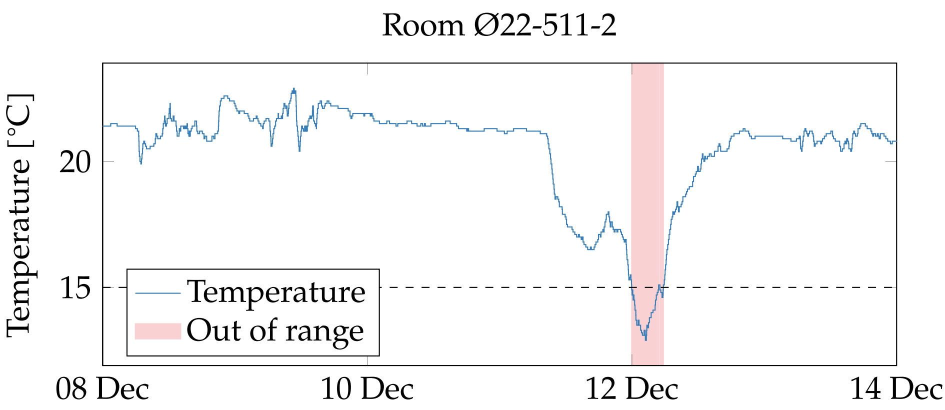

Figure 5.7: Range test violations for temperature.

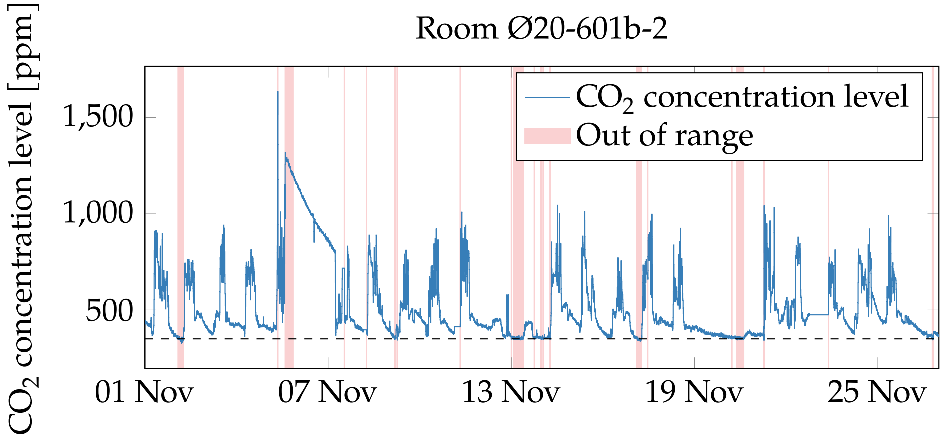

Figure 5.8: Range test violations for CO_2 concentration level after sensors were replaced.

5.4.2 Implications and Discussion

After the CO_2 sensors' faults were detected, the supplier replaced the sensors. The range test violations for CO_2 after the replacement are shown in Figure 5.8 (in the same room of Figure 5.5). There were still few intervals below the minimum range, but they happened when the CO_2 concentration level was close to the minimum, and they are therefore expected due to the sensor accuracy ($\pm$125 ppm).

To summarize, running the tests on the building led us to detecting two particular faults:

- many CO_2 sensors were often below their minimal range,
- for some temperature sensors new readings occurred with a very high latency.

Further investigation showed that CO_2 sensors were biased, mostly to a lower value. The BMS uses these records to determine when to turn on ventilation, therefore, this fault resulted in bad air quality in the building and lower ventilation energy consumption.

Additional exploration helped in discovering that for some temperature sensors (especially the ones in the weather station) the sensor-threshold is too large, up to 1 °C, which means that the accuracy of temperature is 2 °C and smaller variations are not recorded. This configuration is probably due to constraints on the KNX bus bandwidth, which supports a relatively small number of simultaneously transmitting sensors. The effects of decreasing the sensor-threshold need to be investigated.

5.5 CONCLUSIONS

We presented a sensor data validation approach that was designed and developed driven by the real need of our research project, aimed at improving smart buildings' energy efficiency. The data validation tests performed on the project's case-study building exposed at least two faults: CO_2 sensors "out of range" and temperature sensors showing high latency. CO_2 sensors bias impacted the ventilation system and other ongoing research projects regarding model calibration and parameter estimation on the building [85], and the issue was detected only after few months. Temperature sensors' high latency might have impacted model calibration as well. If those tests had being running earlier during the commissioning phase, a long period of ventilation issues could have been avoided and bus limitations could have been addressed.

This experience showed that validating sensors values in new buildings is indeed an essential necessity for a correct building operation. Expert knowledge based practical tests for single streams are sufficient to expose issues with sensors and meters that greatly affect building's performance.

Automated testing of single streams is the first step in data validation. Such tests, however, can only detect a subset of faults and issues. Moreover, expert knowledge is necessary to set the appropriate test parameters. Our future work in this problem domain anticipates utilizing peer validation and exploiting interactions between different data streams with more advanced data validation methods. Once quality of historical data is ensured, our focus will be on data-driven methods.

ACKNOWLEDGEMENT

This work is supported by the Innovation Fund Denmark for the project COORDICY.

REFERENCES

[1] **Claudio Giovanni Mattera**, Sanja Lazarova-Molnar, Hamid Reza Shaker and Bo Nørregaard Jørgensen. 'A Practical Approach to Validation of Buildings' Sensor Data: a Commissioning Experience Report'. In: *Third International Conference on Big Data Computing Service and Applications* (*BigDataService*) (San Francisco, CA, USA, 6th–9th Apr. 2017). IEEE. 12th June 2017, pp. 287–292. DOI: 10.1109/BigDataService.2017.48.

[39] Srinivas Katipamula and Michael R. Brambley. 'Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems–A Review, Part I'. In: *HVAC&R Research* 11.1 (Jan. 2005), pp. 3–25. DOI: 10.1080/10789669.2005.10391123.

[58] KNX Association. *What Is KNX?* URL: https://www.knx.org/knx-en/for-professionals/What-is-KNX/A-brief-introduction// (visited on 26th Dec. 2018).

[59] Sanja Lazarova-Molnar, Mikkel Baun Kjærgaard, Hamid Reza Shaker and Bo Nørregaard Jørgensen. 'Commercial Buildings Energy Performance within Context Occupants in Spotlight'. In: *Proceedings of the 4th International Conference on Smart Cities and Green ICT Systems - Volume 1: SMARTGREENS* (Lisbon, Portugal, 20th–22nd May 2015). SciTePress, 2015, pp. 306–312. DOI: 10.5220/0005495203060312.

[60] Muhyiddine Jradi, Christian T. Veje and Bo Nørregaard Jørgensen. 'Towards Energy Efficient Office Buildings in Denmark: the Maersk Building Case Study'. In: *29th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems* (Portorož, Slovenia, 19th–

23rd June 2016). Ed. by Andrej Kitanovski and Alojz Poredoš. University of Ljubljana. 2016. ISBN: 978-961-6980-15-9.

[61] Bo Nørregaard Jørgensen, Mikkel Baun Kjærgaard, Sanja Lazarova-Molnar, Hamid Reza Shaker and Christian T. Veje. 'Challenge: Advancing Energy Informatics to Enable Assessable Improvements of Energy Performance in Buildings'. In: *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems* (Bangalore, India, 14th–17th July 2015). ACM. New York, NY, USA, 2015, pp. 77–82. DOI: 10.1145/2768510.2770935.

[62] R. Jeyanthi and K. Anwamsha. 'Fuzzy-based Sensor Validation for a Nonlinear Bench-mark Boiler under MPC'. In: *10th International Conference on Intelligent Systems and Control (ISCO)* (Coimbatore, India, 7th–8th Jan. 2016). IEEE. 3rd Nov. 2016, pp. 1–6. DOI: 10.1109/isco.2016.7727133.

[63] Nithya Kancherla and R. Jeyanthi. 'Design of a Generic Fuzzy-Based Sensor Data Validation Algorithm for a Chemical Process'. English. In: *International conference on Circuits, Controls and Communications (CCUBE)* (Bengaluru, India, 27th–28th Dec. 2013). IEEE. 23rd Jan. 2014, pp. 1–6. DOI: 10.1109/ccube.2013.6718559.

[64] Ihab Samy, Ian Postlethwaite and Dawei Gu. 'Neural Network Based Sensor Validation Scheme Demonstrated on an Unmanned Air Vehicle (UAV) Model'. In: *47th Conference on Decision and Control* (Cancun, Mexico, 9th–11th Dec. 2008). IEEE. 6th Jan. 2009, pp. 1237–1242. DOI: 10.1109/CDC.2008.4738703.

[65] Shengwei Wang and Youming Chen. 'Sensor Validation and Reconstruction for Building Central Chilling Systems Based on Principal Component Analysis'. English. In: *Energy Conversion and Management* 45.5 (Mar. 2004), pp. 673–695. ISSN: 0196-8904. DOI: 10.1016/S0196-8904(03)00180-8.

[66] Shengwei Wang and Fu Xiao. 'AHU Sensor Fault Diagnosis Using Principal Pomponent Analysis Method'. In: *Energy and Buildings* 36.2 (Feb. 2004), pp. 147–160. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2003.10.002.

[67] Shengwei Wang and Jianying Qin. 'Sensor Fault Detection and Validation of VAV Terminals in Air Conditioning Systems'. English. In: *Energy Conversion and Management* 46.15–16 (Sept. 2005), pp. 2482–2500. ISSN: 0196-8904. DOI: 10.1016/j.enconman.2004.11.011.

[68] Zhijian Hou, Zhiwei Lian, Ye Yao and Xinjian Yuan. 'Data Mining Based Sensor Fault Diagnosis and Validation for Building Air Conditioning System'. In: *Energy Conversion and Management* 47.15 (Sept. 2006), pp. 2479–2490. ISSN: 0196-8904. DOI: 10.1016/j.enconman.2005.11.010.

[69] MinJeong Kim, Hongbin Liu, Jeong Tai Kim and ChangKyoo Yoo. 'Sensor Fault Identification and Reconstruction of Indoor Air Quality (IAQ) Data Using a Multivariate Non-Gaussian Model in Underground Building Space'. In: *Energy and Buildings* 66 (Nov. 2013), pp. 384–394. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2013.07.002.

[70] Alexey Kozionov, Mikhail Kalinkin, Alexey Natekin and Alexander Loginov. 'Wavelet-based Sensor Validation: Differentiating Abrupt Sensor Faults from System Dynamics'. In: *7th International Symposium on Intelligent Signal Processing* (Floriana, Malta, 19th–21st Sept. 2011). IEEE. 20th Oct. 2011, pp. 1–5. DOI: 10.1109/wisp.2011.6051716.

[71] J. Kullaa. 'Sensor Validation Using Minimum Mean Square Error Estimation'. In: *Mechanical Systems and Signal Processing* 24.5 (July 2010), pp. 1444–1457. ISSN: 0888-3270. DOI: 10.1016/j.ymssp.2009.12.001.

[72] Z. G. Shen and Q. Wang. 'Data Validation and Confidence of Self-validating Multifunctional Sensor'. English. In: *Proceedings of Sensors* (Taipei, Taiwan, 28th–31st Oct. 2012). IEEE. 17th Jan.

2013, pp. 1045–1048. DOI: 10.1109/icsens.2012.6411058.

[73] K. M. Tsang and W. L. Chan. 'Data Validation of Intelligent Sensor Using Predictive Filters and Fuzzy Logic'. In: *Sensors and Actuators A: Physical* 159.2 (May 2010), pp. 149–156. ISSN: 0924-4247. DOI: 10.1016/j.sna.2010.03.013.

[74] R. Sharifi and R. Langari. 'Nonlinear Sensor Fault Diagnosis Using Mixture of Probabilistic PCA Models'. English. In: *Mechanical Systems and Signal Processing* 85 (15th Feb. 2017), pp. 638–650. ISSN: 0888-3270. DOI: 10.1016/j.ymssp.2016.08.028.

[75] Sanja Lazarova-Molnar, Hamid Reza Shaker, Nader Mohamed and Bo Nørregaard Jørgensen. 'Fault Detection and Diagnosis for Smart Buildings: State of the Art, Trends and Challenges'. In: *3rd MEC International Conference on Big Data and Smart City* (*ICBDSC*) (Muscat, Oman, 15th–16th Mar. 2016). IEEE. 28th Apr. 2016, pp. 1–7. DOI: 10.1109/ICBDSC.2016.7460392.

[76] Nemanja Branisavljević, Zoran Kapelan and Dušan Prodanović. 'Improved Real-time Data Anomaly Detection Using Context Classification'. In: *Journal of Hydroinformatics* 13.3 (July 2011), pp. 307–323. ISSN: 1464-7141. DOI: 10.2166/hydro.2011.042.

[77] Charles C. Castello, Jibonananda Sanyal, Jeffrey Rossiter, Zachary Hensley and Joshua R. New. 'Sensor Data Management, Validation, Correction, and Provenance for Building Technologies'. In: *ASHRAE Transactions* 120 (July 2014), pp. 370–382. ISSN: 0001-2505.

[78] Miquel À. Cugueró-Escofet, Diego García, Joseba Quevedo, Vicenç Puig, Santiago Espin and Jaume Roquet. 'A Methodology and a Software Tool for Sensor Data Validation/reconstruction: Application to the Catalonia Regional Water Network'. In: *Control Engineering Practice* 49 (Apr. 2016), pp. 159–172. ISSN: 0967-0661. DOI: 10.1016/j.conengprac.2015.11.005.

[79] Jari Näsi and Aki Sorsa. *On-line Measurement Validation through Confidence Level Based Optimal Estimation of a Process Variable*. Tech. rep. A No 25. Oulu, Finland: University of Oulu, Dec. 2004.

[80] Jari Näsi, Aki Sorsa and Kauko Leiviskä. 'Sensor Validation and Outlier Detection Using Fuzzy Limits'. In: *44th Conference on Decision and Control, and the European Control Conference, CDC-ECC '05* (Seville, Spain, 15th Dec. 2015). IEEE. 30th Jan. 2006, pp. 7828–7833. DOI: 10.1109/cdc.2005.1583427.

[81] K. M. Tsang. 'Sensor Data Validation Using Gray Models'. English. In: *ISA Transactions* 42.1 (Jan. 2003), pp. 9–17. ISSN: 0019-0578. DOI: 10.1016/S0019-0578(07)60109-8.

[82] Hector Rodriguez, Vicenç Puig, Juan J. Flores and Rodrigo Lopez. 'Combined Holt-Winters and GA Trained ANN Approach for Sensor Validation and Reconstruction: Application to Water Demand Flowmeters'. In: *3rd Conference on Control and Fault-Tolerant Systems (SysTol)* (Barcelona, Spain, 7th–9th Sept. 2016). IEEE. 10th Nov. 2016, pp. 202–207. DOI: 10.1109/SYSTOL.2016.7739751.

[83] Sun Siao, Jean-LucBertrand Krajwewski, Anders Lynggaard-Jensen, Joep Van den Broeke, Florian Edthofer, Maria do Ceu Almeida, Ribeiro Alvaro Silva and Jose Menaia. *Literature Review for Data Validation Methods*. 2011.

[84] Ivan Miguel Pires, Nuno M. Garcia, Nuno Pombo, Francisco Flórez-revuelta and Natalia Díaz Rodríguez. 'Validation Techniques for Sensor Data in Mobile Health Applications'. English. In: *Journal of Sensors* 2016 (2016), pp. 1–9. ISSN: 1687-725X. DOI: 10.1155/2016/2839372.

[85] Krzysztof Arendt, Ana Ionesi, Muhyiddine Jradi, Ashok Kumar Singh, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'A Building Model Framework for a Genetic Algorithm Multi-objective Model Predictive Control'. In: *CLIMA 2016 - 12th REHVA World Congress* (Aalborg, Denmark, 22nd–25th May 2016). Vol. 8. Aalborg University, Department

of Civil Engineering, 2016. ISBN: 87-91606-33-0. URL: http://vbn.aau.dk/en/publications/clima-2016--proceedings-of-the-12th-rehva-world-congress(cd42481c-d63a-4aba-9ac7-f779e332efd5).html.

CHAPTER 6

IDENTIFYING FAULTY SUBSYSTEMS WITH ONLINE ENERGY SIMULATOR

This chapter is a cosmetic adaptation of the following journal paper.

Claudio Giovanni Mattera, Muhyiddine Jradi and Hamid Reza Shaker. 'Online Energy Simulator for Building Fault Detection and Diagnostics Using Dynamic Energy Performance Model'. In: *International Journal of Low-Carbon Technologies* 13.3 (17th May 2018), pp. 231–239. ISSN: 1748-1325. DOI: 10.1093/ijlct/cty019

ABSTRACT

Faults in building systems affect energy efficiency and occupancy comfort. Simulating building behaviour and comparing it with measured data allows to detect discrepancies due to faults.

We propose a methodology to recursively compare actual data with dynamic energy simulations at different layers of aggregation to reduce the scope in searching for faults through the development the online energy simulator, a tool to set up automated simulations using standard interfaces usable with different building systems and simulation engines.

We test our simulator on a real building at the University of Southern Denmark, showing how continuous monitoring allows to quickly detect and identify buildings faults.

6.1 INTRODUCTION

Buildings are responsible for a large portion of energy consumption. In the U. S. A. they accounted for 7% of primary energy consumption in 2010, which is more than transportation and industrial sector. Buildings energy consumption is also rapidly increasing over time, doubling from 1290 TWh in 1980 to 2784 TWh in 2010 [16]. In the European Union buildings account for 40% of the total energy used and 36% of the total CO_2 emissions [12]. Thus, the focus on buildings is fundamental to achieve the energy efficiency and environmental objectives, such as the European goal of saving 20% of primary energy consumption by 2020 compared to projections [86], and 30% by 2030 [22].

Modern buildings have complex control systems that monitor the current status and manage heating, cooling, ventilation and lighting. Each one of these subsystems has also increasing complexity, and can, therefore, suffer from faults and malfunctions. Faults can impact occupancy comfort, e.g. a broken radiator would result in a cold room, but can also yield higher energy consumption. It is estimated that in 2009 the most common faults in U. S. A. commercial buildings were responsible for over $ 3.3 billion in energy waste [36].

Without a continuous monitoring of the building, faults can happen and go undetected for a long time. Moreover, many fault detection methods rely on detecting changes from previous behaviour, and are, therefore, ineffective in detecting faults present since the construction of the building. *Energy models* of the buildings can be developed and used to assess that the actual energy consumption follows the design goals by simulating the building's behaviour. *Static* energy models are simpler and require low computational power but assume steady-state conditions and require simplifications. *Dynamic* energy models are instead more complex both to develop and to simulate but can accurately capture interactions between components and changes over time.

In this paper we propose a methodology for fault detection and diagnostics (FDD) in buildings using energy models simulations and comparing with real building at different aggregation layers. We present a software solution to automate simulations without relying on any manual procedure. Our tool uses industry standard interfaces to support different simulation engines and automatic data retrieval from the building. We then report the application and testing of our method and tool on a real case-study building.

Our tool was developed under the COORDICY Project, a strategic DK-US interdisciplinary research project for advancing ICT-driven research and innovation in energy efficiency of public and commercial buildings [87]. We use our tool to monitor the daily energy usage of our case-study building at several aggregation layers, such as whole building, by subsystem or by floor.

The rest of the paper is organized as follows. The state of the art is reviewed in Section 6.2. The FDD methodology is introduced in Section 6.3 and the online energy simulator in Section 6.4. Section 6.5 presents the case study and discusses results and implications. Finally, conclusions are drawn in Section 6.6.

6.2 State of the Art

6.2.1 Fault detection and diagnostics in buildings

Kim et al. present a comprehensive review of FDD for building systems in recent years. FDD studies are classified using two different schemes: based on building equipment/size, such as large/small buildings, heating, ventilation and air conditioning (HVAC) systems, lighting, water heaters and ventilation units, and based on method. FDD methods can be divided in history based and qualitative or quantitative model based [41].

History-based methods rely on the availability of historical data for a building. Such data is used to create black-box or gray-box models, often using machine learning techniques such as artificial neural networks, for the system under analysis. Faults impact the system's behaviour so that it does no longer match the model's predictions. Historical-based models can be used when little or no information about the physical system under test is available and can in general represent complex interactions. However, they require good quality fault-free training data and can only make accurate predictions within its range. Moreover, they are specific to the system used for training and cannot easily be used on other ones.

Qualitative model-based methods rely on a priori knowledge of the system under investigation. Such knowledge, provided by documentation or expert knowledge, is used to create rule-based or qualitative physical systems. Qualitative model-based methods are simple to implement and can usually be validated by field experts. They are also usually robust to numerical uncertainty in input data. However, they often result in rigid models that cannot be applied to different systems or easily extended.

Quantitative model-based methods rely on explicit mathematical models of system under investigation. Such models, which accurately represent the system's physical function, are used to simulate the system's expected behaviour, which can be compared with the actual one. Quantitative model-based methods provide the most accurate results, and are usually able to simulate transients in dynamic systems, and even faulty behaviour. However, such models are often complex and are both difficult to develop and computationally heavy. They also require validation and parameter estimation with experimental data before their results can be trusted, and cannot easily be used with different systems.

Methods from each category have different trade-offs and are suitable for different kinds of systems. Hybrid approaches that make use of multiple methods are also common, in order to exploit advantages and reduce disadvantages of individual methods. Using multiple methods also increases robustness and reliability.

6.2.2 Building simulation

Many simulation engines are available for simulating buildings energy performance, some explicitly oriented to this field, such as EnergyPlus [52], some more generic, such as Modelica [53].

Clarke et al. describe the overall topic of building performance simulation, its aims and achievements both at the present and in the future. The authors analyze the current state of the art for building performance simulation tools with respect to different aspects, such as subsystems modeling, control, occupants representation, computation time and economic considerations [54].

Costa et al. discuss the advantages brought by monitoring buildings and comparing with energy performance simulations. The authors describe some of the available visualization techniques to display information obtained from building monitoring in a way to facilitate FDD. They also describe how results from monitoring can be used to improve model calibration and operations optimization [55].

Maile et al. propose a new methodology to compare results from simulations using energy models to actual measured data. They consider the importance of multiple hierarchies, such as by component and by location, which can be used to better evaluate the results. An assessor should gather measurement and simulation assumption, perform simulation and collect data, and finally compare the results. All differences between simulated and measured data must be categorized in either: measurements problems, simulation problems and operational problems. Not all differences are actually performance problems, some may be due to measurement or simulation assumptions. Models should be iteratively adapted to reflect the actual building [88].

Wetter propose a framework to connect several simulation engines together using Ptolemy II modeling environment as middleware to manage communication. The authors define an interface for communication between the engines and implement it for several engines such as EnergyPlus, Modelica, Matlab and Simulink. The authors test their framework by performing a co-simulation between EnergyPlus and Modelica, exploiting the advantages of each engine in a particular domain [89].

Pang et al. present a framework for real-time simulation synchronized with the actual building using the simulation engine EnergyPlus. The simulation is managed using Ptolemy II actors and a BACnet interface is used to exchange data with the building management system (BMS). The authors proceed to test their methodology on a real test bed and observe large differences between measures and simulated total energy consumption. However, when looking at disaggregate plots it is possible to figure out what are the causes. Difference of cooling energy consumptions have similar peaks of difference of total energy consumptions, and they are caused by mismatch in chilling strategies between the model and the actual equipment. The same was noted in the case of lights left on overnight [90].

This framework support only few selected simulation engines and only BMSs that publish data over a BACnet interface. In order to overcome these restrictions Pang et al. revise their work and re-implement their framework by using functional mock-up interface (FMI), which is a standard interface supported by many simulation engines. They also use the simple measurement and actuation profile (sMAP) to exchange data, which is an open protocol for data publication [91].

Sharmin et al. present a methodology for sensor-based monitoring of buildings and apply it to two residential buildings and run data analysis on the results. The authors show how monitoring reveals non-obvious information and insights about energy consumption, e.g. heating loss was higher for units on middle floors, which suggests the need for better insulation. The authors also observe that users react by improving their energy usage when introducing feedback from monitoring, but only short term. Automated control is necessary to achieve long-term results [92].

With most engines, users must perform repetitive, time-consuming and error-prone operations to setup and run a simulation. First they have to fetch the input data, optionally preprocess it, and convert it to the expected format (e.g. many engines expect data at fixed intervals corresponding to the simulation step). Then the model must be modified to point to the correct input data files. Then the user must manually start the simulation. Finally the user can access the results usually from a comma separated value (CSV) file.

Often simulation results are interesting for multiple users. Either such users must each independently go through all the mentioned steps, or one user usually shares the results by *unstructured* ways, such as sending files by email. The former option multiplies the necessary time (and the potential for errors), while the latter presents other problems, such as misunderstandings with respect to successive versions of results and possibly authorization issues.

Finally, models in quantitative model-based methods are complex and strictly related to the equipment under test and, therefore, are difficult to generalize and apply them with different equipment. Different simulation engines are optimized for certain systems and users need to learn the details of each of them. Thus, it appears evident that a solution able to automate simulations from different engines in a transparent way and make real-time results easily available online to multiple users is valuable.

6.3 METHODOLOGY FOR FAULT DETECTION AND DIAGNOSTICS IN BUILDINGS

Faults in buildings impact either occupants comfort or energy consumption. We use a dynamic energy performance model to simulate the building's behaviour and compute the expected energy consumption. Thus, any deviation of the actual energy consumption data compacted to the simulated results will highlight faults and anomalies to be investigated.

Buildings record energy consumption at different layers. There is a main meter for electricity that measures the entire building consumption and sub-meters for every system, such as HVAC and lighting. Some buildings also have individual sub-meters for floors, other zones or other components. Separate energy distribution trees can be available for hot water and district heating systems, depending on the building. Figure 6.1 shows an example of electrical energy distribution tree for a building. Sub-meters allow to split the aggregate data from the main meter and to understand how different systems use energy in the building in a more clear and detailed manner. Building energy models are able to provide results at different granularity, therefore, it is possible to compare actual and simulated values for sub-meters.

In this study we develop and implement a top-down approach for FDD as shown in Figure 6.2: when a deviation between actual and

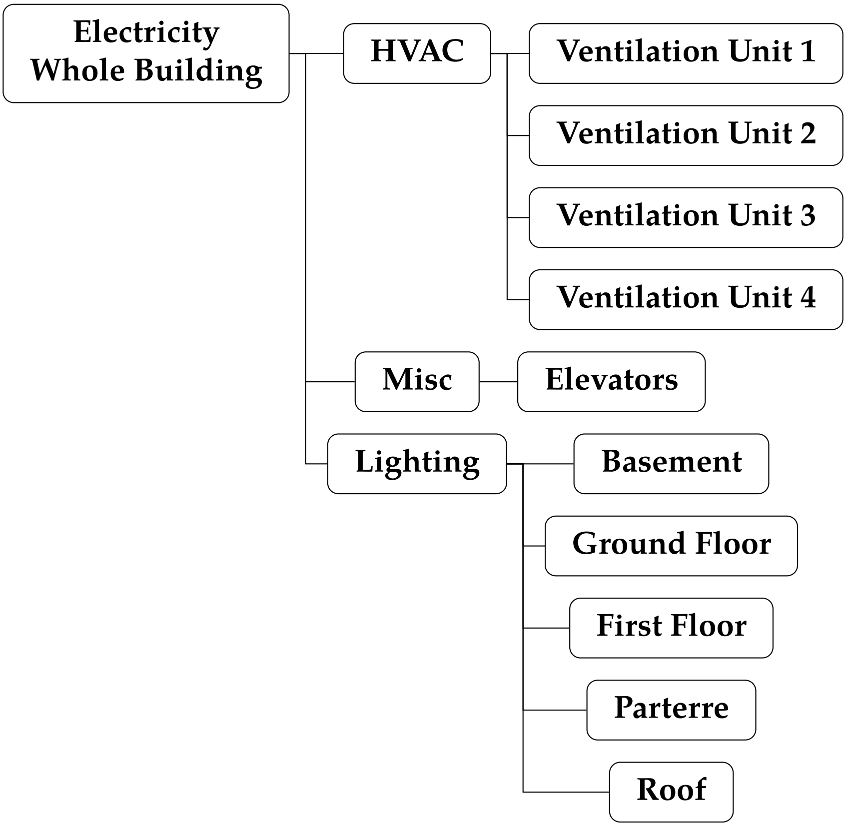

Figure 6.1: Distribution tree in a building for electrical energy. The main meter can be decomposed in HVAC and its ventilation units, and in lighting, which can be in turn decomposed by floor, and miscellaneous.

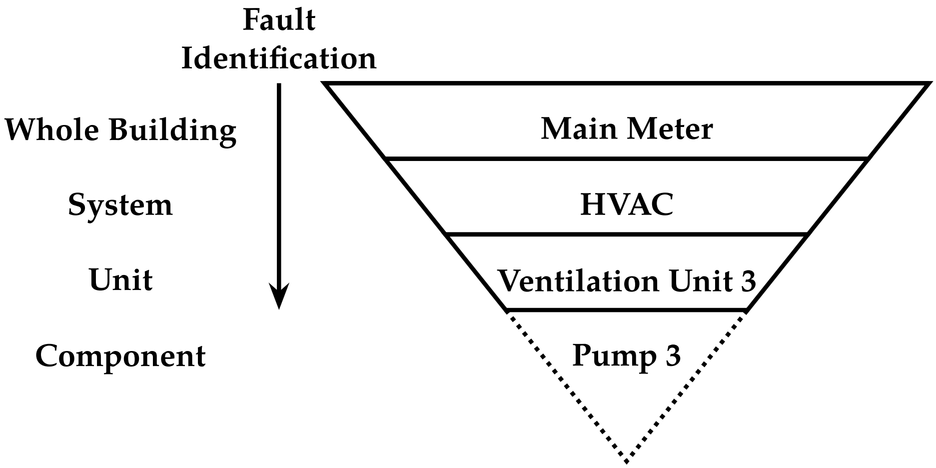

Figure 6.2: Top-down approach in fault detection and diagnostics. Comparing recursively different layers of the building's distribution tree allows to reduce the scope of faults.

simulated values is detected at the main meter, the next sub-meters layer are compared to understand which system is affected by the fault. This recursive investigation continues until reaching the leaves of the energy distribution tree. At this point the smallest unit or zone where the fault is located was identified. After the scope was reduced, it is possible to use a more focused FDD method to completely isolate the fault.

Let's assume, for instance, that we have detected a higher consumption of the building with respect to the district heating distribution tree. Hot water coming from the district heating pipes is used to heat up air in the ventilation units and water in radiators. In our next step the simulated and actual values for the respective sub-meters are compared. If the radiators are found responsible for the deviation, the ventilation units are then excluded from the investigation and labeled as not faulty. Depending on the granularity of sub-meters, we could go deeper in the distribution tree and isolate the exact areas responsible for higher energy consumption, and from there perform specific FDD for radiators.

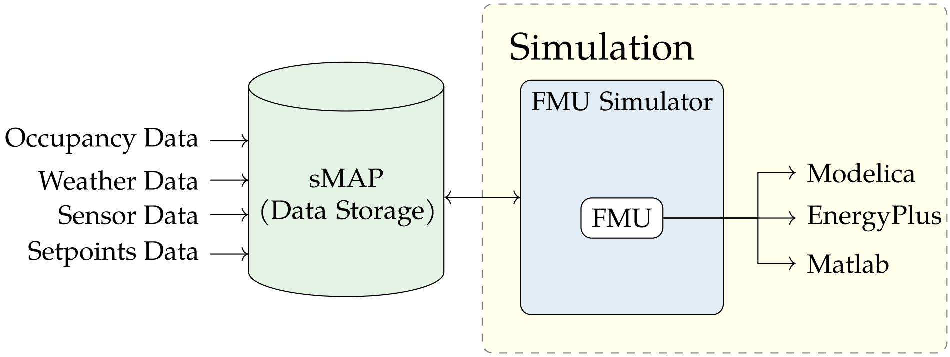

Figure 6.3: Architecture of online energy simulator. All data are accessed through sMAP and the simulation engine is embedded in a FMU and operated through FMI.

6.4 ONLINE ENERGY SIMULATOR

The online energy simulator is a tool that

- fetches required data for the simulation (e.g. weather conditions or occupancy count) from time series on the data storage;
- maps such time series to a model's input variables;
- runs the simulation for a specified number of steps / period of time;
- collects results from model's output variables;
- posts results to the data storage.

All these operations are automated and the online energy simulator can be run without any manual intervention. The high-level architecture is shown in Figure 6.3.

The online energy simulator uses the simple measurement and actuation profile (sMAP) for accessing building data, a protocol common for building systems [93]. The protocol supports reading and writing time series. It also supports time series metadata in form of key-value pairs. Metadata can be used to query the data storage for the correct time series. The protocol is independent of the underlying storage system. In order to add support for sMAP to a system it is enough to develop a *driver*, i.e. an application that forwards data from such system over sMAP.

In order to support different simulation engines, the online energy simulator uses the functional mock-up interface (FMI). FMI is an interface to perform model exchange and co-simulation of dynamic models [94]. It allows to wrap an existing model in a self-contained functional mock-up unit (FMU) and to make it available to other programs. A program can run simulations through FMUs without any information about the actual simulation engine.

6.4.1 Configuration

FMUs expose input and output variable through the FMI. The online energy simulator uses a set of configuration files to map such variables to time series. Input variables can be provided in three different ways

- Explicitly: the variable's value is constant over the whole simulation period and set in the configuration file;
- From a CSV file;
- From a time series on sMAP, identified by its universally unique identifier (UUID).

Basic arithmetic operations are also supported to allow unit conversion. For each input variable the online energy simulator will either prepare a constant time series, load it from the CSV file or fetch it from the data storage. Then it will pass it to the FMU and start the simulation.

Output variables are mapped to sMAP time series by *source name*, *path* and *UUID*. The online energy simulator also supports setting metadata of output time series, e.g. its unit or its location. An example of mapping configuration is shown in Listing 6.1.

Besides input / output mappings the online energy simulator reads from configuration files the path to FMU, simulation start / end time, simulation step size and sMAP connection details.

The FMU and configuration files completely define the behaviour of the online energy simulator. Therefore, it is simple to replace the model when a new more accurate version is available, or even to switch

Listing 6.1: Example Mapping Configuration File.

```
# Input variables
CO2_Setpoint_Zone_1_UUID=12345678-abcd-...
Heating_Setpoint_Zone_1_VALUE=21
Cooling_Setpoint_Zone_1_CSV=cooling_1.csv

# Output variables
CO2_Level_Zone_1_PATH=/simulations/zone_1/co2
CO2_Level_Zone_1_Metadata/SourceName=Simulation
CO2_Level_Zone_1_Metadata/Unit=ppm
CO2_Level_Zone_1_Metadata/Room=Room 1
CO2_Level_Zone_1_Metadata/Floor=0
```

to a different simulation engine, as long as the new one supports the FMI.

BATCH AND REAL-TIME SIMULATIONS

Necessary input data for the whole simulation period must be available at the beginning of simulation. This assumption holds for simulations over historical data, but not for simulations over present or future time, where data become available during the simulation itself. A naive solution would be to divide the simulation period in single iterations and run independent simulations in sequence. For instance, the online energy simulator could simulate one day at the time over a week. However, some engines such as EnergyPlus perform a certain amount of initial *warm-up* steps to compute initial values for room temperature and other measurements. This would result in discontinuities at the boundaries of each iteration.

To account for this use case, the online energy simulator supports a special kind of execution. The simulation period is again divided in single iterations, but the online energy simulator stops at the end of each iteration and waits for user input. Then it fetches input data only for the *next iteration period* (with the exception of weather data), and runs the next iteration. The warm-up phase is only performed at the beginning of the first iteration, and all the measurements are continuous over the entire simulation period. User input for iteration start is deterministic and, therefore, the user can be replaced by another program.

6.4.2 ENERGYPLUS SIMULATION ENGINE

EnergyPlus is a whole building energy simulation tool developed by the U. S. National Renewable Energy Laboratory [52]. It is used to simulate the building's behaviour and energy consumption over time, both at whole building level but also at room and subsystem level. It can simulate large variety of buildings subsystems such as HVAC, water and hot water distribution and lighting.

The model describing the building is contained in a single Energy Plus input (IDF) file. This file contains information about the whole building envelope, such as walls, pavements and windows, their geometry, material and thermal properties, and about the building subsystems such as ventilation units and lights. The building is divided in independent thermal zones that interacts between each other over time.

EnergyPlus supports wrapping its models to FMUs and to expose a machine-friendly interface usable by the online energy simulator [95].

WEATHER FILE UPDATE

Due to using the FMI the online energy simulator is engine-agnostic, i.e. it supports EnergyPlus models but also models from other simulation engines, as long as they expose the correct interface. There is one exception, however, because EnergyPlus has limited support for weather data as input. Instead, weather data must be provided in the form of an Energy Plus weather (EPW) file, and it needs to be available at FMU *creation time*.

Since providing updated weather data at execution time is a useful use case, the online energy simulator supports this EnergyPlus-specific feature. FMUs are in practice renamed ZIP files containing the simulation engine (or a wrapper to call the actual engine) in form of a shared library. FMUs created from EnergyPlus contains also additional files, i.e. the model IDF file and an EPW file.

When the online energy simulator loads an FMU it decompresses its ZIP file, replaces the interesting columns of its EPW file with weather data provided as input and re-compresses as a new ZIP file. In this way it is possible to provide weather data at the beginning of a simulation. Providing weather data as input *during the simulation*, such as for occupancy data or setpoints, is not possible due to limitations of EnergyPlus engine.

6.5 CASE STUDY: BUILDING OU44

In this paper we present building Odense undervisning 44 (OU44) as case study [7]. The building, shown in Figure 6.4, is located at University of Southern Denmark, campus Odense and was built in 2015. It has four floors and is mainly used for teaching and it consists of classrooms, study rooms and offices. Regarding the HVAC system, there are four ventilation units, each serving one of the corners of the building. In addition, the building is heated using a district heating loop and, partially, through the ventilation system.

Every room has the following sensors:

- Temperature [celsius];
- CO_2 [ppm];
- PIR [boolean];
- Light [lux].

Some rooms have additional sensors or meters. For instance some have separate meters for plug load or sensors for humidity. Four test rooms are equipped with occupancy counting cameras that provide an estimate of people in the room. In addition to that, the building has a weather station that records outdoor temperature, wind speed, rain and solar radiation. There are also several energy meters: for heating, ventilation, hot water, lighting, plug load, usually aggregate by floor

Figure 6.4: Building OU44 at University of Southern Denmark, campus Odense.

or area. Finally, occupancy counting cameras are also located at every entrance of the building, providing an estimate of people in the entire building.

All sensors are accessible through a KNX bus [58] and broadcast records to the BMS according to their configuration. All energy meters are accessible through an EnergyKey system. Custom drivers fetch data from the BMS and EnergyKey system and publish it to a centralized data storage using sMAP, so that it is available to other applications, such as occupancy prediction [59] and model development and calibration [60].

6.5.1 MONITORING BUILDING PERFORMANCE WITH ONLINE ENERGY SIMULATOR

An overall dynamic energy performance model for the OU44 model was developed by Jradi et al. considering various building characteristics and specifications including physical envelope, energy supply systems and operational parameters [7]. The building model is continuously re-calibrated within the developed framework, considering a

3 months timeframe. The model was prepared for export by exposing selected input/output variables in the interface. This step is automated using the EPQuery tool [96], which helps to modify EnergyPlus IDF files using Python scripts. Employing the developed dynamic model, the online energy simulator was configured and deployed to the case-study building OU44 to monitor its energy performance. Once per day a simulation is run over the previous 24 h providing the following input data:

– Weather data from the local weather station: outdoor temperature, wind speed and solar radiation;
– Whole building occupancy data, obtained from occupancy counting cameras;
– Single room occupancy data for the four test rooms that have occupancy counting cameras.

We focused on the four test rooms because having an estimate of the occupants count helps understanding their dynamics. These rooms also have additional room level energy meters and higher resolution sensors.

The following output variables were collected at each simulation step, i.e. 10 min, and posted to data storage:

– Whole building electrical energy consumption;
– Whole building heating energy consumption;
– Whole building lighting energy consumption;
– Electricity consumption for the four ventilation units;
– Room temperature for the four test rooms;
– CO_2 level for the four test rooms.

An overall building occupancy profile was generated using input from the different camera counts around the building [97]. The model assumes that occupants spread uniformly over the entire building. For the four test rooms, however, specific occupancy count estimates are provided to improve simulation accuracy.

Once results are posted to data storage, they are available to every other application. In particular, simulation results can be compared with the actual measured values. This allows to detect any deviation or differences between the actual and predicted performance of the building.

6.5.2 RESULTS

In this section we show the results obtained by running the online energy simulator on OU44. We used an EnergyPlus model and we ran simulations for 8 months from Thursday 1 September 2016 to Sunday 14 May 2017. We provided whole building occupancy count, room level occupancy counts for four test rooms, outdoor temperature, wind speed and solar radiation as simulation input. We show charts for selected time periods.

RESULTS FOR ENERGY PERFORMANCE

Figure 6.5 shows the simulated and measured electrical energy consumption over a week for building OU44. Cumulative energy consumption over time is shown on the left column and energy consumed every 2 h is shown on the right column. We chose this value because some of the sub-meters have low time resolution, which resulted in spikes using shorter values. The last row shows the total occupants in the building, estimated through the occupancy counting cameras.

Energy performance at the whole building level is on par with the simulation results, with a small deviation toward the end of the week. We consider the next sub-meters layer, i.e. the ventilation system and lighting. The rest of energy consumption is due to building operations, such as elevators and plugs load. We observe two distinct phenomena: the ventilation system performs consistently worse than the model, and energy consumption for lighting deviates significantly during the weekend.

We can explain the anomaly for lighting by looking at occupancy over time. During the weekend, occupants count drops but the building is not completely empty. It is possible that a small number of students come to study on weekends and spread to different rooms. In this case the lights would be turned on for many rooms even with a small number of occupants, while the model assumes a proportional lighting energy consumption.

We continue our investigation of the ventilation system and examine the sub-meters in the next layer, i.e. at the individual ventilation units. Unit 1 follows closely the simulation, but the other three deviate. Units 2 and 4 consume less energy than expected, while unit 3 consumes significantly more. There are no more meters in the ventilation units, therefore, we cannot further compare simulated and measured performance. We succeeded in reducing the scope to ventilation unit 3, which has a large deviation from the expected performance and now we can run specific FDD techniques to completely isolate the faulty component. Further investigation should also be performed to understand why ventilation units 2 and 4 have a lower energy consumption than expected.

RESULTS FOR INDOOR CONDITIONS

In addition to energy meters, we compared the room level indoor conditions measured by building sensors with the ones from the simulation.

Figure 6.6 shows the simulated and measured room temperature for one of the four test rooms. Although the dynamic EnergyPlus model was calibrated based on the overall energy consumption of the building, actual room indoor air temperature were found to be in line with the model predictions, with the two values following the same trend. However, it is noticed that room temperature measured by the building sensor quickly drops during the night of Tuesday 4 April 2017, deviating from the simulated value.

We can explain this anomaly by noticing that the indoor temperature follows closely the outdoor temperature recorded by the building's weather station. The most likely cause was that the room windows were left open during the night.

