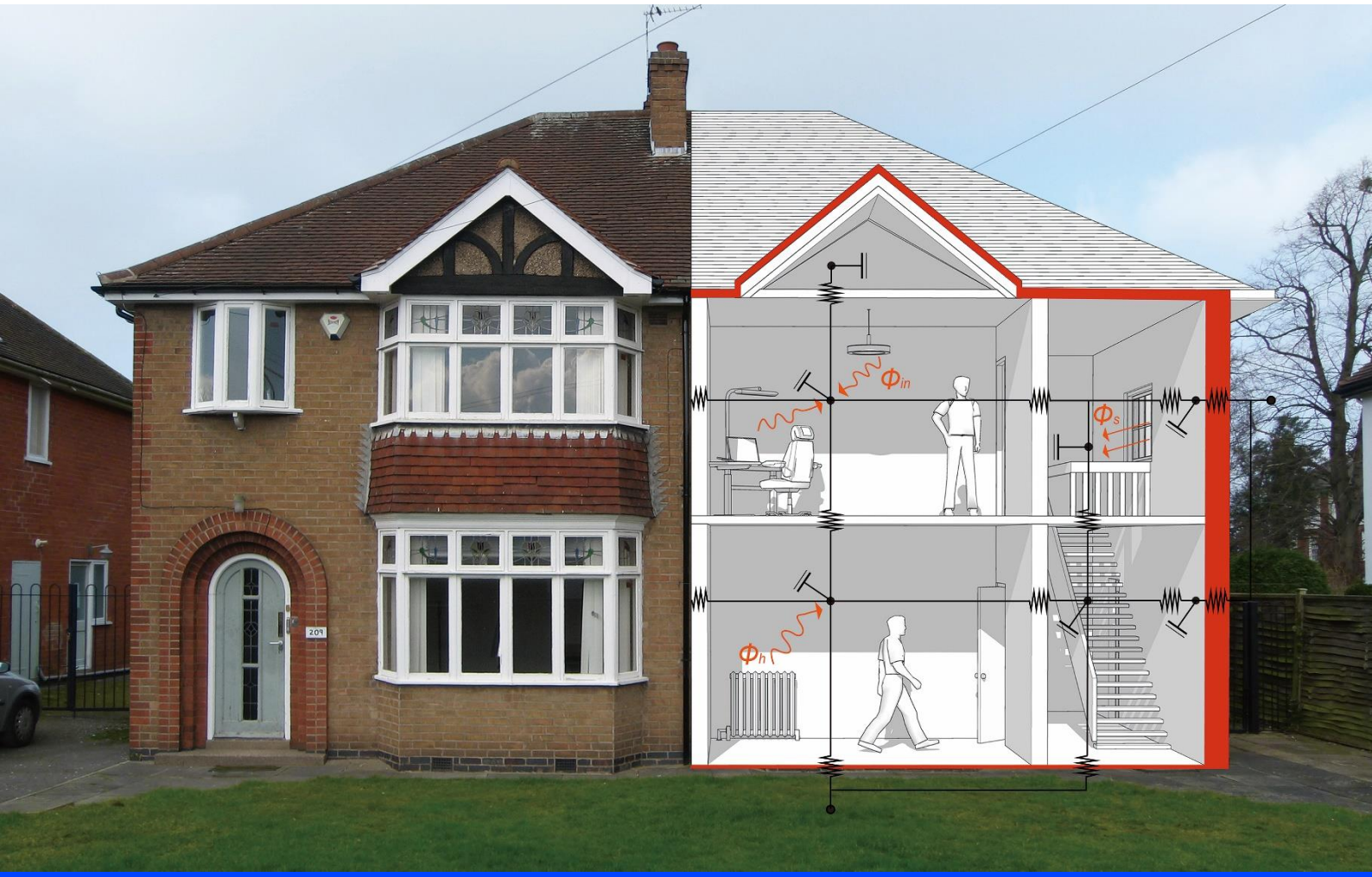
ANNEX 71

EBC
Energy in Building and Communities Programme

Building energy performance assessment based on in-situ measurements

Building Behaviour Identification

August 2021



Technology Collaboration Programme
by IEA

Published by KU Leuven, Belgium

ISBN **9789075741087**

Participating countries in the EBC: Australia, Austria, Belgium, Brazil, Canada, P.R. China, Czech Republic, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Republic of Korea, the Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States of America.
Additional copies of this report may be obtained from: EBC Executive Committee Support Services Unit (ESSU), C/o AECOM Ltd, The Colmore Building, Colmore Circus Queensway, Birmingham B4 6AT, United Kingdom
www.iea-ebc.org
essu@iea-ebc.org

Cover picture: drawing of Xiang Zhang (Jason) (KU Leuven, Belgium) on a picture of the Loughborough test houses (Loughborough University, UK)

ANNEX 71

EBC
Energy in Buildings and Communities Programme

Building energy performance assessment based on in-situ measurements

Building behaviour identification

August 2021

Main Authors

Glenn Reynders, EnergyVille-VITO, Belgium (glenn.reynders@vito.be)

Arash Erfani, KU Leuven, Belgium (arash.erfanibeyzaee@kuleuven.be)

Dirk Saelens, KU Leuven, Belgium (dirk.saelens@kuleuven.be)

Contributing Authors

Balsam Ajib, IMT Lille Douai, France
Peder Bacher, DTU, Denmark
Max Blöchle, AIT, Austria
Alexandra Cosseron, Saint-Gobain, France
Arash Erfani, Dirk Saelens, KU Leuven, Belgium
Tianyun Gao, CTSB, France
Jordan Gauvrit, Antoine Caucheteux, Cerema
Adam Hardy, Leeds Beckett, United Kindom
Maria José Jiménez, PSA, Spain
Matthias Kersken, Fraunhofer, Germany
Gerard Laguna, Florencia Lazzari, Gerard Mor, CIMNE, Spain
Tuule Mall Kull, Karl-Villem Vosa, TalTech, Estonia
Eirini Mantesi, Loughborough University, United Kindom
Gabrielle Masy, UC Louvain, Belgium
Michael Parzinger, TH Rosenheim, Germany
Glenn Reynders, EnergyVille-VITO, Belgium
Gabriel Rojas-Kopeinig, UIBK, Austria
Simon Rouchier,Université Savoie Mont-Blanc, France
Xingji Yu, NTNU, Norway

Preface

The International Energy Agency

The International Energy Agency (IEA) was established in 1974 within the framework of the Organisation for Economic Co-operation and Development (OECD) to implement an international energy programme. A basic aim of the IEA is to foster international cooperation among the 30 IEA participating countries and to increase energy security through energy research, development and demonstration in the fields of technologies for energy efficiency and renewable energy sources.

The IEA Energy in Buildings and Communities Programme

The IEA co-ordinates international energy research and development (R&D) activities through a comprehensive portfolio of Technology Collaboration Programmes (TCPs). The mission of the IEA Energy in Buildings and Communities (IEA EBC) TCP is to support the acceleration of the transformation of the built environment towards more energy efficient and sustainable buildings and communities, by the development and dissemination of knowledge, technologies and processes and other solutions through international collaborative research and open innovation. (Until 2013, the IEA EBC Programme was known as the IEA Energy Conservation in Buildings and Community Systems Programme, ECBCS.)

The high priority research themes in the EBC Strategic Plan 2019-2024 are based on research drivers, national programmes within the EBC participating countries, the Future Buildings Forum (FBF) Think Tank Workshop held in Singapore in October 2017 and a Strategy Planning Workshop held at the EBC Executive Committee Meeting in November 2017. The research themes represent a collective input of the Executive Committee members and Operating Agents to exploit technological and other opportunities to save energy in the buildings sector, and to remove technical obstacles to market penetration of new energy technologies, systems and processes. Future EBC collaborative research and innovation work should have its focus on these themes.

At the Strategy Planning Workshop in 2017, some 40 research themes were developed. From those 40 themes, 10 themes of special high priority have been extracted, taking into consideration a score that was given to each theme at the workshop. The 10 high priority themes can be separated in two types namely ‘Objectives’ and ‘Means’. These two groups are distinguished for a better understanding of the different themes.

Objectives: The strategic objectives of the EBC TCP are as follows:

– reinforcing the technical and economic basis for refurbishment of existing buildings, including financing, engagement of stakeholders and promotion of co-benefits;
– improvement of planning, construction and management processes to reduce the performance gap between design stage assessments and real-world operation;
– the creation of ‘low tech’, robust and affordable technologies;
– the further development of energy efficient cooling in hot and humid, or dry climates, avoiding mechanical cooling if possible;– the creation of holistic solution sets for district level systems taking into account energy grids, overall performance, business models, engagement of stakeholders, and transport energy system implications.

Means: The strategic objectives of the EBC TCP will be achieved by the means listed below:

– the creation of tools for supporting design and construction through to operations and maintenance, including building energy standards and life cycle analysis (LCA);
– benefitting from ‘living labs’ to provide experience of and overcome barriers to adoption of energy efficiency measures;
– improving smart control of building services technical installations, including occupant and operator interfaces;
– addressing data issues in buildings, including non-intrusive and secure data collection;
– the development of building information modelling (BIM) as a game changer, from design and construction through to operations and maintenance.

The themes in both groups can be the subject for new Annexes, but what distinguishes them is that the ‘objectives’ themes are final goals or solutions (or part of) for an energy efficient built environment, while the ‘means’ themes are instruments or enablers to reach such a goal. These themes are explained in more detail in the EBC Strategic Plan 2019-2024.

The Executive Committee

Overall control of the IEA EBC Programme is maintained by an Executive Committee, which not only monitors existing projects, but also identifies new strategic areas in which collaborative efforts may be beneficial. As the Programme is based on a contract with the IEA, the projects are legally established as Annexes to the IEA EBC Implementing Agreement. At the present time, the following projects have been initiated by the IEA EBC Executive Committee, with completed projects identified by (*) and joint projects with the IEA Solar Heating and Cooling Technology Collaboration Programme by (☼):

Annex 1: Load Energy Determination of Buildings (*)

Annex 2: Ekistics and Advanced Community Energy Systems (*)
Annex 3: Energy Conservation in Residential Buildings (*)
Annex 4: Glasgow Commercial Building Monitoring (*)
Annex 5: Air Infiltration and Ventilation Centre
Annex 6: Energy Systems and Design of Communities (*)
Annex 7: Local Government Energy Planning (*)
Annex 8: Inhabitants Behaviour with Regard to Ventilation (*)
Annex 9: Minimum Ventilation Rates (*)
Annex 10: Building HVAC System Simulation (*)
Annex 11: Energy Auditing (*)
Annex 12: Windows and Fenestration (*)
Annex 13: Energy Management in Hospitals (*)
Annex 14: Condensation and Energy (*)
Annex 15: Energy Efficiency in Schools (*)
Annex 16: BEMS 1- User Interfaces and System Integration (*)
Annex 17: BEMS 2- Evaluation and Emulation Techniques (*)
Annex 18: Demand Controlled Ventilation Systems (*)
Annex 19: Low Slope Roof Systems (*)
Annex 20: Air Flow Patterns within Buildings (*)
Annex 21: Thermal Modelling (*)
Annex 22: Energy Efficient Communities (*)
Annex 23: Multi Zone Air Flow Modelling (COMIS) (*)
Annex 24: Heat, Air and Moisture Transfer in Envelopes (*)
Annex 25: Real time HVAC Simulation (*)
Annex 26: Energy Efficient Ventilation of Large Enclosures (*)
Annex 27: Evaluation and Demonstration of Domestic Ventilation Systems (*)
Annex 28: Low Energy Cooling Systems (*)
Annex 29: ☼ Daylight in Buildings (*)
Annex 30: Bringing Simulation to Application (*)
Annex 31: Energy-Related Environmental Impact of Buildings (*)
Annex 32: Integral Building Envelope Performance Assessment (*)
Annex 33: Advanced Local Energy Planning (*)
Annex 34: Computer-Aided Evaluation of HVAC System Performance (*)
Annex 35: Design of Energy Efficient Hybrid Ventilation (HYBVENT) (*)
Annex 36: Retrofitting of Educational Buildings (*)
Annex 37: Low Exergy Systems for Heating and Cooling of Buildings (LowEx) (*)
Annex 38: ☼ Solar Sustainable Housing (*)
Annex 39: High Performance Insulation Systems (*)
Annex 40: Building Commissioning to Improve Energy Performance (*)
Annex 41: Whole Building Heat, Air and Moisture Response (MOIST-ENG) (*)
Annex 42: The Simulation of Building-Integrated Fuel Cell and Other Cogeneration Systems (FC+COGEN-SIM) (*)
Annex 43: ☼ Testing and Validation of Building Energy Simulation Tools (*)
Annex 44: Integrating Environmentally Responsive Elements in Buildings (*)
Annex 45: Energy Efficient Electric Lighting for Buildings (*)
Annex 46: Holistic Assessment Tool-kit on Energy Efficient Retrofit Measures for Government Buildings (EnERGo) (*)
Annex 47: Cost-Effective Commissioning for Existing and Low Energy Buildings (*)
Annex 48: Heat Pumping and Reversible Air Conditioning (*)
Annex 49: Low Exergy Systems for High Performance Buildings and Communities (*)
Annex 50: Prefabricated Systems for Low Energy Renovation of Residential Buildings (*)
Annex 51: Energy Efficient Communities (*)
Annex 52: ☼ Towards Net Zero Energy Solar Buildings (*)
Annex 53: Total Energy Use in Buildings: Analysis and Evaluation Methods (*)
Annex 54: Integration of Micro-Generation and Related Energy Technologies in Buildings (*)
Annex 55: Reliability of Energy Efficient Building Retrofitting - Probability Assessment of Performance and Cost (RAP-RETRO) (*)
Annex 56: Cost Effective Energy and CO_2 Emissions Optimization in Building Renovation (*)

Annex 57: Evaluation of Embodied Energy and CO_2 Equivalent Emissions for Building Construction (*)
Annex 58: Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements (*)
Annex 59: High Temperature Cooling and Low Temperature Heating in Buildings (*)
Annex 60: New Generation Computational Tools for Building and Community Energy Systems (*)
Annex 61: Business and Technical Concepts for Deep Energy Retrofit of Public Buildings (*)
Annex 62: Ventilative Cooling (*)
Annex 63: Implementation of Energy Strategies in Communities (*)
Annex 64: LowEx Communities - Optimised Performance of Energy Supply Systems with Exergy Principles (*)
Annex 65: Long-Term Performance of Super-Insulating Materials in Building Components and Systems (*)
Annex 66: Definition and Simulation of Occupant Behavior in Buildings (*)
Annex 67: Energy Flexible Buildings (*)
Annex 68: Indoor Air Quality Design and Control in Low Energy Residential Buildings (*)
Annex 69: Strategy and Practice of Adaptive Thermal Comfort in Low Energy Buildings
Annex 70: Energy Epidemiology: Analysis of Real Building Energy Use at Scale
Annex 71: Building Energy Performance Assessment Based on In-situ Measurements
Annex 72: Assessing Life Cycle Related Environmental Impacts Caused by Buildings
Annex 73: Towards Net Zero Energy Resilient Public Communities
Annex 74: Competition and Living Lab Platform
Annex 75: Cost-effective Building Renovation at District Level Combining Energy Efficiency and Renewables
Annex 76: ☼ Deep Renovation of Historic Buildings Towards Lowest Possible Energy Demand and CO_2 Emissions
Annex 77: ☼ Integrated Solutions for Daylight and Electric Lighting
Annex 78: Supplementing Ventilation with Gas-phase Air Cleaning, Implementation and Energy Implications
Annex 79: Occupant-Centric Building Design and Operation
Annex 80: Resilient Cooling
Annex 81: Data-Driven Smart Buildings
Annex 82: Energy Flexible Buildings Towards Resilient Low Carbon Energy Systems
Annex 83: Positive Energy Districts
Annex 84: Demand Management of Buildings in Thermal Networks
Annex 85: Indirect Evaporative Cooling
Annex 86: Energy Efficient Indoor Air Quality Management in Residential Buildings