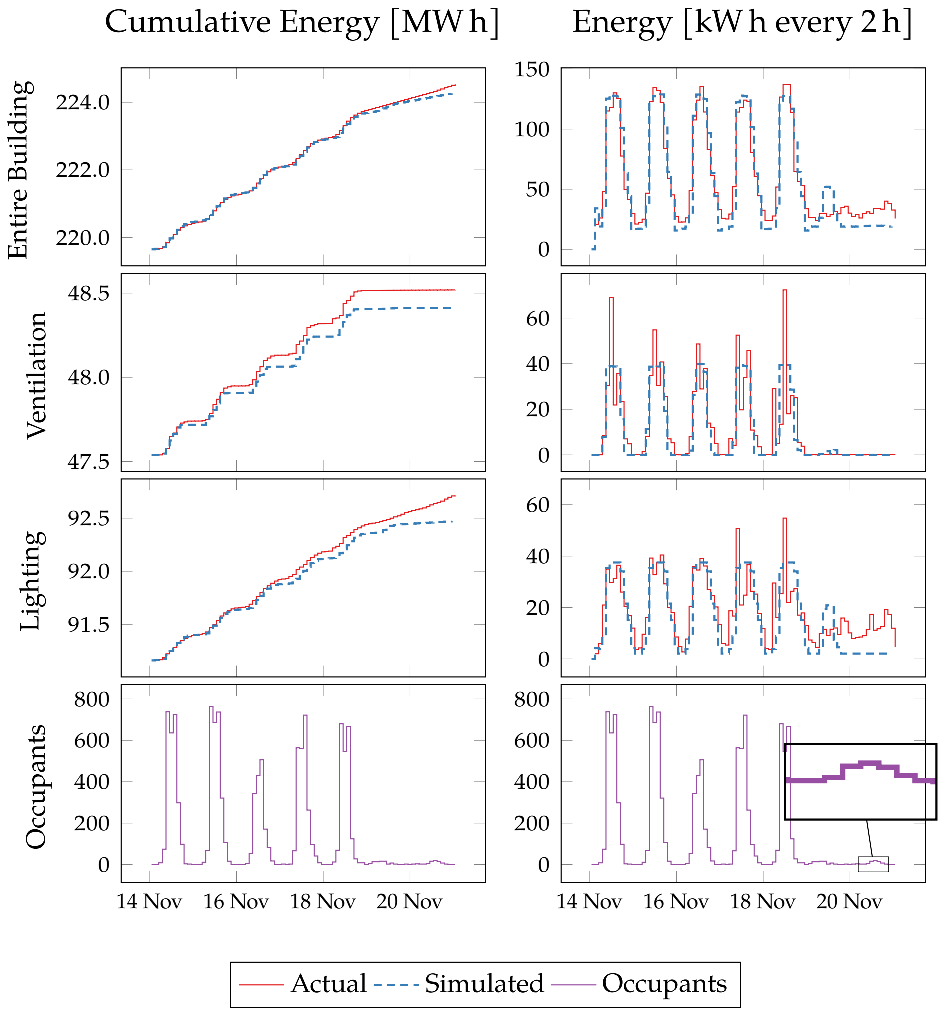

Figure 6.5 (first): Data from energy meters and simulation results for building OU44. Cumulative energy consumption over time is shown on the left column, energy consumed every 2 h on the right one. Total occupants in the buildings are shown on the last row.

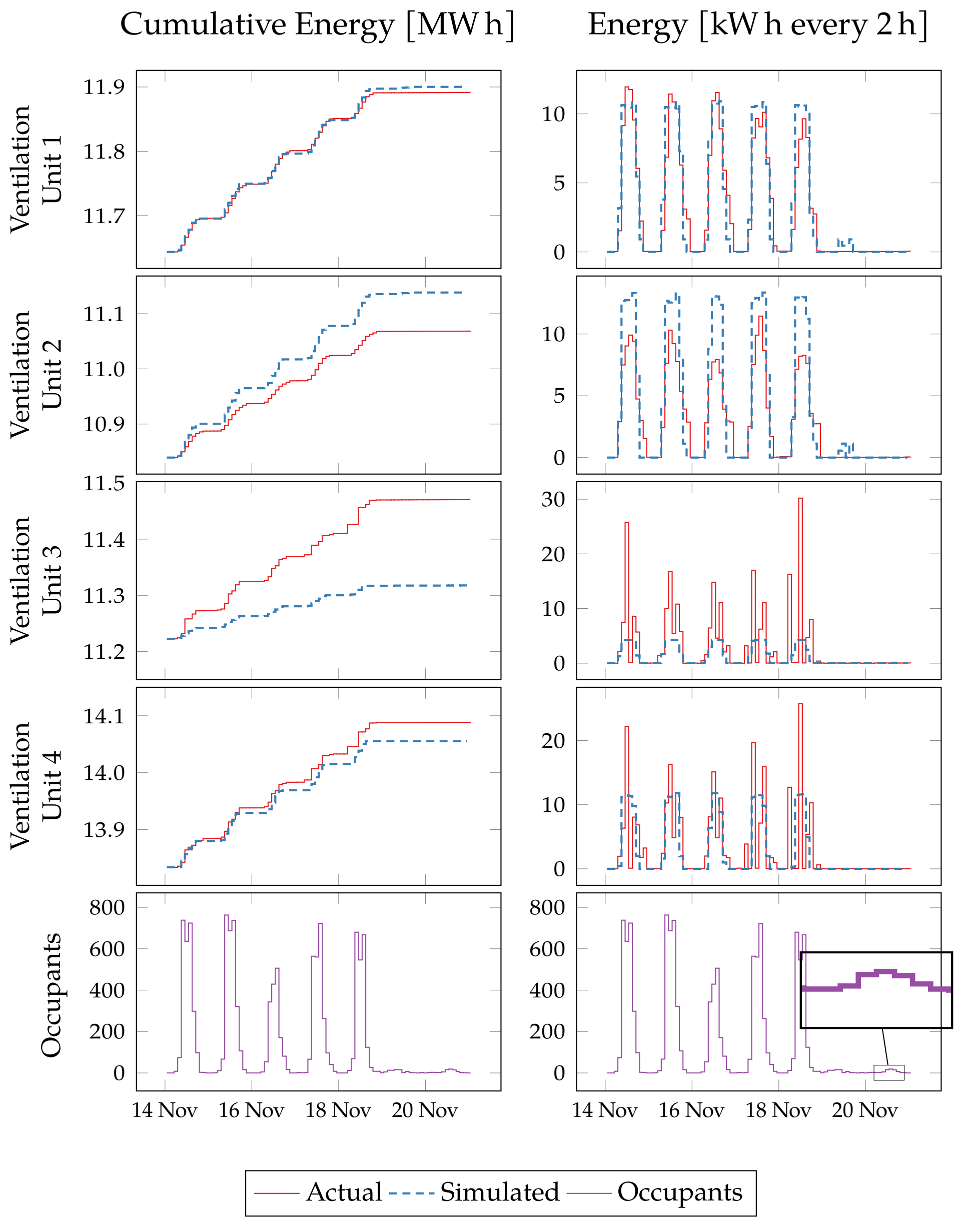

Figure 6.5 (continued): Data from energy meters and simulation results for building OU44. Cumulative energy consumption over time is shown on the left column, energy consumed every 2 h on the right one. Total occupants in the buildings are shown on the last row.

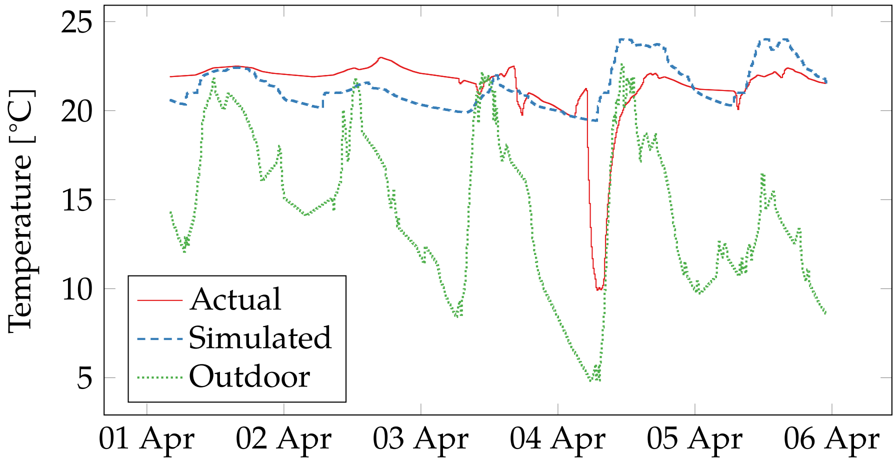

Figure 6.6: Comparison between simulated and measured room temperature. Temperature dropped sharply during one night, following the outdoor temperature.

COMPUTATIONAL LOAD OF SIMULATIONS

In order to estimate the computational load of simulations we ran the online energy simulator over periods of different lengths and recorded the elapsed time. The results are shown in Table 6.1. Simulating an entire day or even an entire month only takes few minutes. The elapsed times are very similar even for very different simulation periods because EnergyPlus spends long time during the warm-up phase, which is the same for every simulation.

6.6 CONCLUSIONS

We proposed a method for FDD in building systems using dynamic energy models to simulate the expected behaviour of the building and compare it with the actual one at different layers. We presented a tool for scheduling and automatically running simulations without user interaction, using industry standard interfaces to support many simulation engines and building systems. Finally, we tested our method

Table 6.1: Elapsed time for different simulation periods.

Simulation Period	Elapsed Time
1 d	279 s
7 d	349 s
30 d	518 s
60 d	683 s

and tool on a real buildings, identifying anomalies in energy consumption of lighting and ventilation units, and in room temperature. As the tool was implemented for a short time for validation in the case-study building, the savings due to the implementation were not evaluated, but major expected savings include less operational costs, higher maintenance process response, lower energy consumption and higher thermal comfort.

Splitting energy consumption in sub-meters allowed us to understand how different subsystems use energy inside our building. We were able to follow the energy distribution tree from its root to its leaves, ruling out branches where measured values were on par with simulation results and exploring the ones where the they deviate. We succeeded in identifying the ventilation unit responsible for higher energy consumption and gained insights about the lighting system.

We also showed how using an automated solution to schedule simulations can reduce the risk for human errors. The online energy simulator developed and presented in this study has been running automatically for several months in the OU44 building within the 'ObepME Tool', Online Building Energy Performance Monitoring and Evaluation, for automatic and continuous energy monitoring and evaluation of the overall building energy performance aiming to reduce energy performance gaps and forming a backbone for fault detection and diagnostics [8]. Thanks to a configuration-based approach, we are able to easily upgrade and calibrate the dynamic model to newer versions and repeat simulations over any period with any functional changes.

6.6.1 FUTURE WORK

The methodology proposed in this paper covers the high level identification of a faulty subsystem, and represents an important intermediate block of a complete FDD solution for building systems. In order to perform a full FDD it is first necessary to ensure validation of input data—which we previously approached in [1]—and then to use specific methods to completely isolate the faulty component inside the identified subsystem. Those methods should exploit the characteristics of the considered systems, such as individual ventilation units or room lighting, to reach the best FDD performance. Moreover, simulated and measured data are both available on our data storage for client applications, but they are not accessible in a user-friendly way. A dashboard application would enable non-technical users to assess the building status and performance.

Furthermore, we are extending the online energy simulator to play an important role as component of a new *virtual building*. The virtual building behaves as closely as possible to a real building, also with respect to control input. It waits for new actuation commands to be posted to our data storage and simulate the outcome. A BMS can then be deployed on the virtual building making possible to test our control strategies before deploying it on a real one.

ACKNOWLEDGEMENT

This work is supported by the Innovation Fund Denmark for the project COORDICY.

REFERENCES

[1] **Claudio Giovanni Mattera**, Sanja Lazarova-Molnar, Hamid Reza Shaker and Bo Nørregaard Jørgensen. 'A Practical Approach to Validation of Buildings' Sensor Data: a Commissioning Experience Report'. In: *Third International Conference on Big*

Data Computing Service and Applications (*BigDataService*) (San Francisco, CA, USA, 6th–9th Apr. 2017). IEEE. 12th June 2017, pp. 287–292. DOI: 10.1109/BigDataService.2017.48.

[2] **Claudio Giovanni Mattera**, Muhyiddine Jradi and Hamid Reza Shaker. 'Online Energy Simulator for Building Fault Detection and Diagnostics Using Dynamic Energy Performance Model'. In: *International Journal of Low-Carbon Technologies* 13.3 (17th May 2018), pp. 231–239. ISSN: 1748-1325. DOI: 10.1093/ijlct/cty019.

[7] Muhyiddine Jradi, Fisayo Caleb Sangogboye, **Claudio Giovanni Mattera**, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'A World Class Energy Efficient University Building by Danish 2020 Standards'. In: *Energy Procedia* 132 (Oct. 2017): *11th Nordic Symposium on Building Physics*, pp. 21–26. ISSN: 1876-6102. DOI: 10.1016/j.egypro.2017.09.625.

[8] Muhyiddine Jradi, Krzysztof Arendt, Fisayo Caleb Sangogboye, **Claudio Giovanni Mattera**, Elena Markoska, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'ObepME: An Online Building Energy Performance Monitoring and Evaluation Tool to Reduce Energy Performance Gaps'. In: *Energy and Buildings* 166 (1st May 2018), pp. 196–209. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2018.02.005.

[12] European Parliament and Council of the European Union. 'Directive 2010/31/EU of the European Parliament and of the Council on the Energy Performance of Buildings'. In: *Official Journal of the European Union* L153 (18th June 2010), pp. 13–35. EU DIRECTIVE: 2010/31.

[16] *2011 Buildings Energy Data Book*. Tech. rep. U. S. A. Department of Energy, Mar. 2012. URL: https://openei.org/doe-opendata/dataset/buildings-energy-data-book (visited on 20th Dec. 2018).

[22] European Commission. *Proposal for a Directive of the European Parliament and of the Council Amending Directive 2012/27/EU on Energy Efficiency*. 30th Nov. 2016. EU DOCUMENT: 52016PC0761.

[36] Evan Mills. 'Building Commissioning: A Golden Opportunity for Reducing Energy Costs and Greenhouse Gas Emissions in the United States'. In: *Energy Efficiency* 4.2 (May 2011), pp. 145–173. ISSN: 1570-6478. DOI: 10.1007/s12053-011-9116-8.

[41] Woohyun Kim and Srinivas Katipamula. 'A Review of Fault Detection and Diagnostics Methods for Building Systems'. In: *Science and Technology for the Built Environment* 24.1 (8th May 2018), pp. 3–21. ISSN: 2374-4731. DOI: 10.1080/23744731.2017.1318008.

[52] Drury B. Crawley, Linda K. Lawrie, Frederick C. Winkelmann, W.F. Buhl, Y.Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte and Jason Glazer. 'EnergyPlus: Creating a New-generation Building Energy Simulation Program'. In: *Energy and Buildings* 33.4 (Apr. 2001): *Building Simulation '99*, pp. 319–331. ISSN: 0378-7788. DOI: 10.1016/S0378-7788(00)00114-6.

[53] Peter Fritzson. *Introduction to Modeling and Simulation of Technical and Physical Systems with Modelica*. John Wiley & Sons, Inc, 17th Apr. 2012. DOI: 10.1002/9781118094259.

[54] J. A. Clarke and J. L. M. Hensen. 'Integrated Building Performance Simulation: Progress, Prospects and Requirements'. In: *Building and Environment* 91 (Sept. 2015): *Fifty Year Anniversary for Building and Environment*, pp. 294–306. ISSN: 0360-1323. DOI: 10.1016/j.buildenv.2015.04.002.

[55] Andrea Costa, Marcus M. Keane, J. Ignacio Torrens and Edward Corry. 'Building Operation and Energy Performance: Monitoring, Analysis and Optimisation Toolkit'. In: *Applied Energy* 101 (Jan. 2013): *Sustainable Development of Energy, Water and Environment Systems*, pp. 310–316. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2011.10.037.

[58] KNX Association. *What Is KNX?* URL: https://www.knx.org/knx-en/for-professionals/What-is-KNX/A-brief-introduction// (visited on 26th Dec. 2018).

[59] Sanja Lazarova-Molnar, Mikkel Baun Kjærgaard, Hamid Reza Shaker and Bo Nørregaard Jørgensen. 'Commercial Buildings Energy Performance within Context Occupants in Spotlight'. In: *Proceedings of the 4th International Conference on Smart Cities and Green ICT Systems - Volume 1: SMARTGREENS* (Lisbon, Portugal, 20th–22nd May 2015). SciTePress, 2015, pp. 306–312. DOI: 10.5220/0005495203060312.

[60] Muhyiddine Jradi, Christian T. Veje and Bo Nørregaard Jørgensen. 'Towards Energy Efficient Office Buildings in Denmark: the Maersk Building Case Study'. In: *29th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems* (Portorož, Slovenia, 19th–23rd June 2016). Ed. by Andrej Kitanovski and Alojz Poredoš. University of Ljubljana. 2016. ISBN: 978-961-6980-15-9.

[86] European Parliament and Council of the European Union. 'Directive 2012/27/EU of the European Parliament and of the Council on on energy efficiency'. In: *Official Journal of the European Union* L315 (14th Nov. 2012), pp. 1–56. EU DIRECTIVE: 2012/27.

[87] Center for Energy Informatics, University of Southern Denmark. *COORDICY*. URL: https://www.sdu.dk/en/om_sdu/institutter_centre/centreforenergyinformatics/research+projects/coordicy (visited on 3rd Dec. 2018).

[88] Tobias Maile, Vladimir Bazjanac and Martin Fischer. 'A Method to Compare Simulated and Measured Data to Assess Building Energy Performance'. In: *Building and Environment* 56 (Oct. 2012), pp. 241–251. ISSN: 0360-1323. DOI: 10.1016/j.buildenv.2012.03.012.

[89] Michael Wetter. 'Co-simulation of Building Energy and Control Systems with the Building Controls Virtual Test Bed'. In: *Journal*

of Building Performance Simulation 4.3 (5th Nov. 2011), pp. 185–203. DOI: 10.1080/19401493.2010.518631.

[90] Xiufeng Pang, Michael Wetter, Prajesh Bhattacharya and Philip Haves. 'A Framework for Simulation-based Real-time Whole Building Performance Assessment'. In: *Building and Environment* 54 (Aug. 2012), pp. 100–108. ISSN: 0360-1323. DOI: 10.1016/j.buildenv.2012.02.003.

[91] Xiufeng Pang, Thierry S. Nouidui, Michael Wetter, Daniel Fuller, Anna Liao and Philip Haves. 'Building Energy Simulation in Real Time Through an Open Standard Interface'. In: *Energy and Buildings* 117 (1st Apr. 2016), pp. 282–289. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2015.10.025.

[92] Tanzia Sharmin, Mustafa Gül, Xinming Li, Veselin Ganev, Ioanis Nikolaidis and Mohamed Al-Hussein. 'Monitoring Building Energy Consumption, Thermal Performance, and Indoor Air Quality in a Cold Climate Region'. In: *Sustainable Cities and Society* 13 (Oct. 2014), pp. 57–68. ISSN: 2210-6707. DOI: 10.1016/j.scs.2014.04.009.

[93] Stephen Dawson-Haggerty, Xiaofan Jiang, Gilman Tolle, Jorge Ortiz and David Culler. 'sMAP: A Simple Measurement and Actuation Profile for Physical Information'. In: *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems* (Zürich, Switzerland, 3rd–5th Nov. 2010). New York, NY, USA: ACM, Nov. 2010, pp. 197–210. DOI: 10.1145/1869983.1870003.

[94] Torsten Schwan, René Unger and Jörg Pipiorke. 'Aspects of FMI in Building Simulation'. In: *Proceedings of the 12th International Modelica Conference*. 132. Linköpings, Sweden: Linköping University Electronic Press, May 2017, pp. 73–78. DOI: 10.3384/ecp1713273.

[95] U. S. A. Department of Energy. *EnergyPlusToFMU*. URL: https://simulationresearch.lbl.gov/projects/energyplustofmu (visited on 3rd Dec. 2018).

[96] Center for Energy Informatics, University of Southern Denmark. *EPQuery*. URL: https://sdu-cfei.github.io/epquery/ (visited on 3rd Dec. 2018).

[97] Fisayo Caleb Sangoboye and Mikkel Baun Kjærgaard. 'PLCount: A Probabilistic Fusion Algorithm for Accurately Estimating Occupancy from 3D Camera Counts'. In: *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments* (Palo Alto, CA, USA, 16th–17th Nov. 2016). BuildSys '16. ACM. New York, NY, USA, 2016, pp. 147–156. DOI: 10.1145/2993422.2993575.

CHAPTER 7

INTRODUCING REDUNDANCY THROUGH LINEAR REGRESSION VIRTUAL SENSORS

This chapter is a cosmetic adaptation of the following conference paper.

Claudio Giovanni Mattera, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'Fault Detection and Diagnostics in Ventilation Units Using Linear Regression Virtual Sensors'. In: *International Symposium on Advanced Electrical and Communication Technologies (ISAECT)* (Kenitra, Morocco, 21st–23rd Nov. 2018). IEEE. 24th Jan. 2019. DOI: 10.1109/ISAECT.2018.8618755

This work was the result of a collaboration with the Center for Supervision, Security and Automatic Control (CS^2AC), at the Polytechnic University of Catalonia (UPC), Barcelona, Spain.

The paper was presented at the IEEE International Symposium on Advanced Electrical and Communication Technologies in Kenitra, Morocco, on Wednesday 21 November 2018.

ABSTRACT

Buildings represent a significant portion of global energy consumption. Ventilation units are one of the largest components in building systems and are responsible for large part of energy consumption.

Ventilation units are complex components, often customized for the specific building. Their faults impact buildings' energy efficiency and occupancy comfort. In order to ensure their correct operation, proper

fault detection and diagnostics methods must be applied. Hardware redundancy, an effective approach to detect faults, leads to increased costs and space requirements.

We propose to exploit physical relations inside the unit to create virtual sensors from other sensors' readings, introducing redundancy in the system. We create linear regression models for three sensors using other sensors related through physical laws as inputs. We use two different measures to detect when a virtual sensor deviates from the actual one: coefficient of determination and acceptable range.

We test our method on a real building at the University of Southern Denmark. Our method detects a fault in temperature sensor, where its readings have an abnormal trend and fall outside acceptable range for one day.

7.1 INTRODUCTION

In Europe, buildings account for 40 % of the total energy used and 36 % of the total CO_2 emissions [12]. In the United States, the buildings' sector accounted for about 41 % of primary energy consumption in 2010, 44 % more than the transportation sector and 36 % more than the industrial sector. Total building primary energy consumption in 2009 was about 48 % higher than consumption in 1980, going from 1290 TW h to 2784 TW h [16].

Modern buildings consist of different subsystems such as heating, ventilation and air conditioning (HVAC) units, lighting and heating. Each subsystem contains in turn several components such as pumps, fans, ducts, sensors, lamps, wires etc. monitored and managed by a building management system (BMS). All these components are subject to faults, due to damage, wearing over time, misconfiguration and communication issues. Faults impact occupancy comfort, maintenance cost and particularly energy efficiency. It is estimated that in 2009 just 13 of the most common faults were responsible for over $3.3 billions in energy loss [36].

HVAC load varies depending on building type and location, but they are one of the heaviest subsystems and can make up to 50 % of the total energy consumption and, therefore, faults involving them cause large energy loss [98]. Research suggests that between 20 % to 30 % energy saving could be achieved by re-commissioning malfunctioning HVAC systems [99]. HVAC systems are often customized for their specific building and, therefore, lack quality system integration [100].

7.1.1 Problem statement

Building energy efficiency cannot be achieved without fault detection and diagnostics (FDD) methods applied to ventilation units. Hardware redundancy is an effective approach to high quality FDD, however, duplicating sensors and other components inside every unit increases deployment and maintenance costs, necessary space and complexity. Commercial ventilation units are rarely shipped with hardware redundancy.

In this paper we propose a mixed model-based and data-driven technique to exploit spatial relations among different components in ventilation units to create *virtual sensors* and introduce redundancy in the system, which can be used to detect and diagnose faults. For each considered sensor we train a linear regression model to estimate it given other sensors in the unit. This allows us to detect and diagnose faults that cause actual and virtual sensors to deviate from each other. We apply this technique to a real world building and report the results.

The rest of the paper is organized as follows. The state of the art is reviewed in Section 7.2. The proposed technique is introduced in Section 7.3. Section 7.4 presents the case study and discusses results and implications. Finally, conclusions are drawn in Section 7.5.

7.2 State of the art

7.2.1 Fault detection and diagnostics

Kim et al. present a comprehensive review of recent FDD methods for building systems [41]. FDD methods are categorized in three groups depending on the approach: data-driven methods, model-based methods and rule-based methods.

In data-driven methods a black-box model of the system under test is trained over historical data using techniques such as artificial neural network or regression models. These methods require no detailed knowledge of the system and can be easily generalized. Historical labeled faulty and fault-free data are necessary to improve fault detection and to perform fault isolation and diagnostics.

In model-based methods a model of the system under test is created from first principles. These techniques are often accurate and can detect and diagnose unknown faults. Models can become complex, and detailed knowledge of the physical characteristics and relations of the system's components is required to create them.

In rule-based methods expert knowledge is used to design a set of rules describing the system's behaviour. No historical data or detailed knowledge from the system are necessary. Large rules sets are necessary to describe complex behaviours, which lead to conflicts and maintenance effort.

Yu et al. present a review of FDD techniques for ventilation units [100]. The authors focus on software redundancy techniques, classifying them in model-based, data-driven and rules-based categories as in general FDD methods, and define a list of desirable characteristics: 1. Quick detection and diagnostics, 2. Isolability, 3. Robustness, 4. Novel identifiability, 5. Classification error estimate, 6. Adaptability, 7. Explanation facility, 8. Modeling requirements, 9. Storage and computational requirements, 10. Multiple fault identifiability.

7.2.2 Virtual sensors

Li et al. present a review of virtual sensing techniques in the context of building systems [51]. Virtual sensors have been successfully applied to other fields such as process control and automotive for more than two decades, and their usage would be advantageous in building systems. Virtual sensing techniques are categorized according to three criteria. Measurement characteristics, i.e. whether the sensors represents sensor at steady state or during transients. Modeling method, i. e., model-based or data driven, a similar characterization as general FDD techniques. Application purposes, i. e., whether the sensors are used for redundancy and FDD, or for monitoring additional unknown quantities.

Li et al. propose a method for FDD in air conditioners using features decoupling and virtual sensors. The authors create virtual sensors for several quantities, such as compressor power consumption, refrigerant flow, condenser exit pressure, exit air humidity and evaporation temperature. Virtual sensor performances are tested both at steady state and under transients [101].

Cugueró-Escofet et al. present an approach for sensors data validation and reconstruction and apply it to urban water distribution systems. Raw data undergoes several tests, from low-level tests checking elementary properties of signals to high-level tests exploiting *spatial consistency* between different sensors [78].

Cotrufo et al. develop a virtual sensor modeling exhaust airflow in ventilation units. Airflow sensors for exhaust duct are rarely present in ventilation units due to initial cost. They use energy balance equation to relate other sensors in the system with the airflow and propose two different models. While the local errors can be large, the authors show how the cumulative residuals are small and, therefore, the virtual sensor can be used to estimate daily averages [102].

Kusiak et al. propose data-driven models for virtual sensors for room level indoor air conditions, i.e. temperature, CO_2 level and relative humidity. The authors develop four data mining techniques, including artificial neural networks, support vector machines regression and Pace regression. The obtained virtual sensors can be used for validation and calibration of physical sensors [103].

Verbert et al. propose a multi-model FDD method for HVAC systems that exploits components interdependencies. They develop Bayesian networks for multiple operating modes, using both actual and virtual sensors created from system knowledge and historical data. The authors show how using virtual sensors significantly improves FDD performance [104].

7.3 VIRTUAL SENSORS IN VENTILATION UNITS

A ventilation unit is an aggregate of several components, integrated together to provide air exchange to the building. It is important that every component works correctly, otherwise performance of the unit will deteriorate, causing energy loss and reducing comfort level in the building.

Since all components work together they exhibit common patterns and shared phenomena. Even if there is no explicit redundancy in the system, i.e. no duplicated sensor or meter, many of the quantities in the unit are strongly correlated. In this paper we propose to exploit these relations and create models to predict a quantity from the surrounding ones, generating a set of *virtual sensors*. Given actual sensors available in the ventilation unit $S_1, S_2, \dots, S_n$, a virtual sensor S'_i measuring the same quantity of S_i is created using a model $f(\cdot)$ that takes other sensors as input, i. e.,

$$\begin{aligned} S'_i &= f(\overline{\mathcal{S}}) \\ \overline{\mathcal{S}} &\subsetneq \{S_1, S_2, \dots, S_{i-1}, S_{i+1}, \dots, S_n\}. \end{aligned}$$

For instance, consider a heating system where the following quantities are measured with sensors or meters: initial temperature T_0, heater energy M and final temperature T_f. A virtual sensor for final temperature could be created using a model of initial temperature and heater energy $T'_f = f(M, T_0)$.

Different methods can be used to compute the value of a virtual sensor. When detailed knowledge about the unit is available it is possible to use physical models, e.g. computing airflow using fan speed and duct size and shape. Otherwise, it is possible to train black box models using data-driven techniques.

7.3.1 FAULT DIAGNOSTICS

When two sensors, either actual or virtual, deviate, the only possible inference is that a fault is affecting one of them. In order to diagnose the faulty one a third sensor is necessary. Under the assumption of single simultaneous fault, when in a group of three sensors one deviates from the other twos, the former is identified as faulty.

Due to cost and space constraints, duplicated sensors are rarely available in ventilation units, and even less so are triplicated sensors. However, these constraints do not impact virtual sensors, which can be created without cost using data from other components. Some care is necessary when choosing the inputs: different virtual sensors should share as few inputs as possible, because a fault in an input impacts all its related virtual sensors.

For instance, consider a heating system with two initial temperature sensors T_0, T_1, a heater energy meter M and a final temperature sensor T_f, where two additional virtual sensors for final temperature were created as

$$T_f' = f(M, T_0), \qquad T_f'' = f(M, T_1).$$

Assuming a single fault scenario, if T_f' and T_f'' agree on their readings and T_f deviates from them there are two possible causes:

– Sensor T_f is faulty;
– Heater energy meter M is faulty.

This is due to the fact that heater energy meter M is used as input in both virtual sensors T_f' and T_f'', therefore, its fault impacts both their output.

7.3.2 Measuring deviations from actual sensors

In order to automatically detect a fault, a measure of how much the virtual sensors deviate from the actual one is necessary. Several tools are available from statistical analysis, e.g. the maximal error or the norm of residuals. In this paper we use the coefficient of determination, or R^2 score, which gives an estimate of how much a model fit the data [105]. An R^2 score close to 1 indicates that the model is a good fit for the data, while values close to zero indicates the opposite. Negative values indicate that the model predicts data worse than a constant horizontal line.

We use the R^2 score both to verify that the trained models fit the testing data, i.e. that the designed model accurately follows the actual sensor, and to validate real-time data from the ventilation unit. For each period of interest, e.g. every day, the R^2 score for each virtual sensor against the actual sensor is recorded. When the measure is lower than a given threshold the pair virtual / actual sensors are flagged as faulty.

Another option for detecting deviations between actual and virtual sensor is to make the latter output an *acceptable range*. E.g. the predicted value plus the largest training error, or a confidence interval based on another training error statistics. When actual readings fall outside the acceptable range the two sensors are flagged as faulty.

With both approaches, labeled faulty testing data would be necessary to obtain accurate thresholds.

7.4 CASE STUDY

7.4.1 BUILDING OU44

In this paper we present building Odense undervisning 44 (OU44) as a case study [7]. It was built in 2015 at the University of Southern Denmark, campus Odense, and it is mainly used for teaching. It has three floors plus a basement and it contains classrooms, study zones, offices and auditoriums. It has four nearly identical ventilation units, each serving one corner of the building, or roughly 20 thermal zones.

A ventilation unit consists in a large air loop, as shown in Figure 7.1. Inlet air enters the building, goes through a heat exchanger (HX), then is heated to appropriate indoor temperature and pushed to the supply shaft, which is connected by variable air volume (VAV) units to individual rooms. In the same way, exhaust air is collected from individual rooms in the extract shaft, it goes through the heat exchanger and it is pushed outside. The heat exchanger recovers heat from exhaust air and transfers it to inlet air, reducing the energy required by the heater. Air pressures in supply and extract shafts are kept at constant values 130 Pa and 40 Pa, which cause air to flow in the rooms. Two

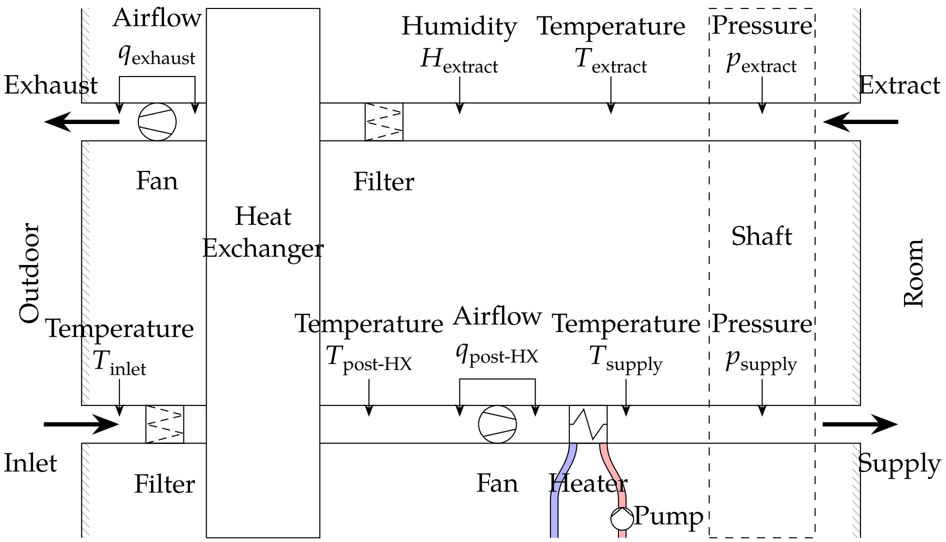

Figure 7.1: Diagram of a ventilation unit in building OU44. Inlet air enters the unit from bottom-left, passes through the heat exchanger and through the heater, before entering the main shaft and supplying individual rooms. From the rooms it enters again the main shaft, goes through the heat exchanger to heat up inlet air, and finally is pushed outside the building. Several sensors, shown by arrows, are available in the unit.

fans in the ventilation unit generate the required airflows to maintain the pressure setpoints.

Several sensors, shown as arrows in Figure 7.1 are available inside ventilation units and heating loops: air temperature at several positions, airflows trough the two fans, supply and extract pressure, incoming and outgoing water temperature, and water flow through the pump. In addition to that, several meters measures the activity of fans and water pump: fan speed, fan current and voltage, fan power and electrical consumption, and pump electrical consumption.

Ventilation units are only working during working hours, i.e. from Monday to Friday from 7am to 6pm in local time, and are shut down at night and during the weekends.

Table 7.1: Virtual Sensors Definitions.

Model Name	Output	Inputs
Model A	post-HX temperature	Inlet temperature, extract temperature, airflow
Model B	post-HX temperature	Inlet temperature, water flow, water loop temperature difference
Model C	Airflow	Effort
Model D	Airflow	Fan speed
Model E	Fan speed	Airflow
Model F	Fan speed	Fan current, fan voltage

7.4.2 RESULTS

Three sensors were considered for monitoring in a ventilation unit: post-heat exchanger temperature, airflow and fan speed. For each of them two different models were constructed using other sensors as inputs, as shown in Table 7.1. Linear regression models were used under the assumption that inputs and outputs obey linear relations, at least locally [106]. Models were trained over a week long historical data from Monday 13 March 2017 to Sunday 19 March 2017, and tested over two weeks from Monday 27 March 2017 to Sunday 9 April 2017.

Another virtual sensor was also constructed, i.e. *Effort*, which is proportional to an estimate of the power requested to the ventilation unit and, therefore, to the airflow. By design fans produce airflow to maintain constant shaft pressure, which in turn depends on how many VAV units are open in the building. Effort is an aggregate count of those units, which makes it effectively a virtual sensor for an unknown quantity in the ventilation unit.

For each sensor two models were used, in order to perform fault diagnostics and not only fault detection. Table 7.2 shows the R^2 score of the models' predictions over each day, which measure how much

Table 7.2: Prediction R^2 score for virtual sensors.

	Temperature		Airflow		Fan Speed	
Date	**Model A**	**Model B**	**Model C**	**Model D**	**Model E**	**Model F**
2017-03-27	0.955	0.782	0.371	0.987	0.988	0.997
2017-03-28	0.989	0.804	**0.04**	0.98	0.977	0.997
2017-03-29	0.839	0.217	0.368	0.992	0.992	0.995
2017-03-30	0.894	0.729	0.681	0.956	0.956	0.996
2017-03-31	$-$**1.162**	$-$**1.995**	0.572	0.852	0.908	0.996
2017-04-03	0.86	0.442	0.87	0.967	0.968	0.997
2017-04-04	0.886	$-$**0.474**	0.644	0.983	0.984	0.997
2017-04-05	0.774	0.57	0.8	0.944	0.953	0.996
2017-04-06	0.73	0.654	0.622	0.988	0.989	0.997
2017-04-07	0.802	0.537	0.772	0.904	0.932	0.996

actual and virtual sensors agree. Low R^2 scores, indicating that models deviate from the actual sensors, are highlighted in boldface.

For temperature two models are used, one exploiting knowledge about the heat exchanger interactions (Model A), and another one which relies on similar but less structured relations between air temperature and heater (Model B). The former predicts temperature value much more accurately than the latter. Table 7.2 shows that both models deviate significantly from the actual sensor on 31 March 2017, and Model B deviates also on 4 April 2017. Readings from the actual sensors are shown in Figure 7.2 with respect to the two models' error ranges, which corresponds to the predictions plus the maximal training error.

On 31 March 2017 the actual sensor's readings oscillate strongly, in contrast with the two virtual sensors which have a smoother behaviour, and fall outside the models' error ranges. Since the two models share an input variable, i.e. inlet temperature, this situation could be caused by a fault in the actual post heat exchanger temperature sensor or in the inlet temperature sensor.

The situation on 4 April 2017 is less extreme. Model B consistently overestimate the actual sensor's readings, but the overall trend is

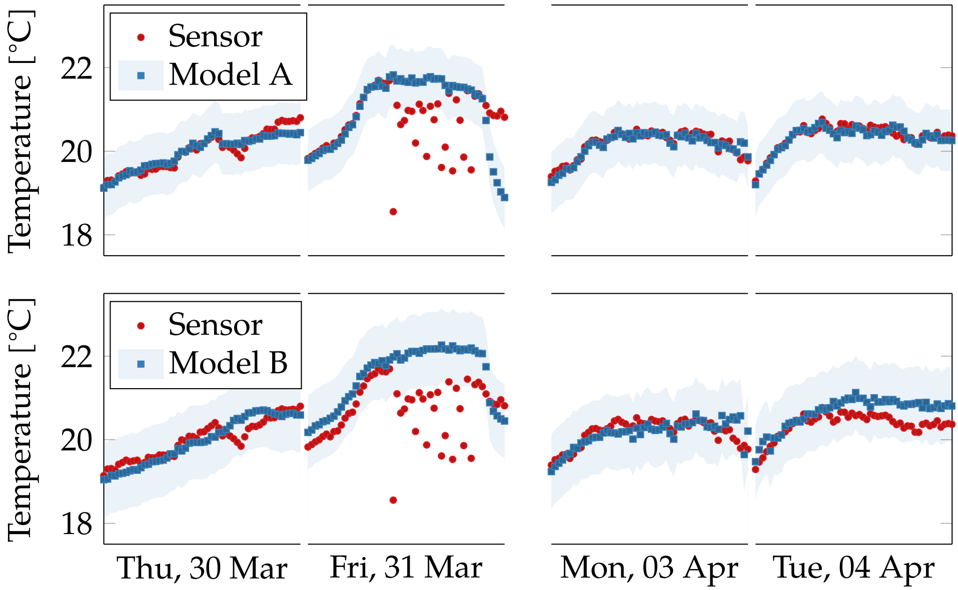

Figure 7.2: Comparison between actual sensors and acceptable ranges obtained from model-based virtual sensors for post heat exchanger temperature during working hours (from 8am to 5pm) for selected days. The sensors readings fall inside the acceptable ranges except on Friday 31 March 2017, when they deviate significantly. On Tuesday 4 April 2017 Model B consistently overestimates the actual sensors, but their trends are similar.

similar and, moreover, all the readings fall inside the model's error range. Therefore, this event could be classified as a false alarm. Using a more accurate model instead of Model B could reduce the frequency of false alarms.

For airflow two models are used, one using only effort as input (Model C) and one using only fan speed as input (Model D). Airflow and fan speed follow the fan laws and are proportional to each other [107], and as expected predictions for this model are nearly exact.

Model C is less accurate, and its R^2 score on Tuesday 28 March 2017 is very low, which suggests a fault in the virtual sensor's input, i.e. ventilation effort, since Model D agrees with the actual sensor on the same day. Ventilation effort is produced by aggregating several independent streams with frequent periods of missing data, which can indeed cause the model to deviate from the actual sensor. Moreover, ventilation effort does not take into account the size of each room and the corresponding VAV dampers, which reduces the model's accuracy. Readings from the actual sensors are shown Figure 7.3 with respect to the two models' error ranges, which corresponds to the predictions plus the maximal training error.

For fan speed two models are used, one using airflow as input (Model E) and one using fan current and voltage as inputs (Model F). Fan speed is proportional to airflow due to fan laws, and also to the fan power consumption, which in turn depends on current and voltage. Both models predict the actual sensor nearly exactly.

7.5 CONCLUSIONS

We proposed a technique to exploit relations between physical quantities inside a ventilation unit to create virtual sensors, introducing, therefore, redundancy, which can be used to perform FDD. We applied our technique to ventilation units in a real building, creating two virtual sensors for each of three existing sensors: temperature, airflow and fan speed, using linear regression models. We noted how on a particular day both virtual sensors for temperature deviated from the actual sensors, which suggests a fault has happened.

We used simple linear regression model to generate virtual sensors and predict physical quantities based on other sensors. Some virtual sensors were accurate, but some others were not. Better performance could be achieved by using more advanced methods, such as artificial neural networks, statistical machine learning algorithms or energy models of the ventilation units [2]. Assuming to have a training period of fault-free historical data it would also be possible to adopt methods from time-series analysis, such as autoregressive moving average with exogenous variables (ARMAX) predictors, to create a virtual sensor using its past actual sensor as input.

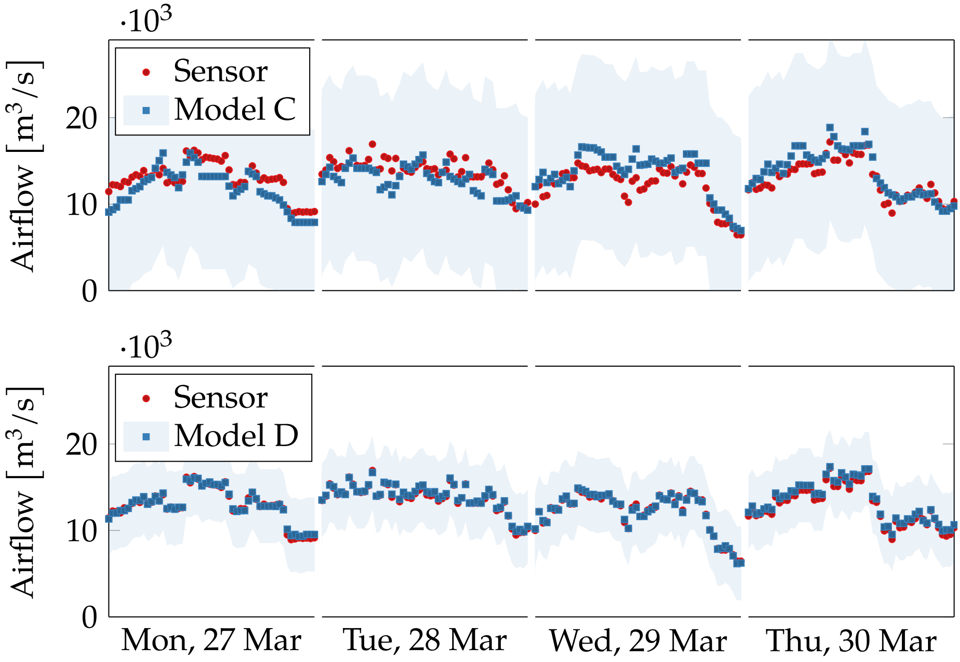

Figure 7.3: Comparison between actual sensors and acceptable ranges obtained from model-based virtual sensors for post heat exchanger airflow during working hours (from 8am to 5pm) for selected days. On Tuesday 28 March 2017 Model C deviates significantly from the actual sensor, but readings always fall inside the acceptable range for the entire period.

We highlighted how during one day the R^2 score between actual and virtual temperature sensors changed abruptly and significantly and actual sensors' readings fell outside the acceptable range, which suggested a fault. However, a proper threshold system must be set up to achieve automatic FDD. This can be achieved by using expert knowledge and a training set of labeled faulty historical data or by generating faulty data using simulations. Moreover, the temperature sensor exhibited faulty behaviour only for a single day during the first week, while it appeared to work correctly in during the second one. Therefore, a threshold system should also be used to decide whether a significant but short-lived deviation is a fault.

Finally, we used regression models to predict data during a period close to the one used for training, under the assumption that the system's behaviour did not change significantly. When extending the prediction to other periods, this assumption might not hold anymore, and seasonal variations must be taken into account.

ACKNOWLEDGEMENT

This work is supported by the Innovation Fund Denmark for the project COORDICY and partly supported by the Spanish MINECO ad FEDER research project SCAV (ref.DPI2017-88403-R).

REFERENCES

[2] **Claudio Giovanni Mattera**, Muhyiddine Jradi and Hamid Reza Shaker. 'Online Energy Simulator for Building Fault Detection and Diagnostics Using Dynamic Energy Performance Model'. In: *International Journal of Low-Carbon Technologies* 13.3 (17th May 2018), pp. 231–239. ISSN: 1748-1325. DOI: 10.1093/ijlct/cty019.

[3] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'Fault Detection and Diagnostics in Ventilation Units Using Linear Regression Virtual Sensors'. In: *International Symposium on Advanced Electrical and Communication Technologies* (*ISAECT*) (Kenitra, Morocco, 21st–23rd Nov. 2018). IEEE. 24th Jan. 2019. DOI: 10.1109/ISAECT.2018.8618755.

[7] Muhyiddine Jradi, Fisayo Caleb Sangogboye, **Claudio Giovanni Mattera**, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'A World Class Energy Efficient University Building by Danish 2020 Standards'. In: *Energy Procedia* 132 (Oct. 2017): *11th Nordic Symposium on Building Physics*, pp. 21–26. ISSN: 1876-6102. DOI: 10.1016/j.egypro.2017.09.625.

[12] European Parliament and Council of the European Union. 'Directive 2010/31/EU of the European Parliament and of the Council on the Energy Performance of Buildings'. In: *Official Journal of the European Union* L153 (18th June 2010), pp. 13–35. EU DIRECTIVE: 2010/31.

[16] *2011 Buildings Energy Data Book*. Tech. rep. U. S. A. Department of Energy, Mar. 2012. URL: https://openei.org/doe-opendata/dataset/buildings-energy-data-book (visited on 20th Dec. 2018).

[36] Evan Mills. 'Building Commissioning: A Golden Opportunity for Reducing Energy Costs and Greenhouse Gas Emissions in the United States'. In: *Energy Efficiency* 4.2 (May 2011), pp. 145–173. ISSN: 1570-6478. DOI: 10.1007/s12053-011-9116-8.

[41] Woohyun Kim and Srinivas Katipamula. 'A Review of Fault Detection and Diagnostics Methods for Building Systems'. In: *Science and Technology for the Built Environment* 24.1 (8th May 2018), pp. 3–21. ISSN: 2374-4731. DOI: 10.1080/23744731.2017.1318008.

[51] Haorong Li, Daihong Yu and James E. Braun. 'A Review of Virtual Sensing Technology and Application in Building Systems'. In: *HVAC&R Research* 17.5 (3rd Oct. 2011), pp. 619–645. DOI: 10.1080/10789669.2011.573051.

[78] Miquel À. Cugueró-Escofet, Diego García, Joseba Quevedo, Vicenç Puig, Santiago Espin and Jaume Roquet. 'A Methodology and a Software Tool for Sensor Data Validation/reconstruction: Application to the Catalonia Regional Water Network'. In: *Control Engineering Practice* 49 (Apr. 2016), pp. 159–172. ISSN: 0967-0661. DOI: 10.1016/j.conengprac.2015.11.005.

[98] Luis Pérez-Lombard, José Ortiz and Christine Pout. 'A Review on Buildings Energy Consumption Information'. In: *Energy and Buildings* 40.3 (2008), pp. 394–398. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2007.03.007.

[99] *Technical Synthesis Report Annex 34. Computer Aided Evaluation of HVAC System Performance*. Tech. rep. Birmingham, United Kingdom: International Energy Agency - Energy Conservation in Buildings and Community Systems Programme, 2006.

[100] Yuebin Yu, Denchai Woradechjumroen and Daihong Yu. 'A Review of Fault Detection and Diagnosis Methodologies on Air-handling Units'. In: *Energy and Buildings* 82 (Oct. 2014), pp. 550–562. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2014.06.042.

[101] Haorong Li and James E. Braun. 'Decoupling Features and Virtual Sensors for Diagnosis of Faults in Vapor Compression Air Conditioners'. In: *International Journal of Refrigeration* 30.3 (May 2007), pp. 546–564. ISSN: 0140-7007. DOI: 10.1016/j.ijrefrig.2006.07.024.

[102] Nunzio Cotrufo and Radu Zmeureanu. 'Virtual Outdoor Air Flow Meter for an Existing HVAC System in Heating Mode'. In: *Automation in Construction* 92 (Aug. 2018), pp. 166–172. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2018.03.036.

[103] Andrew Kusiak, Mingyang Li and Haiyang Zheng. 'Virtual Models of Indoor-air-quality Sensors'. In: *Applied Energy* 87.6 (June 2010), pp. 2087–2094. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2009.12.008.

[104] K. Verbert, R. Babuška and B. De Schutter. 'Combining Knowledge and Historical Data for System-level Fault Diagnosis of HVAC Systems'. In: *Engineering Applications of Artificial Intelligence* 59 (Mar. 2017), pp. 260–273. ISSN: 0952-1976. DOI: 10.1016/j.engappai.2016.12.021.

[105] Olivier Renaud and Maria-Pia Victoria-Feser. 'A Robust Coefficient of Determination for Regression'. In: *Journal of Statistical Planning and Inference* 140.7 (July 2010), pp. 1852–1862. ISSN: 0378-3758. DOI: 10.1016/j.jspi.2010.01.008.

[106] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel,

Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot and Édouard Duchesnay. 'Scikit-learn: Machine Learning in Python'. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830. URL: http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.

[107] *The Basics of Fan Performance Tables, Fan Curves, System Resistance Curves and Fan Laws*. FA/100-99. Greenheck. Schofield, WI, USA, Dec. 1999. 4 pp. URL: http://www.greenheck.com/resources/library/application-articles (visited on 26th Dec. 2018).

CHAPTER 8

IMPROVING VIRTUAL SENSORS WITH NON-LINEAR AND STATISTICAL MODELS

This chapter is a cosmetic adaptation of the following journal paper.

Claudio Giovanni Mattera, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'A Method for Fault Detection and Diagnostics in Ventilation Units Using Virtual Sensors'. In: *Sensors* 18.11 (14th Nov. 2018). ISSN: 1424-8220. DOI: 10.3390/s18113931

This work was the result of a collaboration with the Center for Supervision, Security and Automatic Control (CS^{2}AC), at the Polytechnic University of Catalonia (UPC), Barcelona, Spain.

ABSTRACT

Buildings represent a significant portion of global energy consumption. Ventilation units are complex components, often customized for the specific building, responsible for a large part of energy consumption. Their faults impact buildings' energy efficiency and occupancy comfort. In order to ensure their correct operation, proper fault detection and diagnostics methods must be applied. Hardware redundancy, an effective approach to detect faults, leads to increased costs and space requirements.

We propose exploiting physical relations inside ventilation units to create virtual sensors from other sensors' readings, introducing redundancy in the system. We use two different measures to detect

when a virtual sensor deviates from the physical one: coefficient of determination for linear models, and acceptable range.

We tested our method on a real building at the University of Southern Denmark, developing three virtual sensors: temperature, airflow, and fan speed. We employed linear regression models, statistical models, and non-linear regression models. All models detected an anomalous strong oscillation in the temperature sensors. Readings fell outside the acceptable range and the coefficient of determination dropped.

Our method showed promising results by introducing redundancy in the system, which can benefit several applications, such as fault detection and diagnostics and fault-tolerant control. Future work will be necessary to discover thresholds and set up automatic fault detection and diagnostics.

8.1 INTRODUCTION

In Europe, buildings account for 40 % of the total energy used and 36 % of the total CO_2 emissions [12]. In the United States, the building sector accounted for about 41 % of primary energy consumption in 2010, 44 % more than the transportation sector and 36 % more than the industrial sector. Total building primary energy consumption in 2009 was about 48 % higher than consumption in 1980, going from 1290 TW h to 2784 TW h [16].

Modern buildings consist of different subsystems such as heating, ventilation and air conditioning (HVAC) and lighting. Each subsystem contains, in turn, several components such as pumps, fans, ducts, sensors, lamps, wires etc. monitored and managed by a building management system. All these components are subject to faults, due to damage, wearing over time, misconfiguration, and communication issues. Faults impact occupancy, maintenance cost and particularly energy efficiency. It is estimated that in 2009 just 13 of the most common faults were responsible for over $3.3 billions in energy loss [36, 41].

HVAC load varies depending on building type and location, but they are one of the critical subsystems and can make up to 50 % of the total energy consumption and, therefore, faults involving them cause large energy loss [98, 41]. Research suggests that between 20 % to 30 % energy saving could be achieved by re-commissioning malfunctioning HVAC systems [99]. HVAC systems are often customized for their specific building and, therefore, lack quality system integration [100].

Fault detection and diagnostics (FDD) techniques can be used to monitor building systems and to detect and diagnose anomalies and faults. FDD has been an active research area for many decades in fields such as process operations [13], avionics [108] or water distribution [109, 110], and in the past few years has caught the interest in the field of buildings technology [39, 40, 41].

8.1.1 PROBLEM STATEMENT

Building energy efficiency and safety cannot be achieved without FDD methods applied to ventilation units. Hardware redundancy is an effective approach to high-quality FDD; however, duplicating sensors and other components inside every unit increases deployment and maintenance costs, necessary space, and complexity. Commercial ventilation units are rarely shipped with hardware redundancy.

In this paper, we propose a mixed model-based and data-driven technique to exploit spatial relations among different components in ventilation units to create *virtual sensors* and introduce redundancy in the system, which can be used to detect and diagnose faults. For each considered sensor, we train a model to estimate its readings given other sensors in the unit. This allows us to detect and diagnose faults that cause physical and virtual sensors to deviate from each other. In addition to linear regression models, covered in previous work [3], in this paper we consider also autoregressive moving average with exogenous variables (ARMAX) models from statistical analysis and non-linear models such as support vector machine (SVM) regression and artificial neural network (ANN). We define two measures to detect when physical and virtual sensors deviate. We apply this technique to a real-world building and report the results.

The rest of the paper is organized as follows. The state of the art

FDD methods for buildings

Data-driven	Model-based	Rule-based
A model of the system is built from historical data and used to predict/validate data from the system itself. Common techniques: – Machine learning – Artificial neural networks – Support vector machines	A model of the system is built from first principles and used to predict/validate data from the system itself. Common techniques: – Parity equations – Observers – Kalman filters	A set of rules are defined from expert knowledge and used to determine whether the system is working as expected. Common techniques: – Expert systems – Pattern classifications – Limits and alarms

Figure 8.1: Categorization of FDD methods for buildings adapted from Kim et al. [41].

is reviewed in Section 8.2. The proposed technique is introduced in Section 8.3. Section 8.4 presents the case study and discusses results and implications. Finally, conclusions are drawn in Section 8.5.

8.2 STATE OF THE ART

8.2.1 FAULT DETECTION AND DIAGNOSTICS

Kim et al. present a comprehensive review of recent FDD methods for building systems [41]. FDD methods are categorized into three groups depending on the approach: data-driven methods, model-based methods, and rule-based methods, as shown in Figure 8.1.

In data-driven methods, a model of the system under test is trained from historical data and it is used to validate current data from the system. Several techniques exist, such as machine learning, artificial neural network (ANN) or support vector machine (SVM). Little to no physical knowledge of the system is required and the resulting models can be treated as black-box components. For this reason, data-driven methods are easily applicable to several types of systems with small effort. However, historical data are necessary to create the model, which rules out the possibility to apply these techniques to newly deployed systems. These methods require fault-free training data, otherwise the generated models would recognize faults as correct behavior. To perform proper diagnostics and identify the precise fault, labelled faulty historical data is usually necessary.

In model-based methods, a physical model of the system under test is created from first principles and it is used to validate current data from the system. This approach does not require training data and often predictions are more accurate than black-box models. However, accurate models can be complex and require in-depth knowledge of the system and large effort to be created. Often it is necessary to perform parameter estimation to improve accuracy, which might require historical data and, therefore, prevent to use the model with newly deployed systems.

In rule-based methods, expert knowledge gathered from field experts is used to design a set of rules describing the system's behavior. No historical data and no detailed physical knowledge of the system are necessary. Moreover, some faults have effects that can be described by rules, which makes it possible to precisely identify and diagnose the problem. However, rules can only describe behaviors up to a certain complexity and they can only cover simple cases. As the number of rules grows, the possibility of conflicting rules increases and so does the effort to maintain the set of rules.

Yu et al. present a review of FDD techniques for ventilation units [100]. In this case, the authors classify FDD techniques into four groups: hardware redundancy, software redundancy, signal analysis and plausibility tests, as shown in Figure 8.2. Multiple identical sensors and actuators lead to hardware redundancy, which allows high accuracy and precision, but also to higher deployment and maintenance costs. In software redundancy, multiple physical sensors are replaced by models obtained by other sensors in the system. In signal analysis and plausibility tests methods, the steady-state characteristics and other physical laws in the system are investigated. Software redundancy methods are further classified in model-based, data-driven, and rule-based, as in general FDD methods.

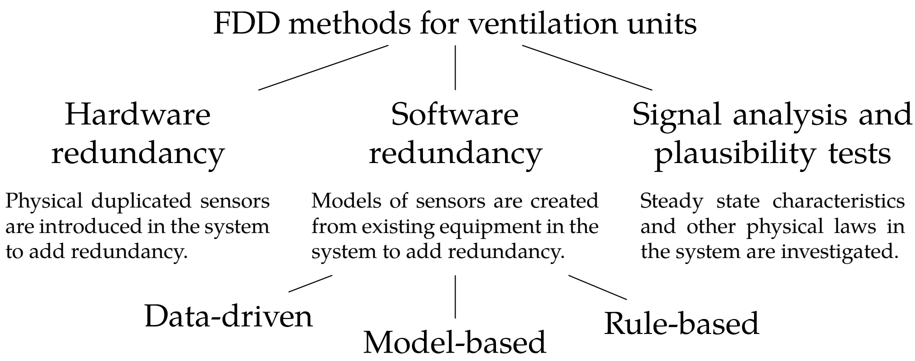

Figure 8.2: Classification of FDD techniques for ventilation units according to Yu et al. [100].

The authors also define a list of desirable characteristics of FDD methods:

1. Quick detection and diagnostics: faults should be identified as soon as possible;
2. Isolability: the ability to distinguish between multiple faults, i.e., performing diagnostics;
3. Robustness: the method should be insensitive to noise and model uncertainties;
4. Novel identifiability: the ability to detect unknown faults;
5. Classification error estimate: the method should make its accuracy explicit, e.g., by having a confidence range as output;
6. Adaptability: the ability to automatically adapt to changes in the system under test;
7. Explanation facility: the ability to identify the precise location and cause of faults;
8. Modeling requirements: lower modeling requirements ease implementation and application on real-time processes;
9. Storage and computational requirements: minimal storage and computational requirements are necessary for an easy implementation and application on real-time processes;
10. Multiple fault identifiability: the ability to diagnose multiple simultaneous faults.

8.2.2 VIRTUAL SENSORS

Virtual sensors have been used successfully in various fields, both for observing hidden unmeasured quantities in the system and for validating the system's status. An example of the former can be found in [111], where the authors study spark-ignition engines in avionics. They develop virtual sensors for quantities for which a physical sensor would have been too expensive to deploy, or too slow at collecting data. They use artificial neural networks to predict measurements from other sensors' readings. Other authors use virtual sensors to estimate tire forces in automotive systems [112]. They use Kalman filter, ANN and physical relations between measurable quantities in the system such as wheel speed.

An example of virtual sensors used for data validation can be found in [78], where the authors present an approach for sensors data validation and reconstruction and apply it to urban water distribution systems. Raw data undergoes several tests, from low-level tests checking elementary properties of signals to high-level tests exploiting *spatial consistency* between different sensors.

In complex systems, it is not trivial to design effective virtual sensors, due to the large combination of available inputs but also to the diversity of modeling techniques. While a popular approach is to use general purpose simulation software, there is research effort to produce software tools able to create and parametrize modular virtual sensors [113].

Li et al. present a review of virtual sensing techniques in the context of building systems [51]. Virtual sensors have been successfully applied to fields such as process control and the automotive sector for more than two decades, and building systems could benefit from their application. e.g., many of the FDD techniques proposed for buildings cannot be applied in practice due to sensors not available in real buildings or not accurate enough. Virtual sensors can be used to overcome these difficulties and generate high-quality measurements.

Virtual sensing techniques are categorized according to three different criteria as shown in Figure 8.3: measurement characteristics-based, modeling methods-based and application purpose-based. In the meas-

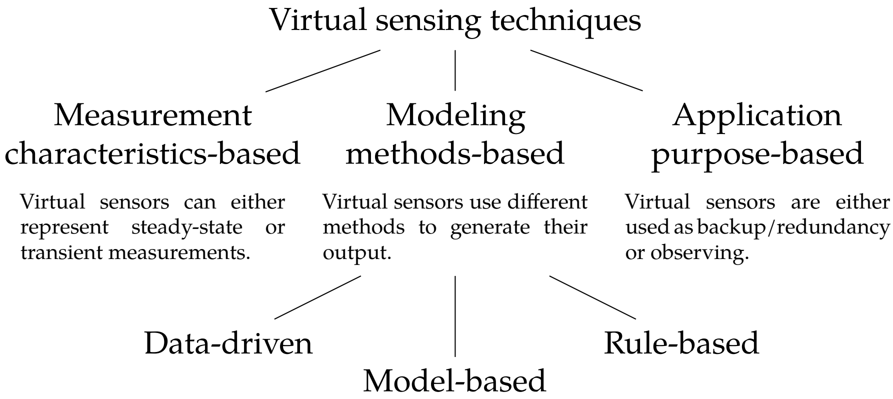

Figure 8.3: Categorization of virtual sensing techniques according to Li et al. [51].

urement characteristics category, virtual sensors can either represent steady-state or transient measurements. In the former case, the model is based on the assumption that the system responds instantaneously to input variables, or that the measured quantities change slowly compared to the system's dynamics. In the latter case, slower reactions and faster variating input variables are taken into account.

In modeling method category virtual sensors techniques can be divided into model-based and data-driven, similarly to FDD methods. In model-based techniques, detailed knowledge about the system such as mathematical relations between sensors is used to create a model of the sensor. In data-driven techniques, historical data is used to train a black-box model of the system. Methods that are based both on physical models and data trained models are called gray-box models.

With respect to application purposes, virtual sensors are either used as backup/redundancy or observing. In the former case, virtual sensors measure quantities for which other physical sensors exist. They can be used to validate such physical sensors' readings together with FDD methods or to replace them if they fail. In the latter case, virtual sensors measure quantities unknown or even non-measurable in the system, such as performance or efficiency, and make them available to client applications.

On ventilation units specifically, virtual sensors have been used both to measure unknown quantities and to perform FDD. An example of system monitoring can be found in [102], where the authors develop a virtual sensor modeling exhaust airflow. Airflow sensors for exhaust duct are rarely present in ventilation units due to their cost. They use energy balance equation to relate other sensors in the system with the airflow and propose two different models. While the local errors can be large, the authors show how the cumulative residuals are small and, therefore, the virtual sensor can be used to estimate daily averages.

In [104] the authors report how using virtual sensors significantly improves FDD performance for HVAC systems. They propose a multi-model FDD method that exploits components interdependencies. They develop Bayesian networks for multiple operating modes, using both physical and virtual sensors created from system knowledge and historical data.

Other buildings subsystems have been considered for FDD using virtual sensors. The method proposed in [101] is applied to air conditioners using features decoupling and virtual sensors. The authors create virtual sensors for several quantities, such as compressor power consumption, refrigerant flow, condenser exit pressure, exit air humidity and evaporation temperature. Virtual sensor performances are tested both at steady state and under transients.

A method for FDD on air conditioners is proposed in [69]. The author develops three different virtual sensors for virtual refrigerant charge sensors using different techniques. Information from laboratory tests and manufacturers' data was used to assess the impact of faults on system performance. A complete implementation was provided for a rooftop air-conditioning unit.

While not part of ventilation units themselves, room-level sensors, i.e., temperature, CO_2 level and relative humidity, are essential to their correct operation. In [103] a data-driven model for virtual sensors for room-level indoor air conditions is proposed. The authors develop four data mining techniques, including artificial neural network, support vector machine regression and Pace regression. The obtained virtual sensors can be used for validation and calibration of physical sensors.

The reviewed state of the art shows that virtual sensors are popular in the field of building systems; however, to our knowledge there is no work so far on employing data-driven virtual sensors for fault detection and diagnostics application on ventilation units. Most of the work reviewed covers other buildings subsystems, such as chillers and air-conditioning units [101, 51, 69], boilers [104], heat pumps [51] and room-level components [103]. Moreover, in ventilation units, virtual sensors are usually developed to provide readings for unmeasured quantities [102], and when they are considered for explicit application for fault detection and diagnostics they are designed using first principles methods [51]. Other approaches focus on a higher level of diagnostics and require significant expert knowledge to define fault and symptoms [104]. *Therefore, the main contribution of this paper is a specific fault detection and diagnostics application for ventilation units based on virtual sensors created using a data-driven approach.*

8.3 MATERIAL AND METHODS

In this section, we describe the proposed method for FDD on ventilation units based on virtual sensors.

A ventilation unit is an aggregate of several components, integrated together to provide air exchange for the building. It is important that every component works correctly, otherwise the performance of the unit will deteriorate, causing energy loss and reducing comfort level in the building.

Since all components work together, they exhibit common patterns and shared phenomena. Even if there is no explicit redundancy in the system, i.e., no duplicated sensor or meter, many of the quantities in the unit are strongly correlated. In this paper, we propose to exploit these relations and create models to predict a quantity from the surrounding ones, generating a set of *virtual sensors*. Given physical sensors available in the ventilation unit $S_1, S_2, \dots, S_n$, a virtual sensor S'_i measuring the same quantity of S_i is created using a model $f(\cdot)$ that takes other sensors as input, i.e.,

$$\begin{aligned} S_i' &= f(\overline{\mathcal{S}}) \\ \overline{\mathcal{S}} &\subsetneq \{S_1, S_2, \ldots, S_{i-1}, S_{i+1}, \ldots, S_n\}. \end{aligned} \tag{8.1}$$

For instance, consider a heating system where the following quantities are measured with sensors or meters: initial temperature T_0, heater energy M and final temperature T_f. A virtual sensor for final temperature could be created using a model of initial temperature and heater energy $T_f' = f(M, T_0)$. In principle, virtual sensors can be created for any measure inside the system under test, it is not a requirement that a real sensor exists.

Different methods can be used to compute the value of a virtual sensor. When detailed knowledge about the unit is available it is possible to use physical models, e.g., computing airflow using fan speed and duct size and shape. Otherwise, it is possible to train black-box models using data-driven techniques such as regression models, artificial neural network or support vector machine.

A ventilation unit contains several sensors necessary to its functions, such as temperature sensors at various locations, airflow and fan speed at each fan and pump, and energy meters for different components. However, not all of them are closely related to each other and, therefore, it is important to carefully design each virtual sensor by choosing quantities that are correlated. e.g., as shown in Figure 8.4, fan speed and airflow through the same fan are obviously highly correlated, while inlet air temperature and extract air temperature are independent on each other.

8.3.1 FAULT DIAGNOSTICS

When two correlated sensors, either physical or virtual, deviate, the only possible inference is that a fault is affecting one of them. To diagnose the faulty one, a third sensor is necessary. Under the assumption of single simultaneous fault, when in a group of three sensors one deviates from the other twos, the former is identified as faulty.

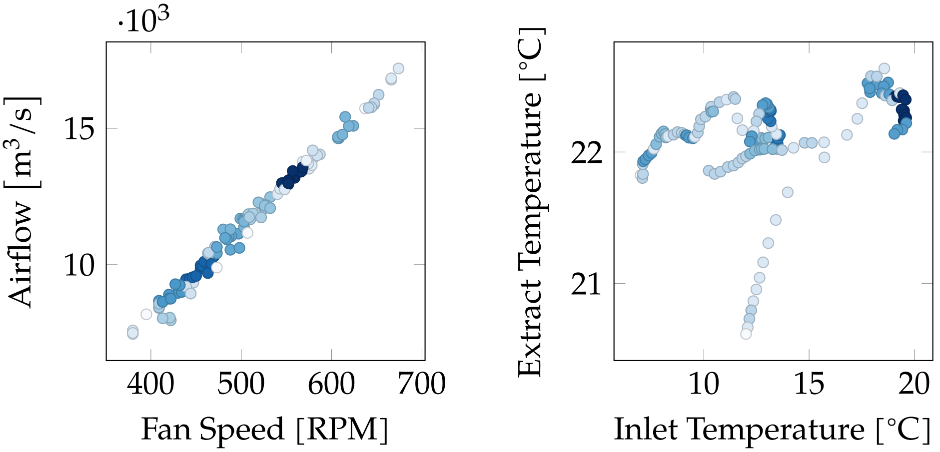

Figure 8.4: Density plots showing correlations between fan speed and airflow, and between inlet and extract temperatures. Darker colors correspond to more frequent readings. The quantities on the left plot are highly correlated, while the ones on the right one are essentially independent on each other.

Due to cost and space constraints, duplicated sensors are rarely available in ventilation units, and even less so are triplicated sensors. However, these constraints do not impact virtual sensors, which can be created without cost using data from other components. Some care is necessary when choosing the inputs: different virtual sensors should share as few inputs as possible because a fault in an input impacts all its related virtual sensors.

For instance, consider a heating system with two initial temperature sensors T_0, T_1, a heater energy meter M and a final temperature sensor T_f, where two additional virtual sensors for final temperature were created as

$$T_f' = f(M, T_0), \qquad T_f'' = f(M, T_1). \tag{8.2}$$

Assuming a single fault scenario, if T_f' and T_f'' agree on their readings and T_f deviates from them there are two possible causes:

- sensor T_f is faulty;
- heater energy meter M is faulty.

This is because heater energy meter M is used as input in both virtual sensors T'_f and T''_f, therefore, its fault impacts both their output.

8.3.2 Measuring deviations from physical sensors

To automatically detect a fault, a measure of how much the virtual sensors deviate from the physical one is necessary. Several tools are available from statistical analysis, e.g., the maximal error or the norm of residuals. For the first part of the case study, where we use linear regression models to create virtual sensors, we use the coefficient of determination, or R^2 score, which gives an estimate of how much a linear regression model fits the data [105]. Given a signal $y_i, i \in [1, n]$ with mean $\overline{y}$ and its predictions $\hat{y}_i$ the R^2 score is defined as

$$\begin{aligned} R^2 &= 1 - \frac{\text{Sum of squares}_{\text{residual}}}{\text{Sum of squares}_{\text{total}}} \\ \text{Sum of squares}_{\text{residual}} &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad (8.3) \\ \text{Sum of squares}_{\text{total}} &= \sum_{i=1}^{n} (y_i - \overline{y})^2. \end{aligned}$$

An R^2 score close to 1 indicates that the model is a good fit for the data, while values close to zero indicates the opposite. Negative values indicate that the model predicts data worse than a constant horizontal line.

We use the R^2 score both to verify that the trained models fit the testing data, i.e., that the designed model accurately follows the physical sensor, and to validate real-time data from the ventilation unit. For each period of interest, e.g., every day, the R^2 score for each virtual sensor against the physical sensor is recorded. When the measure is lower than a given threshold the pair virtual/physical sensors, are flagged as anomalous or faulty.

R^2 score is only meaningful for linear regression models and does not yield useful value for non-linear ones. An alternative option for detecting deviations from the physical sensor is to make the virtual

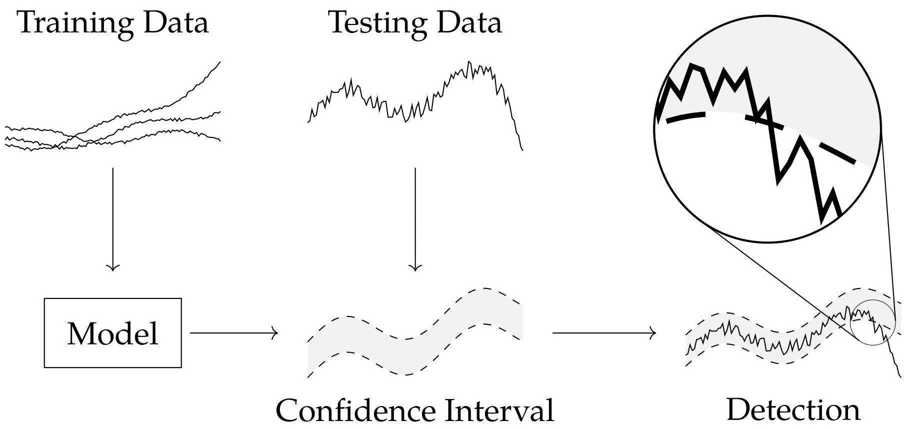

Figure 8.5: Virtual sensors can generate an expected confidence interval. When readings from the physical sensor fall outside such interval the sensors pair is flagged as anomalous or faulty.

sensor generate an *acceptable range* of values. e.g., the acceptable error could be as large as the largest error obtained when predicting the original training data, or a confidence interval could be built from training data. When readings from the physical sensor fall outside the acceptable range the sensors pair is flagged as anomalous or faulty. This approach is illustrated in Figure 8.5.

With both approaches, labelled faulty testing data would be necessary to obtain accurate thresholds.

8.4 RESULTS AND DISCUSSION

In this section, we implement the method presented in Section 8.3 on a ventilation unit of an existing building. We detail the ventilation unit structure and its sensors and components (Figures 8.6 and 8.7). Afterwards, we design three virtual sensors based on linear regression models to duplicate the readings of physical sensors, and we compare physical and virtual readings to detect anomalous behaviors. Finally, we design additional virtual sensors based on statistical and non-linear regression models.

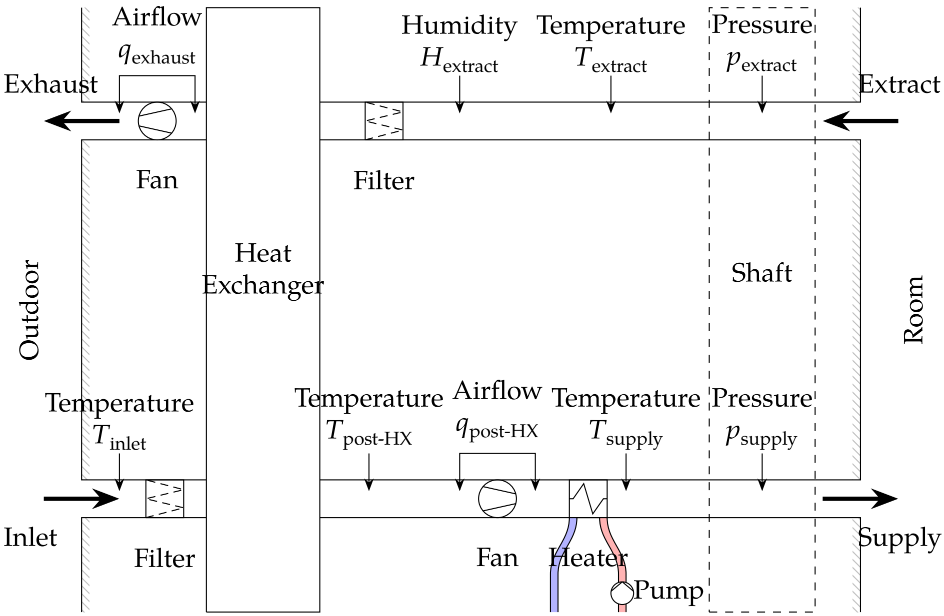

Figure 8.6: Diagram of a ventilation unit in building OU44. Inlet air enters the unit from bottom-left, passes through the heat exchanger and through the heater, before entering the main shaft and supplying individual rooms. From the rooms it enters again the main shaft, goes through the heat exchanger to heat up inlet air, and finally is pushed outside the building. Several sensors, shown by arrows, are available in the unit.

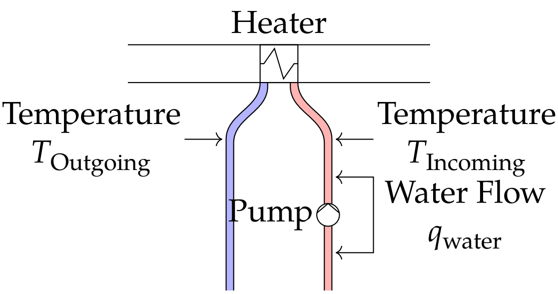

Figure 8.7: Diagram of a heating loop in building OU44. Hot water is used to heat up the air before it enters the main shaft. Several sensors, shown by arrows, are available in the loop.

8.4.1 BUILDING OU44

In this paper, we present building Odense undervisning 44 (OU44) as a case study [7]. It was built in 2015 at the University of Southern Denmark, campus Odense, and it is mainly used for teaching. It has three floors plus a basement and it contains classrooms, study zones, offices, and auditoriums. It has four nearly identical ventilation units, each serving one corner of the building, or roughly 20 thermal zones.

A ventilation unit consists of a large air loop, as shown in Figure 8.6. Inlet air enters the building, goes through a heat exchanger (HX), then is heated to an appropriate indoor temperature and pushed to the supply shaft, which is connected by variable air volume (VAV) units to individual rooms. In the same way, exhaust air is collected from individual rooms in the extract shaft, it goes through the heat exchanger and it is pushed outside. The heat exchanger recovers heat from exhaust air and transfers it to inlet air, reducing the energy required by the heater. Air pressures in supply and extract shafts are kept at constant values 130 Pa and 40 Pa, which cause air to flow in the rooms. Two fans in the ventilation unit generate the required airflows to maintain the pressure setpoints.

Heaters, shown in Figure 8.7, use a hot-water loop, provided by a district-heating system, to heat air inside the ventilation unit.

Several sensors, shown as arrows in Figures 8.6 and 8.7 are available inside ventilation units and heating loops: air temperature at several positions, airflows through the two fans, supply and extract pressure, incoming and outgoing water temperature, and water flow through the pump. In addition to that, several meters measure the activity of fans and water pump: fan speed $\omega_{\text{exhaust/post-HX}}$, fan current $i_{\text{exhaust/post-HX}}$ and voltage $V_{\text{exhaust/post-HX}}$, fan power and electrical consumption, and pump electrical consumption.

Ventilation units only function during working hours, i.e., from Monday to Friday from 7am to 6pm in local time. At night and during weekends they are shut down.

8.4.2 Results using linear regression models

Three sensors were considered for monitoring in a ventilation unit: post-heat exchanger temperature, airflow, and fan speed. For each of them, two different models were constructed using other sensors as inputs, as shown in Table 8.1. Linear regression models were used under the assumption that inputs and outputs obey linear relations, at least locally [106]. Since the periodicity of the system's behavior is one week, models were trained over a week-long historical data from Monday 13 March 2017 to Sunday 19 March 2017 and tested over two weeks from Monday 27 March 2017 to Sunday 9 April 2017. This period was one of the longest ones with continuously available data for every sensor in each ventilation unit. Training and testing periods were within the same month, therefore, no significant seasonal variation that could influence the models was expected. Additional care should be taken when this assumption does not hold, e.g., in this particular case a teaching building could be configured to operate differently during summer vacations.

For both training and testing phases, raw data from the building management system was resampled to a common, fixed period of 10 min. This step was necessary because the various sensors inside the ventilation unit do not report at the same exact time. Regression models, on the other hand, require readings from different time-series to be simultaneous. No other preprocessing operations were performed. In particular, no faults were artificially added to data.

Another virtual sensor was also constructed, i.e., *Effort* (*eff*), which is proportional to an estimate of the power requested to the ventilation unit and, therefore, to the airflow. By design, fans produce airflow to maintain constant shaft pressure, which in turn depends on how many VAV units are open in the building. When a VAV unit is open it makes air flowing from the supply shaft through the room to the extract shaft, which results in pressure loss. Fans will then increase their speed to make up for such loss. Effort is an aggregate count of those units, which makes it effectively a virtual sensor for an unknown

Table 8.1: Virtual sensors definitions for linear regression models.

Model Name	Output	Inputs
Model A	$T_{\text{post-HX}}$	T_{inlet}, T_{extract}, $q_{\text{post-HX}}$
Model B	$T_{\text{post-HX}}$	T_{inlet}, q_{water}, T_{incoming}, T_{outgoing}
Model C	$q_{\text{post-HX}}$	*eff* Equation (8.4)
Model D	$q_{\text{post-HX}}$	$\omega_{\text{post-HX}}$
Model E	$\omega_{\text{post-HX}}$	$q_{\text{post-HX}}$
Model F	$\omega_{\text{post-HX}}$	$i_{\text{post-HX}}$, $V_{\text{post-HX}}$

quantity in the ventilation unit, and is defined as

$$\begin{aligned} \mathit{eff} &= \sum_{i \in \text{VAV units}} \tau_i \\ \tau_i &= \text{openness ratio of VAV unit } i \\ q &\propto \mathit{eff}\ \Delta p. \end{aligned} \tag{8.4}$$

Table 8.2 shows the coefficients obtained for models with multiple input variables. Most variables have coefficients significantly larger than their standard deviation, therefore, they are significant in their relative models. Two exceptions are water flow and incoming water temperature in Model B, whose contributions are smaller.

The three charts in Figure 8.8 show results for $T_{\text{post-HX}}$, $q_{\text{post-HX}}$ and $\omega_{\text{post-HX}}$ virtual sensors. Data obtained from physical sensors are plotted against data obtained from the two corresponding linear regression virtual sensors defined in Table 8.1. Deviation from a single virtual sensor is enough to detect a fault but not to isolate and identify the faulty source, therefore, two virtual sensors were used for each physical one. R^2 scores between physical and predicted readings, which measure how much physical and virtual sensors agree, were computed over daily data as defined in Equation (8.3) and are shown in Table 8.3. Low R^2 scores, indicating that models deviate from the physical sensors, are highlighted in boldface.

For temperature two models are used, one (Model A) exploiting knowledge about the heat exchanger interactions, using inlet temper-

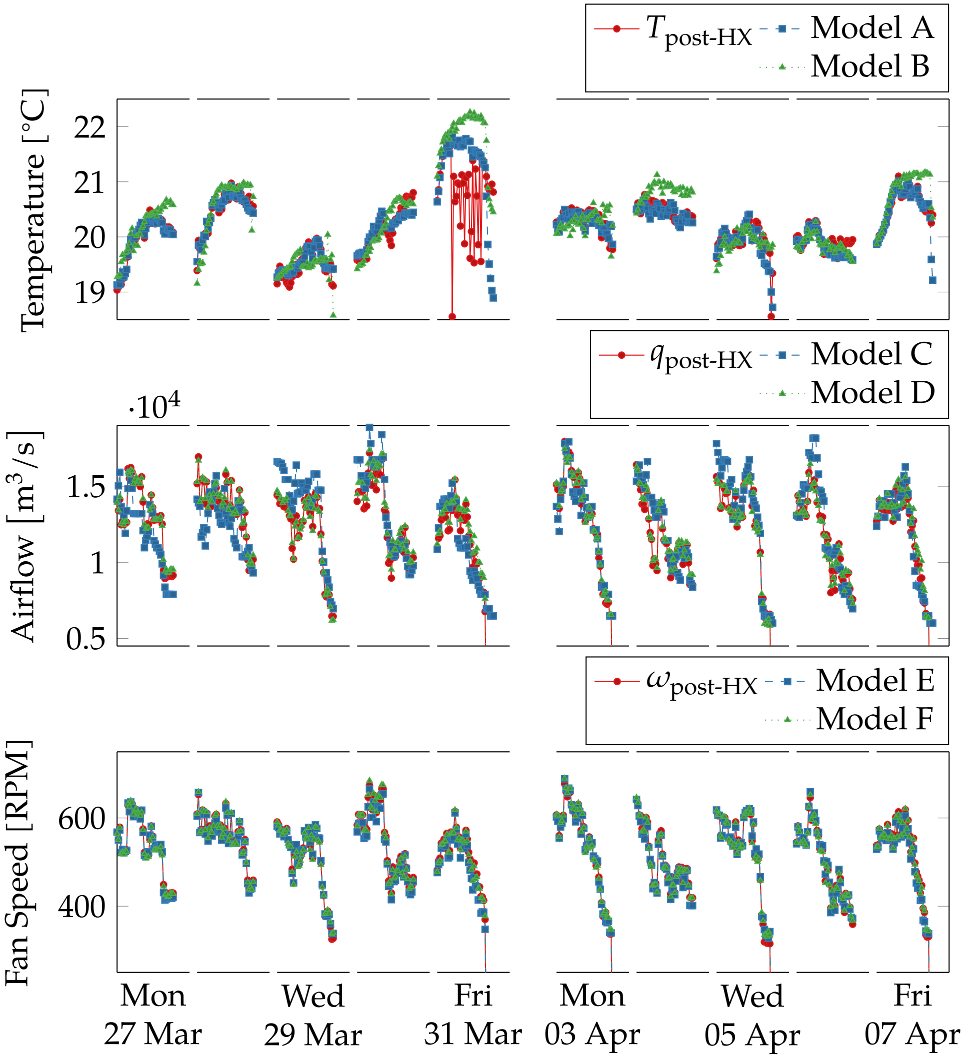

Figure 8.8: Comparison between physical sensors and linear regression model-based virtual sensors for post heat exchanger temperature, airflow, and fan speed during working hours (from 8am to 5pm) over two weeks. Outside working hours and during weekends the ventilation system is shut down. The virtual sensors follow the physical ones except in two cases. On Friday in the first week the temperature sensor oscillates strongly and deviates from the two virtual sensors. On Tuesday in the second week the virtual sensors Model B consistently overestimates the sensors readings.