Working Group - Energy Efficiency in Educational Buildings (*)
Working Group - Indicators of Energy Efficiency in Cold Climate Buildings (*)
Working Group - Annex 36 Extension: The Energy Concept Adviser (*)
Working Group - HVAC Energy Calculation Methodologies for Non-residential Buildings (*)
Working Group - Cities and Communities
Working Group – Building Energy Codes

IEA EBC Annex 71: Building energy performance assessment based on in-situ measurements

Annex 71 in general

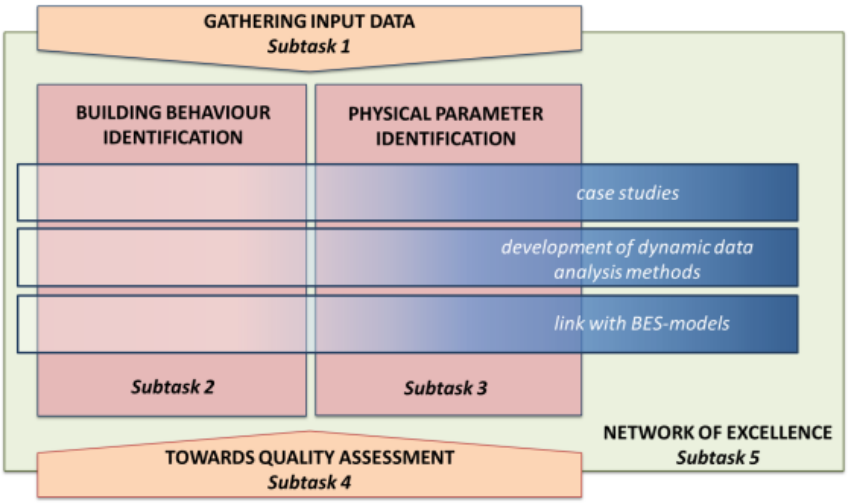
Decreasing the energy use in buildings can only be achieved by an accurate characterization of the as-built energy performance of buildings. This is mainly for two reasons. First of all, despite the ever more stringent energy legislation for new and renovated buildings, monitoring the actual energy performances reveals in many cases a significant performance gap compared to the theoretically designed targets. Secondly, the increasing need for integration of renewable energy stresses on the existing energy systems. This can be remedied by using intelligent systems and energy grids that are aware of the actual status of the buildings.

Within IEA EBC Annex 58, a first step was taken to characterize the actual energy performance of buildings based on full scale dynamic measurements. The onsite assessment methods applied within this project mainly focused on the thermal performance of the building fabric. By investigating the possibilities and limitations of black and grey box system identification models, guidelines were developed on how to assess the overall heat transfer coefficient of a building starting from dynamic measured data instead of static co-heating tests. Notwithstanding Annex 58 showed that onsite quality checks are feasible, the project highlighted at the same time the need of non-intrusive methods. Annex 71 progressed with the achievements of IEA EBC Annex 58, but aimed to make the step towards monitoring in-use buildings. The IEA EBC Annex 71 project focused on the **development of replicable methodologies embedded in a statistical and building physical framework to characterize and assess the actual energy performance of buildings starting from on board monitored data of in-use buildings**.

Structure of the project

The IEA EBC Annex 71-project was limited to residential buildings, for which the development of characterisation methods as well as of quality assurance methods have been explored. Characterisation methods aim to translate the (dynamic) behaviour of a building into a simplified model that can inform predictive control, fault detection, optimisation of district energy systems,… Within Annex 71 we refered to this as building behaviour identification. Quality assurance methods aim to pinpoint some of the most relevant actual building performance metrics. This part is referred to as physical parameter identification.

A reliable characterisation and quality assurance is strongly dependent on the availability and quality of the input data. At the same time, the expected quality and reliability of the outcome will be determined by the required accuracy to perform a quality assurance. As a result, the analysis of potential methods was steered by both the possibilities and limitations of the available input data as well as by the requested outcome to perform real quality checks. Therefore, the research project was organised as illustrated in the figure below and five subtasks were defined:



Subtask 1 investigated the possibilities and limitations of common data bases and monitoring systems. This subtask is strongly related to subtasks 2 and 3 by linking the available input data – as much as possible based on existing (non-intrusive) monitoring systems and data bases – to the accuracy of the predicted outcome. A state of the art survey of existing methods, their costs, timeframe and typical accuracy was made. In a second part the step from monitoring to current on board measuring methods was reviewed. Finally, the application of an on-site measured heat transfer coefficient within the global energy efficiency framework was proposed.

Subtask 2 focused on the development of dynamic data analysis methods suitable for describing the energy dynamics of buildings. Based on in-situ monitored data, prediction models were applied and optimised that can be used in model predictive control, fault detection, and design, control and optimisation of district energy systems,… Necessary data acquisition, development of methodologies and accuracy and reliability of the building behaviour identification models was investigated.

The focus of **Subtask 3** was on development of dynamic data analysis methods suitable for physical parameter identification of buildings. Contrary to Subtask 2, in which the identified parameters do not necessarily have a physical meaning (or do not correspond to the actual value), parameter identification aims to characterize the actual physical parameter. Subtask 3 hence investigated which methodologies are most suitable to determine the actual energy performance indicators of buildings, such as the overall heat loss coefficient, solar aperture,… As in subtask 2, the focus was on methodologies that can be used on occupied buildings, making use of (limited) monitored data.

Subtask 4 investigated to what extent the methodologies developed in ST2 and ST3 can be used in a quality assessment framework. A large survey was performed amongst possible stakeholders on interest and expectations of quality assessment methods based on in-situ measured data. The main focus was on the determination of the actual heat loss coefficient of a building in an easy, cheap and reliable way, so that it can replace the calculated design value in energy performance certifications. That way, subtask 4 made the link between the annex-participants and certification bodies, government, practitioners in the field. At the same time, subtask 4 gave the necessary boundary conditions (reliability, accuracy, cost,…) the methodologies have to fulfil to be applicable in real life quality checks.

Subtask 5 continued the collaboration with DYNASTEE (www.dynastee.info), started within Annex 58. This collaboration showed to be extremely fruitful in dissemination of the results, collecting and distributing research outcomes, and organizing conferences, workshops and training courses.

The **BES-validation exercise i**nvestigated the reliability of common building energy simulation programs. There has been significant work undertaken in past IEA EBC Annexes on validation, particularly inter-program comparisons (e.g BESTEST) and empirical validation on test cells. In Annex 58, empirical validation was extended to full-scale buildings, namely the Twin Houses at Fraunhofer IBP's test site in Holzkirchen, Germany. In this research, the focus was on fabric performance with simple internal heat

gain schedules. The empirical validation undertaken in IEA Annex 71 extended the scope of the experiments in the Twin Houses by including underfloor heating systems and realistic occupancy schedules.

Overview of the working meetings

The preparation and working phase of the project encompassed nine working meetings:

Meeting	Place, date	Attended by
Kick off meeting	Leuven, Belgium, October 2016	49 participants
Second preparation meeting	Loughborough, UK, April 2017	61 participants
First working meeting	Chambéry, France, October 2017	62 participants
Second working meeting	Brussels, Belgium, April 2018	56 participants
Third working meeting	Innsbruck, Austria, October 2018	55 participants
Fourth working meeting	Bilbao, Spain, April 2019	59 participants
Fifth working meeting	Rosenheim, Germany, October 2019	56 participants
Sixth working meeting	On-line meeting, April 2020	50 participants
Seventh working meeting	On-line meeting, October 2020	50 participants
Eighth working meeting	On-line meeting, April 2021	56 participants
Closing event	Salford, UK, September 2021	

During these meetings, working papers on different subjects related to full scale testing and data analysis were presented and discussed. Over the course of the Annex, different experiments on characterisation and quality assessment were undertaken, and several common exercises on data analysis methods were introduced and solved.

Outcome of the project

The IEA EBC Annex 71-project worked closely together with the Dynastee-network (www.dynastee.info). One of the deliverables of the Annex project was the enhancement of this network and promoting of actual building performance characterization based on full scale measurements and the appropriate data analysis techniques. This network of excellence on full scale testing and dynamic data analysis organizes on a regular basis events such as international workshops, annual training, with outputs that support organisations interested in full scale testing campaigns.

In addition to the network of excellence, the outcome of the Annex 71-project has been described in a set of reports, including:

IEA EBC Annex 71 – Building energy performance assessment based on in-situ measurements: challenges and general framework (joint report of Subtasks 1 and 4)

IEA EBC Annex 71 – Building energy performance assessment based on in-situ measurements: building behaviour identification (report of Subtask 2)

IEA EBC Annex 71 – Building energy performance assessment based on in-situ measurements: physical parameter identification (report of Subtask 3)

IEA EBC Annex 71 – Building energy performance assessment based on in-situ measurements: design, description and results of the validation of building energy simulation programs (report of the BES-validation exercise)

IEA EBC Annex 71- Building energy performance assessment based on in-situ measurements: project summary report

List of participants and coregroup

In total 42 institutes from 11 countries participated in the IEA EBC Annex 71-project. The different participants are listed below:

Austria:	Max Blöchle, Austrian Institute of Technology
	A. Susanne Metzger, Vienna University of Technology
	Gabriel Rojas, University of Innsbruck
Belgium:	**Geert Bauwens, KU Leuven (subtask 3 co-leader)**
	Hans Bloem, INIVE – Dynastee (subtask 5 co-leader)
	Karel De Sloover, Belgian Building Research Institute
	Jade Deltour, Belgian Building Research Institute

	Arash Erfani, KU Leuven
	Nicolas Heijmans, Belgian Building Research Institute
	Eline Himpe, Ghent University
	Evi Lambie, KU Leuven / EnergyVille
	Gabrielle Masy, UCLouvain
	Glenn Reynders, VITO EnergyVille (subtask 2 co-leader)
	Katia Ritosa, KU Leuven (secretary)
	Staf Roels, KU Leuven (operating agent)
	Dirk Saelens, KU Leuven / EnergyVille (subtask 2 co-leader)
	Marieline Senave, KU Leuven / EnergyVille (secretary)
	Liesje Van Gelder, BCCA (subtask 4 leader)
	Matthias Van Hove, Ghent University
	Luk Vandaele, INIVE – Dynastee (subtask 5 co-leader)
	Xiang Zhang (Jason), KU Leuven
Denmark:	Peder Bacher, Technical University of Denmark
	Morten Brøgger , Danish Building Research Institute (SBI) , Aalborg University
	Christoffer Rasmussen, Technical University of Denmark
	Kim B. Wittchen, Danish Building Research Institute (SBI) , Aalborg University
Estonia:	Anti Hamburg, Tallinn University of Technology
	Tuule Mall Kull, Tallinn University of Technology
	Karl-Villem Võsa, Tallinn University of Technology
France:	Balsam Ajib, Ecole des Mines de Douai
	Rémi Bouchie, CSTB
	U Bramkamp, Groupe Atlantic
	Antoine Caucheteux, Cerema
	Lorena de Carvalho Araujo, CSTB
	Tianyun Gao, CSTB
	Jordan Gauvrit, Cerema
	Christian Ghiaus, INSA Lyon
	Myriam Humbert, Cerema/ DterOuest/ DLRB
	Sarah Juricic, LOCIE, Université Savoie Mont-Blanc
	Guillaume Pandraud, Saint-Gobain Isover
	Simon Rouchier, Université Savoie Mont-Blanc
	Simon Thebault, CSTB
Germany:	Witta Ebel, Passive House Institute

	Philippe Gorzalka , German Aerospace Center
	Lucia Hanfstaengl, Rosenheim Technical University of Applied Sciences
	Matthias Kersken, Fraunhofer Institute for Building Physics IBP (BES validation co-leader)
	Michael Parzinger, Rosenheim Technical University of Applied Sciences
	Jacob Schmiedt, German Aerospace Center
	Uli Spindler, Rosenheim Technical University of Applied Sciences
	Markus Wirnsberger , Rosenheim Technical University of Applied Sciences
Netherlands:	Twan Rovers, Saxion Hogeschool Enschede
	Marleen Spiekman, TNO
	Christian Struck, Saxion Hogeschool Enschede
Norway:	Arild Gustavsen, Research Centre on Zero Emission Buildings, Norwegian University of Science and Technology
	Kristian Stenerud Skeie, Norwegian University of Science and Technology
	Xingji Yu, Norwegian University of Science and Technology
Spain:	Pablo Eguia, Vigo University
	Aitor Erkoreka, University of the Basque Country (UPV/EHU)
	Catalina Giraldo-Soto, University of the Basque Country
	Enrique Granada, Vigo University
	Hector Herrarda, CIEMAT
	Maria Jose Jimenez, CIEMAT (subtask 5 co-leader)
	Sandra Martínez Mariño, Vigo University
	Gerard Mor Martinez, JRC / CIMNE
	Irati Uriarte Perez de Nanclares, University of the Basque Country (UPV/EHU)
Switzerland:	Chirag Deb, ETH Zurich
UK:	David Allinson, Loughborough University
	Zoe De Grussa, BBSA (ESSO)
	Cliff Elwell, UCL Energy Institute
	David Farmer, Leeds Beckett University
	Richard Fitton, University of Salford (subtask 1 leader)
	Martin Fletcher, Leeds Beckett University
	Virginia Gori, UCL Energy Institute
	Chris Gorse, Leeds Beckett University
	Rajat Gupta, Oxford Brookes University
	Adam Hardy, Leeds Beckett University
	Ross Holleron, Knauf Insulation

Frances Hollick, UCL

Richard Jack, Build Test Solutions (subtask 4 co-leader)

Matthew Li, Loughborough University

Kevin Lomas, Loughborough University

Eirini Mantesi, Loughborough University

Behzad Sodagar, University of Lincoln

Paul Strachan, University of Strathclyde (BES validation co-leader)

Table of content

Abbreviations

List of frequently used abbreviations

Abbreviation	Meaning
AFDD	Automatic fault detection and diagnosis
AHU	Air handling unit
ANN	Artificial neural network
API	Application programming interface
AR	Autoregressive
ARIMAX	Autoregressive integrated moving average with exogenous variables
ARMAX	Autoregressive, moving average model with exogenous inputs
ARX	Autoregressive with exogenous inputs
AT	Austria
BBRI	Belgian building research institute
Bed1	Bedroom 1 of the house 1 (H1)
BES	Building energy simulation
BITS	Building integrated technical systems
CDR	Combined decision rule
CED	Cumulated energy demand
CI	Confidence interval
DHI	Diffuse horizontal irradiance
DHW	Domestic hot water
DK	Denmark
DM	Default method
DNI	Direct normal irradiance
EN	European norm
EPBD	Energy performance of buildings directive
ES	Spain
FDD	Fault detection & diagnosis
FIR	Finite impulse response
FIR-RLS	Finite impulse response - recursive least squares
FMI	Functional mock-up interface

FMU	Functional mock-up unit
GA	Genetic algorithm
GAM	Generalized additive models
GB	Grey-box
GBT	Gradient boosted trees
GHG	Greenhouse gas
GWP	Global warming potential
HMI	Human-machine interfaces
House 1 (H1)	south-facing end-terrace house in Gainsborough
HP	Heat pump
HTC	Heat transfer coefficient W/K
HVAC	Heating, ventilation, and air-conditioning
IEA EBC	Energy in Buildings and Communities Programme of the International Energy Agency
IHG	Internal heat gain
IT	Information technology
KPI	Key performance indicator
kWh	Kilowatt hours: 1 kWh = 3.6 MJ
LC	Life cycle
LCI	Life cycle impact
LCIA	Life cycle impact analysis
LR	Linear regression
L	Lounge
LSTM	Long-short term memory
MA	Moving average
MJ	Mega joule; 1 kWh = 3.6 MJ
MLP	Multilayer perceptron
MLR	Multi-variate linear regeression
MPC	Model predictive control
MRI	Model predictive control relevant identification
MSL	Mean sea level
MSPE	Multi-step ahead prediction error
MVHR	Mechanical ventilation system with heat recovery

N	Control horizon
NARX	Nonlinear autoregressive with exogenous input
NNAR	Neural Network Auto-Regressive integrated moving average
NO	Norway
NRE	Non-renewable energy (fossil, nuclear, wood from primary forests)
NZEB	Nearly zero energy building or nearly zero emissions building
OCP	Optimal control formulation
OSPE	One-step ahead prediction error
PCA	Principle component analysis
PE	Primary energy
PRBS	Pseudo-random binary sequence
PSA	Plataforma solar de Almeria
PT	Portugal
PV	Photovoltaic
PWARX	Piece-wise ARX method
RBC	Rule-based controller
RC	Resistance capacitance
RES	Renewable energy sources
RF	Random forest
RMSE	Root mean square error
ROLBS	Randomly ordered logarithmic binary sequence
RR	Round Robin
SH	Space heating
SIP	Structural insulated panels
SRI	Smart readiness indicator
SS	State space
ST	Subtask
SVM	Support vector machine
U-value	Thermal transmittance of a building element
VFD	Variable frequency drives

Definitions

Explanation of frequently used definitions:

Building behaviour characterization: translation of the dynamic behaviour of a building into a simplified model that can be used in applications such as model predictive control, fault detection, optimisation and design of buildings

Building Energy Simulation (BES): computer modelling based on building physics, used in the evaluation of energy and environmental aspects of building performance.

Building stock: a collection of buildings in a country, region, municipal area or estate. It can include dwellings, offices, factories, shops, educational establishments, agricultural buildings and so on.

Building thermal envelope: defined in ISO EN 52016 as "*total area of all elements of a building that enclose thermally conditioned spaces through which thermal energy is transferred, directly or indirectly, to or from the external environment".*

B-splines: B-spline is piecewise polynomial, where the m^{th} order B-splines signify series of polynomials of degree m-1. The key feature of B-splines is that the point-wise sum of infinitely B-spline series for the entire range of interest is always equal to one.

Delivered energy: energy, expressed per energy carrier, supplied to the technical building systems through the system boundary, to satisfy the uses taken into account (heating, cooling, ventilation, domestic hot water, lighting, appliances, etc.).

Embodied energy: Comprises the cumulated primary energy demand for production, transportation and disposal of building components, appliances, renewable energy generation units and building construction measures within building renovation.

Energy carrier: substance or phenomenon that can be used to produce mechanical work or heat or to operate chemical or physical processes. **g-value (Total Solar Energy Transmittance):** The total energy transmittance of a glazing, indicates the proportion of the incident radiation which is transmitted by the glazing, based on EN 410:2011.

Energy need for domestic hot water: heat to be delivered to the needed amount of domestic hot water to raise its temperature from the cold network temperature to the prefixed delivery temperature at the delivery point.

Energy need for heating or cooling: heat to be delivered to or extracted from a conditioned space to maintain intended temperature conditions during a given period of time.

Energy source: source from which useful energy can be extracted or recovered either directly or by means of a conversion or transformation process.

Energy use for lighting: electrical energy input to the lighting system.

Energy use for space heating or cooling or domestic hot water: energy input to the heating, cooling or hot water system to satisfy the energy need for heating, cooling or hot water respectively.

Energy use for ventilation: electrical energy input to the ventilation system for air transport and heat recovery (not including the energy input for preheating the air).

Exported energy: Energy, expressed per energy carrier, delivered by the technical building systems through the system boundary and used outside the system boundary.

HTC (Heat Transfer Coefficient): defined in ISO 13789 as the "*heat flow rate divided by temperature difference between two environments*." It represents the steady-state aggregate total fabric and ventilation heat transfer from the entire thermal envelope in Watts per kelvin of temperature difference (ΔT) between the internal and external environments, and is expressed in Watts/Kelvin (W/K) (BSI, 2017). In this document, HTC typically refers to the fabric heat transfer by conduction and air infiltration, unless explicitly stated otherwise.

Primary energy: Energy found in the nature that has not been subject to any conversion or transformation process. It is energy contained in raw fuels and other forms of energy received as input. It can be non-renewable or renewable.

Renewable energy Sources (RES): energy from sources that are not depleted by extraction, such as solar energy (thermal and photovoltaic), wind, water power, renewed biomass. (definition different from the one used in Directive 2010/31/EU).

Solar aperture (solar transmittance, gA value): The solar transmittance of an observed transparent building element as a function of window properties, window orientation, shading obstacles, and other variables which are infeasible to observe alone.

System boundary: boundary that includes within it all areas associated with the building (both inside and outside the building) where energy is consumed or produced.

Test case: In the context of this Annex a test case is a dwelling subjected to (extensive) monitoring campaigns in order to get detailed measurement data, which are used, in HTC assessment.

Thermal zone: defined in ISO EN 52016 as **"***internal environment with assumed sufficiently uniform thermal conditions to enable a thermal balance calculation".* In the zoning procedure neighbouring spaces with similar services and comfort settings are merged in thermal zones. A dwelling is often treated as a two-zone building, consisting of a day and night zone.

U-value: U-value, also known as thermal transmittance, is defined in ISO 7345 as the "*heat flow rate in the steady state divided by area and by the temperature difference between the surroundings on both sides of a flat uniform system*" in unit W/(m²·K).

List of symbols

Explanation of frequently used simbols:

Symbol	Description	Unit
C	Heat capacity	kJ/K
C_{el}	Electricity price	€/kWh
CO_2	Concentration of CO_2 in the air	ppm
DHI	Diffuse horizontal irradiance	W/m^2
dir	Direction	
DNI	Direct Normal irradiance	W/m^2
dT_a	change of temperature	°C/h
GHI	Global horizontal irradiance	W/m^2
HTC	Heat transfer coefficient	W/K
I_sol	global solar irradiance on the vertical surface	W/m^2
IHG	Internal heat gains	W
L	Weight between terms of objective function	
N	Prediction horizon	h
P_{el}	Heat pump's electricity consumption	kWh
q_{tres}	separation threshold	
R^2	Coefficient of determination	
RH	Relative (air) humidity	%
S_{hp} (u_1)	Status oft he heat pump (on/off)	
sp	Speed	m/s
$\hat{T}_{in,k+1}$	Estimation of indoor air temperature for the next time step	°C
T_i	Indoor air temperature	°C
T_e	Ambient temperature	°C
T_{low}	Lower band of thermal comfort for air temperature	°C
T_s (u_2)	Supply water temperature	°C
T_{up}	Upper band of thermal comfort for air temperature	°C
VFR	Volumetric flow rate of ventilation system	m^3/h
v_k	Slack variable for defining soft constraints	
x	States of the model	
$\hat{y}$	Estimated output	

y_m	Measured output	
γ	threshold for bisquare robust estimation	MJ/h
λ	Lambda-value (value for the insulating capacity of a material)	W/m·K
Φ_h	Heat injected to building by the heating system	W

1. Introduction

The building sector accounts for about 40% of total final energy use and harbours enormous potential to save energy and reduce CO_2-emissions in a cost effective way. In order to reduce the energy use of buildings, policy makers impose ever more stringent requirements with regard to energy performance of new buildings and renovated buildings, and the use of renewable resource. Most compliance checks and labelling of the energy performances of buildings are done in the design phase by calculating the theoretical energy use. But, despite regulation and policy enforcements, monitoring of actual energy performances reveals in many cases a significant gap with theoretically designed targets. Sources of deviation between the actual and expected performances can be attributed to the design phase (limitations, inaccuracies and assumptions in the numerical models used to predict the energy performance), the construction phase (quality of workmanship and differences between assumed and actual installed materials, components and systems) and the operation phase (malfunctioning of systems and/or no match between assumed and actual building usage). The currently observed performance gap, in combination with the increasing integration of innovative systems such as intelligent elements, low energy technologies, active solar systems, etc., accentuate the need to develop reliable methods and procedures that can be applied on site to assess the actual performance of buildings.

Within IEA EBC Annex 58 a first step has been taken to characterize the actual energy performance of buildings based on full scale dynamic measurements. Annex 58, however, was mainly restricted to the thermal performance of the building envelope, making use of rather intrusive tests and focusing on scale models or test buildings. The current Annex 71 project makes the step towards monitoring in-use buildings to obtain reliable characterization of the actual performance of buildings. This means that the intrusive, dedicated tests are avoided in favour of assessment methods based on on-board monitoring systems.

Specifically, in the scope of Subtask 2 – presented in this report - characterisation methods are studied for their ability to translate the (dynamic) behaviour of a building into a simplified model that can be used in applications such as model predictive control, fault detection, optimisation of district energy systems,... In contrast, Subtask 3 of the Annex 71 project focusses on quality assurance methods that aim to pinpoint some of the most relevant actual building performances, such as the overall heat loss coefficient of a building, the energy efficiency of the heating (cooling) system, air tightness and solar absorption,...

1.1. Context and ambitions for Subtask 2

Subtask 2 focuses on the development of data analysis methods suitable for describing and predicting the energy dynamics of buildings. Such modelling techniques are gaining importance in the ongoing energy transition. Striving towards a complete decarbonisation of the building – and energy – sector, the large-scale adoption or renewable energy sources introduces a new paradigm where not only the amount of energy use but also the time of usage becomes significant. If society is to move towards a renewable based energy system, we will have to take care of matching the energy demand with the intermittent production of renewable energy sources.

Many studies have already showed that buildings can play a significant role in this changing energy context by transitioning from passive consumers to be active prosumers, which are able to adjust their energy use according to the actual level of energy in the energy networks. They need to consume more during periods with more renewable energy in the networks e.g. by storing energy, and/or reduce the energy use during shortages of energy in the networks (Patteeuw et al., 2015). Buildings needs to become energy flexible (Jensen, 2019). To unlock this flexibility (model) predictive control strategies have been proven to facilitate the harvesting of flexibility sources (e.g. batteries, domestic hot water tanks and building thermal mass...) to optimize the use of renewable energy by taking into account the forecasted energy demand of buildings. A critical step in the success of such optimal control frameworks however resides in the ability of the controller to predict the future energy demand of the building and the response of the building towards changes in the control signal.

Providing flexibility services towards the energy system is however only one of the evolutions – and business cases – related to smart buildings. The European Commission recently introduced the Smart Readiness Indicator (SRI) for buildings which focuses on 9 main services provided by smart buildings, as illustrated in

Figure 1 (Verbeke, Aerts, Reynders, & Waide, 2020)

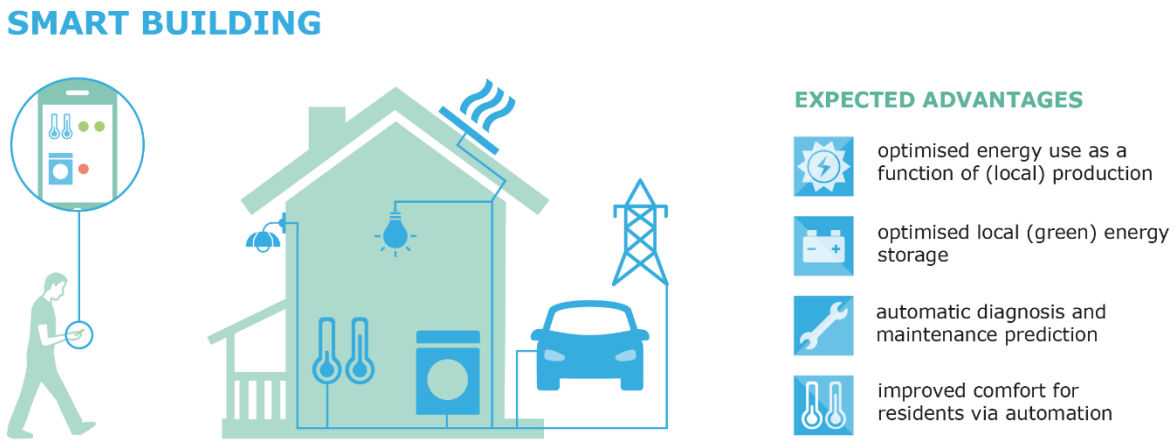


Figure 1. Illustration of a smart building as proposed by the European Commission's second study on the smart readiness indicator (Verbeke et al., 2020)

One of the goals of this Smart Readiness Indicator is to increase the uptake of monitoring and control technologies in buildings, making sure that data on the actual behaviour and energy performance of the building and its technical systems becomes more readily available to optimize the overall operation of the building. In addition to unlocking grid services, smarter buildings should primarily alleviate the energy efficiency losses and comfort issues that generally result from suboptimal control and operation of the building technical systems. Heating, ventilation, and air-conditioning (HVAC) equipment faults and operational errors result in comfort issues and waste of energy in buildings. In order to help the facility managers to identify and fix faults more efficiently, it is essential to have an Automatic Fault Detection and Diagnosis (AFDD) tool, able to automatically detect comfort and energy issues and identify the root faults.

Existing AFDD methods mostly focus on equipment-level fault detection and diagnostics. Almost no attention is given to building level fault diagnosis, considering inter-dependency between equipment through the energy distribution chain (Gao, 2020). While researche carried out exercises in IEA EBC Annex 58 already indicated that data-driven modelling work are able to capture the dynamics in the building behaviour and link this behaviour to physical processes, e.g. by the use of grey-box models, this Annex aims to take that work a step further by exploring innovative data-analysis techniques and by transitioning towards in-use buildings.

1.2. Structure of the report

The present document reports on the main results that were obtained from the activities of Subtask 2. The activities were organized through setting up common exercises in which participants could contribute on a particular topic. The common exercises first explored the existing modelling techniques for building behaviour identification. The results of this analysis are reported in Chapter 2. Subsequently we focused on two applications in which building behaviour identification plays an important role. In Chapter 3 the focus is on the identification of models for Fault-Detection and Diagnostics (FDD) while Chapter 4 focuses on models for Model Predictive Control (MPC) applications. Finally Chapter 5, summarizes the main findings.

2. Modelling techniques for building behaviour identification

Throughout the course of Annex 71 different modelling techniques for building behaviour identification have been evaluated. This evaluation has primarily focussed on predicting the dynamic indoor temperature in single-family houses and on the two applications FDD and MPC.