Table 8.2: Coefficients for linear regression models.

Variable	Coefficient
Model A ($T_{\text{post-HX}}$)	
T_{inlet}	0.49 ± 0.012
T_{extract}	0.23 ± 0.017
$q_{\text{post-HX}}$	$9.86 \times 10^{-2} \pm 1.404 \times 10^{-2}$
Model B ($T_{\text{post-HX}}$)	
T_{inlet}	0.68 ± 0.021
T_{incoming}	-0.05 ± 0.027
T_{outgoing}	-0.16 ± 0.014
q_{water}	0.03 ± 0.026
Model C ($q_{\text{post-HX}}$)	
eff	2375 ± 90.2
Model D ($q_{\text{post-HX}}$)	
$\omega_{\text{post-HX}}$	2766 ± 25.0
Model E ($\omega_{\text{post-HX}}$)	
$q_{\text{post-HX}}$	85.2 ± 0.770
Model F ($\omega_{\text{post-HX}}$)	
$i_{\text{post-HX}}$	14.84 ± 1.308
$V_{\text{post-HX}}$	71.58 ± 1.308

ature, extract temperature and airflow, i.e.,

$$\begin{aligned}\text{Heat} &= c\,(T_{\text{post-HX}} - T_{\text{Inlet}})(\rho\ \Delta t\ \text{q}_{\text{post-HX}}) \\ &= c\,(T_{\text{Exhaust}} - T_{\text{Extract}})(\rho\ \Delta t\ \text{q}_{\text{Exhaust}}),\end{aligned} \tag{8.5}$$

where c, ρ and Δt are respectively air specific heat, air density and time step, and other symbols indicate quantities measured by sensors as shown in Figure 8.6. The other one (Model B) relies on similar but less structured relations between inlet temperature, water flow and temperature difference in the heater. The former predicts temperature value much more accurately than the latter. Table 8.3 shows that both

Table 8.3: Prediction R^2 score for virtual sensors. Low scores are highlighted in boldface.

	$T_{\text{post-HX}}$		$q_{\text{post-HX}}$		$\omega_{\text{post-HX}}$	
Date	**Model A**	**Model B**	**Model C**	**Model D**	**Model E**	**Model F**
2017-03-27	0.955	0.782	0.371	0.987	0.988	0.997
2017-03-28	0.989	0.804	**0.04**	0.98	0.977	0.997
2017-03-29	0.839	0.217	0.368	0.992	0.992	0.995
2017-03-30	0.894	0.729	0.681	0.956	0.956	0.996
2017-03-31	**−1.162**	**−1.995**	0.572	0.852	0.908	0.996
2017-04-03	0.86	0.442	0.87	0.967	0.968	0.997
2017-04-04	0.886	**−0.474**	0.644	0.983	0.984	0.997
2017-04-05	0.774	0.57	0.8	0.944	0.953	0.996
2017-04-06	0.73	0.654	0.622	0.988	0.989	0.997
2017-04-07	0.802	0.537	0.772	0.904	0.932	0.996

models deviate significantly from the physical sensor on 31 March 2017, and Model B deviates also on 4 April 2017. Readings from the physical sensors are shown in Figure 8.9 with respect to the two models' error ranges, which corresponds to the predictions plus the maximal training error.

On 31 March 2017, the physical sensor's readings oscillate strongly, in contrast with the two virtual sensors which have a smoother behavior and fall outside the models' error ranges. Since the two models share an input variable, i.e., inlet temperature, this situation could be caused by a fault in the physical post heat exchanger temperature sensor or in the inlet temperature sensor. Figure 8.10 shows the readings for all involved sensors over the faulty period. All measures except post heat exchanger temperature have smooth trends and behave similarly to the previous day. Inlet temperature rises more than the first day, but it is consistent with outdoor temperature measurements from the local weather station. This suggests that post heat exchanger temperature is indeed the faulty sensor. The anomalous behavior only lasts for a single day; therefore, this event cannot be classified as a sensor failure, and it could be due to an external dis-

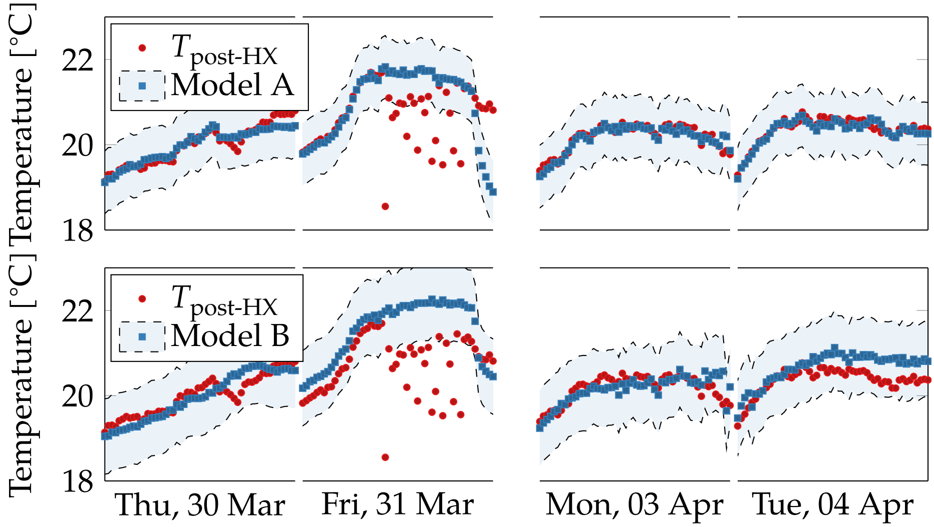

Figure 8.9: Comparison between physical sensors and acceptable ranges obtained from linear regression model-based virtual sensors for post heat exchanger temperature during working hours (from 8am to 5pm) for selected days. The sensors readings fall inside the acceptable ranges except on 31 March 2017, when they deviate significantly. The anomalous trend is not present neither in previous or following days. On 4 April 2017, most models consistently overestimate the physical sensors, but their trends are similar.

turbance. A further on-site investigation would be necessary to finally identify the precise nature of this event.

The situation on 4 April 2017 is less extreme. Model B consistently overestimate the physical sensor's readings, but the overall trend is similar and, moreover, all the readings fall inside the model's error range. Therefore, this event could be classified as a false alarm. Using a more accurate model instead of Model B could reduce the frequency of false alarms.

For airflow two models are used, one using only effort as input (Model C) and one using only fan speed as input (Model D). Airflow and fan speed follow the fan laws and are proportional to each other [107], which could also be inferred from Figure 8.4, and as

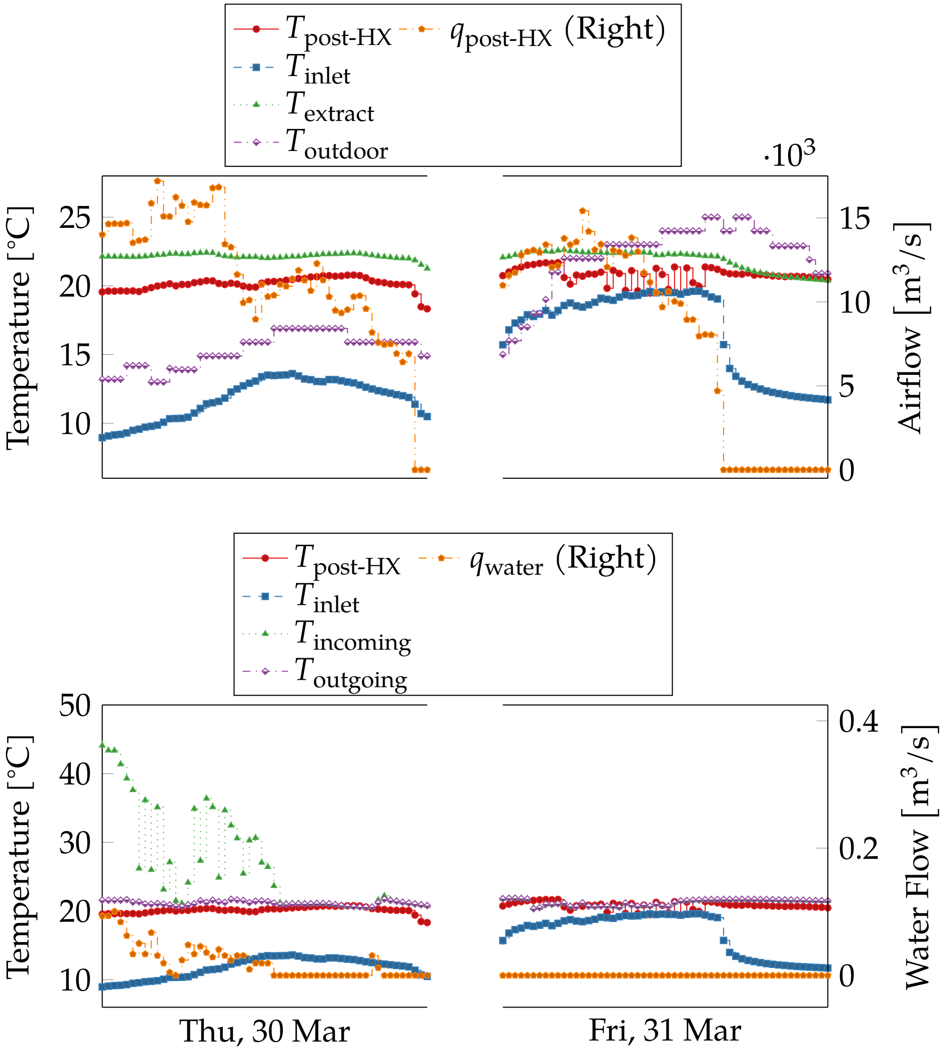

Figure 8.10: Trends of input and output variables in models A and B on Thursday 30 March 2017 and Friday 31 March 2017. Input variables have similar trends over the two days, but the output variable, post heat exchanger temperature, exhibits fast oscillation during the second day. Inlet temperature, the shared input variable between the two models, behave similarly over the two days, following the outdoor temperature measured at the local weather station. During the second day the hot water flow is zero, and incoming temperature is equal to outgoing temperature.

expected predictions for this model are nearly exact.

Model C is less accurate, and its R^2 score on Tuesday 28 March 2017 is very low, which suggests a fault in the virtual sensor's input, i.e., ventilation effort, since Model D agrees with the physical sensor on the same day. Ventilation effort is produced by aggregating several independent streams with frequent periods of missing data, which can indeed cause the model to deviate from the physical sensor. Moreover, ventilation effort does not take into account the size of each room and the corresponding VAV dampers, which reduces the model's accuracy. Readings from the physical sensors are shown in Figure 8.11 with respect to the two models' error ranges, which corresponds to the predictions plus the maximal training error.

For fan speed two models are used, one using airflow as input (Model E) and one using fan current and voltage as inputs (Model F). Fan speed is proportional to airflow due to fan laws, and also proportional to the fan power consumption, which in turn depends on current and voltage $W = VI$. The former model is nearly exact, for the same reasons explained when discussing Model C. The latter model estimates the power used by the fan, which in turn is correlated with the fan speed, and produces accurate results as well.

8.4.3 Results using other models

While linear regression models were able to detect unusual behavior of post-heat exchanger temperature sensor, in some cases they did not accurately predict the values of physical sensors. Four additional models were created, as shown in Table 8.4: two using ARMAX method from statistical analysis [114], and two using non-linear regression methods support vector machine (SVM) [115] and artificial neural network (ANN) [116]. The two approaches augmented linear regression models along two different directions: ARMAX models are linear models over exogenous variables, but they take the endogenous variable's recent trend into account; ANN and SVM models can instead perform non-linear regression by projecting input data to higher dimensional spaces through non-linear transformations and then performing lin-

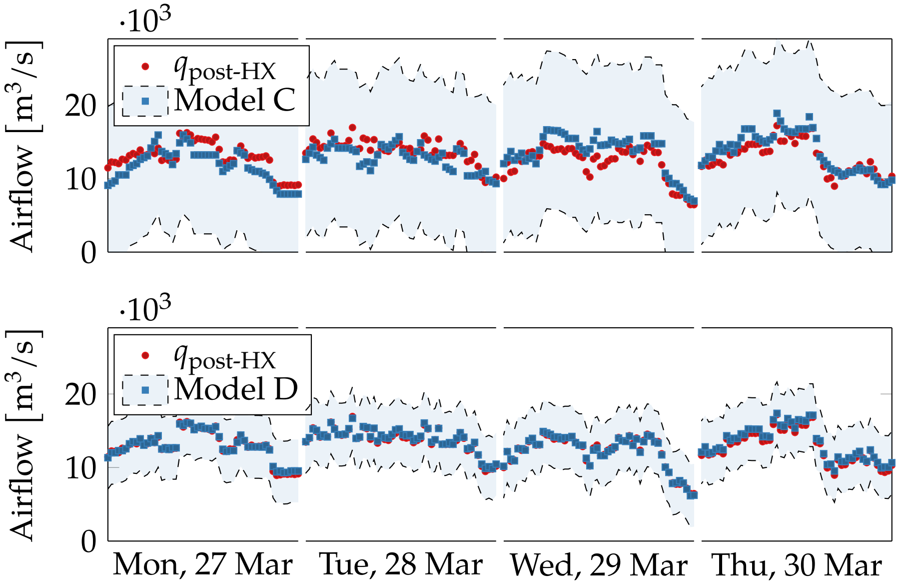

Figure 8.11: Comparison between physical sensors and acceptable ranges obtained from model-based virtual sensors for post heat exchanger airflow during working hours (from 8am to 5pm) for selected days. On Tuesday 28 March 2017 Model C deviates significantly from the physical sensor, but readings always fall inside the acceptable range for the entire period.

ear regression. ANN and SVM have both been successfully used in FDD [41, 39, 40].

ARMAX models

Models ARMAX A and ARMAX B were trained using post-heat exchanger temperature as endogenous variable and input sensors from respectively models A and B as exogenous variables. Models SVM and ANN were trained using the same inputs as Model B. As for linear regression models, they were trained over a week-long historical data from Monday 13 March 2017 to Sunday 19 April 2017 and tested over two weeks from Monday 27 March 2017 to Sunday 9 April 2017. As for the experiment with linear regression models, raw data was

Table 8.4: Virtual sensors definitions for other models.

Model Name	Output	Inputs
SVM	$T_{\text{post-HX}}$	$T_{\text{inlet}}, q_{\text{water}}, T_{\text{incoming}}, T_{\text{outgoing}}$
ANN	$T_{\text{post-HX}}$	$T_{\text{inlet}}, q_{\text{water}}, T_{\text{incoming}}, T_{\text{outgoing}}$
ARMAX A	$T_{\text{post-HX}}$	$T_{\text{post-HX}}, T_{\text{inlet}}, T_{\text{extract}}, q_{\text{post-HX}}$
ARMAX B	$T_{\text{post-HX}}$	$T_{\text{post-HX}}, T_{\text{inlet}}, q_{\text{water}}, T_{\text{incoming}}, T_{\text{outgoing}}$

resampled to a common, fixed period of 10 min.

In ARMAX models data belonging to nights and weekends were removed, i.e., the dataset consisted of continuous working hours. Working and non-working hours correspond to significantly different operation profiles, and since ARMAX methods predict future values based on recent history, they would not perform well when predicting across both. Two different model should instead be created, one for each profile. Since the ventilation system is turned off during non-working hours, in this paper we ignored this case, but in more complex situations where working hours are not fixed, e.g., they depend on the weekday, it would be necessary to split the dataset into distinct parts corresponding to each profile.

Data sampling period was 10 min, model orders were set to $(p, q, d) = (20, 2, 0)$ and prediction horizon was set to one working day, i.e., 10 h. Virtual sensors readings are shown against physical sensors readings in Figure 8.12. The virtual sensors follow closely the physical sensor, except on Friday 31 March 2017 and on Monday 3 April 2017. During the former day, the physical sensor strongly oscillates while the virtual sensors predict a regular trend, in agreement the linear regression virtual sensors. On the latter day, the virtual sensors seem to fail to capture the rising and falling trend from the physical sensor, predicting a straighter line.

Readings from the physical sensor are shown in Figure 8.13 with respect to the two models' error ranges, which corresponds to the predictions plus the maximal training error. On Friday 31 March 2017, the physical sensor's readings fall far outside the acceptable range,

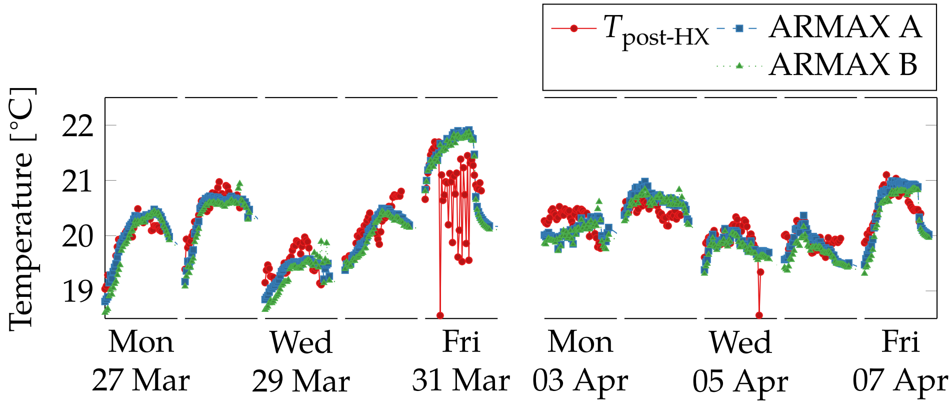

Figure 8.12: Comparison between physical sensors and ARMAX model-based virtual sensors for post heat exchanger temperature, airflow, and fan speed during working hours (from 8am to 5pm) over two weeks. Outside working hours and during weekends the ventilation system is shut down. The virtual sensors follow the physical ones except in one case. On Friday in the first week the temperature sensor oscillates strongly and deviates from the two virtual sensors.

which suggests a fault in the sensors pair. On Monday 3 April 2017, despite the trends being different, all readings fall inside the acceptable range.

NON-LINEAR REGRESSION MODELS

Model SVM uses support vector machine regression with radial basis function kernels and parameters set to $C = 100, \gamma = 0.04$. Model ANN uses an artificial neural network with 200 hidden layer neurons. Parameters for both models were optimized over the training periods. Only working hours were considered, as with the other models. Both models use the inputs as the Model B described in Table 8.1.

Virtual sensors readings are shown against physical sensors readings in Figure 8.14. The virtual sensors follow closely the physical sensor, except on Friday 31 March 2017 and on Tuesday 4 April 2017. During the former, day the physical sensor strongly oscillates while the virtual sensors predict a more regular trend, in agreement the

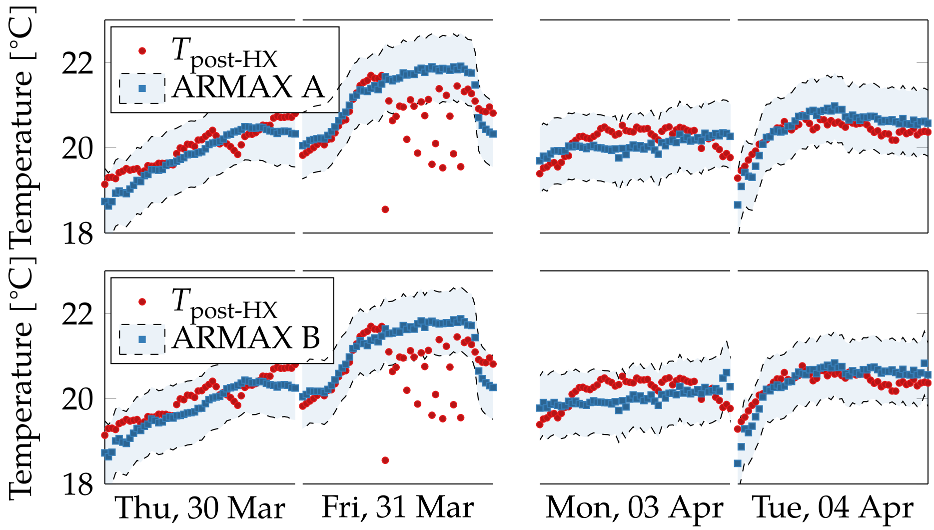

Figure 8.13: Comparison between physical sensors and acceptable ranges obtained from ARMAX model-based virtual sensors for post heat exchanger temperature during working hours (from 8am to 5pm) for selected days. The sensors readings fall inside the acceptable ranges except on Friday 31 March 2017, when they deviate significantly. The anomalous trend is not present neither in previous or following days. On Tuesday 4 April 2017, most models consistently overestimate the physical sensors, but their trends are similar.

linear regression virtual sensors. Model SVM also predict oscillations, but significantly weaker than the physical sensor. On the latter day, the virtual sensors consistently overestimate the physical one, as it happens with Model B.

Readings from the physical sensor are shown in Figure 8.15 with respect to the two models' error ranges, which corresponds to the predictions plus the maximal training error. On Friday 31 March 2017, the physical sensor's readings fall far outside the acceptable range, which suggests a fault in the sensors pair. On Monday 3 April 2017, despite virtual sensors overestimate the physical ones, all readings fall inside the acceptable range.

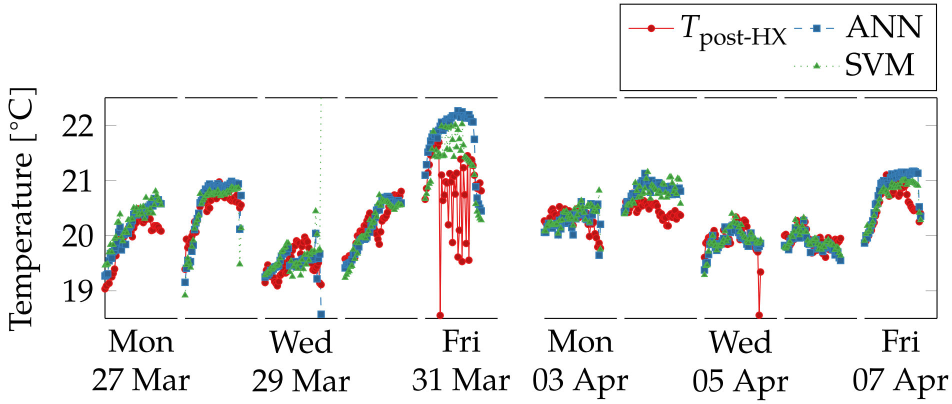

Figure 8.14: Comparison between physical sensors and non-linear regression model-based virtual sensors for post heat exchanger temperature, airflow, and fan speed during working hours (from 8am to 5pm) over two weeks. Outside working hours and during weekends the ventilation system is shut down. The virtual sensors follow the physical ones except in two cases. On Friday in the first week the temperature sensor oscillates strongly and deviates from the two virtual sensors. On Tuesday in the second week the virtual sensors ANN and SVM consistently overestimate the sensors readings.

8.5 Conclusions and Future Directions

8.5.1 Conclusions

We proposed a technique to exploit relations between physical quantities inside a ventilation unit to create virtual sensors, introducing, therefore, virtual redundancy. We applied this technique to ventilation units in a real building, creating virtual sensors for each of three existing sensors: temperature, airflow, and fan speed. We applied our method to one of the ventilation units in an existing building and we noted how on a particular day all virtual sensors for temperature, regardless of the model and input sensors used, deviated from the physical sensor. Its trend was, therefore, detected as anomalous.

Virtual sensors can be developed using a multitude of diverse models, with varying accuracy in predicting physical quantities in the

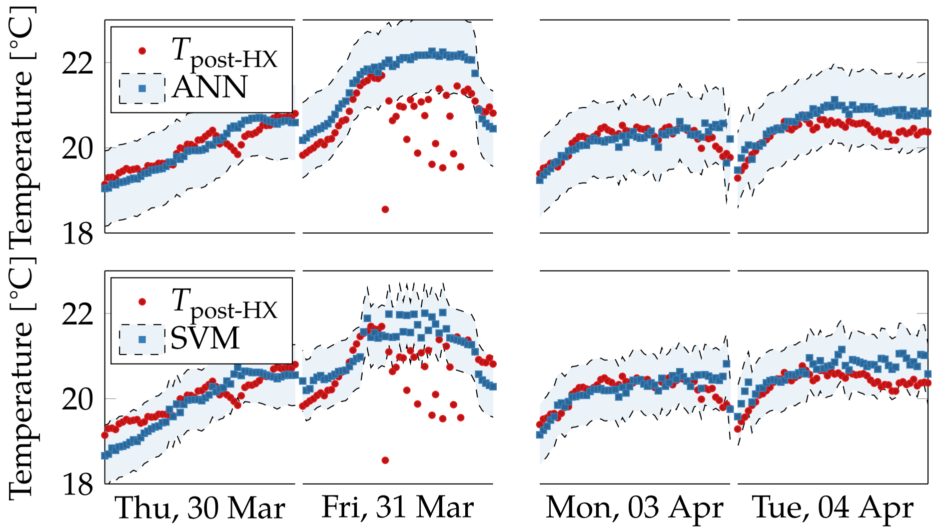

Figure 8.15: Comparison between physical sensors and acceptable ranges obtained from non-linear regression model-based virtual sensors for post heat exchanger temperature during working hours (from 8am to 5pm) for selected days. The sensors readings fall inside the acceptable ranges except on Friday 31 March 2017, when they deviate significantly. The anomalous trend is not present neither in previous or following days. On Tuesday 4 April 2017, most models consistently overestimate the physical sensors, but their trends are similar.

system. At first, we employed linear regression models, under the assumption that the related quantities obey linear relations, at least locally. Afterwards, we used ARMAX methods from statistical analysis, where the current value of a sensor was predicted from its history together with the input sensors. Finally, we developed two virtual sensors using non-linear models such as SVM regression and ANN.

We proposed two different techniques to measure deviations between physical and virtual sensors. R^2 score estimates how good a linear model fits some data. For non-linear models, the R^2 score is meaningless, therefore, we also used acceptable ranges obtained from the maximal training error. The virtual sensors predicted the values of physical sensors with satisfactory accuracy, and large deviations corresponded to actual anomalous behavior.

Contrary to physical redundancy, virtual redundancy does not increase cost and complexity but carries similar advantages, and several applications can profit from it. e.g., fault detection and diagnostics (FDD) methods, such as the one proposed in this paper, and automatic FDD methods can compare duplicated signals and detect when they diverge from each other. Fault-tolerant control can be achieved by duplicating a physical sensor with a virtual one, so that the system can continue functioning even if it fails. Sensors fusion enhances readings from a physical sensor with readings from other ones, improving measurement accuracy. Expensive physical sensors can be replaced by virtual ones in constrained systems, reducing costs and complexity.

In modern buildings, what sensors should be included in a ventilation unit is currently an open question. Sensors can be expensive and increase the construction complexity of a ventilation unit; however, they are necessary for its correct operation and useful for diagnostics. Virtual sensors are a promising technique that can decrease cost and complexity without compromising functionality or decreasing reliability.

The proposed methodology suffers, however, from some limitations. Data must be available both to create the virtual sensors' models and to monitor the ventilation unit. Therefore, a system for data collection and storage must be set in place, which could be difficult for older buildings. Data collection should be reliable, i.e., periods of missing data, or 'data holes' should be rare, and readings should be validated to ensure the models correctly represent the system. Choosing inputs for virtual sensors model is challenging, and so is choosing the type of model. Complex models can be accurate, but also difficult to develop and can have parameters to estimate, while simple models may not be able to reproduce the entire dynamics of the system.

8.5.2 FUTURE DIRECTIONS

While the application of the presented method for FDD on ventilation units using virtual sensors yielded promising results, more work is necessary to design and implement an automatic FDD framework. Automatic FDD is necessary to reduce operation cost and increase energy efficiency of buildings [75]. Moreover, comprehensive experiments should be set up to assess the actual benefits of this method [61].

We performed manual FDD by noticing how for one day the R^2 score between physical and virtual temperature sensors changed abruptly and significantly, and physical sensors' readings fell outside the acceptable range, which suggested a fault. However, a proper threshold system must be set up to achieve automatic FDD. This can be achieved by using expert knowledge and a training set of labelled faulty historical data, or by generating faulty data using simulations. Moreover, the temperature sensor exhibited faulty behavior only for a single day during the first week, while it appeared to work correctly for the rest of the testing period. Therefore, a threshold system should also be used to decide whether a significant but short-lived deviation is a fault.

We developed virtual sensors using linear and non-linear regression models, together with statistical analysis techniques. Better performance could be achieved by using more advanced methods, such as simulation using energy models of the ventilation units [2]. Moreover, to decide what inputs to use for virtual sensors, we reasoned about the physical relations between quantities inside the ventilation unit. While this approach might lead to accurate results, it could be ineffective for more complex systems. An automatic method could be employed to automatically select inputs and design effective virtual sensors, such as the one presented in [113].

We used regression models to predict data during a period close to the one used for training, under the assumption that the system's behavior did not change significantly. When extending the prediction to other periods, this assumption might not hold anymore, and seasonal variations must be taken into account.

Finally, in this paper we applied the proposed methodology to sensors in a ventilation unit. Other buildings subsystems could benefit from virtual sensors, e.g., heating loops, lighting, or room-level equipment. Additional work would be necessary to identify inputs and models and to extend the methodology to each of such subsystems.

ACKNOWLEDGEMENT

This work is supported by the Innovation Fund Denmark for the project COORDICY and partly supported by the Spanish MINECO ad FEDER research project SCAV (ref.DPI2017-88403-R).

REFERENCES

[2] **Claudio Giovanni Mattera**, Muhyiddine Jradi and Hamid Reza Shaker. 'Online Energy Simulator for Building Fault Detection and Diagnostics Using Dynamic Energy Performance Model'. In: *International Journal of Low-Carbon Technologies* 13.3 (17th May 2018), pp. 231–239. ISSN: 1748-1325. DOI: 10.1093/ijlct/cty019.

[3] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'Fault Detection and Diagnostics in Ventilation Units Using Linear Regression Virtual Sensors'. In: *International Symposium on Advanced Electrical and Communication Technologies (ISAECT)* (Kenitra, Morocco, 21st–23rd Nov. 2018). IEEE. 24th Jan. 2019. DOI: 10.1109/ISAECT.2018.8618755.

[4] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'A Method for Fault Detection and Diagnostics in Ventilation Units Using Virtual Sensors'. In: *Sensors* 18.11 (14th Nov. 2018). ISSN: 1424-8220. DOI: 10.3390/s18113931.

[7] Muhyiddine Jradi, Fisayo Caleb Sangogboye, **Claudio Giovanni Mattera**, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'A World Class Energy Efficient University Building by Danish 2020 Standards'. In: *Energy Procedia* 132 (Oct. 2017): *11th Nordic Symposium on Building Physics*, pp. 21–26. ISSN: 1876-6102. DOI: 10.1016/j.egypro.2017.09.625.

[12] European Parliament and Council of the European Union. 'Directive 2010/31/EU of the European Parliament and of the Council on the Energy Performance of Buildings'. In: *Official Journal of the European Union* L153 (18th June 2010), pp. 13–35. EU DIRECTIVE: 2010/31.

[13] R. Isermann. 'Supervision, Fault-detection and Fault-diagnosis Methods — an Introduction'. In: *Control Engineering Practice* 5.5 (May 1997), pp. 639–652. ISSN: 0967-0661. DOI: 10.1016/S0967-0661(97)00046-4.

[16] *2011 Buildings Energy Data Book*. Tech. rep. U. S. A. Department of Energy, Mar. 2012. URL: https://openei.org/doe-opendata/dataset/buildings-energy-data-book (visited on 20th Dec. 2018).

[36] Evan Mills. 'Building Commissioning: A Golden Opportunity for Reducing Energy Costs and Greenhouse Gas Emissions in the United States'. In: *Energy Efficiency* 4.2 (May 2011), pp. 145–173. ISSN: 1570-6478. DOI: 10.1007/s12053-011-9116-8.

[39] Srinivas Katipamula and Michael R. Brambley. 'Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems–A Review, Part I'. In: *HVAC&R Research* 11.1 (Jan. 2005), pp. 3–25. DOI: 10.1080/10789669.2005.10391123.

[40] Srinivas Katipamula and Michael R. Brambley. 'Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems–A Review, Part II'. In: *HVAC&R Research* 11.2 (Apr. 2005), pp. 169–187. DOI: 10.1080/10789669.2005.10391133.

[41] Woohyun Kim and Srinivas Katipamula. 'A Review of Fault Detection and Diagnostics Methods for Building Systems'. In: *Science and Technology for the Built Environment* 24.1 (8th May 2018), pp. 3–21. ISSN: 2374-4731. DOI: 10.1080/23744731.2017.1318008.

[51] Haorong Li, Daihong Yu and James E. Braun. 'A Review of Virtual Sensing Technology and Application in Building Systems'.

In: *HVAC&R Research* 17.5 (3rd Oct. 2011), pp. 619–645. DOI: 10.1080/10789669.2011.573051.

[61] Bo Nørregaard Jørgensen, Mikkel Baun Kjærgaard, Sanja Lazarova-Molnar, Hamid Reza Shaker and Christian T. Veje. 'Challenge: Advancing Energy Informatics to Enable Assessable Improvements of Energy Performance in Buildings'. In: *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems* (Bangalore, India, 14th–17th July 2015). ACM. New York, NY, USA, 2015, pp. 77–82. DOI: 10.1145/2768510.2770935.

[69] MinJeong Kim, Hongbin Liu, Jeong Tai Kim and ChangKyoo Yoo. 'Sensor Fault Identification and Reconstruction of Indoor Air Quality (IAQ) Data Using a Multivariate Non-Gaussian Model in Underground Building Space'. In: *Energy and Buildings* 66 (Nov. 2013), pp. 384–394. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2013.07.002.

[75] Sanja Lazarova-Molnar, Hamid Reza Shaker, Nader Mohamed and Bo Nørregaard Jørgensen. 'Fault Detection and Diagnosis for Smart Buildings: State of the Art, Trends and Challenges'. In: *3rd MEC International Conference on Big Data and Smart City (ICBDSC)* (Muscat, Oman, 15th–16th Mar. 2016). IEEE. 28th Apr. 2016, pp. 1–7. DOI: 10.1109/ICBDSC.2016.7460392.

[78] Miquel À. Cugueró-Escofet, Diego García, Joseba Quevedo, Vicenç Puig, Santiago Espin and Jaume Roquet. 'A Methodology and a Software Tool for Sensor Data Validation/reconstruction: Application to the Catalonia Regional Water Network'. In: *Control Engineering Practice* 49 (Apr. 2016), pp. 159–172. ISSN: 0967-0661. DOI: 10.1016/j.conengprac.2015.11.005.

[98] Luis Pérez-Lombard, José Ortiz and Christine Pout. 'A Review on Buildings Energy Consumption Information'. In: *Energy and Buildings* 40.3 (2008), pp. 394–398. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2007.03.007.

[99] *Technical Synthesis Report Annex 34. Computer Aided Evaluation of HVAC System Performance*. Tech. rep. Birmingham, United Kingdom: International Energy Agency - Energy Conservation in Buildings and Community Systems Programme, 2006.

[100] Yuebin Yu, Denchai Woradechjumroen and Daihong Yu. 'A Review of Fault Detection and Diagnosis Methodologies on Air-handling Units'. In: *Energy and Buildings* 82 (Oct. 2014), pp. 550–562. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2014.06.042.

[101] Haorong Li and James E. Braun. 'Decoupling Features and Virtual Sensors for Diagnosis of Faults in Vapor Compression Air Conditioners'. In: *International Journal of Refrigeration* 30.3 (May 2007), pp. 546–564. ISSN: 0140-7007. DOI: 10.1016/j.ijrefrig.2006.07.024.

[102] Nunzio Cotrufo and Radu Zmeureanu. 'Virtual Outdoor Air Flow Meter for an Existing HVAC System in Heating Mode'. In: *Automation in Construction* 92 (Aug. 2018), pp. 166–172. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2018.03.036.

[103] Andrew Kusiak, Mingyang Li and Haiyang Zheng. 'Virtual Models of Indoor-air-quality Sensors'. In: *Applied Energy* 87.6 (June 2010), pp. 2087–2094. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2009.12.008.

[104] K. Verbert, R. Babuška and B. De Schutter. 'Combining Knowledge and Historical Data for System-level Fault Diagnosis of HVAC Systems'. In: *Engineering Applications of Artificial Intelligence* 59 (Mar. 2017), pp. 260–273. ISSN: 0952-1976. DOI: 10.1016/j.engappai.2016.12.021.

[105] Olivier Renaud and Maria-Pia Victoria-Feser. 'A Robust Coefficient of Determination for Regression'. In: *Journal of Statistical Planning and Inference* 140.7 (July 2010), pp. 1852–1862. ISSN: 0378-3758. DOI: 10.1016/j.jspi.2010.01.008.

[106] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel,

Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot and Édouard Duchesnay. 'Scikit-learn: Machine Learning in Python'. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830. URL: http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.

[107] *The Basics of Fan Performance Tables, Fan Curves, System Resistance Curves and Fan Laws*. FA/100-99. Greenheck. Schofield, WI, USA, Dec. 1999. 4 pp. URL: http://www.greenheck.com/resources/library/application-articles (visited on 26th Dec. 2018).

[108] Rodney A. Martin, M. Schwabacher, N. Oza and A. Srivastava. 'Comparison of Unsupervised Anomaly Detection Methods for Systems Health Management Using Space Shuttle'. In: *Proceedings of the Joint Army Navy NASA Air Force Conference on Propulsion, JANNAF*. 2007.

[109] Miquel À. Cugueró-Escofet, Joseba Quevedo, C. Alippi, M. Roveri, Vicenç Puig, Diego García and F. Trovò. 'Model- vs. Data-based Approaches Applied to Fault Diagnosis in Potable Water Supply Networks'. In: *Journal of Hydroinformatics* 18.5 (10th May 2016), pp. 831–850. DOI: 10.2166/hydro.2016.218.

[110] Eduardo Trutié-Carrero, Damian Valdés-Santiago, Ángela León-Mecías and Jorge Ramírez-Beltrán. 'Burst Detection and Localization using Discrete Wavelet Transform and Cross-Correlation'. In: *Revista Iberoamericana de Automática e Informática industrial* 15.2 (2018), pp. 211–216. ISSN: 1697-7920. DOI: 10.4995/riai.2017.8738.

[111] Emil L. Hanzevack, Theresa W. Long, Chris M. Atkinson and M. Traver. 'Virtual Sensors for Spark Ignition Engines Using Neural Networks'. In: *American Control Conference* (Albuquerque, NM, USA, 6th June 1997). IEEE. 6th Aug. 2002, pp. 669–673. DOI: 10.1109/ACC.1997.611885.

[112] Manuel Acosta, Stratis Kanarachos and Michael E. Fitzpatrick. 'A Virtual Sensor for Integral Tire Force Estimation using Tire Model-less Approaches and Adaptive Unscented Kalman Filter'. In: *Proceedings of the 14th International Conference on Informatics in Control, Automation and Robotics* (Madrid, Spain, 26th–28th July 2017). Vol. 1. INSTICC. SciTePress, 2017, pp. 386–397. DOI: 10.5220/0006394103860397.

[113] Burkhard Hensel and Klaus Kabitzsch. 'Generator for Modular Virtual Sensors'. In: *21st International Conference on Emerging Technologies and Factory Automation (ETFA)* (Berlin, Germany, 6th–9th Sept. 2016). IEEE. 7th Nov. 2016, pp. 1–8. DOI: 10.1109/ETFA.2016.7733726.

[114] Pierre Borne, Dumitru Popescu, Floria G. Filip, Dan Stefanoiu and Florin Gh. Filip. *Optimization in Engineering Sciences: Exact Methods*. Ed. by Bernard Dubuisson. 1st ed. iSTE. Hoboken, NJ, USA: John Wiley & Sons, Incorporated, 20th Feb. 2013. DOI: 10.1002/9781118577899.

[115] Alex J. Smola and Bernhard Schölkopf. 'A Tutorial on Support Vector Regression'. In: *Statistics and Computing* 14.3 (Aug. 2004), pp. 199–222. ISSN: 1573-1375. DOI: 10.1023/B:STCO.0000035301.49549.88.

[116] Kurt Hornik, Maxwell Stinchcombe and Halbert White. 'Multilayer Feedforward Networks Are Universal Approximators'. In: *Neural Networks* 2.5 (1989), pp. 359–366. ISSN: 0893-6080. DOI: 10.1016/0893-6080(89)90020-8.

CHAPTER 9

DETECTING ANOMALOUS COMPONENTS USING CONSENSUS AMONG MULTIPLE PEERS

This chapter is a cosmetic adaptation of the following journal paper.

Claudio Giovanni Mattera, Hamid Reza Shaker and Muhyiddine Jradi. 'Consensus-based Method for Anomaly Detection in VAV Units'. In: *Energies* 12.3 (1st Feb. 2019). ISSN: 1996-1073. DOI: 10.3390/en12030468

ABSTRACT

Buildings account for large part of global energy consumption. Besides energy consumed due to normal operation, a large amount of energy can be wasted due to faults in buildings subsystems.

Fault detection and diagnostics techniques aim to identify faults and prevent energy waste, but are often difficult to apply in practice. Data-driven methods, in particular, require an adequate amount of fault-free training data, which is rarely available.

In this paper, we propose a method for anomaly detection that exploits consensus among multiple identical components. Even if some of the components are faulty, their aggregate behaviour is overall correct, and it can be used to train a data-driven model.

We test our method on variable-air-volume units in an existing building, executing two experiments grouping the components according to ventilation unit, and according to room type. The two experiments

identified the same set of anomalous components, i.e. their behaviour was different from the rest of the group in both cases, and this suggests that the anomaly was not due to wrong group assignment.

The proposed method shows the potential of exploiting consensus among multiple identical systems to detect anomalous ones.

9.1 INTRODUCTION

Nowadays, buildings have a large impact on both energy consumption and other environmental effects such as carbon emissions. In the European Union they are responsible for 40 % of the total energy usage and 36 % of CO_2 emissions [12, 26]. Similarly, in the United States, they are responsible for about 41 % of primary energy consumption in 2010, which was 44 % more than transports and 36 % more than industry. Total building primary energy consumption in 2009 was about 48 % higher than in 1980, going from 1290 TW h to 2784 TW h [16]. It is, therefore, evident that buildings are a key sector for achieving environment and climate targets such as 20 20 by 2020, i.e. 20 % reduction in greenhouse gases and 20 % share of renewable energy sources by year 2020 [21], and the more recent 30 % energy efficiency by year 2030 [22].