2.1. State of the art

Building behaviour characterization is defined within this Annex project as the process of identifying data driven models that are able to predict a dynamic output signal (e.g. temperature, heating power, CO2 concentration, etc.) based on times series data on the disturbances (e.g. weather data) and controlled inputs. As such, building behaviour characterization results in a data driven model of the building that can be used for example for AFFD or MPC

The activities presented in this report continue on the work conducted in Subtask 4 of the IEA EBC Annex 58 project. In that subtask the use of in situ testing and smart meter data for building performance characterization was explored. Yet, in contrast to this Annex, the modelling efforts – and common exercises – were primarily focused on controlled test environments. Based on a study carried out in detailed building energy simulations Reynders et. al. demonstrated the ability of grey-box models – i.e. continuous time stochastic differential equations - to obtain a reduced-order building model of which the parameters can be physically interpreted and which can be used for simulating the building thermal behaviour in a robust and accurate manner (Reynders, Diriken, & Saelens, 2014). Yet, in order to obtain robust and accurate results, generally higher-order models (3rd or 4th order models) were needed to represent the dynamics of a building. This step towards higher-order models required the use of more advanced measurement setups (including heat flux measurements) and to shift towards tailored dynamic excitation of the building using PRBS signals. It was concluded that a dedicated testing period was therefore needed in order to apply this modelling approach in practice, when such higher-order grey-box models were needed. In absence of such detailed measurement setups, it is advised to limit the complexity of the model or use physical prior knowledge about the building to limit the amount of unknown parameters.

In (Reynders, Nuytten, & Saelens, 2013), the modellers however show that for short-term predictions up to 1-day ahead predictions, which are typically used for model predictive control applications, less stringent requirements towards the model order and the corresponding data requirements could be made. In that paper, the authors show how both grey-box models and ARX models can be trained to accurately predict the day-ahead indoor temperature. Yet, when focussing on prediction accuracy only, the models do not guarantee physical interpretability of the model parameters and may suffer from long-term temperature drifts when applied in simulations. In that context of short term predictions– where physical interpretability of the model parameters are of lesser or no importance - black-box models are therefore often considered instead of grey-box or white-box models. For example, machine learning and genetic algorithm have been shown to predict the flexibility of the buildings internal mass (Kristensen, Madsen, & Jørgensen, 2004; Wang & Xu, 2006; Xue, Wang, Sun, & Xiao, 2014) for optimal control purposes. Finally, it was noted that whereas previous example shows the application of grey box modelling for common building components using linear time invariant models, the methodology could also be extended to non-linear models. As such, the methodology also allows the characterisation of innovative building concepts such as building integrated PV systems or ventilated facades (Jiménez, Madsen, Bloem, & Dammann, 2008; Lodi, Bacher, Cipriano, & Madsen, 2012).

2.1.1. Round Robin Test Box benchmark case study (Almería, Spain)

To establish the state of play at the start of the Annex project, a group of researchers from different countries and institutions exemplified the capabilities of different building behaviour modelling techniques on a controlled experiment. More precise, they collaborated in a common exercise to develop prediction models for the temperature and the amount of overheating hours of the Round Robin Test Box. The investigated test box has a cubic form, with exterior dimensions of 120x120x120cm^3. The floor, roof, and wall components are identical and exhibit a thickness of 12cm, leaving an inner volume of 96x96x96 cm^3. One wall contains a window component with dimensions 60x60cm^2, inside window frame (glazed part 52x52 cm^2, outside frame 71x71cm^2). A structure is provided around the box, which allows the box to remain free from the thermal influence of the ground. Hence, the box can be considered as floating in free air. A more detailed description is included in part 1 of the Subtask 3 report of IEA EBC Annex 58, where the test box was used as a central test case for developing and testing different measurements setups and system identification techniques (Jimenez, 2014).

Within Subtask 2 of this Annex, the data of one of the Round Robin (RR) test was used by the researches in a first common exercise. The general objective of this common exercise was to identify prediction models for the prediction of the indoor air temperature and the number of overheating hours. Based on measurements obtained in the Round Robin experiment, models were identified that are able to predict the evolution of the indoor temperature during a blind validation period. In addition to the temperature profile, participants are asked to predict the number of overheating hours for the building during this validation period. The latter challenge is included as a link with Subtask 3. The common exercise is based on experimental data obtained during a measurement campaign performed by CIEMAT at PSA in Almería, Spain. This campaign is a second in a series of campaigns organised in different climatic conditions and using different acquisition equipment depending on which institute is performing the test, compared to a previous campaign at BBRI (Belgium). The conducted experiments and hence, the exercise specifically aim to:

- Evaluate the capability to model and obtain accurate energy performance indicators of the test box under sunny weather conditions
- Predict the indoor temperature variations in the Round Robin box for a blind validation data set.

Figure 2. Picture of the Round Robin box set up at the CIEMAT test site

The RR test box was tested at the CIEMAT's LECE laboratory at Plataforma Solar de Almeria (PSA), in the South East of Spain (37.1°N, 2.4°W). The weather at this test site is dry and extremely hot in summer and cold in winter. Temperature swings largely between day and night. Global solar radiation on the south vertical surfaces is very strong in winter, and on the horizontal surfaces, it is very strong in summer. Sky is usually very clear. The experiments used in this common exercise extended over a period of 43 days, starting December 6, 2013 and ending January 17 th, 2014. Testing is done under real outdoor weather conditions. The following outdoor climate sensors installed near the test box are included in the supplied data:

- air temperature (with a solar radiation shield and ventilated),
- vertical global solar radiation (parallel and next to the glazing)
- wind speed
- wind direction (North 0°, East 90°)

Additional meteorological sensors installed at the test site (named as meteo 1, 80 m east from the test box) are also included in the data sets:

- horizontal global solar radiation
- beam solar radiation.
- diffuse solar radiation.
- vertical long wave radiation.
- relative humidity.
- atmospheric pressure

Other sensors installed at the test site (named as meteo 2, 325 m north from the test box):

- Horizontal long wave radiation from the sky.
- vertical global solar radiation facing north

The data for this common exercise comprises 4 test periods, which were particularly designed to provide different levels of dynamic excitation of the test box, promoting the identifiability of models:

Period 1: 6/12/2013 to 17/12/2013. A controlled dynamic heating experiment is conducted using a ROLBS signal to activate a 100W incandescent lamp. The ROLBS signal is mixed with longer periods during which the heating was off.

Period 2: 18/12/2013 to 26/12/2013. A thermostatic control is implemented to control the on/off-setting of the lamp. The thermostat was set to 35°C.

Period 3: 27/12/2013 to 07/01/2014. The set point for the thermostatic control is reduced to 21.5°C to correspond better with normal operating conditions of buildings. Set point and dead band was changed the 27th of December.

Period 4: 08/01/2014 to 14/01/2015. The heating power is switched off resulting in free-floating conditions.

Figure 3 shows a summary of the data shared with the modelling teams depicting the indoor temperature (T_i), the outdoor temperature (T_e), the electrical power used by the heat source (P_h) and the global solar irradiance on the south-facing vertical surface (I_sol).

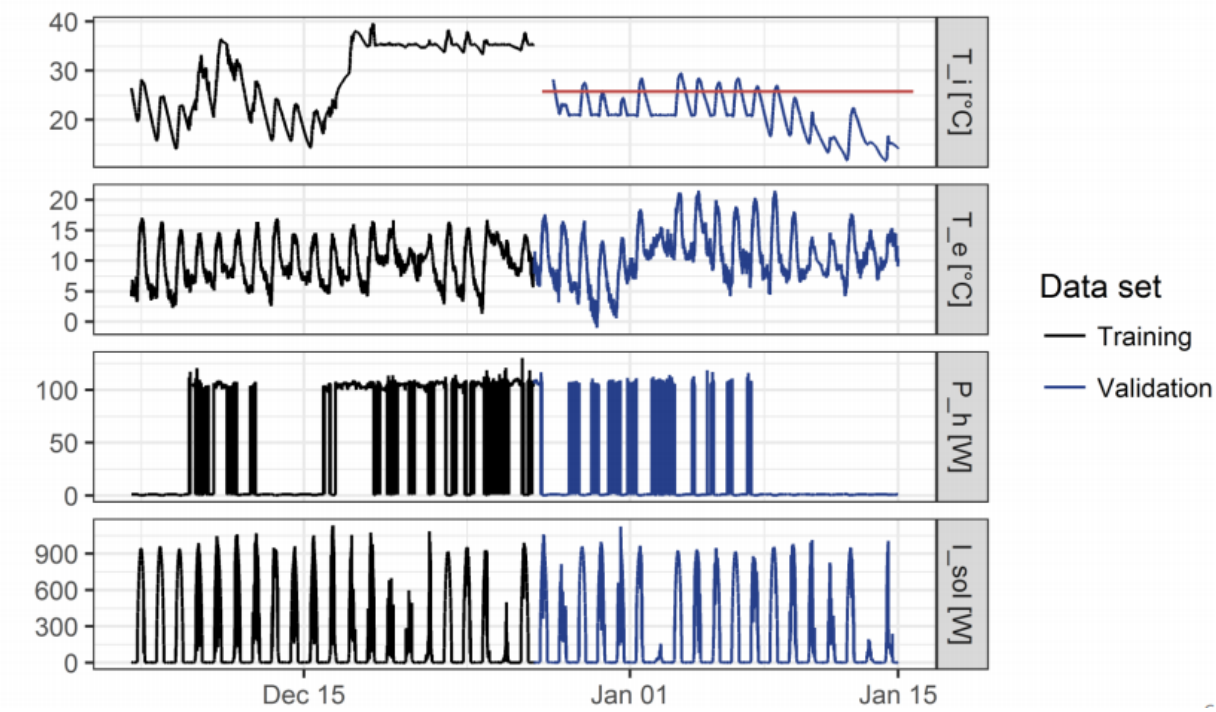


Figure 3. Data for the Round Robin Experiment. The data in black was provided as training data. For the period in blue, the indoor temperature (T_i) was not provided to the modelling teams, but asked for as an output of their modelling exercise.

In total 8 submissions were received covering 3 distinct modelling approaches:

- Calibrated white-box models: 2 modelling teams used the technical documentation of the Round Robin Test box to implement a 'white-box' simulation model in a Building Energy Simulation software (both IDA-ICE and TRNSYS were used). The models were calibrated on the training data.

- Grey-box models: Grey-box models combine a simplified physics based model structure – here in the form of an RC network analogy – and use parameter identification to identify the unknown model parameters from the training data.

3 modelling teams tested a grey-box approach, each using different assumptions in specifying the model equations and different identification techniques.

- Black-box models: Black box models do not rely on any physical insight in formulating the model equations, but rely solely on time series analysis to obtain a prediction model. In this exercise, a piece-wise ARX method (PWARX) was compared against a traditional ARX model.

Figure 4 shows the predicted indoor temperatures of the different modelling teams compared to the measured indoor temperature (in black). Except for model 5, very close agreement between the modelled and measured indoor temperature was observed, especially during the period with thermostatic control between 27/12/2013 to 07/01/2014. During the free-floating period, some models (e.g. model 6) start to drift showing a slower cooling effect. RMSE values vary between 0.5°C and 1.48°C, with an outlier of 5.73°C for model 5. Table 1 also shows that models 2, 7 and 8 can be considered as unbiased, while the other models show some degree of bias towards higher temperatures. Model 3 is the exception, for which the bias error is negative (-0.84°C).

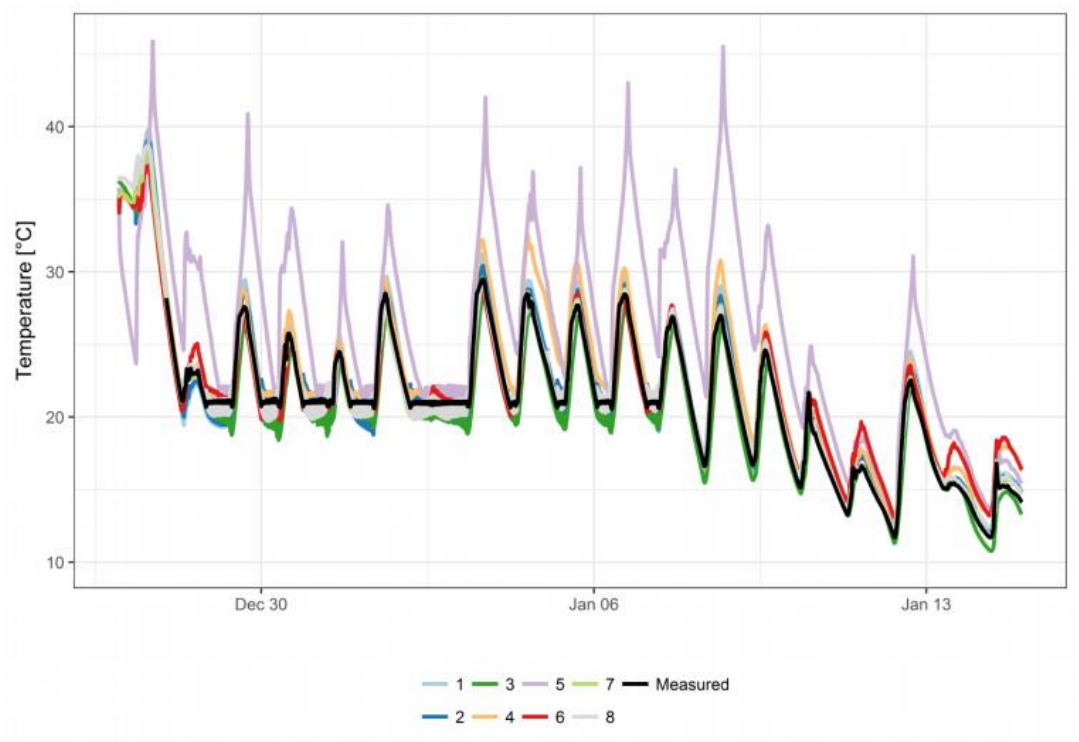


Figure 4. Predicted temperature profile by the different modelling teams (results have been anonymized)

Table 1. Residual analysis for predicted temperatures by the different modelling teams

Team	1	2	3	4	5	6	7	8
RMSE [°C]	0.98	0.83	1.31	1.48	5.73	1.1	0.50	0.63
Bias error [°C]	0.13	0.04	-0.84	1.15	4.64	0.24	-0.01	-0.02
Predicted uncertainty [°C]	0.03	/	3.84	1.47	0.67	6.7	0.01	1.42

While this first exercise in the early stage of the Annex project proved to be valuable to get new participants introduced to the methods and concepts of data-cleaning, model identification and validation (Madsen, 2015), the results largely confirm that for a controlled and well-equipped experiment the tested methods can predict the indoor temperature profile with an acceptable and expected accuracy.

2.2. Data-driven modelling techniques for indoor temperature and heating power prediction in real conditions

This section, the step is made to building behaviour characterization under realistic conditions. In contrast the exercise presented above, the subjects in this section are in-use buildings equipped with high quality – though less extensive – on-board measurement equipment. Under these circumstances, modellers have to face with less ideal excitations of the building (e.g. compared to the PRBS signals typically used in Annex 58), uncontrolled or even unmeasured disturbances by occupants and a higher granularity of data (e.g. only 1 temperature sensor per building or no sub metering on electricity meter).

The remainder of this section summarizes the lessons learned from joined efforts on a building behaviour characterization exercise. More precise, modelling teams worked on training prediction models that were able to predict the indoor temperature and/or heating power for a case study building. The goal was thereby not necessarily to work on the most accurate prediction, but also to explore different (and innovative) modelling techniques and combinations of data sets. As such, the importance of certain measurements and the benefits of certain modelling techniques could be weighed against the costs in measurement equipment and modelling effort.