Modern commercial buildings contain large and complex systems, such as heating, ventilation and air conditioning (HVAC) and lighting, and their operation is controlled by automated building management systems (BMSs), which often require a network of sensors, meters and actuators. Faults in these systems impact building operations, e.g. by causing occupants discomfort, but also increase energy usage. The most common faults in commercial buildings in U. S. are estimated to have caused over 3.3 billion dollars in energy waste in 2009 [36], and over 7 billion dollars in 2017 [37]. It is often difficult to precisely identify faults, and sometimes even to detect them, and a system could operate for a long time before the building management even notices it is not working correctly [1]. Fault detection and diagnostics (FDD) techniques aim to detect faults and identify their precise location and cause. Research and application of FDD techniques, applied successfully in other fields for several decades, gained traction in the buildings sector in the past few years.

Ventilation units are among the largest and most critical systems in buildings, and account for large energy consumption. Their faults, such as incorrect HVAC on/off modes or inappropriate setpoints for thermostats, are responsible for a large share of energy waste [37]. While many FDD techniques have been applied to the large air handling unit (AHU), faults and misconfigurations involving variable air volume (VAV) units at room level are often ignored [117]. However, considering that the VAV units have the main responsibility of direct air supply to each room, and taking into account the importance of attaining good indoor air quality and thermal comfort, proper monitoring and FDD investigations of VAV units seems very sensible.

Many FDD techniques have major limitations when applied in practice, due to non-ideal conditions of the real world. Model-based techniques require detailed knowledge of the system under test, which is often not available. Data-driven techniques, on the other hand, require validated and fault-free historical data to learn the correct behaviour of the system. Historical data is often available, but it is rarely validated, and there is a risk that faulty behaviour is used in the training phase.

Peer validation and consensus-based validation, on the other hand, can be used to mitigate this issue. Multiple identical or similar systems are considered together, under the assumption that the majority of them operate correctly. When their historical data is used to train a model, the contributions from faulty systems are small compared with the ones from healthy ones, and their effect on the model is diluted. Therefore, the requirement for fault-free training data is lifted, and faulty systems are identified as outliers among the healthy ones. This eliminates the need for complex and sophisticated models and large system operation datasets.

The rest of the paper is organized as follows. The state of the art is reviewed in Section 9.2. The proposed technique is introduced in Section 9.3. Section 9.4 presents the case study and discusses results and implications. Finally, conclusions are drawn in Section 9.5.

9.2 STATE OF THE ART

Kim et al. present a comprehensive review of recent FDD methods for building systems [41]. The authors identify three main categories, depending on the approach used: model-based methods, data-driven methods and rules-based methods.

In model-based methods, an explicit model of the system under test is created, using first principles physics, physics and other system and envelope modelling techniques. Results obtained from the model are compared with the ones obtained from the actual system, and, if the two deviate, a fault is detected. Model-based methods have usually high accuracy and can detect faults with smaller impacts, as well as faults absent from historical data. By modifying the model, it is possible to simulate faulty conditions, which makes possible to precisely diagnose faults. On the other hand, such models require extensive knowledge of the system under test and cannot easily be extended to other systems. Correctly estimating the model's parameters is also a challenge [118].

In data-driven–or history-based–methods, a model of the system under test is created from historical data. Several techniques exist, such as artificial neural networks, principal component analysis and statistical machine learning algorithms. The model is treated as a black box and no understanding of the system is necessary. For this reason, these methods can often be easily extended to other systems by simply re-training the model from different data. On the other hand, a relatively large amount of fault-free historical data must be available to train the model. This makes data-driven methods unsuitable for newly deployed systems and for situations where historical data is not provably fault-free. Independent sets of labeled faulty data are often necessary to perform precise fault diagnostics and identification.

In rule-based methods, a set of rules describing the behaviour of the system under test is defined. Rules are usually obtained from expert knowledge and technical documentation, and can describe both correct and faulty behaviour, which makes possible to precisely diagnose faults. No training data is necessary, and only a high-level knowledge of the system is needed. However, rules can only represent relatively simple systems and cannot properly describe complex interactions.

To the best of our knowledge, no previous work has been done on using consensus-based techniques for FDD in building systems, and specifically on VAV units used to control CO_2 level. Narayanaswamy et al. present a model, cluster and compare method for FDD on VAV units, where data from several units are used to detect anomalies [117]. Linear models are trained for each individual VAV unit, and the obtained parameters undergo a clustering procedure. Units that do not belong to any cluster are identified as anomalous and, finally, the results are used to generate a set of expert rules for anomaly detection. The authors deploy and test their method on a real building, and use it to detect anomalies with respect to temperature control in rooms.

Consensus techniques have been used in the field for other purposes, such as features selection. Partially redundant measurements in complex systems such as HVAC systems can make it difficult to apply FDD methods, which are often not designed to handle conflicting inputs or large amounts of inputs. Yuwono et al. present a method for feature selection using swarm intelligence and consensus clustering, which can be used to reduce and aggregate the number of features used in FDD methods. [119]. Consensus clustering has the advantage that the number of clusters is not fixed in advance, instead, clusters are identified automatically.

Consensus-based techniques have also been used for FDD in other fields. FDD methods often use data and findings from models and laboratory tests to validate or predict data for systems in the field. Differences from the model, and different conditions between tests and the real world, can reduce methods accuracy and effectiveness. Byttner et al. present a FDD method for vehicles based on consensus between such models and tests, and on-field systems [120]. Data is first preprocessed on-vehicle and interesting features are identified, which are sent to a central server that collects them for all vehicles. The central server searches for outliers, i.e. features from a single vehicle that do not match the overall distribution across the entire fleet, laboratory tests, or models. The authors prepare two different experiments, one for detecting faults in cooling systems for large vehicles, and one for detecting faults in hard-drives. In the former experiment, only a single real vehicle was used in multiple different driving conditions and paired with a simulated one, however, in the latter experiment, several different hard-drives were used.

Bianchin et al. propose another example of consensus-based techniques: a method for FDD in sensors networks based on clustering and consensus [121]. A token travels across the sensors network, gathering measurement as it visits each node, and computing similarity among them. When a faulty node is present, it is isolated to its own cluster, while connectivity among the other nodes is maintained. The method is shown to be used for static estimation, i.e. when the measured quantity is constant over time, and also for dynamic estimation, i. e., when the measured quantity changes over time and the network must produce a real-time estimation.

Consensus-based techniques are popular in the field of fault-tolerant control, where multiple and partially redundant agents propose concurrent decisions. Such decisions can lead to conflicts due to faults in the system, but also due to noise, missing information or other causes. Multiple agents can then negotiate between each other or be excluded by the majority until a consensus is reached.

Davoodi et al. present a method for consensus control in multi-agent systems and report an experiment on autonomous unmanned underwater vehicles [122]. Zhou et al. present a method for actuator fault estimation in multi-agent systems, where agents can asymptotically converge to a common strategy with bound errors [123].

Consensus-based algorithms are also a popular approach for distribute decision support systems. Lee et al. present a technique to control a multi-microgrid using consensus between peers [124]. Liu et al. present a technique for energy sharing in the context of community energy internet, where a global objective function is optimized through consensus among peers [125].

Table 9.1 summarizes the advantages and disadvantages of categories of FDD methods. Traditional data-driven methods do not require deep knowledge of the system, support complex dynamics, and can be easily generalized to other systems. However, their main disadvantage is to require fault-free historical data to train a model. Consensus-based data-driven methods, on the other hand, replace this requirement with the one for multiple identical systems, while maintaining the other advantages.

9.2.1 Problem statement

Data-driven methods offer several advantages for FDD, however, they have a major drawback of requiring fault-free training data, which is rarely available in practice. If historical data was generated by a faulty system, the resulting model would later recognize similar faults as healthy conditions, reducing its effectiveness in detecting faults. This chicken-and-egg situation is a significant problem in applying FDD techniques: a model is necessary to validate data, but validated data is necessary to construct a model.

In this paper, we propose to solve this problem by training an 'aggregate' model using historical data from a large number of identical or similar systems. Systems whose behaviour significantly deviates from the aggregate behaviour are detected as anomalous. While we cannot ensure that all systems work correctly, we assume that only a small part of them is faulty and that they are not affected by the same fault. Therefore, the individual faults would have a small impact during training, and the resulting model would be largely unaffected.

9.3 Consensus-based method for anomaly detection

The method proposed in this paper analyzes time-series from multiple similar systems. Correct and anomalous conditions are defined based on the consensus from all the systems.

The main intuition of this method, illustrated in Figure 9.1, is to find sequences of events in multiple, related time-series and group them in *episodes*, where each episode represents a qualitative phenomenon. E.g. if CO_2 level rises, then the ventilation flow rate should increase, due to the BMS acting to maintain good air quality. Episodes are, therefore, a sequence of events belonging to a group of time-series.

Table 9.1: Advantages of FDD methods. Consensus-based methods have similar advantages and disadvantages to traditional data-driven methods. Their respective trade-offs, underlined in the table, are the requirements for multiple identical systems, and the requirement for fault-free historical data.

	Data-driven		**Model-based**	**Rule-based**
	Traditional	**Consensus-based**		
Advantage				
Supports complex dynamics	$\checkmark$	$\checkmark$	$\checkmark$	
Has high accuracy			$\checkmark$	
Can be easily generalized to other systems	$\checkmark$	$\checkmark$		
Does not require detailed knowledge	$\checkmark$	$\checkmark$		$\checkmark$
Does not require expert knowledge	$\checkmark$	$\checkmark$		
Does not require fault-free historical data		$\underline{\checkmark}$		$\checkmark$
Does not require multiple identical systems	$\underline{\checkmark}$		$\checkmark$	$\checkmark$

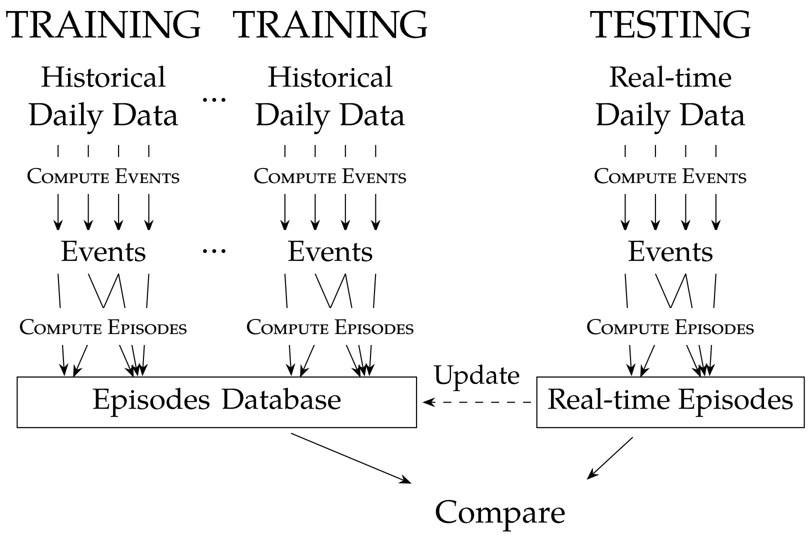

Figure 9.1: Overview of consensus-based FDD method. Events are computed for each time-series from historical data. Episodes, i.e. sequences of events belonging to different time-series, are used to construct a database containing the normal qualitative behaviour. At run-time, events and episodes are computed from real-time data and compared with the database, and optionally added to the database.

A database of episodes is obtained from historical data from several groups of time-series. Frequent episodes are assumed to happen during correct conditions, while rare or unknown episodes are assumed to be symptoms of anomalous behaviour. Episodes are later computed from real-time data and compared with the episodes in the database. When a large part of real-time episodes corresponds to episodes rarely encountered in historical data, i.e., when the current behaviour of the system is qualitatively different from its historical one, the system is flagged as anomalous.

The episode database can optionally be updated with new episodes computed from real-time data. This would allow to track seasonal variations and, moreover, to apply the method on a newly deployed system without using a separate training phase. In that case, the episodes database would be gradually populated over time, and earlier

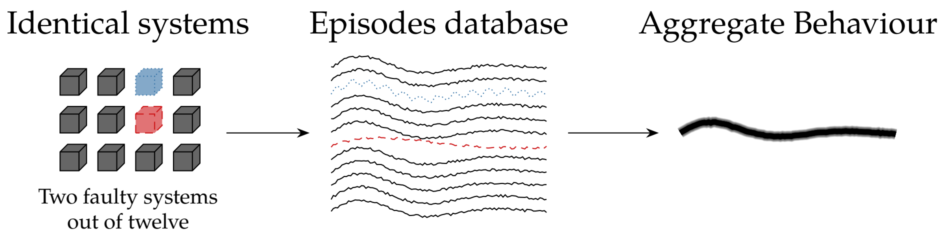

Figure 9.2: Effect of faulty systems in consensus-based methods. If a small number of systems used in training are faulty, their contributions will be diluted among the ones by correctly operating systems.

results could be inaccurate.

In order to avoid the necessity of validated fault-free training data, consensus between multiple similar systems can be exploited. Assuming that only a small part of the systems used in training are faulty or exhibit anomalous behaviours, their episodes would be overwhelmed by the episodes of the rest of the systems, as illustrated in Figure 9.2.

In order to obtain a consistent common behaviour, systems should be grouped by common characteristics. E.g. multiple rooms could be divided by room type, but also by room location, such as by floor number or building side, or by other characteristics. When a room shows anomalous behaviour within its group, it could be due to faulty components, but also to incorrect or insufficient grouping, as shown by Narayanaswamy et al. in [117]. E.g. the only classroom on the top floor might deviate from all other classrooms, which are on the ground floor, due to different thermal loss. Multiple orthogonal characteristics should be used to avoid this possibility, such that systems which are anomalous in several groups are effectively labeled as anomalous.

Compared with a traditional approach of clustering based on model or statistical parameters, such as mean or variance, using episodes allows to represents interactions between different measurements over time. Such interactions, or lack thereof, can be qualitatively linked to physical phenomena within the system, and are learnt from aggregate historical data. Moreover, when updating the episodes database with episodes computed from real-time data, this method adapts to slow seasonal variations in the system's dynamics.

In the rest of this section, we describe the procedure for data preprocessing and preparation, we define events and episodes, and, finally, we describe how to monitor multiple time-series to detect anomalies.

9.3.1 DATA PREPROCESSING AND PREPARATION

Time-series can be divided into two categories depending on the nature of the measured quantity. Time-series with a large number of readings changing gradually over time, such as temperature or CO_2 level, are called *continuous* time-series. Time-series with values defined over a finite and small domain, such as on/off, or a predetermined number of states, which are constant for long periods and change value abruptly, are instead called *discrete* time-series.

In order to extract episodes from a group of time-series, it is first necessary to extract *events* from each of them. An event is a qualitative local trend of a time-series. Events are defined and extracted differently for continuous and discrete time-series, as illustrated in Figure 9.3 and described in the following.

EVENTS FOR CONTINUOUS TIME-SERIES

The following method, based on the one presented by Pisón et al. in [126], is used to extract events from continuous time-series.

Continuous time-series have many readings and are often subject to noise. In order to identify the high-level trend without accounting for small deviations, the time-series are first filtered with a lowpass filter. This operation is necessary to leave out low-order variations that do not impact significantly the system under test.

The next step is to find *important points* in time-series, which are defined as follows. Consider a time-series a_i, where $i \in \mathbb{Z}$ is the time index, and a_i is the value at time index i, e.g. temperature in °C or CO_2 level in ppm. Consider a point $a_m \in$ time-series and a window around it of radius n: $[a_{m-n}, \dots, a_m, \dots, a_{m+n}]$. a_m is an important minimum if

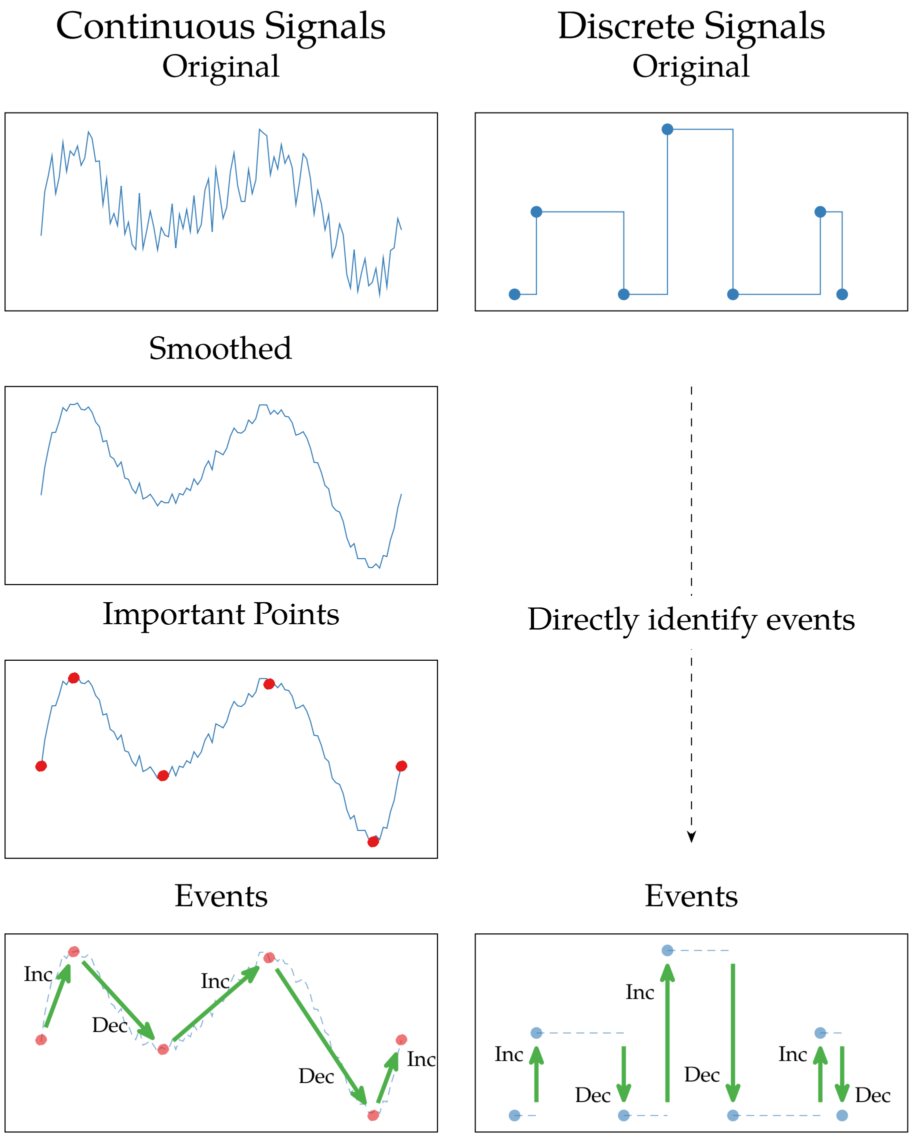

Figure 9.3: Extracting events from continuous and discrete time-series. Continuous time-series often present noise, therefore, they are first preprocessed with resampling, and smoothed using a lowpass filter. Important points are found in the preprocessed time-series, and events are identified from them. Discrete time-series, instead, are not affected by noise and events can be identified directly.

and only if

$$\begin{array}{l} a_m = \min\left[a_{m-n}, \dots, a_{m+n}\right], \\ \frac{a_i}{a_m} \geq r \wedge \frac{a_j}{a_m} \geq r \quad \forall i,j : m-n \leq i \leq m \leq j \leq m+n, \end{array} \tag{9.1}$$

where $r \gtrsim 1$ is a compression factor. The closer is r to 1, the more important points are found. The precise value of r is a parameter that must be tuned for the specific experiment. Similarly, a_m is an important maximum if and only if

$$\begin{array}{l} a_m = \max\left[a_{m-n}, \dots, a_{m+n}\right], \\ \frac{a_m}{a_i} \geq r \wedge \frac{a_m}{a_j} \geq r \quad \forall i,j : m-n \leq i \leq m \leq j \leq m+n. \end{array} \tag{9.2}$$

Once important points have been computed, it is possible to extract *events*. An event is a transition between two consecutive important points a_k and a_ℓ. In this paper, we consider the following event types: increment, decrement and horizontal trend. A transition is labeled as an increment if and only if

$$\begin{array}{l} a_k \text{ is an important minimum} \\ a_\ell \text{ is an important maximum} \\ w_1 \leq \ell - k \leq w_2 \\ h_1 \leq a_\ell - a_k \leq h_2, \end{array} \tag{9.3}$$

where w_1, w_2 are constraints on the length of the transition and h_1, h_2 are constraints on the size of the transition. The constraints w_1, w_2 are measured in number of samples or, equivalently, in length of time intervals, when the time-series has a fixed sampling rate. The constraints h_1, h_2 are measured in the same unit of the time-series values, e.g. °C for time-series recording temperature, or ppm for time-series recording CO_2 level. Similarly, a transition is labeled as a decrement if and only if

$$\begin{array}{l} a_k \text{ is an important maximum} \\ a_\ell \text{ is an important minimum} \\ w_1 \leq \ell - k \leq w_2 \\ h_1 \leq a_k - a_\ell \leq h_2. \end{array} \tag{9.4}$$

A transition is labeled as a horizontal step if and only if

$$\begin{aligned} &a_k, a_\ell \text{ are important points} \\ &w_1 \leq \ell - k \leq w_2 \\ &|a_k - a_\ell| \leq h_2. \end{aligned} \tag{9.5}$$

EVENTS FOR DISCRETE TIME-SERIES

Extracting events from discrete time series is considerably simpler than from continuous ones. Discrete time-series measure logical quantities and are not affected by noise, therefore, no filtering is necessary. Moreover, filtering a discrete time-series would result in a continuous time-series transitioning smoothly from one state to the other, which would not significantly approximate the original signal. Therefore, the filtering step is not performed for discrete time-series.

Since changes of values in discrete time-series represent a logical change in the measured quantity, the values themselves are important points, and the changes themselves are events, as shown in Figure 9.3.

9.3.2 EPISODES INVOLVING MULTIPLE TIME-SERIES

Episodes are ordered chains of events pertaining to multiple time-series, as shown in Figure 9.4. They represent high-level cause-effect transitions, such as (Occupancy increases, CO_2 level increases, Ventilation increases), or (Ventilation increases, CO_2 level decreases). Episodes can contain any number of events but they are limited to a certain window size.

9.3.3 MONITORING MULTIPLE TIME-SERIES

The method consists of an initial training phase and an online detection phase (Figure 9.1). During the training phase, historical data are divided into daily chunks, and episodes are extracted from them and stored to a database. At the end of this phase, the majority of episodes in the database will represent usual behaviour of the system. In the online detection phase, episodes are extracted every day from data

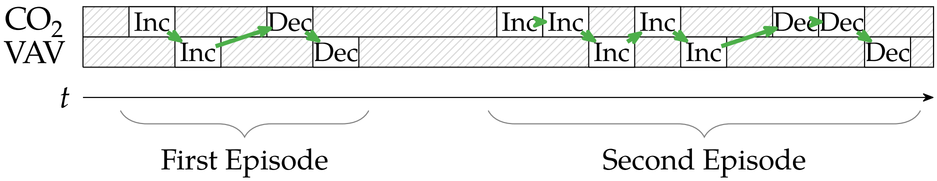

Figure 9.4: Episodes involving multiple time-series. Episodes are sequences of events and are represented with green arrows. The first episode corresponds to a typical threshold-based VAV actuation: when CO_2 level increases, the VAV unit opens, which in turns causes CO_2 level to decrease after some time, and thus the VAV unit to close. The second episode corresponds a more complex dynamics involving the two time-series.

and compared to the ones in the database. All episodes that are absent in the database, or have less than a given probability, e.g. 5 %, are flagged as anomalous. Therefore, when the system behaviour matches the one recorded in the database, it is considered normal, otherwise, it is considered anomalous.

A small number of anomalous episodes are expected even for healthy systems, therefore, a weekly moving average of anomalous episodes is computed. When such moving average exceeds a threshold, i.e. when anomalous episodes become common for a long period, the system as a whole is flagged as anomalous.

The method described so far can be used if validated and fault-free historical data is available. However, for many real-world systems, this might not be the case. If the system was faulty during the training period, the database would contain episodes representing faulty behaviour, and the method would flag such behaviour as correct during the online phase. This problem can be solved by exploiting consensus among several identical or similar systems during the training phase. Assuming most systems work correctly and only a small part are affected by faults, the majority of episodes stored in the database would, therefore, represent correct behaviour. Moreover, if faulty systems were affected by different faults their impact on the whole database would be even more diluted, as illustrated in Figure 9.2.

Figure 9.5: Building OU44 at campus Odense, University of Southern Denmark.

In order to account for slow-varying seasonal changes in the operation of the system, episodes obtained during the detection phase could be added to the database, and older episodes could be removed. In alternative, multiple databases could be created using historical data from different periods.

9.4 CASE STUDY

In this paper we present Odense undervisning 44 (OU44) as a case study. The building, shown in Figure 9.5, is located at the main campus of University of Southern Denmark, in Odense. It was built in 2015 and it is mainly used for teaching and office work. The building contains around 120 rooms of different types, as shown in Table 9.2, spread over three floors, and technical rooms located in the basement.

Data from the building are continuously recorded and stored in a database. Most rooms are equipped with indoor conditions sensors, such as CO_2 level, temperature, humidity and illuminance intensity, and with other meters such as lights status, heating valves and VAV units position, occupancy presence, blinds status and booking status.

Table 9.2: Room types in building OU44.

Room type	Count	Room type	Count
Office	48	Classroom	19
Corridor	21	Study zone	8
Other	7	Stairway	6
Conference room	4	Atrium	4
Copy room	3	Auditorium	2
Kitchen	2		

A selected number of rooms have separate plug load meters and occupancy counting cameras. In total, more than 3500 time-series are recorded for room-level measurements, and more than 1800 for the ventilation system.

The building's ventilation system consists of four identical ventilation units, each of them serving one corner of the building (north-east, south-east, south-west and north-west). They are designed to maintain constant shafts pressures of 130 Pa and 40 Pa in the entire unit, while, at room level, supply flow rates depend on the VAV unit position. When VAV units are open, the pressure difference in the supply and extract shafts induce airflow in the room. The amount of open VAV units can be used as an estimate of the airflow required to maintain a constant pressure in the shafts, as was shown in [4]. The airflow, in turns, is directly related to the energy consumption of the ventilation unit.

The position of VAV units themselves is based on multiple thresholds on CO_2 level: at 600 ppm the VAV unit opens by 45 %, at 750 ppm it opens by 70 % and at 900 ppm it opens by 100 %. When CO_2 level decreases the thresholds are affected by hysteresis of 100 ppm. The ventilation system is used to control room air quality, but also to provide natural cooling using outdoor air. As a result, VAV units can be open due to temperature, even when CO_2 level is low. Heating is provided by radiations, however, inlet air is heated to a setpoint of 20 °C to 22 °C inside the ventilation units before entering the supply shaft.

9.4.1 monitoring VAV units

The BMS opens and closes VAV units to maintain room air quality, which is measured by the CO_2 level. When the VAV unit is working correctly, increasing its position results in higher ventilation, which reduces the CO_2 level in the room. It is difficult to accurately estimate the dependency between CO_2 level and VAV position. Previous attempts using regression models, such as the ones used in [117] for temperature control, lead to unsatisfactory results, perhaps due to the coarseness of VAV position with respect to CO_2 level. However, episodes involving the two time-series can capture the qualitative relation. Ventilation increasing due to cooling rarely occurs in Denmark, and only during summer months. Often, this happens when many occupants are in the room, which results in faster increase due to CO_2 level. Therefore, VAV position is dominated by CO_2 level, and the effect of temperature is small.

Rooms in the building were divided into four groups according to their corresponding ventilation unit. Events were extracted from two time-series, CO_2 level and VAV position ratio. Each group was used to train and generate a database of episodes. Under the assumption that rooms sharing the same ventilation unit have similar behaviour, the episodes are consistent within the group, and the resulting database contains similar episodes.

On the other hand, each ventilation unit serves different types of room, such as offices or classrooms. If a room type is underrepresented in the ventilation unit, the behaviour of such rooms might seem anomalous with respect to its peers. However, it would not be due to a fault, but instead to the room's different shape and usage. To avoid this possibility of false positives, another experiment was performed by grouping rooms according to their type, as shown in Table 9.2.

Therefore, the groupings in the two experiments were defined as it follows.

a) Grouping by ventilation unit: the database was populated with episodes from all rooms belonging to the same ventilation unit. All four ventilation units 1 to 4 were considered.

b) Grouping by room type: the database was populated with episodes from all rooms of the same type. Six room types were considered: classroom, office, corridor, study zone, auditorium and conference room.

For both experiments, all parameters were set to the same values. Two time-series were considered: CO_2 level in the room and VAV position. The database was constructed dynamically, i.e. it was initially empty, and, every day, it was updated with episodes obtained during the online detection phase, and infrequent episodes were recorded. The experiment was performed on data from 20 November 2016 to 27 May 2017. Summer months were excluded, therefore, VAV position was independent of room temperature. Original data was resampled to 5 min and filtered with a Butterworth low-pass filter with cutoff period of 1 h, as outlined in the step 'Smoothed' in Figure 9.3. This type filter was chosen because its monotonously decreasing magnitude, which is flat in passband, does not distort the original signal [127, 128]. The moving window size for episode search was set to 2 h, and its step size was set to 10 min The minimal frequency ratio for anomalous episodes was set to 5 %. Events were obtained using the following parameters. Transitions length constraints (w_1 and w_2 in Equations (9.3) to (9.5)) were set to 15 min and 120 min. Transitions size constraints (h_1 and h_2 in Equations (9.3) to (9.5)) were set to 20 ppm and 30000 ppm.

Table 9.3 shows the most frequent episodes in database for ventilation unit 1. Some episodes represent obvious qualitative behaviour of the ventilation system. E.g. ventilation is turned on for a while, then it is turned off, and CO_2 level decreases as a result (episode 10).

Figures 9.6 to 9.9 show the weekly moving average of anomalous episodes for VAV units in rooms served by ventilation units 1 to 4. The rooms with the largest moving average are plotted separately in the first plots. Table 9.4 summarizes the results of the two experiments: the same rooms were found anomalous whether they were grouped by ventilation unit, or by room type.

Figure 9.6 shows the results for the VAV units in rooms served by ventilation unit 1. The first two rooms, shown separately in the upper plots, have significantly more frequent anomalous episodes, i.e.

Table 9.3: Most frequent episodes in database for ventilation unit 1.

Rank	Episode	Count
1	Ventilation ↘, CO_2 ↘	257
2	Ventilation ↗, CO_2 ↘	252
3	Ventilation ↘, CO_2 ↘, CO_2 ↗	225
4	CO_2 ↗, Ventilation ↘	223
5	CO_2 ↗, Ventilation ↘, CO_2 ↘	222
6	Ventilation ↗, CO_2 ↗	219
7	CO_2 ↘, Ventilation ↗, CO_2 ↗	210
8	CO_2 ↘, Ventilation ↗	203
9	Ventilation ↗, CO_2 ↗, Ventilation ↘, CO_2 ↘	169
10	Ventilation ↗, Ventilation ↘, CO_2 ↘	160

Table 9.4: Rooms flagged as anomalous in both experiments conducted.

	Ventilation unit			
	Unit 1	Unit 2	Unit 3	Unit 4
Room type				
Classroom	Ø22-601b-0	Ø20-601b-0, Ø20-601b-2		Ø20-511-1, Ø20-511-2
Auditorium	Ø22-601b-1		Ø22-511-1	
Conference room		Ø21-606-1		
Office				
Study area				
Corridor				

episodes in these rooms differs more frequently from the episodes commons to all other rooms. Their moving average goes above 15 or it is often above 10, while for all other rooms it is consistently lower, i.e. they behave more similarly among each other.

Figure 9.7 shows the results for the VAV units in rooms served by ventilation unit 2. The moving average for the first two rooms is very large, it goes over 15 and it is often over 10. While never reaching 15, the moving average for the third room is also sometimes larger than 10, while all other rooms have lower values.

Figure 9.8 shows the results for the VAV units in rooms served by ventilation unit 3. In this case, only one room has moving average larger than 10, while all the others have lower values.

Figure 9.9 shows the results for the VAV units in rooms served by ventilation unit 4. The moving average for the first two rooms is very large, it goes over 15 and it is often over 10. All other rooms have lower values.

Figure 9.10 shows the results for the VAV units in classrooms. The five rooms which have the moving average over 10 are also the same that were found anomalous in the first experiment, when rooms were grouped by ventilation unit.

Table 9.4 summarizes the results of the two experiments. The same rooms were found anomalous whether they were grouped by ventilation unit or by room type, i.e. those rooms had a different behaviour compared to other rooms served by the same ventilation unit, and compared to other rooms of the same type. The anomalous rooms are 5 classrooms, two auditoriums and one conference room. Since the building contains only one auditorium and 4 conference rooms, the episode database, when grouping by room type, would contain episodes only for small amount of rooms. Therefore, the sample size is too small to conclude that the rooms have actually anomalous behaviour. Classrooms, however, are numerous in the building, and both experiments independently flagged the same rooms as anomalous.

Figure 9.11 shows the values of CO_2 level and VAV position for room Ø21-606-1 on a day without anomalous episodes. Ventilation in the room follows CO_2 level as expected. The VAV unit opens by, respectively, 45 ppm, 70 ppm and 100 ppm when CO_2 level rises above

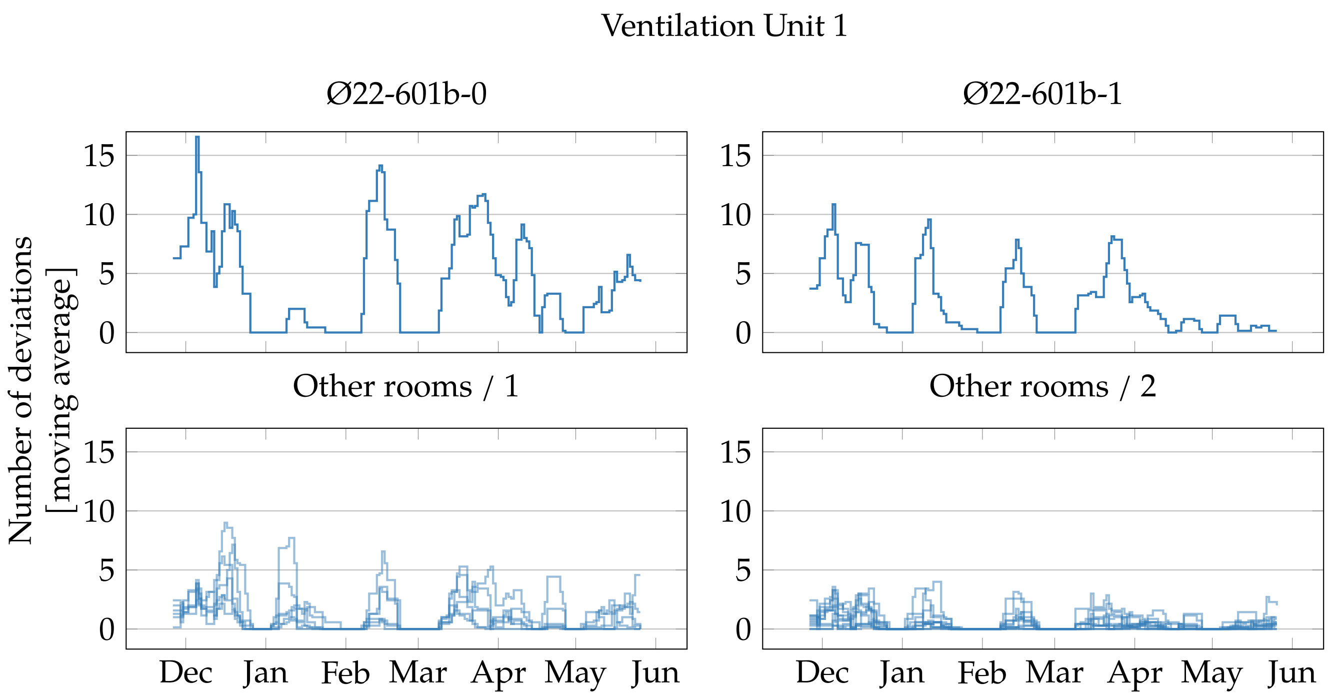

Figure 9.6: Moving average of deviations for ventilation unit 1. In the first two plots, the rooms which frequently deviate from the common behaviour are shown separately. The rest of the rooms are shown together in the last two plots.

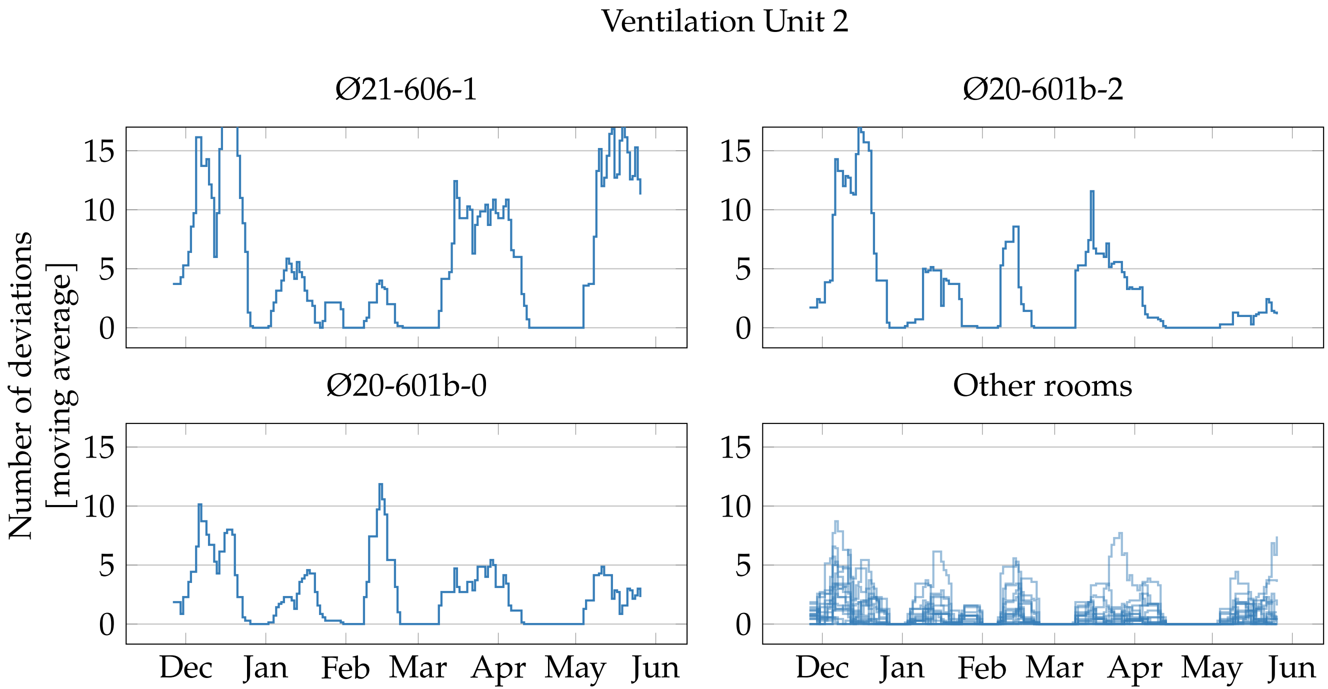

Figure 9.7: Moving average of deviations for ventilation unit 2. In the first three plots, the rooms which frequently deviate from the common behaviour are shown separately. The rest of the rooms are shown together in the last plot.

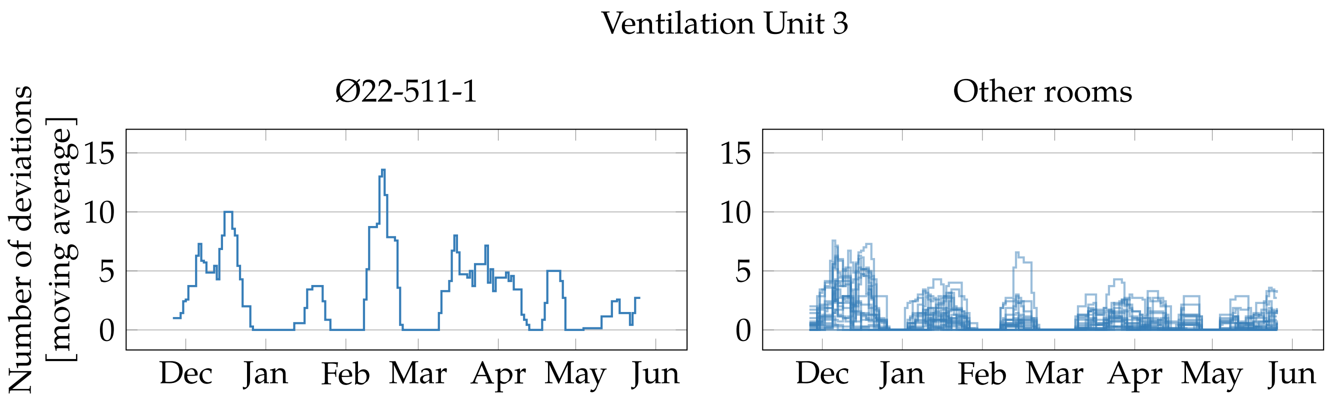

Figure 9.8: Moving average of deviations for ventilation unit 3. In the first plot, the room which frequently deviate from the common behaviour is shown separately. The rest of the rooms are shown together in the second plot.

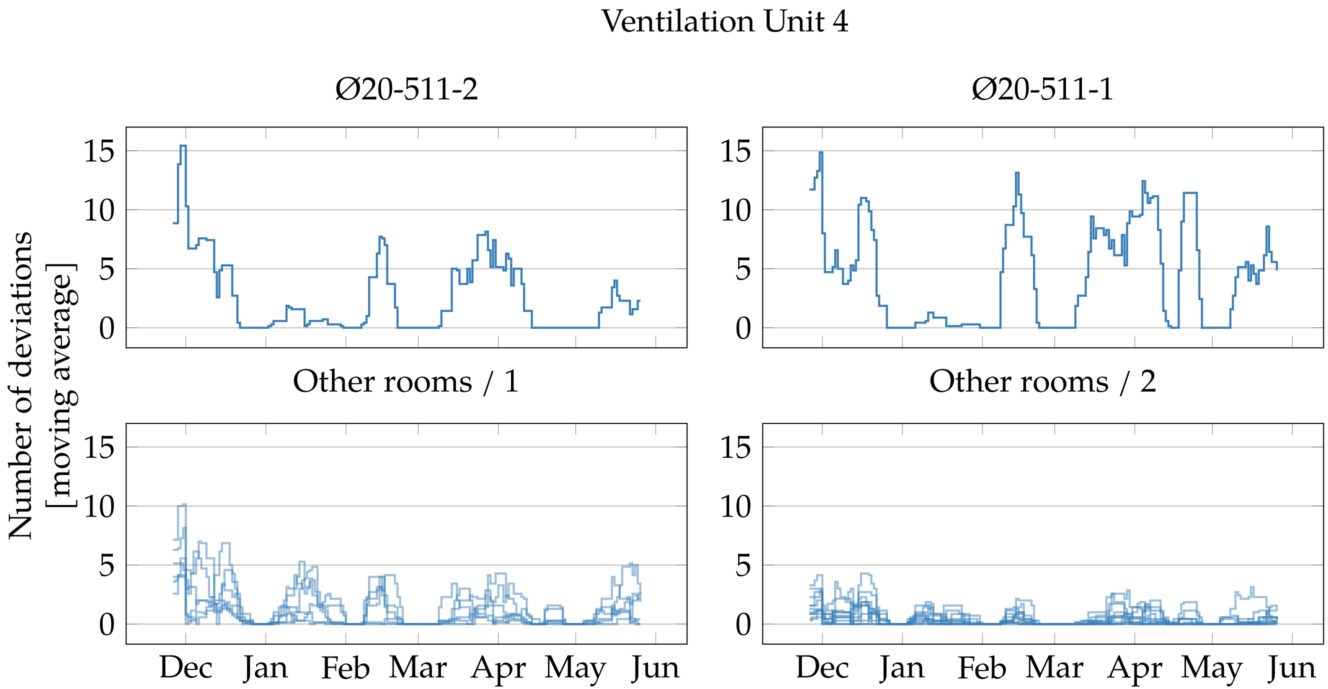

Figure 9.9: Moving average of deviations for ventilation unit 4. In the first two plots, the rooms which frequently deviate from the common behaviour are shown separately. The rest of the rooms are shown together in the last two plots.