2.2.1. An occupied semi-detached house as case study (Gainsborough, UK)

In order to gain insight in the applicability of different modelling techniques for building behaviour characterization using on-board monitoring in an existing building, a common exercise was installed. The case study is a semi-detached dwelling of four social houses built in Gainsborough, UK (53.4° N, 0.77° W). The picture at the front of this document gives an overall view of the houses. The houses have been monitored for 3 years, starting October 2012 until November 2015. The building is used by two adults and one child up to January 2013. In March 2013, new tenants (1 adult and 2 children) have moved in. Due to tenancy change, the house was vacant and unheated in January and February 2013. A detailed description of the houses and monitoring campaign can be found in (Sodagar & Starkey, 2016). This common exercise focuses on the South-facing end-terrace house, referred to as House 1 (H1) in that paper. House 1 is a two-story dwelling with a total floor area of 67 m². The living room, kitchen, toilet and entrance hall are located on the ground floor. Two bedrooms and a bathroom are situated on the first floor. Also, a small technical room is located on the landing of the first floor, housing the metering equipment for the PV system as well as the metering HUB for the ventilation system. Floor plans of the House 1 and the neighbouring houses are given in Figure 5.

The houses are constructed using prefabricated structural insulated panels (SIPs) finished in brick or a render clad. The target U-value of the walls was 0.12 W/(m²K). Heat flux measurement carried out to verify the in-situ U-values using the “Averaging method” as detailed in ISO 9869:1994, confirmed a measured U-value of 0.12 W/(m²K). In addition to U-value measurements, air leakage tests have been conducted following the procedures of the “British Institute of Non-Destructive Testing” using an air depressurization test (ATTMA TS1) on the whole building envelope. The dwellings had a design air permeability of 3.00m³/hr.m². Space heating and domestic hot water (DHW) production are provided by a Potterton Promax boiler with a manufacturer’s quoted efficiency of 91%. Per house, space heating is controlled using two room thermostats, in the hall and bedroom 1. The gas consumption of the heating system is monitored with volumetric gas meters. Sub-meters to differentiate between space heating and DHW production were not installed. Therefore, the energy uses for domestic hot water and space heating need to be estimated based on the gas consumption data. A water meter is available that measures the total mains (cold and warm) consumption of the house. The house has also access to a shared rainwater-harvesting tank that is used for outdoor watering and flushing toilets.

Ventilation is provided using a ‘Lo-Carbon Astra’ mechanical ventilation system with heat recovery (MVHR) from Vent-Axia2. The system is equipped with a monitoring system that registers temperature and relative humidity of supply and return air, as well as the electricity consumption of the ventilation unit. Although this MVHR is promised to be an efficient solution for providing ventilation, the paper [1] discusses how the system efficiency can be undermined by occupants misunderstanding how to operate the system.

In addition to the electricity consumption of the MVHR, the total electricity use of the house is monitored together with the output of the PV system. Apart from the measurements of the MVHR no sub-metering has been installed for the electricity consumption. Hence, the measurements cover everything from lighting to appliances, pumps etc.

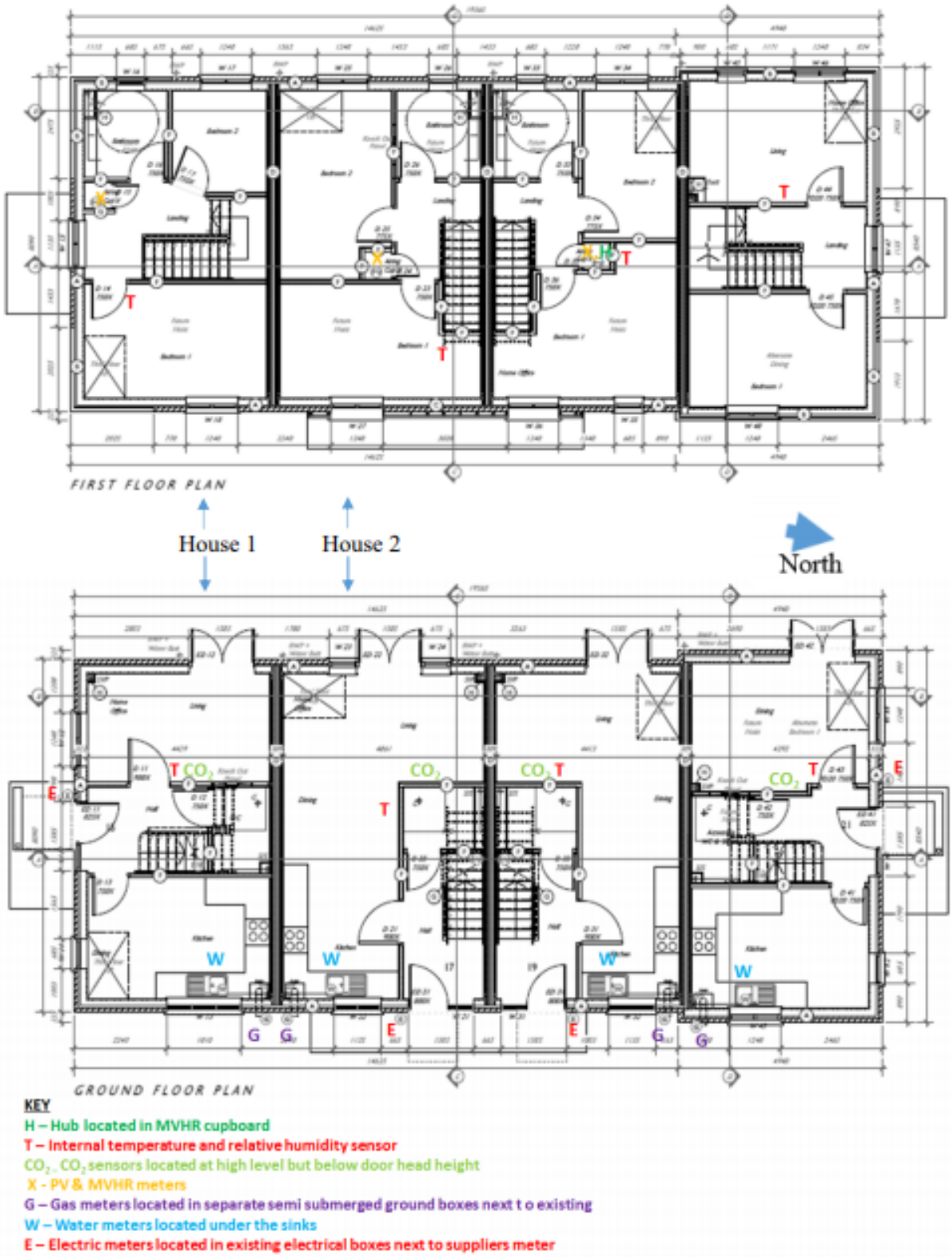


Figure 5. Floor plans of the Gainsborough case study houses (Sodagar & Starkey, 2016).

Time series data is provided for House 1 over the period from October 2012 until November 2015 with a 5 min interval. The dataset covers:

- CO2 concentration (ppm) in the lounge (living room)
- On-site external air temperature (°C) and relative humidity (%)
- Gas consumption meter (m³)
- Total electricity consumption (kWh)
- Supply and return temperature of MVHR (°C)
- Relative humidity supply and return air of MVHR (%)
- Electricity consumption MVHR (kWh)
- PV production (kWh)
- Temperature (°C) and relative humidity (%) in the lounge
- Temperature (°C) and relative humidity (%) in bedroom 1
- Mains water consumption (m³)
- Temperature (°C] and relative humidity (%) in the lounge and bedroom of the adjacent house

In addition to the data for House 1, the indoor temperature profiles of the living room and main bedroom for the adjacent House 2 are also provided. These data can potentially be used by participants as boundary conditions for House 1 when identifying the HTC or the dynamic response of House 1.

As only the outdoor temperature and relative humidity was measured on site, weather data from a nearby weather station is included as an additional dataset. Hourly averaged outdoor temperature, wind speed, wind direction, and global horizontal solar irradiance is collected from a weather station in Waddington (N53.18 W0.55) about 30km from the building site. The weather data is available in a separate data file on the Annex website.

Figure 6 gives an overview of the indoor temperature profiles of House 1. Modelling teams were asked to predict the gas consumption of the house for the period of 16-23 January 2015 – highlighted by the blue box. As explorative first exercise within ST2, the main objective is the exploration of methodologies to obtain an accurate and reliable prediction of the future, hourly energy use for space heating and DHW production. Such predictions are relevant for instance in the context of managing power or district heating networks and the scheduling of peak power units or energy storage. In a second stage this exercise was extended to predict also the indoor temperature – assuming in that case that the heating consumption is known. The following paragraphs present the main findings and encountered challenges as documented by the modelling teams.

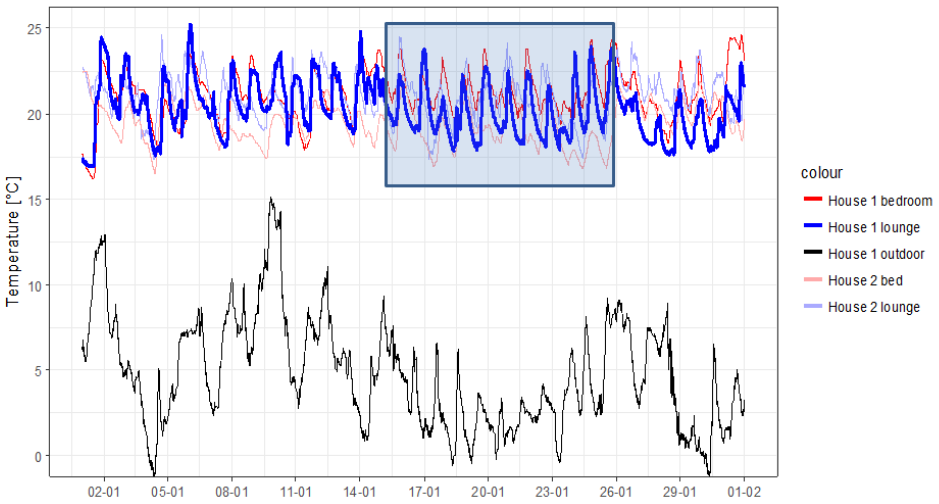


Figure 6. Indoor temperature profiles for the different rooms in House 1, together with the outdoor temperature (black) for that same period) The highlighted period is the week for which the predictions were asked.

2.2.2. Predicting the hourly gas consumption

Four modelling teams have explored different modelling approaches to identify a prediction model capable of predicting the future gas consumption for the Gainsborough case. As asked for in the exercise, they explored a broad range of model types using also different combinations of input data, sampling frequencies etc. The goal of the exercise was therefore not to find the most accurate prediction model, but to explore the range of approaches in their balance between modelling effort and data collection costs against prediction accuracy. The different modelling approaches covered the range from being purely data-driven – black-box – models to physics-based – white-box and grey-box – approaches.

A first modelling team, for example, used the heat balance equations together with the measured indoor temperatures and outdoor climate data to predict the heating power from that the gas consumption. For this method to work construction data were used to specify the heat transfer coefficient of the building and additional assumptions were needed to estimate the level of internal heat gains, the domestic hot water consumption and the efficiency of the boiler. Based on expert input for these model parameters, the modelling team was able to use this approach to predict the monthly averaged and weekly averaged heating consumptions with an acceptable accuracy during cold winter periods. The uncertainty during warmer periods, with lower heating consumptions was found to be higher. Hourly predictions were not deemed feasible with this method, as it does not include any dynamic effects in the model equations nor in the specifications of inputs and outputs.

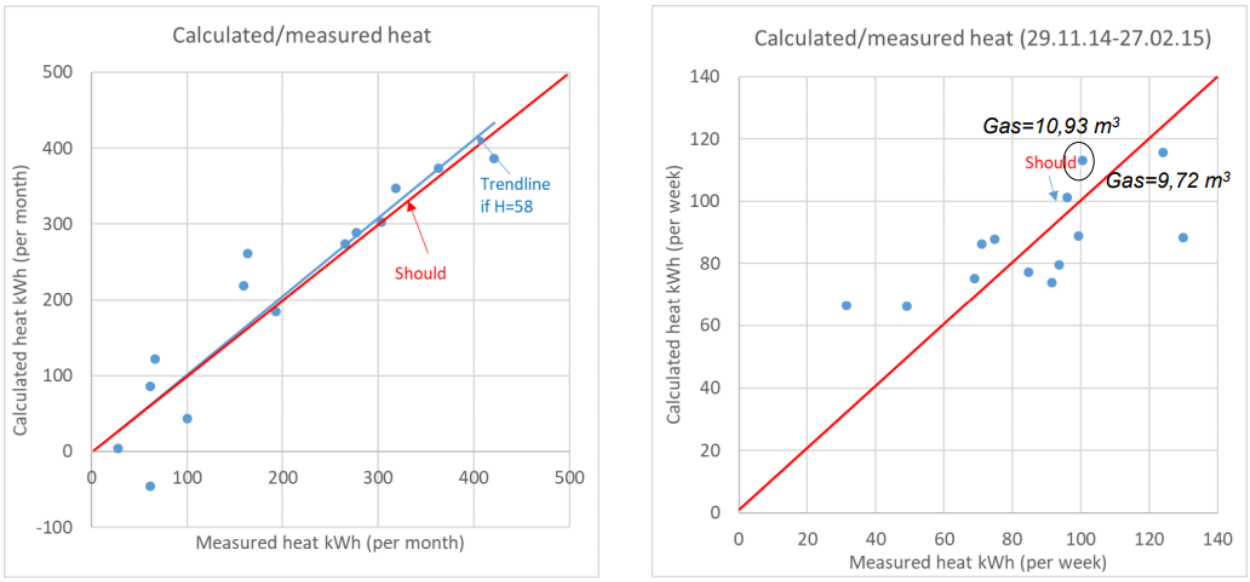


Figure 7. Comparison of measured and predicted heating demand (kWh) per month (left) and per week (right) using a simplified heat balance method.

In contrast to the previous – “physics based” – approach, this modelling team also fitted an ARIMAX model (auto-regressive, moving average model with exogenous inputs). To explore the capabilities of ARIMAX models in this context, the modelling team started by identifying the most explanatory sensor data to use as input and output in the parameter estimation exercise. Hereto, the data was pre-processed by taking time-derivatives of the sensor data to exclude the trend lines (e.g. gas meter data are cumulative signals, taking the time-derivative gives the gas consumption per hour). For the full data set – including the time derivatives – the correlation with gas use (differentiated gas meter data) was calculated for the winter months. The 5 inputs with the highest correlation were taken to predict the gas use:

- Total electricity usage (differentiated total electricity metering data)
- Ventilation supply temperature difference (differentiated ventilation supply temperature)
- Bedroom 1 temperature difference(differentiated temperature data)
- Lounge 1 temperature difference (differentiated temperature data)
- Water usage (differentiated water metering data)

The ARIMAX model was then fitted for different orders and on different training sets, showing the best results – in terms of RMSE and σ^2 were obtained for the ARIMAX(0,1,1) model fitted on 2 weeks of winter data. This model resulted in an RMSE of 0.08 m³/h. Note that the predictions were implemented to set negative values to zero. As the linear ARIMAX model does not handle this non-linear on/off behaviour.

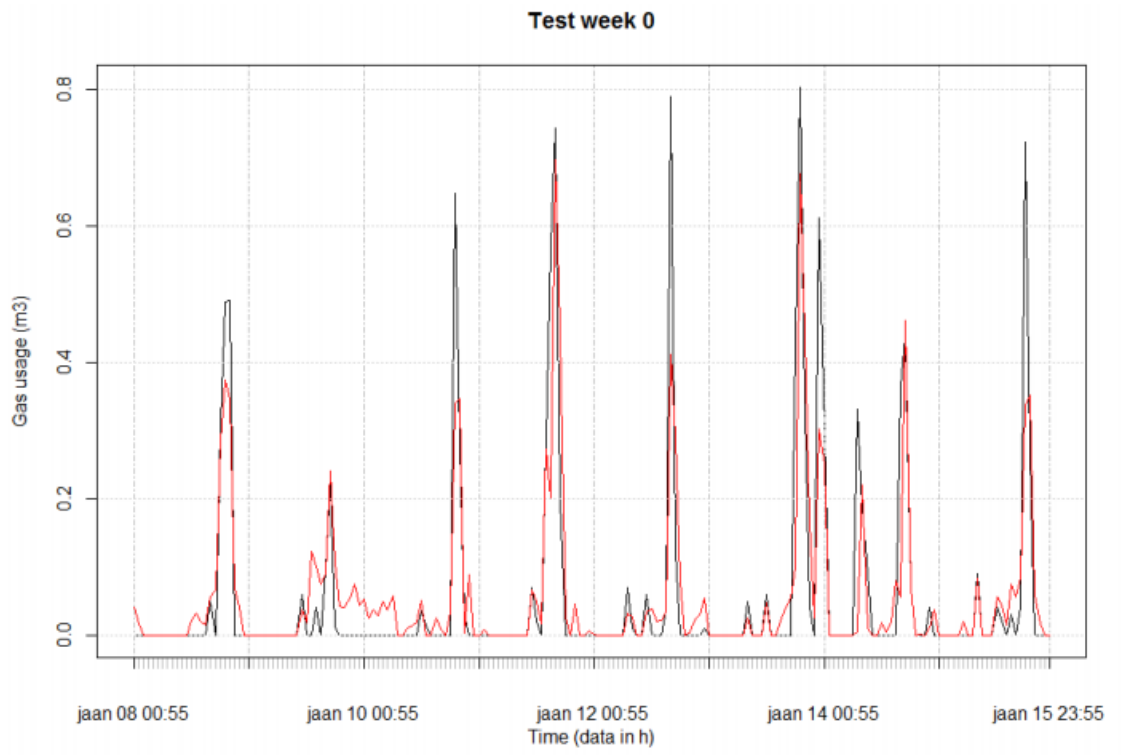


Figure 8. Predicted (red) vs. measured (black) gas usage for the ARIMAX(0,1,1) model for the test period in January

A second modelling team, compared the use of ARX (autoregressive model with exogenous inputs) and the innovative PWARX (piece-wise ARX) models for predicting the hourly gas consumption.

A Piecewise Affine model is a hybrid model; it is dedicated to the modelling of non-linear systems, with more than one functioning mode, by linearizing them around functioning points. It is a collection of linear models that share the same continuous state variables and are related by switches. The PWARX model introduced by (Bako, Boukharouba, Duviella, & Lecoeuche, 2011). The key for validating the PWARX model is to identify the decision rule for switching between one discrete mode and another in order to determine the corresponding sub-model that each new data point belongs to. In this work, a support vector machine (SVM) was used as classification technique for identifying this rule (Ajib, Lefteriu, Caucheteux, Lecoeuche, & Gauvrit, 2018). SVMs are supervised learning methods for classifying new data points based on the acquired labelled data used to train the classifier (Vapnik, 2000). In other words, given labelled training data, the algorithm outputs an optimal hyperplane which separates the regions defined by the labelled data and is used for categorizing new data. The SVM algorithm involves an optimization technique to calculate the values of the parameters and defining the hyperplanes such that the distance between the hyperplane and the closest data point is maximal. For more details about the calculation steps, the reader could refer to (Kim, 2013).

The PWARX method obtained high accuracy during the training period, leading to an R^2 of 0.6. Moreover, while the PWARX with SVM outperformed the ARX method significantly on the validation period, it was found that both models were failing to accurately predict all peaks (eg. 17/01/2015) or predicted high power outputs when the system was actually off (e.g. 21/01/2015). During a work meeting, multiple factors were identified that could result in these high levels of uncertainty, amongst which:

- The heating power does not only follow from the dynamic behaviour of the indoor temperature, but also from the thermostat set point imposed by the users. As this set point was not monitored, it is impossible to tell if users change it;

- two room thermostats are installed in the houses (one on each floor), but the location of the measured indoor temperature is not the same as the location of the thermostats;
- The gas consumption contains both space heating and domestic hot water, making the gas consumption highly dependent on (unmodelled and unmeasured) occupant behaviour.

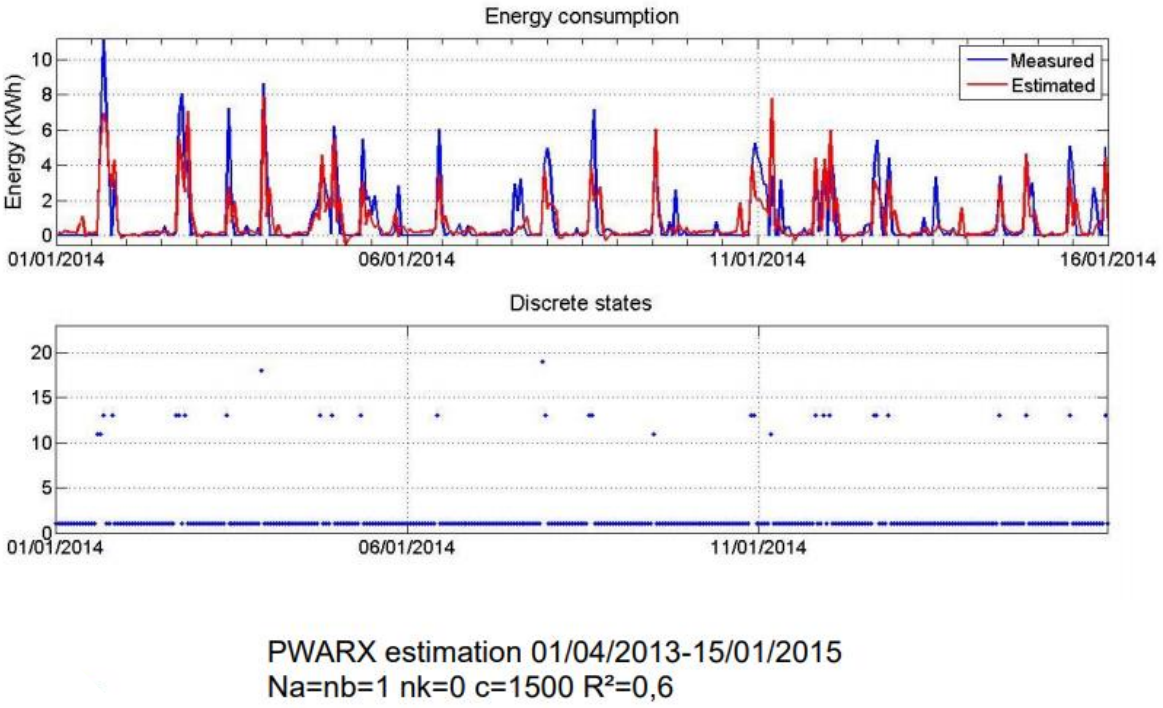


Figure 9. Estimated vs. measured gas consumption profile for the training period using the PWARX method

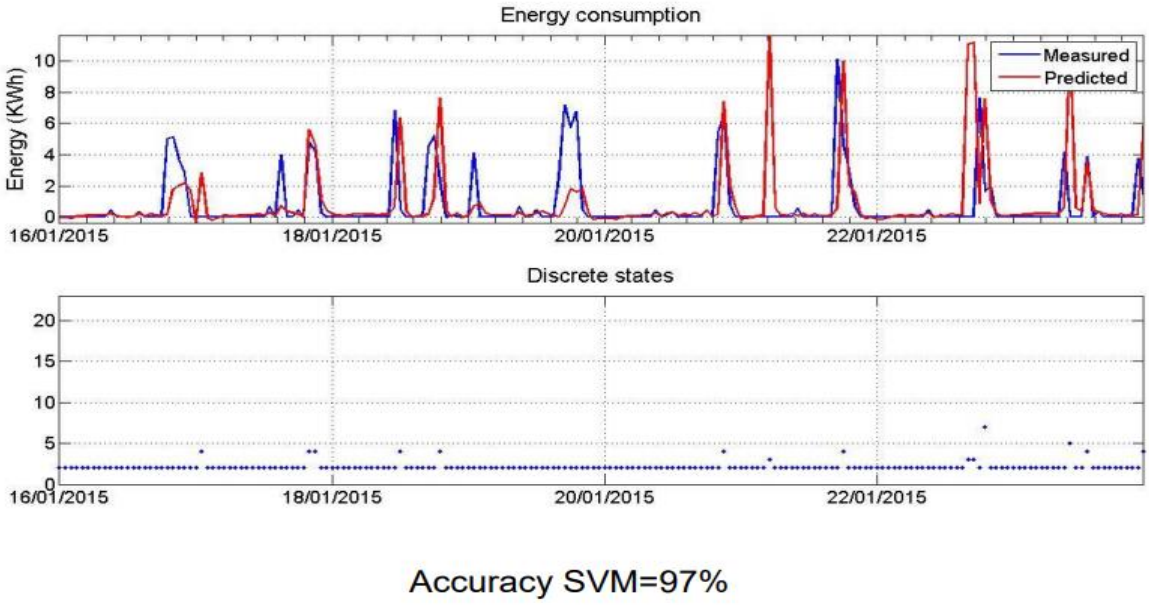


Figure 10. Estimated vs. Predicted gas consumption for the cross-validation period, using the SVM

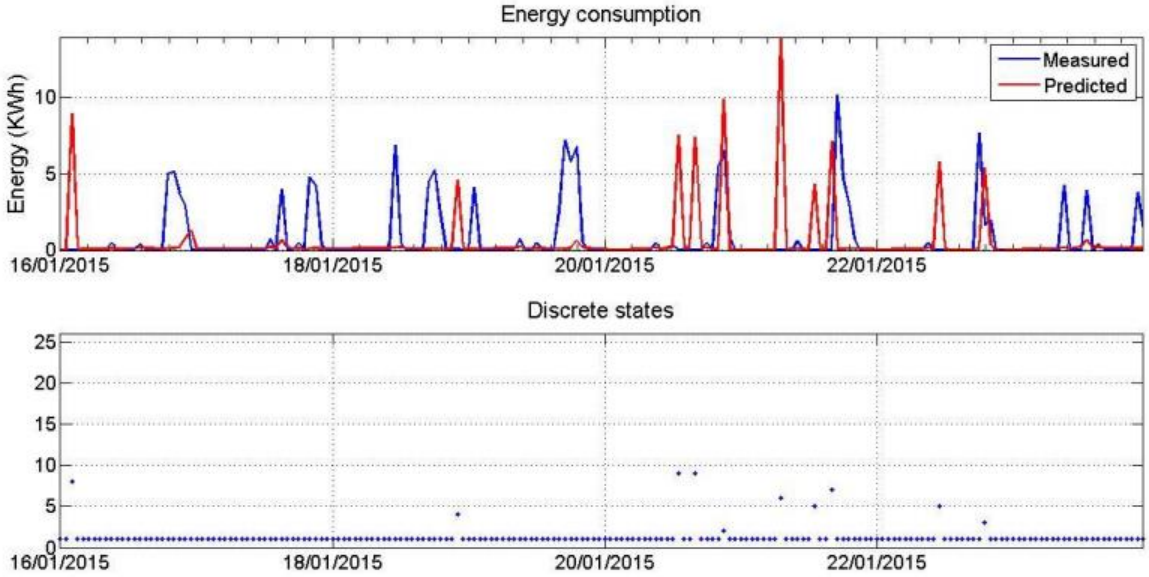


Figure 11. Measured vs. predicted gas consumption for the ARX method

In a second iteration of the common exercise, one of the teams compared 7 modelling approaches for their ability to predict the daily gas consumption:

1. Naïve: Is a simple zero-order-hold function, which assumes the heating power at t_{k+1} is equal to that on t_k.
2. Moving Averaging (MA): Here a simple moving average is fitted using an exponential function based on the consumption of the last 7 days.
3. Multi-variate linear regression (MLR)
4. Support Vector Machine (SVM)
5. Random Forest: 100 models with bootstrapping and a maximum depth of 3-4
6. Gradient Boosted Trees (GBT): 100 models with a maximum depth of 3-4
7. Multilayer Perceptron (MLP): Neural network with a brute force parameter optimization.

The goal was to obtain a prediction of heating consumption with various models and see the effect of reducing the features on the model performance. In pre-processing, the monitoring data of the Gainsborough house was merged with the Waddington weather data, filtered for outliers, and sampled to created daily values. Next, the data was limited to use only the heating season and a feature selection was carried out to avoid multicollinearity. In addition to the raw data, features such as weekend, temperature differences indoor outdoor (Ti-Ta), three outside temperature lags T_a(-1)…, moving averages of outside temperature simple (MA7s) and exponential (MA7e) of 7, 14 and 28 days, change of temperature (dT_a), occupancy depending on daily max CO2 (<700ppm no occupancy) were calculated.

Similar to the previous teams working on the hourly predictions, the resulting predictions of the daily gas consumption were not very reliable as only 49% of the variation was explained by the model. RMSE is at 4,88 kWh for a linear regression (LR) model. However, the models still outperform the simple references (naïve = assuming the last value and MA = moving average 7 days exponential).

Table 2. Comparison of model accuracies for the predicted daily gas consumption of the Gainsborough case

Statistical Measures	Naive	MA	LR	SVM	Tree Ensemble	GBT	MLP
R^2	-0.41	0.07	0.49	0.44	0.37	0.39	0.38
mean absolute error	6.06	5.37	3.50	3.83	3.81	3.83	3.79
mean squared error	66.26	51.28	23.84	26.53	29.88	28.75	29.16
root mean squared error	8.14	7.16	4.88	5.15	5.47	5.36	5.40
mean signed difference	-0.37	0.32	0.36	-0.30	0.40	0.37	-0.56

To analyse the impact of the availability of specific sensors, a feature selection with genetic algorithm and linear regression was carried out to see the effect of the features on the model performance. Table 3 shows the results of the features selection. The table shows how RMSE-values decrease with increasing features to obtain a RMSE of 5,09 kWh with only six features T_a, Wel_HH, RH_S, RH_L, T_L, Range(RH_L), Range(T_L), ext.Wind_sp, MA7s(T_a), MA28s(T_a) and can be further reduced to 4,73 kWh with seven additional features.