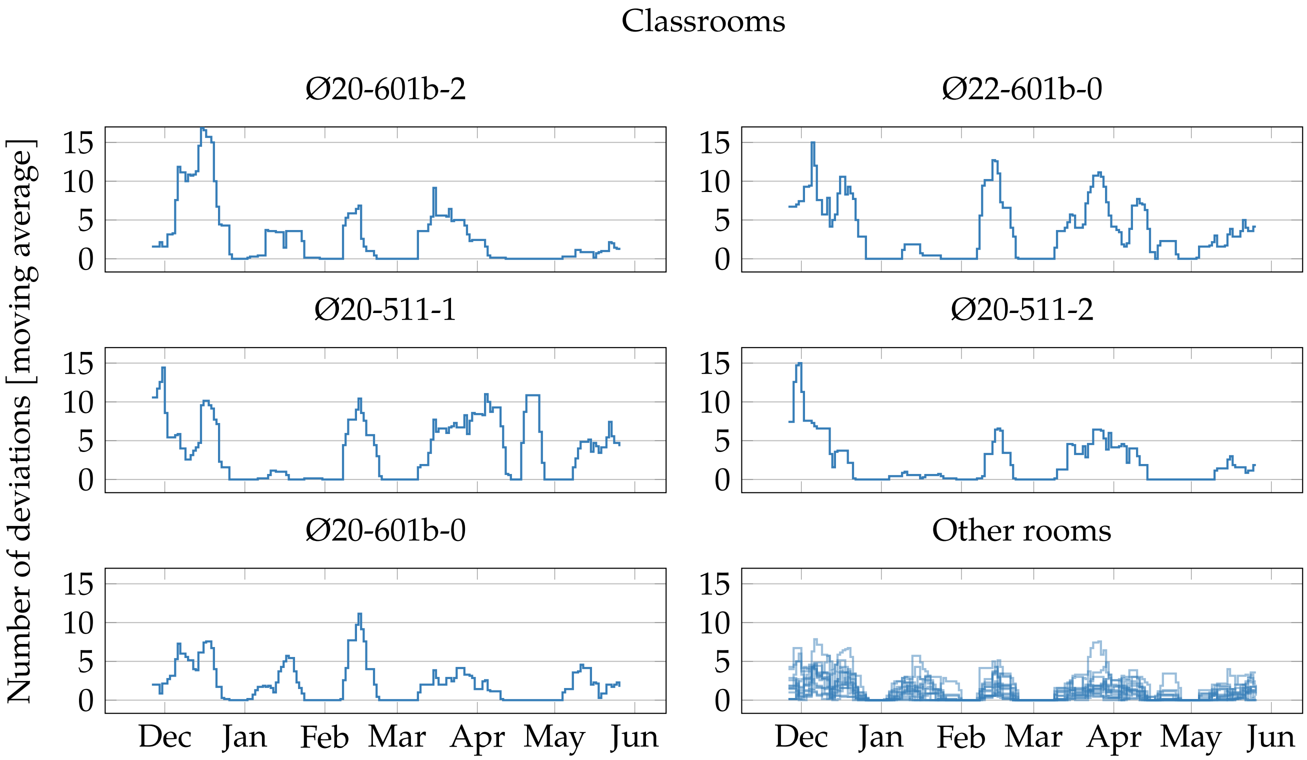

Figure 9.10: Moving average of deviations for classrooms. In the first five plots, the rooms which frequently deviate from the common behaviour are shown separately. The rest of the rooms are shown together in the last plot.

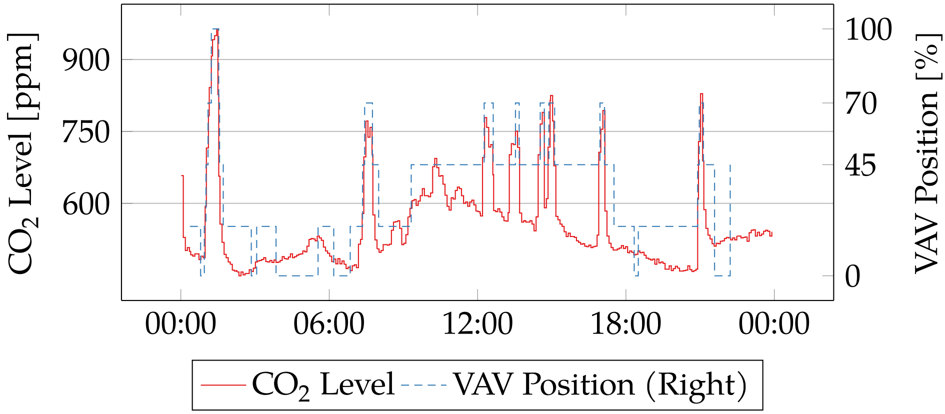

Figure 9.11: Correct behaviour of room ventilation. The VAV unit opens at the expected levels when CO_2 level rises over the thresholds.

600 ppm, 750 ppm and 900 ppm. The VAV unit closes with some delay after the CO_2 level drops below the thresholds, due to hysteresis of 100 ppm.

Figure 9.12 shows the values of CO_2 and VAV position for the same room on an anomalous day. CO_2 level is low during most of the day, and it only rises few times above the first thresholds of 600 ppm. The VAV unit, however, always opens completely. High room temperature could cause ventilation to increase to provide natural cooling through outdoor air. However, room temperature never exceeds 24 °C during the day.

Finally, the moving average of deviations from common behaviour has an irregular trend. Some rooms, however, have a higher deviation at the beginning of the experiment, and, later, align themselves more to the other rooms. E.g. room Ø20-511-2, when clustering by ventilation unit and room type (Figures 9.9 and 9.10), or room Ø20-601b-2, when clustering by ventilation unit and room type (Figures 9.7 and 9.10). This might suggest that, during the first few weeks, the episodes database was not yet fully populated, and deviations during that period should be ignored.

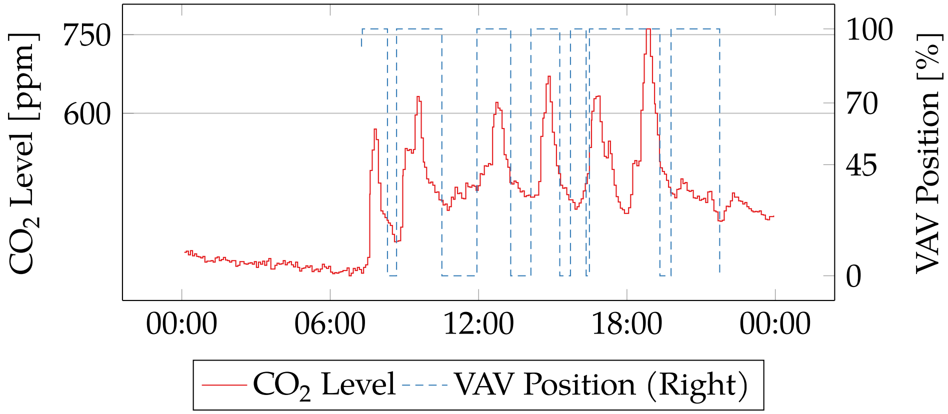

Figure 9.12: Anomalous behaviour of room ventilation. CO_2 level is below the threshold during the day, but the VAV unit is often completely open. Room temperature is steadily below 24 °C during the day, therefore, it cannot cause ventilation to increase.

9.5 CONCLUSIONS

In this paper, we presented a data-driven method for anomaly detection for VAV units based on consensus among several peers. A database of episodes is created from historical data and used to compute the frequency of new episodes. Compared to the majority of data-driven methods in the literature, the method does not need fault-free training data, instead, it relies on a large number of identical or similar systems. The effect of faulty systems during training is diluted over the entire dataset and, therefore, has a small impact on the generated model.

We applied the proposed method to detect anomalous VAV units of an existing building using CO_2 level and VAV position. Each room in the building contains a VAV unit, and all units are identical. We designed two experiments to investigate the behaviour of VAV units. At first, we grouped rooms by ventilation unit. Rooms served by the same ventilation unit are assumed to have the same behaviour, however, this assumption might not hold if their shape and usage are significantly different, and it is possible that they are incorrectly flagged as anomalous. Therefore, we ruled out this possibility by running a second experiment where we grouped the rooms by room type. The two experiments identified the same anomalous rooms, which suggests that their behaviour was, indeed, anomalous.

Some BMSs provide basic FDD capabilities, most often based on simple thresholds-based tests. Some faults at room level can be detected with these tests, e.g. a VAV unit stuck closed will eventually cause CO_2 level to rise above the threshold, however, they are not able to model complex dynamics. Episodes, on the other hand, can model interactions between different measurements, such as CO_2 level and VAV units positions, and, by using consensus, the proposed method can assess whether such interactions are similar to ones observed in their peers.

Consensus-based FDD methods are rarely applied in building systems. The proposed method is used to detect anomalies among interaction between VAV units and CO_2 level in the room. This approach shows the usefulness of using consensus between multiple similar systems to remove the need for fault-free historical data. Additional work would be necessary to decide whether anomalies are due to faults, misconfiguration or other causes, and, furthermore, to precisely diagnose such faults. The proposed method exposes several parameters, such as factor r, windows sizes, and thresholds for anomaly detection. In the experiments they were manually tuned to obtain a reasonable set of episodes, however, for a systematic application a method for self-tuning those parameters should be investigated.

The proposed method relies on the availability of many identical or similar components, and it can also be applied to other systems in buildings, such as heating, by monitoring episodes between radiators and room temperature, or lighting, by monitoring episodes between lights switches and illuminance sensors. More than two time-series can be used for phenomena that influence each other, such as CO_2 level, ventilation, temperature and heating, in order to generate more complex episodes.

Finally, the method presented in this paper was designed in the context of a complete framework for FDD and energy performance monitoring in building systems, aiming at developing a continuous monitoring application [8]. In our previous work, we addressed issues at different levels in building systems. Validation of sensors data through a basic set of rules and tests allows us to trust the status of the building, which is the basis for every advanced method using building's data, and which is not always validated after construction [1]. By monitoring the whole building energy performance with a dynamic energy model, we can assess whether the building respects national regulations and attains its design goals, or if it suffers from unjustified increased energy consumption, and at which level [2]. When one of the building systems does not perform as expected, we can analyse its individual components to detect anomalous behaviour or deviations from past trends [3, 4]. The method presented in this paper fills another area by isolating anomalous systems among multiple peers and, therefore, is another step towards a comprehensive FDD and energy performance monitoring framework for buildings.

ACKNOWLEDGEMENT

This work is supported by the Innovation Fund Denmark for the project COORDICY.

REFERENCES

[1] **Claudio Giovanni Mattera**, Sanja Lazarova-Molnar, Hamid Reza Shaker and Bo Nørregaard Jørgensen. 'A Practical Approach to Validation of Buildings' Sensor Data: a Commissioning Experience Report'. In: *Third International Conference on Big Data Computing Service and Applications* (*BigDataService*) (San Francisco, CA, USA, 6th–9th Apr. 2017). IEEE. 12th June 2017, pp. 287–292. DOI: 10.1109/BigDataService.2017.48.

[2] **Claudio Giovanni Mattera**, Muhyiddine Jradi and Hamid Reza Shaker. 'Online Energy Simulator for Building Fault Detection and Diagnostics Using Dynamic Energy Performance Model'. In: *International Journal of Low-Carbon Technologies* 13.3 (17th May 2018), pp. 231–239. ISSN: 1748-1325. DOI: 10.1093/ijlct/cty019.

[3] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'Fault Detection and Diagnostics in Ventilation Units Using Linear Regression Virtual Sensors'. In: *International Symposium on Advanced Electrical and Communication Technologies (ISAECT)* (Kenitra, Morocco, 21st–23rd Nov. 2018). IEEE. 24th Jan. 2019. DOI: 10.1109/ISAECT.2018.8618755.

[4] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'A Method for Fault Detection and Diagnostics in Ventilation Units Using Virtual Sensors'. In: *Sensors* 18.11 (14th Nov. 2018). ISSN: 1424-8220. DOI: 10.3390/s18113931.

[5] **Claudio Giovanni Mattera**, Hamid Reza Shaker and Muhyiddine Jradi. 'Consensus-based Method for Anomaly Detection in VAV Units'. In: *Energies* 12.3 (1st Feb. 2019). ISSN: 1996-1073. DOI: 10.3390/en12030468.

[8] Muhyiddine Jradi, Krzysztof Arendt, Fisayo Caleb Sangogboye, **Claudio Giovanni Mattera**, Elena Markoska, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'ObepME: An Online Building Energy Performance Monitoring and Evaluation Tool to Reduce Energy Performance Gaps'. In: *Energy and Buildings* 166 (1st May 2018), pp. 196–209. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2018.02.005.

[12] European Parliament and Council of the European Union. 'Directive 2010/31/EU of the European Parliament and of the Council on the Energy Performance of Buildings'. In: *Official Journal of the European Union* L153 (18th June 2010), pp. 13–35. EU DIRECTIVE: 2010/31.

[16] *2011 Buildings Energy Data Book*. Tech. rep. U. S. A. Department of Energy, Mar. 2012. URL: https://openei.org/doe-opendata/dataset/buildings-energy-data-book (visited on 20th Dec. 2018).

[21] Commission of the European Communities. *20 20 by 2020 - Europe's climate change opportunity*. 23rd Jan. 2008. EU DOCUMENT: 52008DC0030.

[22] European Commission. *Proposal for a Directive of the European Parliament and of the Council Amending Directive 2012/27/EU on Energy Efficiency*. 30th Nov. 2016. EU DOCUMENT: 52016PC0761.

[26] European Parliament and Council of the European Union. 'Directive (EU) 2018/844 of the European Parliament and of the Council of 30 May 2018 Amending Directive 2010/31/EU on the Energy Performance of Buildings and Directive 2012/27/EU on Energy Efficiency'. In: *Official Journal of the European Union* L156 (19th June 2018), pp. 75–91. EU DIRECTIVE: 2018/844.

[36] Evan Mills. 'Building Commissioning: A Golden Opportunity for Reducing Energy Costs and Greenhouse Gas Emissions in the United States'. In: *Energy Efficiency* 4.2 (May 2011), pp. 145–173. ISSN: 1570-6478. DOI: 10.1007/s12053-011-9116-8.

[37] Janghyun Kim, Jie Cai and James E. Braun. *Common Faults and Their Prioritization in Small Commercial Buildings: February 2017 – December 2017*. Tech. rep. NREL/SR-5500-70136. Golden, CO, USA: U. S. A. Department of Energy - National Renewable Energy Laboratory, June 2018. URL: https://www.nrel.gov/docs/fy18osti/70136.pdf (visited on 26th Dec. 2018).

[41] Woohyun Kim and Srinivas Katipamula. 'A Review of Fault Detection and Diagnostics Methods for Building Systems'. In: *Science and Technology for the Built Environment* 24.1 (8th May 2018), pp. 3–21. ISSN: 2374-4731. DOI: 10.1080/23744731.2017.1318008.

[117] Balakrishnan Narayanaswamy, Bharathan Balaji, Rajesh Gupta and Yuvraj Agarwal. 'Data Driven Investigation of Faults in HVAC Systems with Model, Cluster and Compare (MCC)'. In: *1st Conference on Embedded Systems for Energy-Efficient Buildings* (Memphis, TN, USA, 5th–6th Nov. 2014). Vol. 1. BuildSys. ACM. New York, NY, USA, 2014, pp. 50–59. DOI: 10.1145/2674061.2674067.

[118] Carlos Fernández Bandera and Germán Ramos Ruiz. 'Towards a New Generation of Building Envelope Calibration'. In: *Energies* 10.12 (2017). ISSN: 1996-1073. DOI: 10.3390/en10122102.

[119] Mitchell Yuwono, Ying Guo, Josh Wall, Jiaming Li, Sam West, Glenn Platt and Steven W. Su. 'Unsupervised Feature Selection Using Swarm Intelligence and Consensus Clustering for Automatic Fault Detection and Diagnosis in Heating Ventilation and Air Conditioning Systems'. In: *Applied Soft Computing* 34 (2015), pp. 402–425. ISSN: 1568-4946. DOI: 10.1016/j.asoc.2015.05.030.

[120] S. Byttner, T. Rögnvaldsson and M. Svensson. 'Consensus Self-organized Models for Fault Detection (COSMO)'. In: *Engineering Applications of Artificial Intelligence* 24.5 (2011), pp. 833–839. ISSN: 0952-1976. DOI: 10.1016/j.engappai.2011.03.002.

[121] G. Bianchin, A. Cenedese, M. Luvisotto and G. Michieletto. 'Distributed Fault Detection in Sensor Networks Via Clustering and Consensus'. In: *54th Conference on Decision and Control (CDC)* (Osaka, Japan, 15th–18th Dec. 2015). IEEE. 11th Feb. 2016, pp. 3828–3833. DOI: 10.1109/CDC.2015.7402814.

[122] Mohammadreza Davoodi, Nader Meskin and Khashayar Khorasani. 'Simultaneous Fault Detection and Consensus Control Design for a Network of Multi-agent Systems'. In: *Automatica* 66 (2016), pp. 185–194. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2015.12.027.

[123] Bo Zhou, Wei Wang and Hao Ye. 'Cooperative Control for Consensus of Multi-agent Systems with Actuator Faults'. In: *Computers & Electrical Engineering* 40.7 (2014), pp. 2154–2166.

ISSN: 0045-7906. DOI: 10.1016/j.compeleceng.2014.04.015.

[124] Sang-Ji Lee, Jin-Young Choi, Hyung-Joo Lee and Dong-Jun Won. 'Distributed Coordination Control Strategy for a Multi-Microgrid Based on a Consensus Algorithm'. In: *Energies* 10.7 (2017). ISSN: 1996-1073. DOI: 10.3390/en10071017.

[125] Nian Liu, Bin Guo, Zifa Liu and Yongli Wang. 'Distributed Energy Sharing for PVT-HP Prosumers in Community Energy Internet: A Consensus Approach'. In: *Energies* 11.7 (2018). ISSN: 1996-1073. DOI: 10.3390/en11071891.

[126] Francisco Javier Martínez de Pisón, Andrés Sanz, Eduardo Martínez de Pisón, Emilio Jiménez and Dante Conti. 'Mining Association Rules from Time Series to Explain Failures in a Hot-dip Galvanizing Steel Line'. In: *Computers & Industrial Engineering* 63.1 (Aug. 2012), pp. 22–36. ISSN: 0360-8352. DOI: 10.1016/j.cie.2012.01.013.

[127] Stephen Butterworth. '"On the Theory of Filter Amplifiers'. In: *Wireless Engineer* 7 (1930), pp. 536–541. URL: http://www.gonascent.com/papers/butter.pdf.

[128] Anatol I. Zverev. *Handbook of Filter Synthesis*. New York, NY, USA: Wiley, 1967. 602 pp. ISBN: 978-0471749424.

CHAPTER 10

EVALUATING FDD METHODS AND ASSESSING THEIR IMPACT WITH FAULTS SIMULATION

This chapter is a cosmetic adaptation of the following paper, under review at the journal *Sustainable Cities and Society*.

Claudio Giovanni Mattera, Hamid Reza Shaker, Muhyiddine Jradi, Mathis Riber Skydt and Sebastian Skals Engelsgaard. 'Fault Detection in Ventilation Units using Dynamic Energy Performance Models'. In: *Sustainable Cities and Society* (2019). ISSN: 2210-6707. **Submitted**

ABSTRACT

Buildings are one of the world's largest energy consumer. Building systems are often affected by faults which cause energy waste and occupants discomfort.

In this paper, we propose a model-based method for fault detection and diagnostics of ventilation units supported by a set of rules. At first, a dynamic energy model of the building is used to obtain the expected energy consumption for each subsystem, and under-performing subsystems are isolated. Afterwards, a set of rules is used to precisely diagnose the faulty component.

The method is tested on the EnergyPlus model of an existing building, where faults are simulated by modifying its parameters. Two types of faults are considered: abrupt faults, where the parameters change abruptly, and gradual faults, where parameters change gradually over time. Different components inside the ventilation unit are

considered, such as heat exchangers, hot-water heating loops, and fans.

Most faults are correctly identified by the proposed method, while others did not result in significant increase in consumption. Moreover, an assessment of the impact of the considered faults is reported. Faults caused up to 8% increase in the energy consumption for the ventilation unit. Finally, detection of gradual faults has promising applications in faults prediction.

10.1 INTRODUCTION

Buildings are one of the world's largest energy consumer. In Europe, they account for 40 % of the total consumed energy and a large part of CO_2 emissions [12]. In the United States of America, they are responsible for over 41 %, more than industry and transportation, and energy consumption has been steadily increasing, doubling from 1290 TW h in 1980 to 2784 TW h in 2010 [16]. Thus, buildings are an important actor with respect to the long-term environmental goals such as Europe 20 20 [21] and its proposed extensions [22], or doubling the U. S. A. energy productivity by 2050 [31].

Modern buildings contain several subsystems, such as lighting, heating, ventilation and air conditioning (HVAC), sensing networks and booking systems, usually coordinated by a central building management system (BMS). Each of these systems can have considerable complexity and, therefore, are more vulnerable to faults. Faults can cause reduced occupants comfort, e.g. a broken thermostat would result in too high or too low temperature in a room, but also increased energy consumption, e.g. simultaneous heating and cooling. Energy waste due to faults can be significant. In 2009, 13 of the most common faults in the U. S. A. were estimated to cause over 3.3 billion $ [36], and this number was supported by other recent reports [37].

Recently, fault detection and diagnostics (FDD) techniques have been studied and applied in the context of building systems. Such techniques have been successfully developed and deployed for decades in several fields, such as avionics and process control. FDD aims for quick and complete identification of faults present in building systems, in order to minimize occupants discomfort and energy waste.

In this paper, we propose a model-based method for FDD in building HVAC systems, supported by a set of rules. Variations in consumed energy are monitored to isolate anomalous components, and a set of rules is further used to validate their operations. We use the dynamic energy model of a real building to simulate faults and test our method, and we report the results. We consider faults affecting heaters, heat exchangers (HXs) and fans.

The work was carried out within project COORDICY, a strategic Danish-American interdisciplinary research project for advancing buildings intelligence.

The rest of the paper is organized as it follows. In Section 10.2, a general review of the current state of the art is presented. The method and the evaluation methodology are presented in Section 10.3. A case study is introduced in Section 10.4, and specific experiments are defined in Section 10.5. Results from the experiments are reported in Section 10.6. Finally, conclusions are drawn in Section 10.7.

10.2 STATE OF THE ART

A comprehensive review of FDD methods for building systems was presented by Kim et al. in 2018 [41]. The authors categorize FDD methods in three large groups: history-based methods, quantitative model-based methods and qualitative model-based methods.

In history-based methods, a black box model of the system under test is created from historical data and used to validate and predict the behaviour of the system itself. Several techniques can be used for the model, such as, linear and non-linear regression, principal component analysis (PCA), artificial neural network (ANN), support vector machine (SVM) and statistical methods. Methods in this category usually consist of two separate phases: an offline training phase, and an online testing phase.

During the training phase, historical data is fed to the model, so that its hidden parameters are estimated. This phase usually occurs only once, or very rarely, only when the system significantly changes structure, and can take considerably long time. The result is a model that predicts or validates the system's behaviour.

During the testing phase, real-time data from the system is fed as input to the model, and its outputs are compared with actual measurements from the system. If no faults are present, the model is assumed to accurately characterize the system. If, on the other hand, the system is affected by a fault, its behaviour is assumed to be different. Therefore, faults are detected when the model does not follow the proper operation of the system.

The typical approach for diagnosing faults is to train multiple models, using historical data from systems affected by each fault. During the testing phase, the system's behaviour is compared with each of the resulting models, and the matching one corresponds to the specific fault, or no fault. If no model matches the current behaviour, an unknown fault might be present.

History-based method do not use any specific information about the system under test, the model is constructed entirely from data. For this reason, such methods are easily portable to different systems, as long as training data is available.

The main disadvantage of history-based methods is the requirement of fault-free historical data. A model trained with unknown faulty historical data does not represent the correct behaviour, and will wrongly flag faults as correct operation. This requirement makes it impossible to deploy history-based methods on newly installed systems, for which historical data has not been collected yet. Moreover, labelled faulty historical data, necessary to perform fault diagnostics, is very rarely available.

An example of history-based method can be found in [129]. The authors present a hybrid method based on PCA and pattern matching for FDD in air handling units (AHUs). Periods in historical data where conditions are similar to the current ones in the system are identified through pattern matching. If the system status is different from the one in those periods, an alarm is raised. The authors test their method on experimental data containing six different faults.

In quantitative model-based methods, a model of the system under test is constructed from first principles and, as in history-based methods, is used to validate and predict the behaviour of the system itself. The main difference from history-based methods is the way the model is constructed. While, in history-based methods, the model was trained using data, in quantitative model-based methods, it is manually constructed by field experts using detailed knowledge of the system. Quantitative models often consists of a set of differential equations modelling the physical relationships of different quantities in the system, and include informations about materials, geometry and specific components.

Quantitative models are usually much more accurate than black box models. They can also be used to detect and diagnose faults without requiring labelled faulty historical data, and sometimes they are even able to detect unknown faults.

Depending on the required accuracy, quantitative models can have large complexity, and it can take extensive work from several field experts to build a model. Complexity also affects the runtime of such models, while history-based methods can validate data in real-time, for quantitative models, this can take several minutes or even hours. Moreover, such models are tailored to a specific system, and they are difficult to adapt to different ones. Finally, sometimes models are too complex to explicitly provide values for all their parameters, and historical data is necessary for parameter estimation.

An example of a model-based method can be found in [130]. The authors present a hierarchical framework for FDD with respect to contaminants across ventilation systems. The first two layers are used to detect local changes in contaminants concentration, and to ensure that such changes have statistical significance. The top layer analyzes the building's layout and determines whether changes are due to actual contaminants or sensors faults, under the assumption that they would propagate through the building in different ways. The proposed method is tested on several artificially generated datasets.

In qualitative model-based methods, the behaviour of the system is encoded as a set of qualitative relationships. The most common approach is to build a set of rules describing the system's correct operation, and use them to validate the current behaviour.

The main advantage of rule-based methods is their simplicity. Rules are defined by field experts without requiring historical data or detailed knowledge of the system's physics. Set of rules are usually simple enough that can be understood and manually validated. Moreover, rules can often be obtained directly from system documentation without any manual intervention.

Set of rules, however, can only represent relatively simple system dynamics. For complex systems, a long list of narrow, non-overlapping rules is necessary. A long list of rules, moreover, is difficult to maintain and to validate. E.g. it becomes likely to add a rule which is inconsistent or contradictory with the rest.

Nevertheless, rule-based systems are effective in representing and validating high-level, expected behaviour of the system. E.g. a simple rule stating that empty rooms have no lighting can detect a broken occupancy sensor. Another example could be a rule stating that temperature must be within setpoints. In case the rule is not satisfied, a fault in the heating and cooling system is present. In fact, many commercial BMSs offer rudimentary FDD using simple alarms.

A recent example of rule-based method can be found in [131]. The authors define a set of rules for detecting simultaneous heating and cooling in a building. Two decision trees are defined, one for winter and one for summer, due to different dynamics and behaviours in the two seasons. The method is tested on a dataset from a real academic building. Moreover, the authors assess the impact of simultaneous heating and cooling over a single month in 400 MW h, which corresponds to a financial loss of more than 24000 $.

10.2.1 Faults simulation

According to Shi et al., modern FDD methods, while providing useful information about existing faults to building maintenance, often lack an assessment of such faults' impact on the overall building energy performance [132]. The authors propose a methodology for faults impact evaluation based on comparing faulty conditions with a baseline, obtained from healthy conditions. They identify three challenges:

identifying symptoms caused by faults, quantifying the severity of such symptoms, and mapping them to simulation inputs and outputs. They test their methodology on a model of a real building, simulating four different faults: low hot-water supply temperature, low AHU supply fan efficiency, increased infiltration rate due to window leakage and stuck closed reheat valve inside a variable air volume (VAV) terminal.

A study using EnergyPlus to simulate HVAC faults was presented in [133]. Four faults were considered: clogging of pipes in the plant loop, fouling of water heating coils, leaking outside air economizer dampers and zone temperature sensor offset. In the first case, a model of the system curve in both healthy and faulty conditions is developed. The authors consider different approaches for faults simulation. EnergyPlus supports an internal scripting engine, however, it is complex and requires reimplementing control logic. Reusable fault model objects, on the other hand, remove that requirement. Finally, the authors test their method on a model of a real building, assessing the energy impact of the four faults.

The custom implementation for faults simulation in EnergyPlus was presented in details in [134], applied to HVAC systems. The authors argue that most of the current research focus on impact at component or subsystem level, and do not assess impact at the whole building level, and, therefore, propose to use EnergyPlus as a whole building simulation engine. The authors define three approaches for faults simulation in EnergyPlus.

- Modifying parameters directly in the model file. This is the simplest approach, however, only a limited number of parameters are exposed, and it cannot be used for simulating faults which result in complex dynamics.
- Using the internal scripting engine. This allows, but also requires, to define complex interactions between the individual components, and it requires extensive knowledge of the system.
- Using native fault objects. With this approach, users can define complex dynamics within the system, however, internal interactions are handled by the engine itself. Moreover, faults objects

> are often generic and reusable across different buildings. The main drawback is that only a limited number of fault objects are available.

The authors consider four faults objects implemented in EnergyPlus: sensor faults with air economizers, thermostat/humidistat offset, heating and cooling coil fouling, and dirty air filters. The impact on the whole building energy consumption of these faults is assessed on the model of an ideal building.

Other engines has been successfully used for faults simulation. A model of a room-level ventilation system was developed using Simulink by Behravan et al., which was used to perform FDD using acyclic graphs [135]. Modelica simulation software was used to extend the building information model (BIM) by Andriamamonjy et al., who also presented a FDD framework for AHUs [57].

10.2.2 CONTRIBUTIONS

FDD for HVAC systems is a popular research field, however, most publications focus on components such as chillers, refrigerators and heating coils, while heat exchangers are often ignored. Moreover, while faults simulation has been successfully used in literature for assessing impact of faults, the impact is rarely reported at whole building level.

The main contribution of this paper is, therefore, two-fold. At first, we present a model-based method for FDD on HVAC systems, supported by a set of rules. We use a detailed whole building dynamic energy model, calibrated using actual data, to isolate under-performing subsystems, and a set of rules to verify the correct operation of individual components. In particular, we consider components such as heat exchangers, hot-water heating loops and fans, and we modify parameters in the model to simulate different faults. Afterwards, we present an assessment of the impact of selected faults on the energy consumption of a real case-study building.

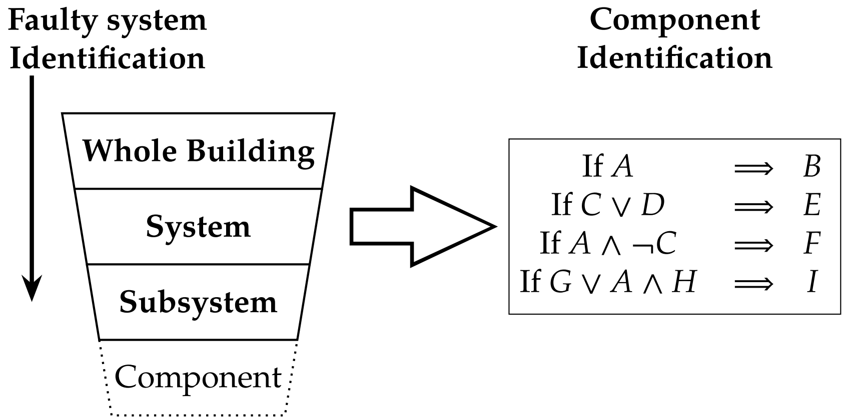

Figure 10.1: Overview.

10.3 METHOD

In this paper, we propose a hybrid quantitative model-based method for FDD in ventilation units, supported by a set of rules. The overview, shown in Figure 10.1, is the following. A simulation of the building in healthy conditions is performed to obtain the expected energy consumption. Under-performing subsystems are identified by monitoring the building's energy distribution tree, as presented in our previous work [2]. Afterwards, a set of rules is defined to identify specific faults.

10.3.1 EVALUATION

In order to test and validate a FDD method, it must be deployed on a known faulty system. We simulate the operation of a real building using its dynamic energy model, developed using the EnergyPlus simulation engine [52].

The models exposes several parameters corresponding to characteristics of many buildings equipment. E.g. operational schedules, building materials, airflow constraints, components efficiency and operational temperature, CO_2 and pressure setpoints. Normally, values for these parameters are selected to match the real building, either manually or using some parameter estimation method. However, they can also be modified to represent faulty components. E.g. the air infiltration rate can be modified in order to simulate an increased thermal loss caused by a leak. Airflow constraints can be changed to simulate oversized or undersized equipment. Components efficiency can be decreased to simulate wearing and damages.

In this paper, we design a set of experiments to evaluate the presented FDD method. In each experiment, we select a fault and we simulate it by modifying the corresponding parameters. The output from the simulation represent the behaviour of the entire building when affected by the fault, not only the single component. Finally, we apply the presented method to the resulting data.

10.4 CASE STUDY: BUILDING OU44

In this paper, we introduce building Odense undervisning 44 (OU44) as case study. The building, located at the Campus Odense of the University of Southern Denmark, has been used as living lab for several studies and experiment. The building is mainly used for teaching and office work, and it contains about 200 rooms, divided among classrooms, offices, and other room types, layered in three floors. Four ventilation units provide air supply to the building, one for each corner (north-east, north-west, south-east, south-west).

In order to provide an accurate prediction of healthy behaviour as well as to estimate accurately the impact of different faults, a detailed whole building energy performance model of building OU44 is used [8]. The dynamic energy model, built in EnergyPlus, was calibrated used actual data from the building on the level of the individual energy supply systems and was found to predict well the building energy performance with acceptable uncertainty.

A diagram of a ventilation unit is shown in Figure 10.2. Inlet air enters from the bottom left, goes through a heat exchanger, which is used for preheating using exhaust air. Preheated air, if still colder than the desired setpoint, is further heated using a hot-water based heating loop, and pushed to the main shaft, from which it will reach

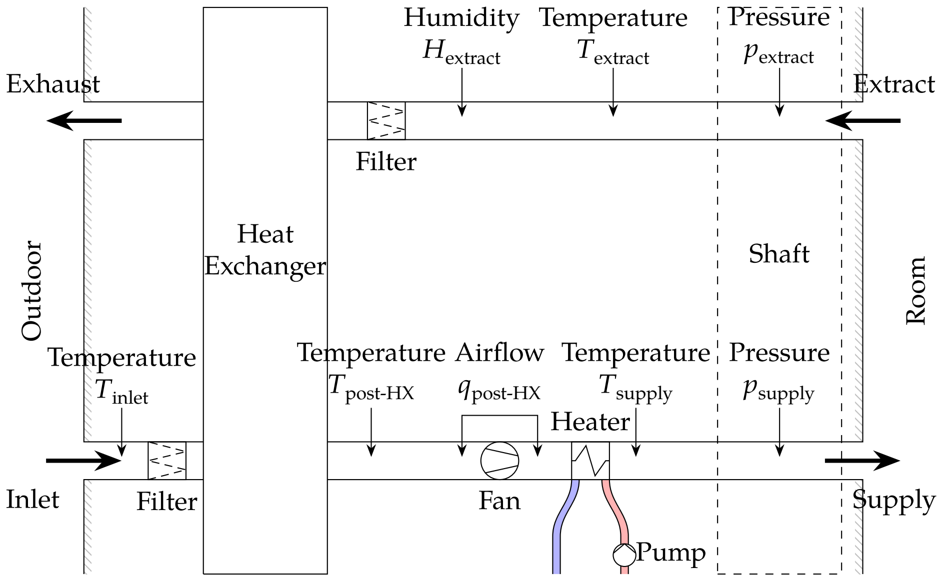

Figure 10.2: Ventilation unit.

individual rooms. On the way out, air exits the rooms through the extract shaft, goes through the heat exchanger, and it finally leaves the building. In the model, one fan is used for pushing air through the unit pipes. In the real building, one additional fan is located on the exhaust pipe, before air exists the building.

A ventilation unit contains several components, and each of them can be affected by faults. Few experiments were defined and performed, in order to evaluate the impact of faults on the final energy consumption, and also to test the FDD method previously defined.

10.5 EXPERIMENTS

In total, five experiments (A, B, C, D and E) were defined in order to conduct simulation of different faults in a ventilation unit. The experiments were divided in two groups: one for abrupt faults and one for gradual faults. In abrupt faults experiments, two simulations were performed over the same period, one with original parameters values, and one with faulty parameters values. In gradual faults experiments, one simulation was performed over the entire period using original parameters values, and several consecutive simulations were performed by gradually changing the parameters values.

In abrupt faults experiments, faults in air temperature loop—i.e. heat exchanger and heater—were simulated. Each experiment was duplicated in both summer and winter, in order to account for difference in heating between warm and cold periods. In gradual faults experiments, on the other hand, faults were simulated in the air supply systems. Since air supply does not have a strong dependence on outdoor temperature, only one experiment was performed. Moreover, gradual faults required a longer simulation period, which span several months during both winter and summer.

10.5.1 Abrupt faults in air temperature loop

The heat exchanger and the heater do not consume a significant amount of electrical energy, i.e. turning them on or off cannot be observed in the electricity meters, where they are dwarfed by the fans' electrical consumption. Instead, a faulty heat exchanger will result in increased heating, which could have a quantifiable effect on the heating meter. A faulty heater might also affect the heating meter, however, it should mainly affect the supply air temperature.

Two faults were considered in these experiments.

- Reduced efficiency of heat exchanger, experiments A and B.
- Reduced heater effect, experiment C.

Since heating depends strongly on outdoor temperature, every experiment is performed twice, once during winter and once summer.

10.5.2 Gradual faults in air supply systems

Fans are responsible for the largest share of electrical consumption in the ventilation unit.

Two faults were considered in these experiments.

- Reduced fan efficiency (gradual, due to wearing), experiment D.
- Obstruct ducts (gradual, due to, e.g. filters accumulating dust), experiment E.

In experiment E, there are two different outcomes.

- The requested airflow is low enough that the fan can achieve it despite the ducts being obstructed. This situation should be indistinguishable from reduced efficiency.
- The requested airflow is above the fan capacity (due to obstruction).

10.5.3 RULES

The following rules are defined to detect the faults introduced in the experiments.

If energy consumed for heating the air is higher than the expected value, either the post-HX temperature is lower than expected, too, or the heater is not operating correctly. This behaviour is encoded in Rule 10.1.

Rule 10.1: Heater energy rule.

if Heater energy > expected **then**
    if $T_{\text{post-HX}} \geq$ setpoint **then**
        Heater is faulty
    end if
end if

If post-HX temperature is lower than the expected setpoint, the heat exchanger is not recovering enough heat from extract air. The heat exchanger efficiency is lower than expected and, therefore, the component is faulty. This behaviour is encoded in Rule 10.2.

Regardless of action of heat exchanger and heating energy consumption, supply air temperature must be above setpoint. If this constraint is not satisfied, the heater is not operating correctly. This behaviour is encoded in Rule 10.3.

Rule 10.2: Post-HX air temperature rule.

if $T_{\text{post-HX}} <$ setpoint **then**
  Heat exchanger is faulty
end if

Rule 10.3: Supply air temperature rule.

if $T_{\text{supply}} <$ setpoint **then**
  Heater is faulty
end if

If fan energy is higher than expected, either the fan is wearing and its efficiency is decreasing, or ducts are obstructed. If, at the same time, provided airflow is lower than expected, ducts are obstructed. This behaviour is encoded in Rule 10.4.

Rule 10.4: Fan energy rule.

if Fan energy < expected **then**
  if Airflow < expected **then**
    Ducts are obstructed
  else
    Fan is faulty $\vee$ ducts are obstructed
  end if
end if

10.5.4 EXPERIMENTS PARAMETERS

EXPERIMENTS A AND B

In these experiments, the heat exchanger efficiency is reduced by 20 % and 50 % from its proper operation calibrated value, in order to simulate wearing. In both cases, two separate instances were simulated, one for warm months and one for cold months. In the former, the experiment runs from 1 May 2016 to 31 July 2016. In the latter, the experiment runs from 1 October 2016 to 31 December 2016. Table 10.1 shows the parameters' values in the EnergyPlus model.

EXPERIMENT C

In this experiment, hot-water temperature in the heating loop is reduced by 3 °C from its proper operation calibrated value, in order to simulate a reduced heater effect. This experiment was divided in two instances, one for warm months and one for cold months. In the former, the experiment runs from 1 May 2016 to 31 July 2016. In the latter, the experiment runs from 1 October 2016 to 31 December 2016. Table 10.2 shows the parameters' values in the EnergyPlus model.

EXPERIMENT D

In this experiment, fan efficiency is gradually reduced over time from its proper operation calibrated value, in order to simulate wearing. The experiment runs from 1 January 2016 to 31 July 2016. A fault is simulated on 1 March 2016, at which point the fan efficiency decreases from its initial value of 0.8 by 2.5 % every week. The relevant parameter in the EnergyPlus model is `Fan Total Efficiency`, and its values are shown in Table 10.3.

EXPERIMENT E

In this experiment, fan pressure rise is gradually increased over time from its proper operation calibrated value, in order to simulate obstructed ducts due to dust and other material. The experiment runs from 1 January 2016 to 31 July 2016. A fault is simulated on 1 March 2016, at which point the fan pressure rise increases from its initial value of 560 kPa by 2.5 % every week. The relevant parameter in the EnergyPlus model is `Pressure Rise`, and its values are shown in Table 10.4.

Table 10.1: Parameters for experiment A and B.

Variable	**Healthy**	**Faulty (A)**	**Faulty (B)**
`Sensible Effectiveness at 100% Heating Air Flow`	0.86	0.688	0.430
`Latent Effectiveness at 100% Heating Air Flow`	0.78	0.624	0.390
`Sensible Effectiveness at 75% Heating Air Flow`	0.91	0.728	0.455
`Latent Effectiveness at 75% Heating Air Flow`	0.83	0.664	0.415

Table 10.2: Parameters for experiment C.

Variable	Healthy	Faulty
`Rated Inlet Water Temperature`	23	20
`Rated Outlet Water Temperature`	21.5	18.5

Table 10.3: Values for parameter `Fan Total Efficiency` in experiment D. Efficiency decreases by 2.5 % every week, i.e. original efficiency is multiplied by 0.975^i, where i is the number of weeks from the fault. Week numbers are formatted according to *ISO8601:2000(E)* [136].

Week	Multiplier	Efficiency
2016-W01 to 2016-W10	1.000	0.800
2016-W11	0.975	0.780
2016-W12	0.951	0.761
2016-W13	0.927	0.741
2016-W14	0.904	0.723
2016-W15	0.881	0.705
2016-W16	0.859	0.687
2016-W17	0.838	0.670
2016-W18	0.817	0.653
2016-W19	0.796	0.637
2016-W20	0.776	0.621
2016-W21	0.757	0.606
2016-W22	0.738	0.590
2016-W23	0.720	0.576
2016-W24	0.702	0.561
2016-W25	0.684	0.547
2016-W26	0.667	0.534
2016-W27	0.650	0.520
2016-W28	0.634	0.507
2016-W29	0.618	0.495

Table 10.4: Values for parameter `Pressure Rise` in experiment E. Pressure rise increases by 2.5 % every week, i.e. original efficiency is multiplied by 1.025^i, where i is the number of weeks from the fault. Week numbers are formatted according to *ISO8601:2000*(*E*) [136].

Week	**Multiplier**	**Pressure Rise** [kPa]
2016-W01 to 2016-W10	1.000	560
2016-W11	1.025	574
2016-W12	1.051	588
2016-W13	1.077	603
2016-W14	1.104	618
2016-W15	1.131	634
2016-W16	1.160	649
2016-W17	1.189	666
2016-W18	1.218	682
2016-W19	1.249	699
2016-W20	1.280	717
2016-W21	1.312	735
2016-W22	1.345	753
2016-W23	1.379	772
2016-W24	1.413	791
2016-W25	1.448	811
2016-W26	1.485	831
2016-W27	1.522	852
2016-W28	1.560	873
2016-W29	1.599	895

10.6 RESULTS

In experiments A and B, the efficiency of the heat exchanger were decreased by 20 % and 50 %. The effect of heat exchanger is to heat up inlet air using extract air, in order to reduce the necessary energy from the heater. During warm months, no visible effect was present, neither in energy consumption, or in temperatures across the ventilation unit. During cold months, on the other hand, post-HX temperature was lower, and energy consumption increased.

The weekly energy consumption for heating is shown in Figure 10.3 for the two experiments, together with the one in the healthy case. In experiment A, the energy consumption increases over time compared with the healthy one, reaching about 2 % higher in the coldest weeks. In experiment B, on the other hand, the increase is steeper and faster. The energy consumption is over 4 % higher than expected already at the beginning of the cold season, and it reaches over 8 % during the coldest weeks.

The first chart in Figure 10.4 shows the daily average post-HX temperature for the two experiments against the one in the healthy case. In the healthy case, temperature is always above setpoint, i.e. the heat exchanger is able to preheat inlet air.

In both experiments A and B, on the other hand, the post-HX temperature is frequently lower than the setpoint, especially during the coldest weeks. In experiment A, where the efficiency was reduced by 20 %, post-HX temperature can fall 1 °C below the setpoint. In experiment B, where the efficiency was reduced by a more drastic 50 %, the temperature consistently falls over 2 °C below the setpoint.

In the second chart in Figure 10.4, the point-wise values of post-HX temperature are shown during one of the coldest weeks in the simulation period. Post-HX temperature falls below the setpoint at the beginning of the day, however, it also rises during the day. This is due to increase in outdoor temperature and indoor temperature, which, in turns, increases the effect of the heat exchanger.

While the daily average captures the overall trend, the point-wise plot shows that the temperature falls over 2 °C below the setpoint for long time during the day in experiment A. In experiment B, where the

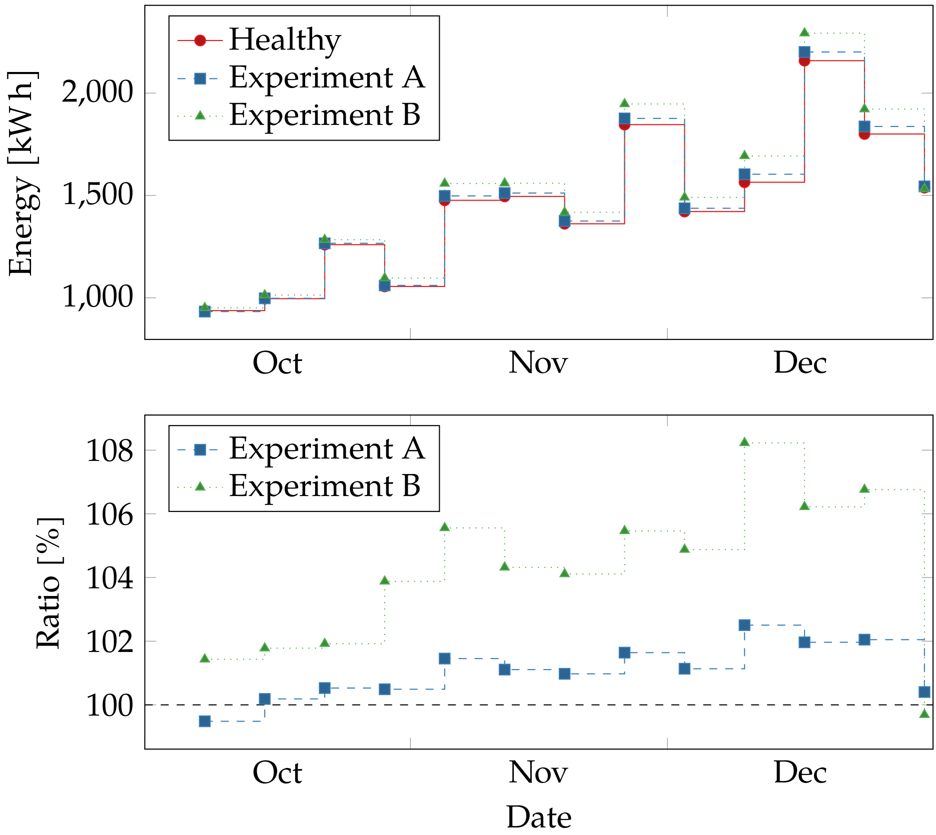

Figure 10.3: Results for experiments A and B during cold months. The first chart shows the weekly heating energy consumption, the second one shows the ratio compared with the healthy case. A small reduction in heat exchanger efficiency causes an increase in heating of around 2 % in the coldest periods. A larger reduction, on the other hand, causes an increase of over 8 %.

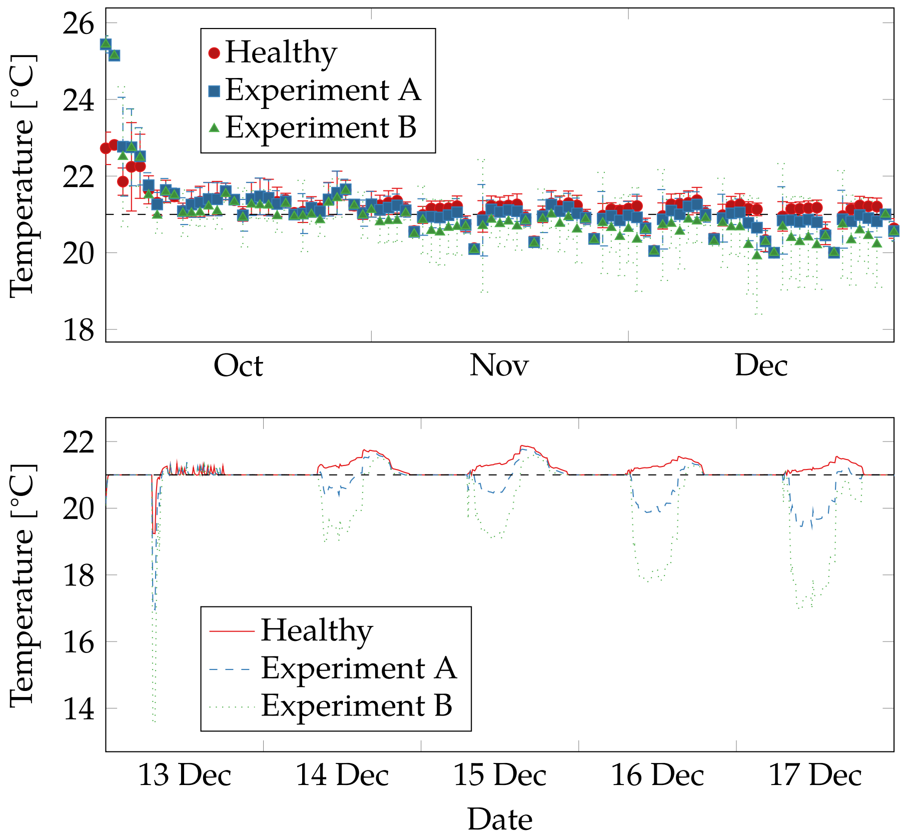

Figure 10.4: Post-HX temperature in experiments A and B. In the first chart, the daily average temperature is visibly lower than the setpoint in experiment A, and even more in experiment B. In the second chart, the temperature falls significantly below the setpoint at the beginning of the day, but often it recovers later.

efficiency was significantly reduced, the temperature falls over 4 °C below the setpoint. The low spike at the beginning of the week is due to the heating system being turned off during the weekend.

In experiment C, the inlet and outlet water temperatures in the hot-water loop were decreased by 3 °C. The heater is used to heat up air when the effect of the heat exchanger is not enough to reach the setpoint. No significant increase in energy consumption was visible, neither during warm months, or during cold months.

The charts in Figure 10.5 shows the daily average supply temperature for the experiment, against the one in the healthy case during cold months, and the point-wise temperature for one of the coldest weeks. Besides some irregular behaviour on Monday, there is no visible difference. Therefore, the decrease in hot-water loop temperature does not affect the heating system.

In experiment D, the efficiency of the fan in the ventilation unit was decreased by 2.5 % every week, starting on 1 March 2016. In experiment E, the pressure rise across the fan was similarly increased by 2.5 % every week. Figure 10.6 shows the energy consumed by the fan in the two experiments against the one in the healthy case. In the healthy case, the fan energy consumption is roughly constant over time. In both experiments, on the other hand, the consumed energy visibly increases over time after the fault happened.

Figure 10.7 shows the efficiency of the fan in the two experiments, computed as daily airflow divided by daily energy consumption, and aggregated by week, against the one in the healthy case. Airflows in experiments D and E coincide and, therefore, it was not possible to distinguish the case of reduced fan efficiency from the case of duct obstruction.

10.6.1 Impact of faults

In experiments A and B, the heating energy consumption increased by up to 2 % and 8 % compared with the one in healthy conditions. In the coldest weeks, heating energy consumption can reach 2000 kW h. Additional heating due to decreased heat exchanger efficiency would

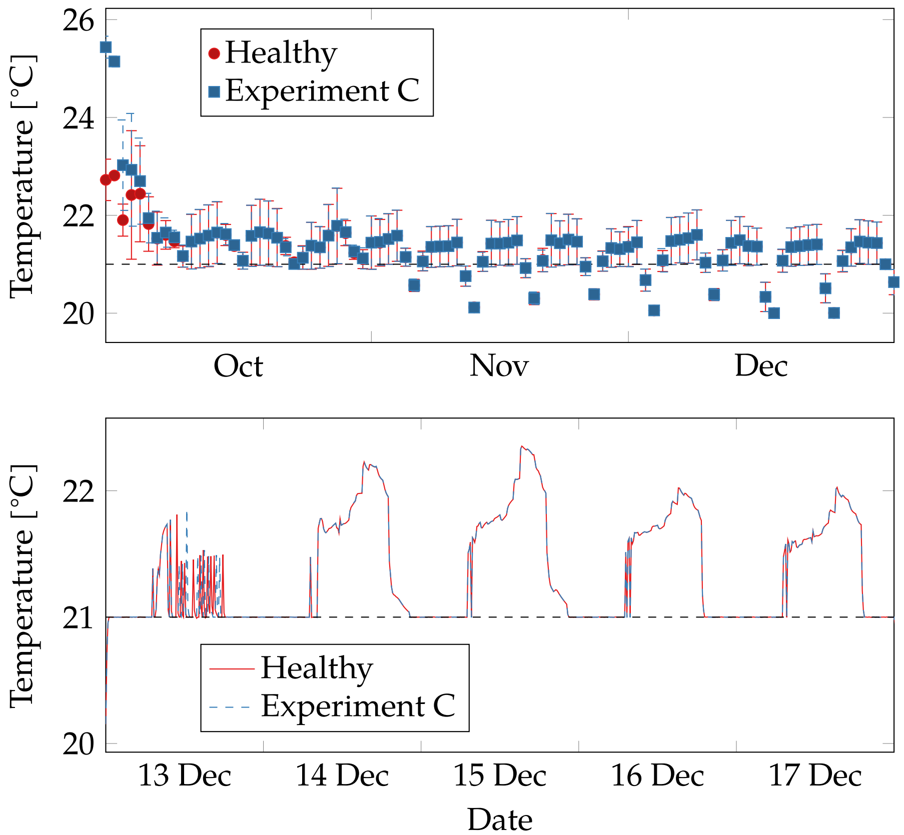

Figure 10.5: Daily average supply temperature in experiment C. The reduced water loop temperature does not have a visible effect on the air temperature..

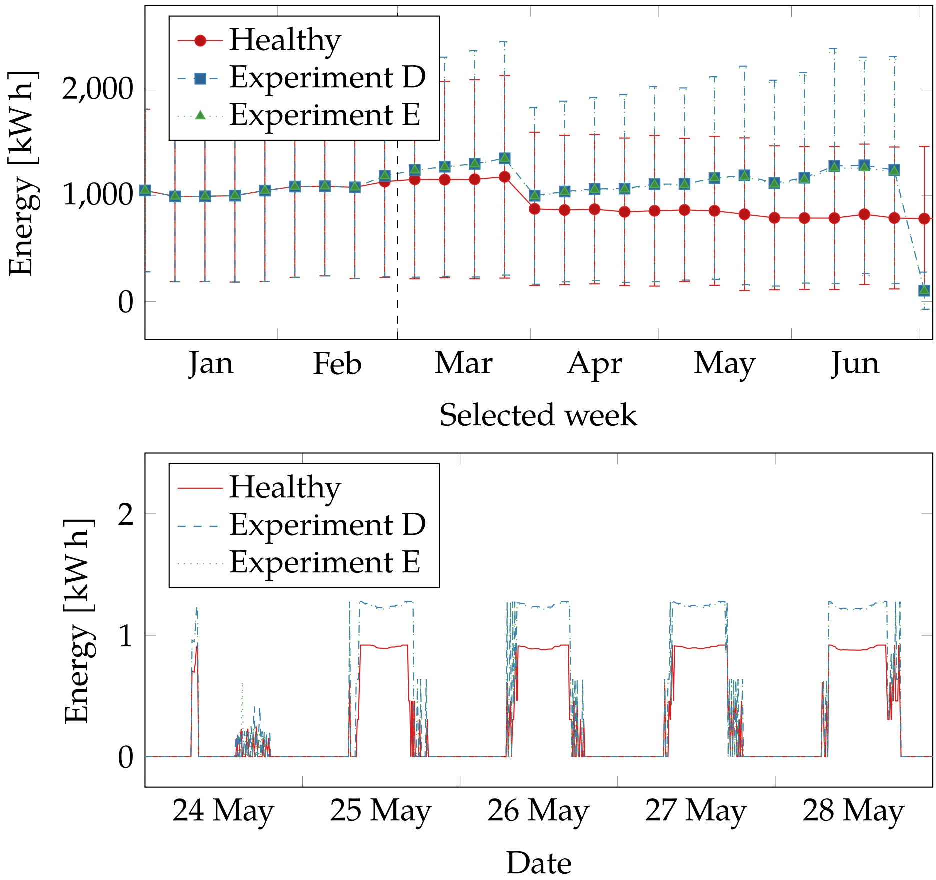

Figure 10.6: Fan energy consumption in experiments D and E. After the fault happens, energy consumption visibly increases over time..

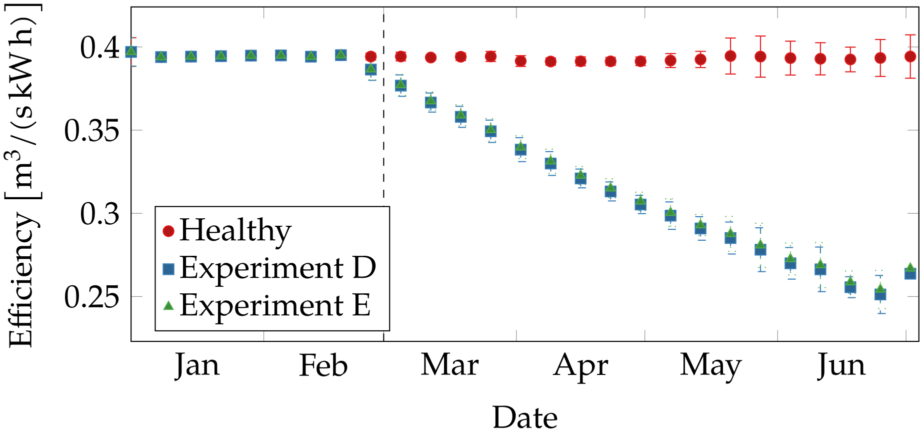

Figure 10.7: Fan efficiency in experiments D and E, computed as daily airflow divided by daily energy consumption. After the faults happen, fan efficiency decreases constantly over time. The visible decrease in efficiency right before the fault is an artifact due to weekly aggregation..

be responsible for 40 kW h and 160 kW h. In experiments D and E, at the end of the simulation, the weekly fan electrical energy consumption was 464 kW h higher than the one in the healthy conditions.

The energy waste caused by the faults simulated in the experiment appears to be relatively small. However, this is due to the high energy efficiency of building OU44. The building is, as a matter of fact, classified as a 2020 class building, according to the Danish Building Regulations [29]. This class includes buildings whose annual primary energy consumption must not exceed 25 kW h/m^2, and was defined in the context of long-term energy savings goals. Therefore, the impact of faults on energy consumption, while significant in relative terms, is small in absolute terms for highly efficient buildings.

The vast majority of existing buildings, on the other hand, are significantly less energy efficiency compared to building OU44. E.g. in Europe, over 80 % of the buildings were constructed before 1990, and over 40 % date back to 1960 [137], where there was not the same focus on energy efficiency as there is today, and, therefore, they consume a considerably higher amount of energy. The same relative impacts, therefore, become significant also in absolute terms for many buildings.

10.6.2 Implications for faults prediction

The method proposed in this paper shows promising implications in the context of faults prediction. One of the limitations of FDD methods is that they can only detect faults after they happen. At that point, faults might disrupt the building use, and, moreover, maintenance operations might be delayed, causing prolonged energy waste and occupants discomfort. FDD methods are, therefore, reactive, since they react to faults, quickly detecting them and identifying their cause.

Faults prediction methods, on the other hand, are proactive, i.e. they aim to predict the occurrence of faults before they even manifest, usually by observing behaviour that are known to lead to malfunctioning. For this reason, they are, in general, less accurate and reliable than FDD methods. However, the information that a fault is going to happen in the near future can be extremely useful, e.g. to avoid service disruption, to schedule maintenance in advance, or to gather replacement equipment before necessary.

In experiments D and E, a decreasing efficiency was observed. In a real building, efficiency might be affected by noise, e.g. due to sensor bias, and seasonal variations. Therefore, the acceptable efficiency might be lower than the usual one. A slow decrease, such as the one observed in the experiments, might be detected before it could have a significant impact on building operations. Maintenance work could then be scheduled to address the fault, e.g. by replacing a wearing component, or cleaning the ducts.

10.7 Conclusion

In this paper, we proposed a model-based method for FDD on ventilation units, supported by a set of rules. At first, under-performing components are identified by comparing the energy consumption recorded by the energy meters with the expected one. Afterwards, a set of rules is used to precisely diagnose the faulty component. We defined five experiments, in which we used an EnergyPlus model of an existing building, calibrated with real data, to simulate its ventilation system. In each experiment, we introduced different faults, and we tested our method on those data. One of the experiments did not yield any measurable effect on the building, while the others caused a measurable increase in energy consumption, and were detected and further diagnosed by the presented method.

Faults were simulated by modifying parameters in the model. This allowed to assess the impact of the faults, both in terms of occupants comfort, and in terms of energy waste. Faults were responsible for significant relative increased energy consumption, up to 8 % increase for heating energy, however, their effect was modest in absolute terms. The cause was the extremely high energy efficiency of the building presented as a case study, which belongs to the highest efficiency class defined in the Danish Building Regulations. Nevertheless, the results give insights on less efficient buildings, which are much more common nowadays, where the same relative energy consumption increase would result in significant absolute amounts.

Finally, the experiments involving gradual faults showed promising implications with respect to faults prediction. By detecting a decreasing term in efficiency early on, maintenance operations can be scheduled before the system reaches problematic conditions. This would reduce disruption and downtime for building operations.

Further development is still necessary for achieving a complete and autonomous FDD application. E.g. proper thresholds need to be defined to automatically determine whether a measurement is too distant from the expected value. Simulating abrupt faults with gradually changing parameters can be a way to determine such thresholds. Other faults can be simulated by changing parameters in the model, or by providing custom schedules, or by employing more advanced techniques such as the ones described in [134].

This method was developed in the context of a complete framework for FDD and energy performance monitoring for buildings [8]. At first, a series of test for time-series data [1] are used to validate the sensing infrastructure, which is fundamental for any building application.

Once the status of the building can be trusted, the energy distribution tree is used to identify under-performing subsystems [2]. Finally, different methods were developed to further investigate individual subsystems and diagnose the precise faulty component [3, 4, 5].

ACKNOWLEDGEMENT

This work is supported by the Innovation Fund Denmark for the project COORDICY.

REFERENCES

[1] **Claudio Giovanni Mattera**, Sanja Lazarova-Molnar, Hamid Reza Shaker and Bo Nørregaard Jørgensen. 'A Practical Approach to Validation of Buildings' Sensor Data: a Commissioning Experience Report'. In: *Third International Conference on Big Data Computing Service and Applications* (*BigDataService*) (San Francisco, CA, USA, 6th–9th Apr. 2017). IEEE. 12th June 2017, pp. 287–292. DOI: 10.1109/BigDataService.2017.48.

[2] **Claudio Giovanni Mattera**, Muhyiddine Jradi and Hamid Reza Shaker. 'Online Energy Simulator for Building Fault Detection and Diagnostics Using Dynamic Energy Performance Model'. In: *International Journal of Low-Carbon Technologies* 13.3 (17th May 2018), pp. 231–239. ISSN: 1748-1325. DOI: 10.1093/ijlct/cty019.

[3] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'Fault Detection and Diagnostics in Ventilation Units Using Linear Regression Virtual Sensors'. In: *International Symposium on Advanced Electrical and Communication Technologies* (*ISAECT*) (Kenitra, Morocco, 21st–23rd Nov. 2018). IEEE. 24th Jan. 2019. DOI: 10.1109/ISAECT.2018.8618755.

[4] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'A Method for Fault Detection and Diagnostics in Ventilation Units Using Virtual Sensors'. In: *Sensors* 18.11 (14th Nov. 2018). ISSN: 1424-8220. DOI: 10.3390/s18113931.

[5] **Claudio Giovanni Mattera**, Hamid Reza Shaker and Muhyiddine Jradi. 'Consensus-based Method for Anomaly Detection in VAV Units'. In: *Energies* 12.3 (1st Feb. 2019). ISSN: 1996-1073. DOI: 10.3390/en12030468.

[6] **Claudio Giovanni Mattera**, Hamid Reza Shaker, Muhyiddine Jradi, Mathis Riber Skydt and Sebastian Skals Engelsgaard. 'Fault Detection in Ventilation Units using Dynamic Energy Performance Models'. In: *Sustainable Cities and Society* (2019). ISSN: 2210-6707. **Submitted**.

[8] Muhyiddine Jradi, Krzysztof Arendt, Fisayo Caleb Sangogboye, **Claudio Giovanni Mattera**, Elena Markoska, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'ObepME: An Online Building Energy Performance Monitoring and Evaluation Tool to Reduce Energy Performance Gaps'. In: *Energy and Buildings* 166 (1st May 2018), pp. 196–209. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2018.02.005.

[12] European Parliament and Council of the European Union. 'Directive 2010/31/EU of the European Parliament and of the Council on the Energy Performance of Buildings'. In: *Official Journal of the European Union* L153 (18th June 2010), pp. 13–35. EU DIRECTIVE: 2010/31.

[16] *2011 Buildings Energy Data Book*. Tech. rep. U. S. A. Department of Energy, Mar. 2012. URL: https://openei.org/doe-opendata/dataset/buildings-energy-data-book (visited on 20th Dec. 2018).

[21] Commission of the European Communities. *20 20 by 2020 - Europe's climate change opportunity*. 23rd Jan. 2008. EU DOCUMENT: 52008DC0030.

[22] European Commission. *Proposal for a Directive of the European Parliament and of the Council Amending Directive 2012/27/EU on Energy Efficiency*. 30th Nov. 2016. EU DOCUMENT: 52016PC0761.

[29] *Danish Building Regulations 2015*. Copenhagen, Denmark: The Danish Transport and Construction Agency, 14th Dec. 2015.

[31] *Doubling U. S. Energy Productivity by 2030*. Tech. rep. Washington, DC, USA: Alliance Commission on National Energy Efficiency Policy, Feb. 2013. URL: https://www.ase.org/policy/energy2030.

[36] Evan Mills. 'Building Commissioning: A Golden Opportunity for Reducing Energy Costs and Greenhouse Gas Emissions in the United States'. In: *Energy Efficiency* 4.2 (May 2011), pp. 145–173. ISSN: 1570-6478. DOI: 10.1007/s12053-011-9116-8.

[37] Janghyun Kim, Jie Cai and James E. Braun. *Common Faults and Their Prioritization in Small Commercial Buildings: February 2017 – December 2017*. Tech. rep. NREL/SR-5500-70136. Golden, CO, USA: U. S. A. Department of Energy - National Renewable Energy Laboratory, June 2018. URL: https://www.nrel.gov/docs/fy18osti/70136.pdf (visited on 26th Dec. 2018).

[41] Woohyun Kim and Srinivas Katipamula. 'A Review of Fault Detection and Diagnostics Methods for Building Systems'. In: *Science and Technology for the Built Environment* 24.1 (8th May 2018), pp. 3–21. ISSN: 2374-4731. DOI: 10.1080/23744731.2017.1318008.

[52] Drury B. Crawley, Linda K. Lawrie, Frederick C. Winkelmann, W.F. Buhl, Y.Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte and Jason Glazer. 'EnergyPlus: Creating a New-generation Building Energy Simulation Program'. In: *Energy and Buildings* 33.4 (Apr. 2001): *Building Simulation '99*, pp. 319–331. ISSN: 0378-7788. DOI: 10.1016/S0378-7788(00)00114-6.

[57] Ando Andriamamonjy, Dirk Saelens and Ralf Klein. 'An Auto-deployed Model-based Fault Detection and Diagnosis Approach for Air Handling Units using BIM and Modelica'. In: *Automation in Construction* 96 (Dec. 2018), pp. 508–526. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2018.09.016.

[129] Shun Li and Jin Wen. 'Application of pattern matching method for detecting faults in air handling unit system'. In: *Automation in Construction* 43 (2014), pp. 49–58. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2014.03.002.

[130] Giacomo Boracchi, Michalis Michaelides and Manuel Roveri. 'A Cognitive Monitoring System for Detecting and Isolating Contaminants and Faults in Intelligent Buildings'. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 48.3 (6th Oct. 2016), pp. 433–447. ISSN: 2168-2216. DOI: 10.1109/TSMC.2016.2608419.

[131] Suhrid Deshmukh, Leon Glicksman and Leslie Norford. 'Case study results: fault detection in air-handling units in buildings'. In: *Advances in Building Energy Research* (15th Nov. 2018). DOI: 10.1080/17512549.2018.1545143.

[132] Zixiao Shi and William O'Brien. 'Using building performance simulation for fault impact evaluation'. In: *eSim 2018* (Montreal, QC, Canada, 9th–10th May 2018). 2018. URL: https://nrc-publications.canada.ca/eng/view/object/?id=e81253ef-c153-4048-9a23-925884825667 (visited on 22nd Jan. 2019). **Pre-published**.

[133] Mangesh Basarkar, Xiufeng Pang, Liping Wang, Philip Haves and Tianzhen Hong. 'Modeling and Simulation of HVAC Faults in EnergyPlus'. In: *12th Conference of International Building Performance Simulation Association* (Sydney, Australia, 14th–16th Nov. 2011). Jan. 2011, pp. 2897–2903. URL: http://www.ibpsa.org/?page_id=105 (visited on 22nd Jan. 2019).

[134] Rongpeng Zhang and Tianzhen Hong. 'Modeling of HVAC operational faults in building performance simulation'. In: *Applied Energy* 202 (2017), pp. 178–188. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2017.05.153.

[135] Ali Behravan, Roman Obermaisser, Deepak Hanike Basavegowda and Simon Meckel. 'Automatic Model-based Fault Detection and Diagnosis Using Diagnostic Directed Acyclic Graph for a Demand-controlled Ventilation and Heating System in Simulink'. In: *2018 Annual IEEE International Systems Conference (SysCon)* (Vancouver, BC, Canada, 23rd–26th Apr. 2018). IEEE. 31st May 2018, pp. 1–7. DOI: 10.1109/SYSCON.2018.8369614.

[136] *ISO8601:2000(E)*. Standard. Geneva, CH: International Organization for Standardization, Dec. 2000.

[137] *Europe's Buildings Under the Microscope. A country-by-country review of the energy performance of buildings*. Buildings Performance Institute Europe, Oct. 2011. 132 pp. ISBN: 9789491143014. URL: http://bpie.eu/publication/europes-buildings-under-the-microscope/.

Part III

Conclusions

This part concludes the thesis. In Chapter 11, future research directions are suggested in the context of the thesis. The findings of the thesis are summarized and elaborated in Chapter 12.

CHAPTER 11
FUTURE RESEARCH

In this chapter, future research directions are discussed. At first, a brief recap of the individual publications is reported and, afterwards, a general summary is presented for the whole thesis.

11.1 INDIVIDUAL PUBLICATIONS

Since the individual publications included in this thesis were laid out in a planned sequence, many of the suggested directions for future research were covered by later publications. Some others were also investigated by other people at the center working within project COORDICY. However, several of them are still unexplored.

In order to achieve a more robust validation of sensors data, an exhaustive battery of tests should be designed and deployed. Thresholds and other parameters could be obtained by system documentation, physical properties, or expert knowledge.

Results from online energy simulator are stored as plain time-series and must be manually inspected to detect faults. An automated system for fault detection and diagnostics (FDD) could be set up and, moreover, advanced visualization techniques could be used to improve the understanding by non-technical personnel. A dashboard application has been developed within the center to work as an information tool [10] and it could be expanded as a decision support tool for longer-term planning of building maintenance and improvement. Moreover, the online energy simulator has the potential to be used for testing control strategies without deploying a new control system on an actual building [11].

Additional virtual sensors can be deployed for other systems and components in buildings, possibly using techniques for automatically design such sensors from historical data or from the system's model. Moreover, the results obtained in Chapters 7 and 8 raised questions about temporary anomalous or faulty behaviours. Further methods should be investigated to decide whether short-lived anomalies are indeed faults, or whether they are only artefacts due to system interferences or data quality issues.

The application of a consensus-based method showed the advantage of this class of techniques, i.e. that fault-free historical training data is no longer required. It was successfully used for finding rooms which exhibited anomalous air distribution patterns, however, more advanced techniques could be used to detect actual faults. Consensus-based methods have a potential wherever a system consists of a multitude of similar components, therefore, it could be applied in other kinds of building equipment such as lighting devices or radiators.

Finally, other techniques are available for simulating different categories of faults, both with EnergyPlus or with alternative simulation engines, especially those modelling specific equipment. Moreover, simulating faults can help in detecting earlier patterns, which can be used for fault prediction.

11.2 GENERAL SUMMARY

The obvious future research direction is toward the integration of the methods developed in this thesis into a self-contained, mature software solution. Such solution would implement the main framework outlined in Chapter 6, allowing pluggable specific methods for specialized FDD methods for specific equipment. Standardized interfaces, both for data access and alarm triggering, would make it possible to implement advanced methods from the literature. They would also make it possible to deploy the framework on different buildings.

One of the recurring issues encountered in designing and implementing FDD methods was determining sensible values for thresholds. Due to inaccuracies in modelling and sensing technology, real systems always slightly deviate from their expected correct behaviour, regardless whether it was obtained from data or quantitative model or if it was encoded in rules. Thresholds are used to ignore small deviations that are probably not caused by faults, reducing the rate of false alarms. However, when set too lax, they might also ignore legitimate faults. These are called type I and type II errors [138]. Effective FDD methods should avoid overlooking actual faults, however, an excessive rate of false alarms might reduce the trust of the management on such techniques. Therefore, proper thresholds should be derived by system documentation, expert knowledge, or even historical data or simulations. Another alternative could be to use 'fuzzy' thresholds, where faults are reported together with a degree of severity.

Finally, two more research directions stem from FDD: 'fully automated' FDD, and faults prediction. The majority of FDD methods in the literature, including the ones presented in this thesis, require manual intervention to work. Input data must be identified and collected, and a detection pipeline must be explicitly designed and implemented. Alarms are generated manually and changes in the system often require a new deployment. In fully automated FDD, on the other hand, the manual intervention is kept at minimal. Building systems are monitored in real-time, FDD methods are applied continuously, and faults alarms are forwarded directly to the building management. Changes in the building's configuration can be picked up automatically, at least up to a certain complexity. The online energy simulator, presented in Chapter 6, which significantly reduces manual operation for simulating a building's behaviour, is a step in the direction of fully automated FDD. The integration of the framework would also factor out many of the manual operations to the framework itself, and the efforts for automating it would reflect also on the specialized methods.

In FDD methods, faults are detected after they appear and impact the system. Faults prediction methods, on the other hand, have a more ambitious goal of detecting faults before they appear. The main advantage is to avoid disruption of service due to faults, which can instead be handled in advance, e.g. when the building is unoccupied at night or during weekends. FDD, however, is more accurate, due to having more information from the system.

CHAPTER

CONCLUSION

12

In this chapter, conclusions about the thesis work are drawn.

This thesis presents a hierarchical framework for fault detection and diagnostics (FDD) on building systems. The framework considers the following hierarchy in building systems:

- whole building;
- system;
- subsystem;
- component.

The following three major steps are taken to perform FDD: 1. data validation, 2. isolation of under-performing subsystems, 3. diagnostics of faulty component.

Data validation is a prerequisite for any building application. Therefore, as a first step, a system for data validation is designed in Chapter 5, to ensure that the sensing and collection infrastructure is working correctly. A set of rules and tests is used to confirm that data is collected, and that physical and logical properties are satisfied.

Once data from the building can be trusted, the status of the building is monitored regularly to detect unexpected energy waste. A dynamic energy performance model of the building is used to simulate its expected behaviour and, specifically, its expected energy consumption for each system and subsystem. If the measurements from the energy meters agree with the simulation, the building is considered performing correctly. Otherwise, one or more subsystems must be responsible for the discrepancy. Measured and simulated energy consumption are then recursively compared at lower hierarchy, traversing the building's energy distribution tree at system and subsystem level, until all the under-performing subsystems are identified.

Once the scope is reduced to a single subsystem, specialized FDD methods can be used for isolating the specific component which is affected by a fault. Ventilation units are one of the largest energy consumers in building systems, therefore, faults affecting them cause major energy waste. For this reason, the rest of the work focuses on ventilation units, and three specialized FDD methods are presented.

Redundancy is introduced inside a system using virtual sensors, i.e. by using models to replicate a sensor's readings using other sensors as input. Redundant readings are used to isolate faulty sensors, either physical or virtual. Moreover, virtual sensors can also be used to obtain values for unmeasured quantities. Linear regression models are first used to design virtual sensors, and they are later augmented with non-linear and statistical models.

A severe drawback of data-driven methods is the requirement for fault-free historical data, which is used to train a model of the system. Using consensus among multiple similar peers this requirement can be removed, assuming that the contributions of faulty peers are diluted among the rest. A consensus-based method is proposed to analyse the patterns in air distribution at room level and to isolate anomalous rooms.

Finally, a set of rules is used to diagnose specific faults on components inside ventilation units, i.e. heat exchangers (HXs), hot-water heating loop heaters and fans. The behaviour of the system is simulated using a dynamic energy model of the building, whose parameters are modified in order to introduce faults. The same results are also used to assess the impact of faults on the energy consumption of the building.

The main advantage of the proposed framework is the reduction in the scope achieved by traversing the energy distribution performance tree. It is not necessary to continuously run FDD methods on every single subsystem in the building, instead, only under-performing subsystems are investigated. This reduces the computational load compared with continuous, extensive monitoring, and it also reduces the chance for false alarms, since subsystems that have good performance cannot generate alarms at all. Moreover, it lays the basis for an integrated system, where specialized methods can be implemented independently of each other.

All the methods presented in this thesis were tested on building Odense undervisning 44 (OU44), a real building at the University of Southern Denmark. The building, which is used as a living lab for different experiments, has a comprehensive sensing infrastructure, and its data over several years was available for testing. Using a real building validated the results of the methods and also forced to deal with limitations in the infrastructure and with implementation details. The methods presented in this thesis, therefore, are not purely theoretical speculation detached from the real world, but they have instead been concretely deployed and used in practice. Moreover, actual faults and anomalies were discovered in the building, which improved both its energy consumption and occupants comfort.

Additional software tools and libraries were developed to aid the implementation of the presented methods, and their deployment on the existing infrastructure of building OU44. Libraries for data access over the simple measurement and actuation profile (sMAP) and other protocols, and software tools such as building drivers and tools for automation of building performance simulations have been successfully used in other programs within project COORDICY and others projects at the center, and some of them have been released to the public.

BIBLIOGRAPHY

Open Access publications [139] are identified by the icon .

[1] **Claudio Giovanni Mattera**, Sanja Lazarova-Molnar, Hamid Reza Shaker and Bo Nørregaard Jørgensen. 'A Practical Approach to Validation of Buildings' Sensor Data: a Commissioning Experience Report'. In: *Third International Conference on Big Data Computing Service and Applications* (*BigDataService*) (San Francisco, CA, USA, 6th–9th Apr. 2017). IEEE. 12th June 2017, pp. 287–292. DOI: 10.1109/BigDataService.2017.48 (cit. on pp. 43–46, 75, 121, 184, 212, 243).

[2] **Claudio Giovanni Mattera**, Muhyiddine Jradi and Hamid Reza Shaker. 'Online Energy Simulator for Building Fault Detection and Diagnostics Using Dynamic Energy Performance Model'. In: *International Journal of Low-Carbon Technologies* 13.3 (17th May 2018), pp. 231–239. ISSN: 1748-1325. DOI: 10.1093/ijlct/cty019 (cit. on pp. 42–44, 46, 47, 99, 139, 176, 212, 225, 244).

[3] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'Fault Detection and Diagnostics in Ventilation Units Using Linear Regression Virtual Sensors'. In: *International Symposium on Advanced Electrical and Communication Technologies* (*ISAECT*) (Kenitra, Morocco, 21st–23rd Nov. 2018). IEEE. 24th Jan. 2019. DOI: 10.1109/ISAECT.2018.8618755 (cit. on pp. 43, 44, 46, 48, 127, 147, 212, 244).

[4] **Claudio Giovanni Mattera**, Joseba Quevedo, Teresa Escobet, Hamid Reza Shaker and Muhyiddine Jradi. 'A Method for Fault Detection and Diagnostics in Ventilation Units Using Virtual Sensors'. In: *Sensors* 18.11 (14th Nov. 2018). ISSN: 1424-8220. DOI: 10.3390/s18113931 (cit. on pp. 43, 44, 46, 48, 145, 199, 212, 244).

[5] **Claudio Giovanni Mattera**, Hamid Reza Shaker and Muhyiddine Jradi. 'Consensus-based Method for Anomaly Detection in VAV Units'. In: *Energies* 12.3 (1st Feb. 2019). ISSN: 1996-1073. DOI: 10.3390/en12030468 (cit. on pp. 43, 44, 46, 48, 183, 244).

[6] **Claudio Giovanni Mattera**, Hamid Reza Shaker, Muhyiddine Jradi, Mathis Riber Skydt and Sebastian Skals Engelsgaard. 'Fault Detection in Ventilation Units using Dynamic Energy Performance Models'. In: *Sustainable Cities and Society* (2019). ISSN: 2210-6707. **Submitted** (cit. on pp. 43, 44, 46, 49, 217).

[7] Muhyiddine Jradi, Fisayo Caleb Sangogboye, **Claudio Giovanni Mattera**, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'A World Class Energy Efficient University Building by Danish 2020 Standards'. In: *Energy Procedia* 132 (Oct. 2017): *11th Nordic Symposium on Building Physics*, pp. 21–26. ISSN: 1876-6102. DOI: 10.1016/j.egypro.2017.09.625 (cit. on pp. 61, 68, 112, 113, 134, 160).

[8] Muhyiddine Jradi, Krzysztof Arendt, Fisayo Caleb Sangogboye, **Claudio Giovanni Mattera**, Elena Markoska, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'ObepME: An Online Building Energy Performance Monitoring and Evaluation Tool to Reduce Energy Performance Gaps'. In: *Energy and Buildings* 166 (1st May 2018), pp. 196–209. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2018.02.005 (cit. on pp. 120, 212, 226, 243).

[9] Krzysztof Arendt, Aslak Johansen, Bo Nørregaard Jørgensen, Mikkel Baun Kjærgaard, **Claudio Giovanni Mattera**, Fisayo Caleb Sangogboye, Jens Hjort Schwee and Christian T. Veje.

'Room-level Occupant Counts, Airflow and CO2 Data from an Office Building'. In: *The 16th ACM Conference on Embedded Networked Sensor Systems*. Proceedings of the First Workshop on Data Acquisition To Analysis (Shenzhen, China, 4th–7th Nov. 2018). ACM. New York, NY, USA, 2018, pp. 13–14. DOI: 10.1145/3277868.3277875.

[10] Muhyiddine Jradi, Na Liu, Aslak Johansen, Krzysztof Arendt, **Claudio Giovanni Mattera**, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'Dynamic Energy Model-Based Automatic Building Performance Testing for Continuous Commissioning'. In: *Proceedings of the 16th IBPSA International Conference Building Simulation 2019* (Rome, Italy, 2nd–4th Sept. 2019). International Building Performance Simulation Association. URL: http://buildingsimulation2019.org/. **Submitted** (cit. on p. 253).

[11] Krzysztof Arendt, Anders Clausen, **Claudio Giovanni Mattera**, Muhyiddine Jradi, Aslak Johansen, Christian T. Veje, Mikkel Baun Kjægaard and Bo Nørregaard Jørgensen. 'Controleum: Multi-Objective Model Predictive Control Framework for Buildings'. In: *Proceedings of the 16th IBPSA International Conference Building Simulation 2019* (Rome, Italy, 2nd–4th Sept. 2019). International Building Performance Simulation Association. URL: http://buildingsimulation2019.org/. **Submitted** (cit. on p. 253).