Table 3. Results of feature selection when predicting daily gas consumptions

RMSE	Filtered Count	Nr. of features	Selected features
4.73	13	18	CO2_L, Wel_HH, RH_Bed1, T_Bed1, RH_L, Q_WW, Range(RH_L), Range(T_L), Max*(CO2_L), ext.Wind_dir, ext.Wind_sp, Weekend, Tl-Ta, T_a(-2), T_a(-3), MA28s(T_a), MA7e(T_a), Occupancy
4.74	12	15	CO2_L, Wel_HH, RH_S, RH_Bed1, RH_L, Q_WW, Range(RH_L), Range(T_L), Max*(CO2_L), ext.Wind_dir, ext.Wind_sp, Weekend, Tl-Ta, T_a(-2), T_a(-3)
4.75	11	14	CO2_L, Wel_HH, RH_Bed1, RH_L, Q_WW, Range(RH_L), Range(T_L), Max*(CO2_L), ext.Wind_dir, ext.Wind_sp, Weekend, Tl-Ta, T_a(-2), T_a(-3)
4.85	10	15	CO2_L, Wel_HH, RH_L, RH_H2_L, Range(RH_L), Range(T_L), Max*(CO2_L), ext.Wind_dir, ext.Wind_sp, Tl-Ta, T_a(-2), T_a(-3), MA28s(T_a), MA7e(T_a), Occupancy
4.87	17	22	CO2_L, Wel_HH, RH_S, RH_Bed1, T_Bed1, RH_L, T_H2_L, RH_H2_L, Q_WW, Range(RH_L), Range(T_L), Max*(CO2_L), ext.Wind_dir, ext.Wind_sp, ext.S_global, Weekend, Tl-Ta, T_a(-2), T_a(-3), MA28s(T_a), MA7e(T_a), Occupancy
4.91	9	11	RH_S, T_Bed1, RH_L, T_L, Q_WW, Range(T_L), Max*(CO2_L), ext.Wind_dir, Weekend, T_a(-1), MA28s(T_a)
4.92	8	15	T_a, CO2_L, Wel_HH, RH_S, RH_L, T_L, Range(RH_L), Range(T_L), Max*(CO2_L), ext.Wind_sp, Weekend, T_a(-3), MA7s(T_a), MA14s(T_a), MA28s(T_a)
5.07	7	13	CO2_L, RH_Bed1, RH_L, T_L, T_H2_L, Range(T_L), Max*(CO2_L), ext.S_global, T_a(-1), T_a(-2), MA7s(T_a), MA14s(T_a), MA7e(T_a)
5.09	6	10	T_a, Wel_HH, RH_S, RH_L, T_L, Range(RH_L), Range(T_L), ext.Wind_sp, MA7s(T_a), MA28s(T_a)
5.75	5	8	T_Bed1, T_L, Range(T_L), ext.Wind_dir, Weekend, T_a(-2), T_a(-3), MA7s(T_a)
5.91	4	6	T_Bed1, Range(RH_L), Tl-Ta, T_a(-3), MA7s(T_a), MA28s(T_a)

While the methods presented above were a first exploration of modelling approaches that are subject to further optimizations few discussion points and lessons learned were drawn from this exercise. A first lesson learn, which is generally applicable, is that data quality, and hence pre-processing of the data is an important factor in accurate building performance characterization. In this case, study data was carefully collected as part of a research project, yet

unavoidable periods with missing data for 1 or more sensors exist. Interpolation is therefore often needed to overcome short sensor malfunctions. The impact of such interpolations on the identification process, but also on the validation test (e.g. auto-correlation of residuals) should be considered carefully.

Second, in real-life setups measurements are often influenced by multiple disturbances. For example, in this case, study gas consumption is measured, but gas is used for both space heating and domestic hot water production. At the same time, the central water measurement does not allow to differentiate between hot water and cold-water consumption.

Senave, Reynders, Sodagar, & Saelens, (2018) explored this issue further by comparing different classification techniques for decomposing the gas consumption. A first decomposition method to disentangle both end uses ('DM1') could therefore be the application of a default distribution. In this case study there will be opted for a 76/24 distribution for the end uses SH/DHW, as reported by Menkveld (2009). A major drawback of this method is that it does not take the actual consumption, SH demand or occupant behaviour into account. The second decomposition method ('DM2') is fully based on the assumptions that (1) in the case of the combi boiler, the production of DHW and SH do not occur at the same time and (2) the gas consumption for DHW production perfectly coincides with the DHW consumption. It involves the implementation of two rules on the 5min-interval-monitoring data. The first rule stipulates that the gas consumption for DHW production must be set to 0 when mains water consumption is 0, else gas consumption for the production of DHW must be set equal to the total monitored gas consumption. The second rule states that the gas consumption for SH must be set to 0 when mains water is consumed, else gas consumption for SH must be set equal to the total monitored gas consumption. This DM is straightforward and easy to implement. However, a number of potential flaws can be identified. First, the assumptions imply that all cold-water tapping's occurring while gas is used for SH are classified as DHW usage. The fact that grey water is used to flush the toilets though makes this assumption more reasonable. Secondly, the hot water tapings could be significantly shorter than the 5min sampling time. Yet, from the moment water, consumption is observed, however small, the full gas consumption for that 5min period is allocated to DHW production. Higher frequency logging could solve this issue. Thirdly, minor delays between starting and stopping of water and gas consumption will create some error. The third approach, 'DM3', which was demonstrated by Bacher et al (2016), uses a robust, zero order, Gaussian kernel smoother to estimate the 'gas consumption for SH'-profile underlying the noisy 5min gas consumption data. Next, all spikes of the total gas consumption significantly above this kernel (smoother) estimate are classified as DHW heating spikes and their values are obtained by subtraction of the kernel estimate. Finally, subtraction of the estimated heat load for the production of DHW from the total heat load gives an estimate for the heat load for SH. The parameters of the kernel smoother procedure were tuned with an eye on limiting the gas consumption classified as 'gas used for space heating' during the summer months. The final model parameter values are as follows: kernel window: 1h, bandwidth: 0.5h, threshold for bisquare robust estimation γ: 7MJ/h, separation threshold q_{tres}: 1.1. Just like DM2, this decomposition method has not been verified on a case study were the total gas consumption and the consumption for the production of DHW and SH were measured separately. In contrast with the previously described approach, this method assumes that gas consumption for both end uses can occur simultaneously. It should furthermore be noted that all peaks are classified as DHW heating, although the start-up of the space heating might also result in a similar peak in the fuel consumption.

Table 4 compares the decomposition of gas consumption for SH and DHW production obtained through the different approaches. The default method (DM1) almost always results in a higher gas consumption for SH than DM2 and DM3. The method with the robust kernel smoother (DM3) uses a certain threshold instead of selectively classifying the gas consumption as either gas consumption for SH or production of DHW as DM2 does. This way it appears to systematically obtain a lower gas consumption for SH. Further research is needed to validate these approaches since no sub-metering data was available for this case study to perform a validation.

Table 4. Total gas consumption [kWh] and gas consumption estimated for space heating using the 4 approaches (as percentage of the total gas consumption)

	Oct 2014	Nov 2014	Dec 2014	Jan 2015	Feb 2015	Mar 2015
Total gas consumption	188 kWh	286 kWh	486 kWh	526 kWh	428 kWh	267 kWh
No decomposition	100%	100%	100%	100%	100%	100%
DM 1	76%	76%	76%	76%	76%	76%
DM 2	58%	69%	74%	77%	74%	59%
DM 3	44%	55%	62%	63%	59%	45%

A third important conclusion from this exercise in predicting the heating power, is that in these buildings the heating power is actually a controller variable: the heating power is controlled to keep the indoor temperature at desired levels.

As the case study consists of a well-insulated building, the gas consumption profile is highly intermittent and highly sensitive for occupant behaviour which is in itself stochastic of nature. As a result, it is shown that both the PWARX and the ARIMAX models are not always able to accurately capture the peaks in the gas consumption.

2.2.3. Predicting indoor temperature

Given the difficulties in predicting the highly intermittent gas consumption, a second exercise was conducted in which the gas consumption is available as one of the inputs to the model in order to predict the indoor temperature.

A first modelling team, explored 4 regression-based models:

- ARX (auto regression models with exogenous input)
- ARMAX (autoregressive moving-average models with exogenous input)
- FIR (finite impulse response model)
- FIR-RLS (finite impulse response with recursive least squares)

For each model, the outdoor temperature, heating power for space heating and global irradiation were used as model inputs, while the temperature of the living room is used as model output. Since the heating power is not a directly measured, but only available through the gas consumptions, the modelling team used a quantile regression method to split the gas consumption between space heating and domestic hot water consumption.

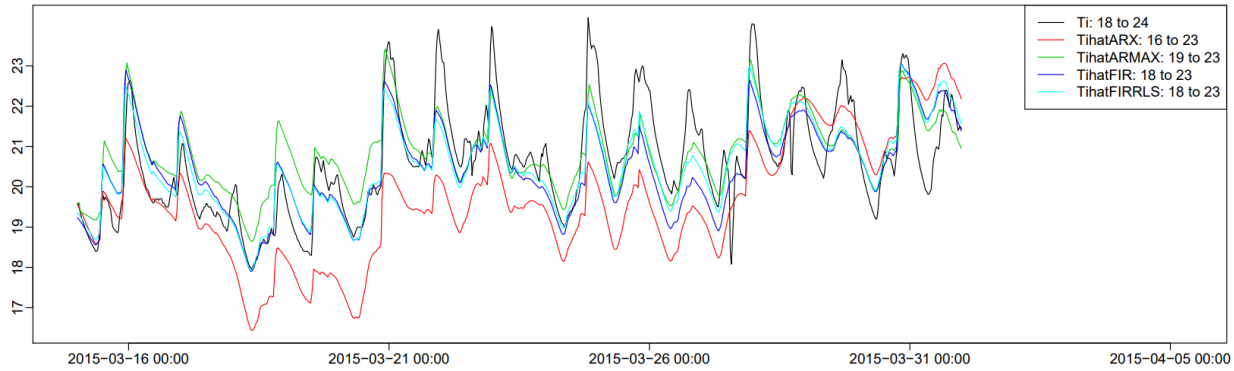


Figure 12. Prediction of the living room temperature comparing the ARX, ARMAX, FIR and FIR-RLS models.

Figure 12 shows the prediction of the indoor temperature over the period between 16-31 March. While further optimization of the model orders was not carried out in this study, the results show that indoor temperature predictions for short prediction horizons (few hours) are manageable with these type of models. For longer prediction periods, up to a few days mostly the ARX models shows significant deviations for the measured temperature profile. These findings should be handled carefully though, as further optimization of the model order or the selection of inputs was not explored here.

One important remark made by the modellers was that the FIR method, not only allows to predict the expected value of the indoor temperature but also the uncertainty on the predicted value when fitted using quantile regression. Predicting this uncertainty is an important input for stochastic optimal control formulations that are able to take into account the uncertainty on the indoor temperature when optimizing the control (Junker, 2019).

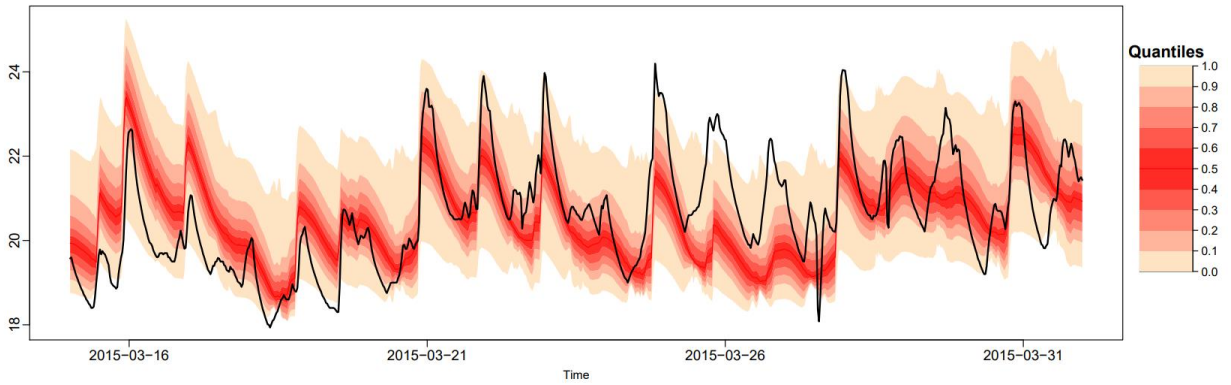


Figure 13. Prediction of indoor temperature profile using the FIR method fitted with quantile regression

A second modelling team abandoned the regression-based models and used a Random Forest method to predict the indoor temperature of the living room. The team primarily focused on the feature selection process in which the optimal set of input signals was composed from the full data set. Thereby, not only the actual measurement data but also time derivatives and lagged data (up to 6h time lag) were considered. Data from March 2013 until end of 2014 was used, in order to predict the indoor temperature for the first 2 weeks of January 2015. The RMSE for the prediction of a cross-validation set was used to evaluate the performance of the model. In the feature selection the found that a minimum of 4 features was needed in order while the lowest RMSE was obtained for 16 features. Thereby, it is noted that these 16 features only rely upon 9 physical properties: the outdoor temperature, the indoor temperature of the house, the indoor temperature of the neighbours, the relative humidity and CO_2-concentration in the lounge, the ventilation supply temperature, the relative humidity in supply and extraction air and the photo-voltaic production. It was argued that except for the temperature of the neighbouring house, all these sensors are readily available today in technical building systems and this data could hence be captured at low cost.

16 features

$T_{indoor\_House1\ (t-1)}$

$MVHR\_Supply\_T_t - MVHR\_Supply\_T_{t-1}$

$MVHR\_Supply\_T_t$

$CO_2\_Lounge_{\ t-1} - CO_2\_Lounge_{\ t}$

$\Delta T_{i-o\ (t)} - \Delta T_{i-o\ (t-1)}$

$RH_{Lounge\_House1\ (t)} - RH_{lounge\_House1\ (t-1)}$

$T_{indoor\_House1\ (t-1)} - T_{indoor\_House1\ (t-2)}$

$PV_t - PV_{t-1}$

$T_{outdoor\ (t)}$

$T_{lounge\_House1\ (t-1)} - T_{lounge\_House1\ (t-2)}$

$MVHR\_Extract\_RH_t - MVHR\_Extract\_RH_{t-1}$

$MVHR\_Supply\_RH_t - MVHR\_Supply\_RH_{t-1}$

$CO_2\_Lounge_t$

$T_{indoor\_House2\ (t)}$

$RH_{Lounge\_House1\ (t)}$

$T_{outdoor\_Waddington(t)}$

Figure 14. Overview of the 16 features selected for the random forest approach

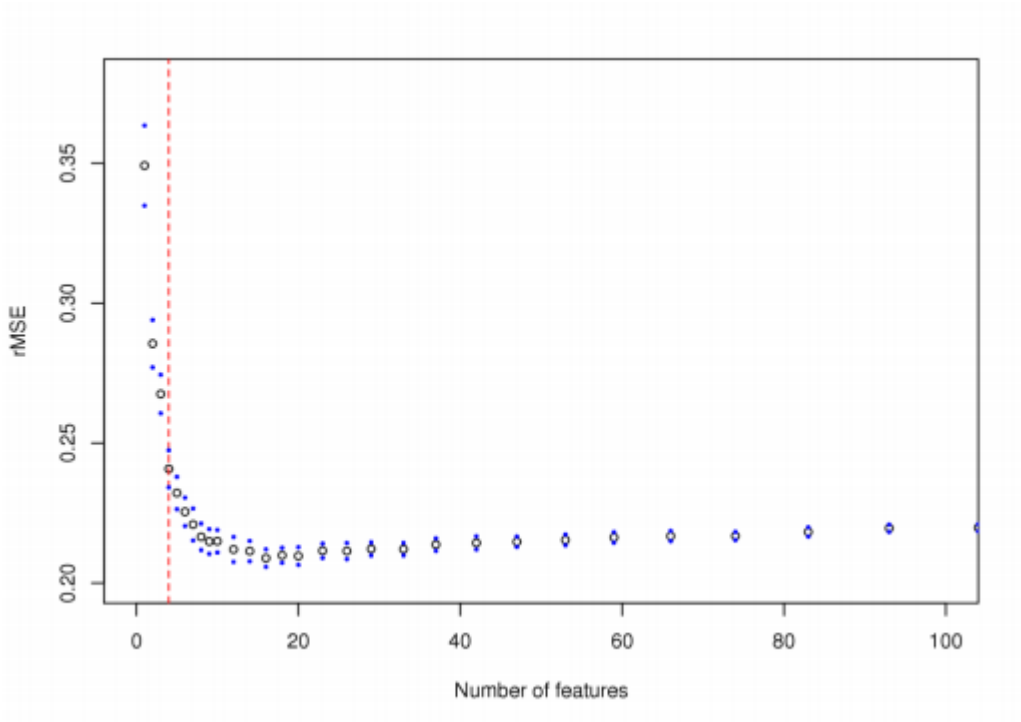


Figure 15. Evolution of RMSE values for increasing number of features for the random forest approach

2.3. Conclusions

The exercises presented above – and the discussions that went with them during the process of the project – indicate the importance of a thorough understanding of the building and the corresponding data is an important factor in generating confident predictions. Understanding the building as a system and correctly handling (controlled) input and outputs showed to be an asset in obtaining reliable prediction models. For example, knowledge on the location (and representativeness) of the temperature sensors, the functioning of the mechanical ventilation system or data on the temperature in the adjacent building were found to be important, but argued to be difficult to obtain when upscaling (and automating) these building behaviour characterization techniques. In this context, the results showed that predicting the hourly heating demand is significantly more difficult as predicting the indoor temperature profile. This was attributed to the actual layout of the building as a system in which the indoor temperature is controlled (output of the system) by modulating the heating power using in this case the room thermostats. The heating power is in this case hence a natural input to the system.