[12] European Parliament and Council of the European Union. 'Directive 2010/31/EU of the European Parliament and of the Council on the Energy Performance of Buildings'. In: *Official Journal of the European Union* L153 (18th June 2010), pp. 13–35. EU DIRECTIVE: 2010/31 (cit. on pp. xxix, xxx, 6, 10, 100, 128, 146, 184, 218).

[13] R. Isermann. 'Supervision, Fault-detection and Fault-diagnosis Methods — an Introduction'. In: *Control Engineering Practice* 5.5 (May 1997), pp. 639–652. ISSN: 0967-0661. DOI: 10.1016/S0967-0661(97)00046-4 (cit. on pp. xxix, 15, 147).

[14] Nicholas Stern. *The Economics of Climate Change. The Stern Review*. Cambridge, United Kingdom: Cambridge University Press, 30th Oct. 2006. ISBN: 9780521700801. URL: https://webarchive.nationalarchives.gov.uk/20100407172811/http://www.hm-treasury.gov.uk/stern_review_report.htm (visited on 3rd Jan. 2019) (cit. on pp. 5, 6).

[15] *Historical CO2 Datasets*. Zürich, Switzerland: Institute for Atmospheric and Climate Science at Eidgenössische Technische Hochschule. URL: https://www.co2.earth/historical-co2-datasets (visited on 3rd Jan. 2019) (cit. on p. 6).

[16] *2011 Buildings Energy Data Book*. Tech. rep. U. S. A. Department of Energy, Mar. 2012. URL: https://openei.org/doe-opendata/dataset/buildings-energy-data-book (visited on 20th Dec. 2018) (cit. on pp. 7, 13, 100, 128, 146, 184, 218).

[17] *Eurostat*. European Union. URL: https://ec.europa.eu/eurostat/ (visited on 26th Dec. 2018) (cit. on pp. 7, 8).

[18] Council of the European Union. 'Council Regulation (EC) No 322/97 of 17 February 1997 on Community Statistics'. In: *Official Journal of the European Union* L52 (22nd Feb. 1997), pp. 1–7. EU REGULATION: 1997/322 (cit. on p. 7).

[19] *Monthly Energy Review - November 2018*. Tech. rep. DOE/EIA-0035(2018/11). Washington, DC, USA: U. S. A. Energy Information Administration, 20th Nov. 2018. URL: https://www.eia.gov/totalenergy/data/monthly/ (visited on 26th Dec. 2018) (cit. on pp. 7–9, 14).

[20] European Commission. *EU Energy in Figures. Statistical Pocketbook*. Luxembourg: Publications Office of the European Union, Aug. 2018. 264 pp. DOI: 10.2833/105297 (cit. on pp. 7, 9).

[21] Commission of the European Communities. *20 20 by 2020 - Europe's climate change opportunity*. 23rd Jan. 2008. EU DOCUMENT: 52008DC0030 (cit. on pp. 10, 21, 61, 184, 218).

[22] European Commission. *Proposal for a Directive of the European Parliament and of the Council Amending Directive 2012/27/EU on Energy Efficiency*. 30th Nov. 2016. EU DOCUMENT: 52016PC0761 (cit. on pp. 10, 100, 184, 218).

[23] Philippe Moseley and Arnold Bruhin. *High Energy Performing Buildings. Support for innovation and market uptake under Horizon 2020 Energy Efficiency*. Luxembourg: Publications Office of the European Union, 3rd Oct. 2018. 15 pp. DOI: 10.2826/495211 (cit. on p. 11).

[24] Delia D'Agostino, Paolo Zangheri, Barbara Cuniberti, Daniele Paci and Paolo Bertoldi. *Synthesis Report on the National Plans for Nearly Zero Energy Buildings (NZEBs)*. Science for Policy Report. Luxembourg: Publications Office of the European Union, 13th Apr. 2016. 94 pp. DOI: 10.2790/659611 (cit. on p. 11).

[25] Max Jamieson, Yann Verstraeten, Mark Allington, Joanne Arbon, Olivia Brajterman and Jonathan Lonsdale. *Energy Performance of Buildings Directive (EPBD) Compliance Study*. 22nd Dec. 2015. 128 pp. DOI: 10.2833/281509 (cit. on pp. 11, 13).

[26] European Parliament and Council of the European Union. 'Directive (EU) 2018/844 of the European Parliament and of the Council of 30 May 2018 Amending Directive 2010/31/EU on the Energy Performance of Buildings and Directive 2012/27/EU on Energy Efficiency'. In: *Official Journal of the European Union* L156 (19th June 2018), pp. 75–91. EU DIRECTIVE: 2018/844 (cit. on pp. 11, 184).

[27] *Energy Strategy 2050*. Copenhagen, Denmark: The Danish Ministry of Climate and Energy, Feb. 2011. ISBN: 978-87-92727-12-1. URL: http://www.danishwaterforum.dk/activities/Climate%20change/Dansk_Energistrategi_2050_febr.2011.pdf (visited on 10th Jan. 2019) (cit. on pp. 12, 20).

[28] *Building Regulations*. Version 1. Copenhagen, Denmark: The Danish Ministry of Economic and Business Affairs, 12th Dec. 2010. ISBN: 978-87-92518-61-3 (cit. on p. 12).

[29] *Danish Building Regulations 2015*. Copenhagen, Denmark: The Danish Transport and Construction Agency, 14th Dec. 2015 (cit. on pp. 12, 241).

[30] *Executive Order on Building Regulations 2018*. BS0200-00353. Copenhagen, Denmark: The Danish Ministry of Transport, Building and Housing, 4th Jan. 2018 (cit. on p. 13).

[31] *Doubling U. S. Energy Productivity by 2030*. Tech. rep. Washington, DC, USA: Alliance Commission on National Energy Efficiency Policy, Feb. 2013. URL: https://www.ase.org/policy/energy2030 (cit. on pp. 13, 20, 218).

[32] *A Common Definition for Zero Energy Buildings*. Tech. rep. DOE/EE-1247. U. S. A. Department of Energy, Sept. 2015. URL: https://www.energy.gov/eere/buildings/downloads/common-definition-zero-energy-buildings (visited on 4th Jan. 2019) (cit. on p. 14).

[33] *Achieving Energy Savings and Emission Reductions from Building Energy Codes. A Primer for State Planning*. Tech. rep. U. S. A. Department of Energy4, Apr. 2015. URL: https://www.energycodes.gov/achieving-energy-savings-and-emission-reductions-building-energy-codes-primer-state-planning (visited on 4th Jan. 2019) (cit. on p. 14).

[34] J Zhang, M Rosenberg, Y. Xie, R Hart, R. Athalye, B. Liu and J. Zhuge. *Energy and Energy Cost Savings Analysis of the 2015 IECC for Commercial Buildings*. Tech. rep. PNNL-24269. U. S. A. Department of Energy, Aug. 2015 (cit. on p. 14).

[35] Kurt W. Roth, Detlef Westphalen, Michael Y. Feng, Patricia Llana and Louis Quartararo. *Energy Impact of Commercial Building Controls and Performance Diagnostics: Market Characterization, Energy Impact of Building Faults and Energy Savings Potential*. Tech. rep. TIAX LLC D0180. Cambridge, MA, USA: U. S. A. Department of Energy, Nov. 2005 (cit. on pp. 16, 17).

[36] Evan Mills. 'Building Commissioning: A Golden Opportunity for Reducing Energy Costs and Greenhouse Gas Emissions in the United States'. In: *Energy Efficiency* 4.2 (May 2011), pp. 145–173. ISSN: 1570-6478. DOI: 10.1007/s12053-011-9116-8 (cit. on pp. 16, 17, 100, 128, 146, 184, 218).

[37] Janghyun Kim, Jie Cai and James E. Braun. *Common Faults and Their Prioritization in Small Commercial Buildings: February 2017 – December 2017*. Tech. rep. NREL/SR-5500-70136. Golden, CO, USA: U. S. A. Department of Energy - National Renewable Energy Laboratory, June 2018. URL: https://www.nrel.gov/docs/fy18osti/70136.pdf (visited on 26th Dec. 2018) (cit. on pp. 16, 18, 184, 185, 218).

[38] N. Fernandez, S. Katipamula, W. Wang, Y. Xie, M. Zhao and C. Corbin. *Impacts of Commercial Building Controls on Energy Savings and Peak Load Reduction*. Tech. rep. PNNL-25985. U. S. A. Department of Energy, May 2017 (cit. on pp. 16, 19).

[39] Srinivas Katipamula and Michael R. Brambley. 'Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems–A Review, Part I'. In: *HVAC&R Research* 11.1 (Jan. 2005), pp. 3–25. DOI: 10.1080/10789669.2005.10391123 (cit. on pp. 19, 29, 78, 147, 169).

[40] Srinivas Katipamula and Michael R. Brambley. 'Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems–A Review, Part II'. In: *HVAC&R Research* 11.2 (Apr. 2005), pp. 169–187. DOI: 10.1080/10789669.2005.10391133 (cit. on pp. 19, 29, 147, 169).

[41] Woohyun Kim and Srinivas Katipamula. 'A Review of Fault Detection and Diagnostics Methods for Building Systems'. In: *Science and Technology for the Built Environment* 24.1 (8th May 2018), pp. 3–21. ISSN: 2374-4731. DOI: 10.1080/23744731.2017.1318008 (cit. on pp. 19, 29, 30, 32, 38, 39, 45, 101, 129, 146–148, 169, 186, 219).

[42] U. S. A. Department of Energy. *Better Buildings Initiative*. URL: https://betterbuildingsinitiative.energy.gov/ (visited on 6th Dec. 2018) (cit. on p. 21).

[43] *EBC Annual Report*. Tech. rep. Birmingham, United Kingdom: AECOM Ltd on behalf of the International Energy Agency, 2017. URL: http://www.iea-ebc.org/publications/annual-reports (visited on 26th Dec. 2018) (cit. on p. 21).

[44] *Summary of the Financial Benefits of Energy Star® Labeled Office Buildings*. Tech. rep. EPA 430-S-06-003. Washington, DC, USA: U. S. A. Environmental Protection Agency, Feb. 2006. URL: https://www.energystar.gov/buildings/tools-and-resources/summary-financial-benefits-energy-star-labeled-office-buildings (visited on 26th Dec. 2018) (cit. on p. 21).

[45] U. S. A. Green Building Council. *LEED Reference Guide for Building Design and Construction*. Version 4. Washington, DC, USA, Nov. 2013. ISBN: 978-1-932444-18-6 (cit. on p. 21).

[46] Guy R. Newsham, Sandra Mancini and Benjamin J. Birt. 'Do LEED-certified buildings save energy? Yes, but...' In: *Energy and Buildings* 41.8 (Aug. 2009), pp. 897–905. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2009.03.014 (cit. on p. 21).

[47] John H. Scofield. 'Do LEED-certified buildings save energy? Not really...' In: *Energy and Buildings* 41.12 (Dec. 2009), pp. 1386–1390. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2009.08.006 (cit. on p. 21).

[48] John H. Scofield. 'Efficacy of LEED-certification in Reducing Energy Consumption and Greenhouse Gas Emission for Large New York City Office Buildings'. In: *Energy and Buildings* 67 (Dec. 2013), pp. 517–524. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2013.08.032 (cit. on p. 21).

[49] Chris van Dronkelaar, Mark Dowson, E. Burman, Catalina Spataru and Dejan Mumovic. 'A Review of the Energy Performance Gap and Its Underlying Causes in Non-Domestic

Buildings'. In: *Frontiers in Mechanical Engineering* 1 (13th Jan. 2016). DOI: 10.3389/fmech.2015.00017 (cit. on p. 21).

[50] Per Runeson and Martin Höst. 'Guidelines for conducting and reporting case study research in software engineering'. In: *Empirical Software Engineering* 14.2 (19th Dec. 2008), pp. 131–164. ISSN: 1573-7616. DOI: 10.1007/s10664-008-9102-8 (cit. on p. 25).

[51] Haorong Li, Daihong Yu and James E. Braun. 'A Review of Virtual Sensing Technology and Application in Building Systems'. In: *HVAC&R Research* 17.5 (3rd Oct. 2011), pp. 619–645. DOI: 10.1080/10789669.2011.573051 (cit. on pp. 40, 130, 151, 152, 154).

[52] Drury B. Crawley, Linda K. Lawrie, Frederick C. Winkelmann, W.F. Buhl, Y.Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte and Jason Glazer. 'EnergyPlus: Creating a New-generation Building Energy Simulation Program'. In: *Energy and Buildings* 33.4 (Apr. 2001): *Building Simulation '99*, pp. 319–331. ISSN: 0378-7788. DOI: 10.1016/S0378-7788(00)00114-6 (cit. on pp. 41, 56, 68, 102, 111, 225).

[53] Peter Fritzson. *Introduction to Modeling and Simulation of Technical and Physical Systems with Modelica*. John Wiley & Sons, Inc, 17th Apr. 2012. DOI: 10.1002/9781118094259 (cit. on pp. 41, 102).

[54] J. A. Clarke and J. L. M. Hensen. 'Integrated Building Performance Simulation: Progress, Prospects and Requirements'. In: *Building and Environment* 91 (Sept. 2015): *Fifty Year Anniversary for Building and Environment*, pp. 294–306. ISSN: 0360-1323. DOI: 10.1016/j.buildenv.2015.04.002 (cit. on pp. 41, 102, 103).

[55] Andrea Costa, Marcus M. Keane, J. Ignacio Torrens and Edward Corry. 'Building Operation and Energy Performance: Monitoring, Analysis and Optimisation Toolkit'. In: *Applied Energy* 101 (Jan. 2013): *Sustainable Development of Energy, Wa-*

ter and Environment Systems, pp. 310–316. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2011.10.037 (cit. on pp. 41, 103).

[56] 'ASHRAE Building EQ'. In: *ASHRAE* (Dec. 2009). URL: https://www.ashrae.org/ (visited on 2nd Feb. 2019) (cit. on p. 41).

[57] Ando Andriamamonjy, Dirk Saelens and Ralf Klein. 'An Auto-deployed Model-based Fault Detection and Diagnosis Approach for Air Handling Units using BIM and Modelica'. In: *Automation in Construction* 96 (Dec. 2018), pp. 508–526. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2018.09.016 (cit. on pp. 56, 224).

[58] KNX Association. *What Is KNX?* URL: https://www.knx.org/knx-en/for-professionals/What-is-KNX/A-brief-introduction// (visited on 26th Dec. 2018) (cit. on pp. 63, 85, 113).

[59] Sanja Lazarova-Molnar, Mikkel Baun Kjærgaard, Hamid Reza Shaker and Bo Nørregaard Jørgensen. 'Commercial Buildings Energy Performance within Context Occupants in Spotlight'. In: *Proceedings of the 4th International Conference on Smart Cities and Green ICT Systems - Volume 1: SMARTGREENS* (Lisbon, Portugal, 20th–22nd May 2015). SciTePress, 2015, pp. 306–312. DOI: 10.5220/0005495203060312 (cit. on pp. 64, 85, 113).

[60] Muhyiddine Jradi, Christian T. Veje and Bo Nørregaard Jørgensen. 'Towards Energy Efficient Office Buildings in Denmark: the Maersk Building Case Study'. In: *29th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems* (Portorož, Slovenia, 19th–23rd June 2016). Ed. by Andrej Kitanovski and Alojz Poredoš. University of Ljubljana. 2016. ISBN: 978-961-6980-15-9 (cit. on pp. 64, 85, 113).

[61] Bo Nørregaard Jørgensen, Mikkel Baun Kjærgaard, Sanja Lazarova-Molnar, Hamid Reza Shaker and Christian T. Veje. 'Challenge: Advancing Energy Informatics to Enable Assessable Improvements of Energy Performance in Buildings'. In:

Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems (Bangalore, India, 14th–17th July 2015). ACM. New York, NY, USA, 2015, pp. 77–82. DOI: 10.1145/2768510.2770935 (cit. on pp. 76, 176).

[62] R. Jeyanthi and K. Anwamsha. 'Fuzzy-based Sensor Validation for a Nonlinear Bench-mark Boiler under MPC'. In: *10th International Conference on Intelligent Systems and Control (ISCO)* (Coimbatore, India, 7th–8th Jan. 2016). IEEE. 3rd Nov. 2016, pp. 1–6. DOI: 10.1109/isco.2016.7727133 (cit. on pp. 77, 79, 81).

[63] Nithya Kancherla and R. Jeyanthi. 'Design of a Generic Fuzzy-Based Sensor Data Validation Algorithm for a Chemical Process'. English. In: *International conference on Circuits, Controls and Communications (CCUBE)* (Bengaluru, India, 27th–28th Dec. 2013). IEEE. 23rd Jan. 2014, pp. 1–6. DOI: 10.1109/ccube.2013.6718559 (cit. on pp. 77, 79, 81).

[64] Ihab Samy, Ian Postlethwaite and Dawei Gu. 'Neural Network Based Sensor Validation Scheme Demonstrated on an Unmanned Air Vehicle (UAV) Model'. In: *47th Conference on Decision and Control* (Cancun, Mexico, 9th–11th Dec. 2008). IEEE. 6th Jan. 2009, pp. 1237–1242. DOI: 10.1109/CDC.2008.4738703 (cit. on pp. 77, 79).

[65] Shengwei Wang and Youming Chen. 'Sensor Validation and Reconstruction for Building Central Chilling Systems Based on Principal Component Analysis'. English. In: *Energy Conversion and Management* 45.5 (Mar. 2004), pp. 673–695. ISSN: 0196-8904. DOI: 10.1016/S0196-8904(03)00180-8 (cit. on pp. 77, 79).

[66] Shengwei Wang and Fu Xiao. 'AHU Sensor Fault Diagnosis Using Principal Pomponent Analysis Method'. In: *Energy and Buildings* 36.2 (Feb. 2004), pp. 147–160. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2003.10.002 (cit. on pp. 77, 79).

[67] Shengwei Wang and Jianying Qin. 'Sensor Fault Detection and Validation of VAV Terminals in Air Conditioning Systems'. English. In: *Energy Conversion and Management* 46.15–16 (Sept. 2005), pp. 2482–2500. ISSN: 0196-8904. DOI: 10.1016/j.enconman.2004.11.011 (cit. on pp. 77, 79).

[68] Zhijian Hou, Zhiwei Lian, Ye Yao and Xinjian Yuan. 'Data Mining Based Sensor Fault Diagnosis and Validation for Building Air Conditioning System'. In: *Energy Conversion and Management* 47.15 (Sept. 2006), pp. 2479–2490. ISSN: 0196-8904. DOI: 10.1016/j.enconman.2005.11.010 (cit. on pp. 77, 79, 80).

[69] MinJeong Kim, Hongbin Liu, Jeong Tai Kim and ChangKyoo Yoo. 'Sensor Fault Identification and Reconstruction of Indoor Air Quality (IAQ) Data Using a Multivariate Non-Gaussian Model in Underground Building Space'. In: *Energy and Buildings* 66 (Nov. 2013), pp. 384–394. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2013.07.002 (cit. on pp. 77, 79, 153, 154).

[70] Alexey Kozionov, Mikhail Kalinkin, Alexey Natekin and Alexander Loginov. 'Wavelet-based Sensor Validation: Differentiating Abrupt Sensor Faults from System Dynamics'. In: *7th International Symposium on Intelligent Signal Processing* (Floriana, Malta, 19th–21st Sept. 2011). IEEE. 20th Oct. 2011, pp. 1–5. DOI: 10.1109/wisp.2011.6051716 (cit. on pp. 77–79).

[71] J. Kullaa. 'Sensor Validation Using Minimum Mean Square Error Estimation'. In: *Mechanical Systems and Signal Processing* 24.5 (July 2010), pp. 1444–1457. ISSN: 0888-3270. DOI: 10.1016/j.ymssp.2009.12.001 (cit. on pp. 77, 79, 81).

[72] Z. G. Shen and Q. Wang. 'Data Validation and Confidence of Self-validating Multifunctional Sensor'. English. In: *Proceedings of Sensors* (Taipei, Taiwan, 28th–31st Oct. 2012). IEEE. 17th Jan. 2013, pp. 1045–1048. DOI: 10.1109/icsens.2012.6411058 (cit. on pp. 77, 79, 81).

[73] K. M. Tsang and W. L. Chan. 'Data Validation of Intelligent Sensor Using Predictive Filters and Fuzzy Logic'. In: *Sensors and Actuators A: Physical* 159.2 (May 2010), pp. 149–156. ISSN: 0924-4247. DOI: 10.1016/j.sna.2010.03.013 (cit. on pp. 77, 79–81).

[74] R. Sharifi and R. Langari. 'Nonlinear Sensor Fault Diagnosis Using Mixture of Probabilistic PCA Models'. English. In: *Mechanical Systems and Signal Processing* 85 (15th Feb. 2017), pp. 638–650. ISSN: 0888-3270. DOI: 10.1016/j.ymssp.2016.08.028 (cit. on pp. 77, 79, 80).

[75] Sanja Lazarova-Molnar, Hamid Reza Shaker, Nader Mohamed and Bo Nørregaard Jørgensen. 'Fault Detection and Diagnosis for Smart Buildings: State of the Art, Trends and Challenges'. In: *3rd MEC International Conference on Big Data and Smart City (ICBDSC)* (Muscat, Oman, 15th–16th Mar. 2016). IEEE. 28th Apr. 2016, pp. 1–7. DOI: 10.1109/ICBDSC.2016.7460392 (cit. on pp. 77, 176).

[76] Nemanja Branisavljević, Zoran Kapelan and Dušan Prodanović. 'Improved Real-time Data Anomaly Detection Using Context Classification'. In: *Journal of Hydroinformatics* 13.3 (July 2011), pp. 307–323. ISSN: 1464-7141. DOI: 10.2166/hydro.2011.042 (cit. on pp. 79–81).

[77] Charles C. Castello, Jibonananda Sanyal, Jeffrey Rossiter, Zachary Hensley and Joshua R. New. 'Sensor Data Management, Validation, Correction, and Provenance for Building Technologies'. In: *ASHRAE Transactions* 120 (July 2014), pp. 370–382. ISSN: 0001-2505 (cit. on pp. 79–81).

[78] Miquel À. Cugueró-Escofet, Diego García, Joseba Quevedo, Vicenç Puig, Santiago Espin and Jaume Roquet. 'A Methodology and a Software Tool for Sensor Data Validation/reconstruction: Application to the Catalonia Regional Water Network'. In: *Control Engineering Practice* 49 (Apr. 2016), pp. 159–172. ISSN: 0967-0661. DOI: 10.1016/j.conengprac.2015.11.005 (cit. on pp. 79–81, 131, 151).

[79] Jari Näsi and Aki Sorsa. *On-line Measurement Validation through Confidence Level Based Optimal Estimation of a Process Variable*. Tech. rep. A No 25. Oulu, Finland: University of Oulu, Dec. 2004 (cit. on pp. 79, 81).

[80] Jari Näsi, Aki Sorsa and Kauko Leiviskä. 'Sensor Validation and Outlier Detection Using Fuzzy Limits'. In: *44th Conference on Decision and Control, and the European Control Conference, CDC-ECC '05* (Seville, Spain, 15th Dec. 2015). IEEE. 30th Jan. 2006, pp. 7828–7833. DOI: 10.1109/cdc.2005.1583427 (cit. on pp. 79, 81).

[81] K. M. Tsang. 'Sensor Data Validation Using Gray Models'. English. In: *ISA Transactions* 42.1 (Jan. 2003), pp. 9–17. ISSN: 0019-0578. DOI: 10.1016/S0019-0578(07)60109-8 (cit. on pp. 79, 81).

[82] Hector Rodriguez, Vicenç Puig, Juan J. Flores and Rodrigo Lopez. 'Combined Holt-Winters and GA Trained ANN Approach for Sensor Validation and Reconstruction: Application to Water Demand Flowmeters'. In: *3rd Conference on Control and Fault-Tolerant Systems (SysTol)* (Barcelona, Spain, 7th–9th Sept. 2016). IEEE. 10th Nov. 2016, pp. 202–207. DOI: 10.1109/SYSTOL.2016.7739751 (cit. on p. 79).

[83] Sun Siao, Jean-LucBertrand Krajwewski, Anders Lynggaard-Jensen, Joep Van den Broeke, Florian Edthofer, Maria do Ceu Almeida, Ribeiro Alvaro Silva and Jose Menaia. *Literature Review for Data Validation Methods*. 2011 (cit. on p. 79).

[84] Ivan Miguel Pires, Nuno M. Garcia, Nuno Pombo, Francisco Flórez-revuelta and Natalia Díaz Rodríguez. 'Validation Techniques for Sensor Data in Mobile Health Applications'. English. In: *Journal of Sensors* 2016 (2016), pp. 1–9. ISSN: 1687-725X. DOI: 10.1155/2016/2839372 (cit. on p. 79).

[85] Krzysztof Arendt, Ana Ionesi, Muhyiddine Jradi, Ashok Kumar Singh, Mikkel Baun Kjærgaard, Christian T. Veje and Bo Nørregaard Jørgensen. 'A Building Model Framework for a Genetic Algorithm Multi-objective Model Predictive Control'. In:

CLIMA 2016 - 12th REHVA World Congress (Aalborg, Denmark, 22nd–25th May 2016). Vol. 8. Aalborg University, Department of Civil Engineering, 2016. ISBN: 87-91606-33-0. URL: http://vbn.aau.dk/en/publications/clima-2016--proceedings-of-the-12th-rehva-world-congress(cd42481c-d63a-4aba-9ac7-f779e332efd5).html (cit. on p. 91).

[86] European Parliament and Council of the European Union. 'Directive 2012/27/EU of the European Parliament and of the Council on on energy efficiency'. In: *Official Journal of the European Union* L315 (14th Nov. 2012), pp. 1–56. EU DIRECTIVE: 2012/27 (cit. on p. 100).

[87] Center for Energy Informatics, University of Southern Denmark. *COORDICY*. URL: https://www.sdu.dk/en/om_sdu/institutter_centre/centreforenergyinformatics/research+projects/coordicy (visited on 3rd Dec. 2018) (cit. on p. 101).

[88] Tobias Maile, Vladimir Bazjanac and Martin Fischer. 'A Method to Compare Simulated and Measured Data to Assess Building Energy Performance'. In: *Building and Environment* 56 (Oct. 2012), pp. 241–251. ISSN: 0360-1323. DOI: 10.1016/j.buildenv.2012.03.012 (cit. on p. 103).

[89] Michael Wetter. 'Co-simulation of Building Energy and Control Systems with the Building Controls Virtual Test Bed'. In: *Journal of Building Performance Simulation* 4.3 (5th Nov. 2011), pp. 185–203. DOI: 10.1080/19401493.2010.518631 (cit. on p. 103).

[90] Xiufeng Pang, Michael Wetter, Prajesh Bhattacharya and Philip Haves. 'A Framework for Simulation-based Real-time Whole Building Performance Assessment'. In: *Building and Environment* 54 (Aug. 2012), pp. 100–108. ISSN: 0360-1323. DOI: 10.1016/j.buildenv.2012.02.003 (cit. on pp. 103, 104).

[91] Xiufeng Pang, Thierry S. Nouidui, Michael Wetter, Daniel Fuller, Anna Liao and Philip Haves. 'Building Energy Simulation in Real Time Through an Open Standard Interface'. In: *Energy and Buildings* 117 (1st Apr. 2016), pp. 282–289. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2015.10.025 (cit. on p. 104).

[92] Tanzia Sharmin, Mustafa Gül, Xinming Li, Veselin Ganev, Ioanis Nikolaidis and Mohamed Al-Hussein. 'Monitoring Building Energy Consumption, Thermal Performance, and Indoor Air Quality in a Cold Climate Region'. In: *Sustainable Cities and Society* 13 (Oct. 2014), pp. 57–68. ISSN: 2210-6707. DOI: 10.1016/j.scs.2014.04.009 (cit. on p. 104).

[93] Stephen Dawson-Haggerty, Xiaofan Jiang, Gilman Tolle, Jorge Ortiz and David Culler. 'sMAP: A Simple Measurement and Actuation Profile for Physical Information'. In: *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems* (Zürich, Switzerland, 3rd–5th Nov. 2010). New York, NY, USA: ACM, Nov. 2010, pp. 197–210. DOI: 10.1145/1869983.1870003 (cit. on p. 108).

[94] Torsten Schwan, René Unger and Jörg Pipiorke. 'Aspects of FMI in Building Simulation'. In: *Proceedings of the 12th International Modelica Conference*. 132. Linköpings, Sweden: Linköping University Electronic Press, May 2017, pp. 73–78. DOI: 10.3384/ecp1713273 (cit. on p. 109).

[95] U. S. A. Department of Energy. *EnergyPlusToFMU*. URL: https://simulationresearch.lbl.gov/projects/energyplustofmu (visited on 3rd Dec. 2018) (cit. on p. 111).

[96] Center for Energy Informatics, University of Southern Denmark. *EPQuery*. URL: https://sdu-cfei.github.io/epquery/ (visited on 3rd Dec. 2018) (cit. on p. 114).

[97] Fisayo Caleb Sangoboye and Mikkel Baun Kjærgaard. 'PLCount: A Probabilistic Fusion Algorithm for Accurately Estimating Occupancy from 3D Camera Counts'. In: *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient*

Built Environments (Palo Alto, CA, USA, 16th–17th Nov. 2016). BuildSys '16. ACM. New York, NY, USA, 2016, pp. 147–156. DOI: 10.1145/2993422.2993575 (cit. on p. 114).

[98] Luis Pérez-Lombard, José Ortiz and Christine Pout. 'A Review on Buildings Energy Consumption Information'. In: *Energy and Buildings* 40.3 (2008), pp. 394–398. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2007.03.007 (cit. on pp. 129, 147).

[99] *Technical Synthesis Report Annex 34. Computer Aided Evaluation of HVAC System Performance*. Tech. rep. Birmingham, United Kingdom: International Energy Agency - Energy Conservation in Buildings and Community Systems Programme, 2006 (cit. on pp. 129, 147).

[100] Yuebin Yu, Denchai Woradechjumroen and Daihong Yu. 'A Review of Fault Detection and Diagnosis Methodologies on Air-handling Units'. In: *Energy and Buildings* 82 (Oct. 2014), pp. 550–562. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2014.06.042 (cit. on pp. 129, 130, 147, 149, 150).

[101] Haorong Li and James E. Braun. 'Decoupling Features and Virtual Sensors for Diagnosis of Faults in Vapor Compression Air Conditioners'. In: *International Journal of Refrigeration* 30.3 (May 2007), pp. 546–564. ISSN: 0140-7007. DOI: 10.1016/j.ijrefrig.2006.07.024 (cit. on pp. 131, 153, 154).

[102] Nunzio Cotrufo and Radu Zmeureanu. 'Virtual Outdoor Air Flow Meter for an Existing HVAC System in Heating Mode'. In: *Automation in Construction* 92 (Aug. 2018), pp. 166–172. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2018.03.036 (cit. on pp. 131, 153, 154).

[103] Andrew Kusiak, Mingyang Li and Haiyang Zheng. 'Virtual Models of Indoor-air-quality Sensors'. In: *Applied Energy* 87.6 (June 2010), pp. 2087–2094. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2009.12.008 (cit. on pp. 131, 153, 154).

[104] K. Verbert, R. Babuška and B. De Schutter. 'Combining Knowledge and Historical Data for System-level Fault Diagnosis of

HVAC Systems'. In: *Engineering Applications of Artificial Intelligence* 59 (Mar. 2017), pp. 260–273. ISSN: 0952-1976. DOI: 10.1016/j.engappai.2016.12.021 (cit. on pp. 131, 153, 154).

[105] Olivier Renaud and Maria-Pia Victoria-Feser. 'A Robust Coefficient of Determination for Regression'. In: *Journal of Statistical Planning and Inference* 140.7 (July 2010), pp. 1852–1862. ISSN: 0378-3758. DOI: 10.1016/j.jspi.2010.01.008 (cit. on pp. 133, 157).

[106] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot and Édouard Duchesnay. 'Scikit-learn: Machine Learning in Python'. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830. URL: http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html (cit. on pp. 136, 161).

[107] *The Basics of Fan Performance Tables, Fan Curves, System Resistance Curves and Fan Laws*. FA/100-99. Greenheck. Schofield, WI, USA, Dec. 1999. 4 pp. URL: http://www.greenheck.com/resources/library/application-articles (visited on 26th Dec. 2018) (cit. on pp. 138, 166).

[108] Rodney A. Martin, M. Schwabacher, N. Oza and A. Srivastava. 'Comparison of Unsupervised Anomaly Detection Methods for Systems Health Management Using Space Shuttle'. In: *Proceedings of the Joint Army Navy NASA Air Force Conference on Propulsion, JANNAF*. 2007 (cit. on p. 147).

[109] Miquel À. Cugueró-Escofet, Joseba Quevedo, C. Alippi, M. Roveri, Vicenç Puig, Diego García and F. Trovò. 'Model- vs. Data-based Approaches Applied to Fault Diagnosis in Potable Water Supply Networks'. In: *Journal of Hydroinformatics* 18.5 (10th May 2016), pp. 831–850. DOI: 10.2166/hydro.2016.218 (cit. on p. 147).

[110] Eduardo Trutié-Carrero, Damian Valdés-Santiago, Ángela León-Mecías and Jorge Ramírez-Beltrán. 'Burst Detection and Localization using Discrete Wavelet Transform and Cross-Correlation'. In: *Revista Iberoamericana de Automática e Informática industrial* 15.2 (2018), pp. 211–216. ISSN: 1697-7920. DOI: 10.4995/riai.2017.8738 (cit. on p. 147).

[111] Emil L. Hanzevack, Theresa W. Long, Chris M. Atkinson and M. Traver. 'Virtual Sensors for Spark Ignition Engines Using Neural Networks'. In: *American Control Conference* (Albuquerque, NM, USA, 6th June 1997). IEEE. 6th Aug. 2002, pp. 669–673. DOI: 10.1109/ACC.1997.611885 (cit. on p. 151).

[112] Manuel Acosta, Stratis Kanarachos and Michael E. Fitzpatrick. 'A Virtual Sensor for Integral Tire Force Estimation using Tire Model-less Approaches and Adaptive Unscented Kalman Filter'. In: *Proceedings of the 14th International Conference on Informatics in Control, Automation and Robotics* (Madrid, Spain, 26th–28th July 2017). Vol. 1. INSTICC. SciTePress, 2017, pp. 386–397. DOI: 10.5220/0006394103860397 (cit. on p. 151).

[113] Burkhard Hensel and Klaus Kabitzsch. 'Generator for Modular Virtual Sensors'. In: *21st International Conference on Emerging Technologies and Factory Automation (ETFA)* (Berlin, Germany, 6th–9th Sept. 2016). IEEE. 7th Nov. 2016, pp. 1–8. DOI: 10.1109/ETFA.2016.7733726 (cit. on pp. 151, 176).

[114] Pierre Borne, Dumitru Popescu, Floria G. Filip, Dan Stefanoiu and Florin Gh. Filip. *Optimization in Engineering Sciences: Exact Methods*. Ed. by Bernard Dubuisson. 1st ed. iSTE. Hoboken, NJ, USA: John Wiley & Sons, Incorporated, 20th Feb. 2013. DOI: 10.1002/9781118577899 (cit. on p. 168).

[115] Alex J. Smola and Bernhard Schölkopf. 'A Tutorial on Support Vector Regression'. In: *Statistics and Computing* 14.3 (Aug. 2004), pp. 199–222. ISSN: 1573-1375. DOI: 10.1023/B:STCO.0000035301.49549.88 (cit. on p. 168).

[116] Kurt Hornik, Maxwell Stinchcombe and Halbert White. 'Multilayer Feedforward Networks Are Universal Approximators'. In: *Neural Networks* 2.5 (1989), pp. 359–366. ISSN: 0893-6080. DOI: 10.1016/0893-6080(89)90020-8 (cit. on p. 168).

[117] Balakrishnan Narayanaswamy, Bharathan Balaji, Rajesh Gupta and Yuvraj Agarwal. 'Data Driven Investigation of Faults in HVAC Systems with Model, Cluster and Compare (MCC)'. In: *1st Conference on Embedded Systems for Energy-Efficient Buildings* (Memphis, TN, USA, 5th–6th Nov. 2014). Vol. 1. BuildSys. ACM. New York, NY, USA, 2014, pp. 50–59. DOI: 10.1145/2674061.2674067 (cit. on pp. 185, 187, 192, 200).

[118] Carlos Fernández Bandera and Germán Ramos Ruiz. 'Towards a New Generation of Building Envelope Calibration'. In: *Energies* 10.12 (2017). ISSN: 1996-1073. DOI: 10.3390/en10122102 (cit. on p. 186).

[119] Mitchell Yuwono, Ying Guo, Josh Wall, Jiaming Li, Sam West, Glenn Platt and Steven W. Su. 'Unsupervised Feature Selection Using Swarm Intelligence and Consensus Clustering for Automatic Fault Detection and Diagnosis in Heating Ventilation and Air Conditioning Systems'. In: *Applied Soft Computing* 34 (2015), pp. 402–425. ISSN: 1568-4946. DOI: 10.1016/j.asoc.2015.05.030 (cit. on p. 187).

[120] S. Byttner, T. Rögnvaldsson and M. Svensson. 'Consensus Self-organized Models for Fault Detection (COSMO)'. In: *Engineering Applications of Artificial Intelligence* 24.5 (2011), pp. 833–839. ISSN: 0952-1976. DOI: 10.1016/j.engappai.2011.03.002 (cit. on p. 187).

[121] G. Bianchin, A. Cenedese, M. Luvisotto and G. Michieletto. 'Distributed Fault Detection in Sensor Networks Via Clustering and Consensus'. In: *54th Conference on Decision and Control (CDC)* (Osaka, Japan, 15th–18th Dec. 2015). IEEE. 11th Feb. 2016, pp. 3828–3833. DOI: 10.1109/CDC.2015.7402814 (cit. on p. 188).

[122] Mohammadreza Davoodi, Nader Meskin and Khashayar Khorasani. 'Simultaneous Fault Detection and Consensus Control Design for a Network of Multi-agent Systems'. In: *Automatica* 66 (2016), pp. 185–194. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2015.12.027 (cit. on p. 188).

[123] Bo Zhou, Wei Wang and Hao Ye. 'Cooperative Control for Consensus of Multi-agent Systems with Actuator Faults'. In: *Computers & Electrical Engineering* 40.7 (2014), pp. 2154–2166. ISSN: 0045-7906. DOI: 10.1016/j.compeleceng.2014.04.015 (cit. on p. 188).

[124] Sang-Ji Lee, Jin-Young Choi, Hyung-Joo Lee and Dong-Jun Won. 'Distributed Coordination Control Strategy for a Multi-Microgrid Based on a Consensus Algorithm'. In: *Energies* 10.7 (2017). ISSN: 1996-1073. DOI: 10.3390/en10071017 (cit. on p. 188).

[125] Nian Liu, Bin Guo, Zifa Liu and Yongli Wang. 'Distributed Energy Sharing for PVT-HP Prosumers in Community Energy Internet: A Consensus Approach'. In: *Energies* 11.7 (2018). ISSN: 1996-1073. DOI: 10.3390/en11071891 (cit. on p. 188).

[126] Francisco Javier Martínez de Pisón, Andrés Sanz, Eduardo Martínez de Pisón, Emilio Jiménez and Dante Conti. 'Mining Association Rules from Time Series to Explain Failures in a Hot-dip Galvanizing Steel Line'. In: *Computers & Industrial Engineering* 63.1 (Aug. 2012), pp. 22–36. ISSN: 0360-8352. DOI: 10.1016/j.cie.2012.01.013 (cit. on p. 193).

[127] Stephen Butterworth. '"On the Theory of Filter Amplifiers'. In: *Wireless Engineer* 7 (1930), pp. 536–541. URL: http://www.gonascent.com/papers/butter.pdf (cit. on p. 201).

[128] Anatol I. Zverev. *Handbook of Filter Synthesis*. New York, NY, USA: Wiley, 1967. 602 pp. ISBN: 978-0471749424 (cit. on p. 201).

[129] Shun Li and Jin Wen. 'Application of pattern matching method for detecting faults in air handling unit system'. In: *Automation in Construction* 43 (2014), pp. 49–58. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2014.03.002 (cit. on p. 220).

[130] Giacomo Boracchi, Michalis Michaelides and Manuel Roveri. 'A Cognitive Monitoring System for Detecting and Isolating Contaminants and Faults in Intelligent Buildings'. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 48.3 (6th Oct. 2016), pp. 433–447. ISSN: 2168-2216. DOI: 10.1109/TSMC.2016.2608419 (cit. on p. 221).

[131] Suhrid Deshmukh, Leon Glicksman and Leslie Norford. 'Case study results: fault detection in air-handling units in buildings'. In: *Advances in Building Energy Research* (15th Nov. 2018). DOI: 10.1080/17512549.2018.1545143 (cit. on p. 222).

[132] Zixiao Shi and William O'Brien. 'Using building performance simulation for fault impact evaluation'. In: *eSim 2018* (Montreal, QC, Canada, 9th–10th May 2018). 2018. URL: https://nrc-publications.canada.ca/eng/view/object/?id=e81253ef-c153-4048-9a23-925884825667 (visited on 22nd Jan. 2019). **Pre-published** (cit. on p. 222).

[133] Mangesh Basarkar, Xiufeng Pang, Liping Wang, Philip Haves and Tianzhen Hong. 'Modeling and Simulation of HVAC Faults in EnergyPlus'. In: *12th Conference of International Building Performance Simulation Association* (Sydney, Australia, 14th–16th Nov. 2011). Jan. 2011, pp. 2897–2903. URL: http://www.ibpsa.org/?page_id=105 (visited on 22nd Jan. 2019) (cit. on p. 223).

[134] Rongpeng Zhang and Tianzhen Hong. 'Modeling of HVAC operational faults in building performance simulation'. In: *Applied Energy* 202 (2017), pp. 178–188. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2017.05.153 (cit. on pp. 223, 243).

[135] Ali Behravan, Roman Obermaisser, Deepak Hanike Basavegowda and Simon Meckel. 'Automatic Model-based Fault Detection and Diagnosis Using Diagnostic Directed Acyclic Graph for a Demand-controlled Ventilation and Heating System in Simulink'. In: *2018 Annual IEEE International Systems Conference (SysCon)* (Vancouver, BC, Canada, 23rd–26th Apr. 2018).

IEEE. 31st May 2018, pp. 1–7. DOI: 10.1109/SYSCON.2018.8369614 (cit. on p. 224).

[136] *ISO8601:2000(E)*. Standard. Geneva, CH: International Organization for Standardization, Dec. 2000 (cit. on pp. 233, 234).

[137] *Europe's Buildings Under the Microscope. A country-by-country review of the energy performance of buildings*. Buildings Performance Institute Europe, Oct. 2011. 132 pp. ISBN: 9789491143014. URL: http://bpie.eu/publication/europes-buildings-under-the-microscope/ (cit. on p. 241).

[138] Parampreet Kaur and Jill Stoltzfus. 'Type I, II, and III Statistical Errors: A Brief Overview'. In: *International Journal of Academic Medicine* 3.2 (2017), pp. 268–270. DOI: 10.4103/IJAM.IJAM_92_17 (cit. on p. 255).

[139] Peter Suber. *Open Access*. Cambridge, MA, USA: The MIT Press, 2012. ISBN: 978-0-262-51763-8. URL: https://mitpress.mit.edu/books/open-access (cit. on p. 261).