Also the lack of sub metering, e.g. to break down the gas consumption in space heating and domestic hot water production, was found to be a major challenge when dealing with real-life measurement setups. Nonetheless, time series analysis of the detailed (15 min resolution) gas consumption data over a longer period (including summer data) showed provide reasonable estimates for the energy breakdown. Yet, further research is needed to confirm and improve these results on a broader dataset.

Finally, comparing the different modelling approaches it was found that as the goal of the exercise was on predicting temperature (or heating power) profiles, black-box models were often considered. Thereby also stochastic prediction methods were explored showing the important benefit of not predicting the most likely single value (or profile) but to also predict the uncertainty. PW-ARX, SVM and FIR-RLS methods were found to be promising, more research is needed to generalize these results on a broader set of buildings, but also to further fine-tune the modelling approaches (e.g. optimal model order selection, filtering input data etc.).

3. Fault-detection and diagnostics

3.1. State of the art

Poor energy performance of the building stock becomes one of the main challenges to the successful implementation of energy renovation and energy efficiency strategies in Europe (de Wilde, 2014),(Galvin, 2014). It is one of the critical issues to be addressed before rolling out a massive strategy to reduce the energy use in buildings. This poor building performance is related to the building design and construction, the building materials, the mechanical and electrical systems, and the control and operation of the buildings. However, this last factor is the only one, which can be addressed in real buildings without large investments or ambitious energy retrofitting strategies. The improvement of the control and management of buildings becomes the first step to be addressed in any strategy towards more sustainable use of the energy. In the case of the residential sector, the energy management is clearly linked to the user behaviour (Baranski & Voss, 2004); (Masoso & Grobler, 2010);(Wilson, 2014) and in most cases, a suitable strategy to reduce the energy use in homes must include the tenants as a central driver if intends to succeed. In the case of commercial and public buildings, although the users also play an important role (Masoso & Grobler, 2010),(Labeodan, Zeiler, Boxem, & Zhao, 2015);(Martani, Lee, Robinson, Britter, & Ratti, 2012), the building manager becomes the key actor instead. The building manager is usually responsible for managing the building energy performance and detecting failures in energy performance. However, at present, there is no standard for the operation of commercial buildings, which can define clear and accepted rules. The release in 2012 of the new standard EN 15232-Impact of building automation, controls and building management in the energy use, does not include any guideline or specification on how to optimize the building control once the building is in real operation. This lack of standardization is worsened because building managers have a wide range of background, knowledge, training, and education. A survey carried out in Ireland (O'Donnell, 2009) revealed that few building managers seldom have relevant work experience or have the skills and training that are required to optimize energy management of buildings. Since no standards have neither been established for analysing and transforming building performance data and information (O'Donnell, Keane, Morrissey, & Bazjanac, 2013), the implementation of effective visualizing and supporting applications, enhanced with advanced anomaly detection methods, is becoming a priority nowadays.

In the European market, there is an increasing number of IT companies commercializing building energy management platforms designed to deliver useful information to the building managers and occupants. In the last years, many authors performed detailed reviews of these available applications and highlighted their strengths and weakness, as well the opportunities for further research and innovation (Lee & Yang, 2017), (Shaikh, Nor, Nallagownden, Elamvazuthi, & Ibrahim, 2014),(Volk, Stengel, & Schultmann, 2014), (Marinakis, Doukas, Karakosta, & Psarras, 2013), (Zhao, Zhang, & Liang, 2013). These studies state that most of these applications require a relatively high number of variables, from multiple sensors and devices; however, they only offer limited real-time information about the energy performance of buildings. They are focused on deep analysis of single buildings and they rarely incorporate more advanced features such as clustering, benchmarking, energy load or thermal comfort forecasting and detection of abnormal energy performance. These limitations prevent building managers to take advantage of trend analysis and anomaly detection techniques, which would help them to identify inefficiencies in the energy use. Only recently some research studies (Taal, Itard, & Zeiler, 2018) made the effort to define algorithms and procedures able to deal with fault detection of HVAC systems that can be accurate and scalable enough to be integrated in computational distributed IT architectures.

In the last years, the rapid penetration of smart meters and the growing digitization of the energy systems, has generated an explosion of available data which entails an increase complexity in the data processing to obtain valuable information (Dibley, Li, Rezgui, & Miles, 2012). This rapid transformation from very specific tailored energy monitoring projects to wide data opened projects able to be benefited from the available data sources coming from multiple devices connected to the internet, requires sophisticated data-driven algorithms capable of combining running in real-time and batch processing procedures (K. Zhou, Fu, & Yang, 2016). In recent years, some developments are following this direction. In (J.-S. Chou, Telaga, Chong, & Gibson, 2017) a mobile application to inform the occupants of office buildings about the energy performance and the detected anomalies is presented. It is specially designed to be used by non-expert users. The anomaly detection method is based on a procedure (J. S. Chou & Telaga, 2014) which follows two stages: consumption prediction and anomaly detection. In this research, the prediction is performed at daily frequency following a hybrid Neural Network Auto-Regressive integrated moving average model (NNAR). Anomalies are then identified by differences between real and predicted consumption by applying the two-sigma rule, which classifies any points outside of 2 standard deviations from the mean as anomalous data. The techniques outlined in this research assume constant data periodicity, which can cause many false positives. Very recently, (Araya, Grolinger, ElYamany, Capretz, & Bitsuamlak, 2017) performed a detailed review of previous anomaly detection research works (Wrinch, El-Fouly, & Wong, 2012), (Paulo, Branco, de Brito, & Silva, 2016), (Peña, Biscarri, Guerrero, Monedero, & León, 2016), (Hill & Minsker, 2010). One of the conclusions of this review was that these previous methodologies lacked of giving

context to the anomaly detection process and didn't consider in detail the effect of different energy performance load patterns (structures) caused by seasonal dependent schedules, external weather effects or changes in use. For example, a value might be anomalous in winter but not in summer or neither in the night hours of one specific day of the week. The research of (Araya et al., 2017) also developed two generic anomaly detection frameworks: a first pattern-based method which uses overlapping sliding windows integrated in an ensemble learning process (Z.-H. Zhou, 2009) and a second method based on the combination of the previous one with prediction-based classifiers. The first framework reconstructs the raw data by using an unsupervised artificial neural network, called auto-encoder, which is trained to reproduce input vectors as output vectors (Sakurada & Yairi, 2014). In the second framework, the prediction models are Support Vector Regression (SVR) and Random Forest (RF) models. This research develops a very precise methodology to detect both single and collective anomalies with a high accuracy. However, this kind of procedures, based on unsupervised learning techniques, do not include the effect of the exogenous factors, such as the climate or the occupancy schedule variation when the models are trained and, therefore, their prediction capacity is limited to buildings without seasonal or external weather conditions dependencies.

To address these limitations, supervised statistical learning methods (Hastie, Tibshirani, & Friedman, 2009) for building energy load forecasting and anomaly detection have been developed in the last years. First attempts started to implement linear temperature dependent change point models for electricity load forecasting. These change point models were firstly implemented in data analytics environments by (Muggeo, 2003). In (Paulus, Claridge, & Culp, 2015) an algorithm for automating the selection of a temperature dependent change model was developed and applied to the forecasting of the building energy load in monthly frequency. A refined of this methodology to use it with hourly frequency data was performed by (Abushakra & Paulus, 2016). The results were satisfactory but limited to buildings with a weak influence of non-linear exogenous variables such as occupants. Later on, a variation of these change point models was developed (Borgeson, 2015; Dyson, 2014) by applying them to every week hour (168 hours) along 1 year of historical hourly data. This approach achieves accurate results when the electricity load is highly correlated with the outdoor temperature; however, their accuracy rapidly decreases when other exogenous variables have higher importance. Further research works moved on step beyond by implementing linear dynamic auto-regressive (AR) or AR extended models to the prediction of both thermal and electricity load of buildings. Starting from the forecasting of the aggregated demand of large number of buildings, (Sumer, Goktas, & Hepsag, 2009) used Auto-Regressive Integrated Moving Average (ARIMA) and seasonal ARIMA models. In (Ferracuti et al., 2017), a comparison of three data-driven models for short-term thermal performance of commercial buildings was performed: a grey-box model of first, second and third order, an Auto-Regressive model with eXogenous inputs (ARX) and a Non-linear Auto-Regressive network with eXogenous inputs (NARX). These three models predicted the ambient temperature using the thermal energy delivered by the district heating system as an input. The anomaly detection criteria considered as anomalous values those data with deviations out the 99% of the confidence interval of the forecasted ambient temperature. The three models showed good accuracy in predicting short-term behaviour. This research contains plenty of details about its application to individual buildings analysis as long as you have the resources to monitor ambient temperature, which is a challenging factor when applying these techniques to a massive scale. Moving beyond linear models, Generalized Additive Models (GAM) have become popular in recent studies since they can evaluate the effect of several exogenous variables and can model complex non-linear features. GAM models enable analysts and building managers to better understand the driving forces behind the building energy use In (Ploennigs, Chen, Schumann, & Brady, 2013), a GAM model for load forecasting, combined with an ARMA model for the correlated residuals, was developed to detect abnormal energy load patterns in buildings with sub-metering systems. The anomaly detection methodology was based on computing the GAM model for the main meter, applying ARMA to compute the upper and lower bounds of the confidence intervals (CI) and then identifying those values out of these CI as abnormal values. Once the abnormal pattern is detected in the main meter, the root cause of this anomaly is obtained through comparison with the residuals of the sub-meters. This methodology considers a long training period (4 years) but do not analyse the influence of seasonality on different daily load curves, for instance due to seasonal changes of the HVAC operation in the model coefficients identification. Moreover, it is applied to only one real building. Hence, a wider validation might be necessary to demonstrate its full validity. In (Pathak, Ba, Ploennigs, & Roy, 2018), an accuracy comparison is performed between a GAM, a GAMAR, GAM+ARMA, and a deep learning variant known as Long-Short Term Memory neural network (LSTM) in predicting the gas usage of 2 buildings. This research concludes that although LSTM outperforms any GAM model in relation to the day ahead prediction accuracy, the GAMAR model is finally preferred to be used in an anomaly detection environment because it is capable of handling random missing data and it helps to gain interpretable insights regarding the data and its covariates. This research did not analyse the effect of training different models for different training data structures and limited the auto-regression terms to daily and weekly lags.

As can be seen, a wide variety of modelling techniques can be used to perform the FDD, alternatively referred as anomaly detection. The main cause of this heterogeneity is the multiple objectives of this detection, the enormous variety of building typologies where it can be implemented, and the unclear methodology to estimate what is normal, and not, once a baseline prediction is obtained. In the framework of the Annex 71 Subtask 2, three major objectives regarding the FDD research line where the following:

1. The theoretical classification of possible faults that can occur on building energy systems, depending the level of data availability and the type of faults to be detected.
2. The definition and implementation of a robust statistical methodology of how a measure can be considered as a fault.
3. And finally, the validation of these methodologies over a case study located in Holzkirchen (Germany), where real faults were implemented and estimated using multiple data-driven models.

Furthermore, in this case study, a detailed building energy simulation model (developed in EnergyPlus) was available. Hence, several data scenarios (real and simulated) have been tested to evaluate the possibility to pre-train models with simulated data and apply them in real environments in the case of buildings without an initial set of data, such as those recently constructed.

3.2. Development of an FDD framework

3.2.1. Components of building energy systems and classification of faults

We divide the building energy system into three parts: a) the human-machine interfaces (HMI); b) the controllers; and c) the physical systems, including actuators, mechanical systems, thermal building, and sensors. Correspondingly there are three categories of faults: a) HMI faults, b) controller faults, and c) physical system faults.

a) HMI (human-machine interface) faults

For the purpose of maintaining specific temperature, humidity, and air quality, many building energy systems are equipped with HMI, where people can specify set points and schedules of the equipment. Sometimes the set points and schedules are different from expected. This kind of faults are usually caused by human mistakes, for example a temporary override which is forgotten to change back.

b) Controller faults

Controllers compare the controlled variable with the set point and outputs the control command to regulate the mechanical system. Room thermostat is a typical on-off controller. PI controllers are used to regulate valves, dampers, and Variable Frequency Drives (VFD). Staging and lead lag controllers are used to realize load control of heating and cooling sources (such as heat pumps, boilers, and chillers), cooling towers, air fans, and pumps. Generally speaking, any unreasonable action which causes big control error and oscillation is regarded as a controller fault. Most common controller faults are failure (no output signal) and inverse action (output on/increase when it's supposed to be off/decrease).

c) Physical systems faults

c-1) Actuators: Actuators take the electrical control signal and transform it to the moving action of valves and dampers to regulate water flow and air flow. Actuators may get stuck at fully closed, fully open, or any other position. Inverse action is another common fault, which is caused by errors in the commissioning phase.

c-2) Mechanical systems: Mechanical systems include valves, dampers, heat exchangers, pumps, air fans, compressors, burners, piping, air ducts, etc. All the equipment output either heating and cooling power, or water and airflow. Most of the faults either reduce the output, or cause the output to be zero. For example: heat exchanger fouling reduces the heating and cooling capacity; a leakage of pipes or air ducts reduces the water flow or air flow; a broken motor starter stops the pump or air fan from running and causes water or air flow to be zero.

c-3) Thermal building: Room temperature, humidity and air quality are given by thermal power and ventilation input from mechanical system, weather, internal heat gains, and the property of thermal building. An undesired change of the thermal building, such as damage of building insulations and opened windows, are regarded as thermal building faults.

c-4) Sensors: All of our observations are based on the measurements collected by sensors. But sensors may also have faults, such as bias, invalid value, and frozen value. Besides that, wrong installation may cause abnormal readings.

3.2.2. Detecting faults on three levels

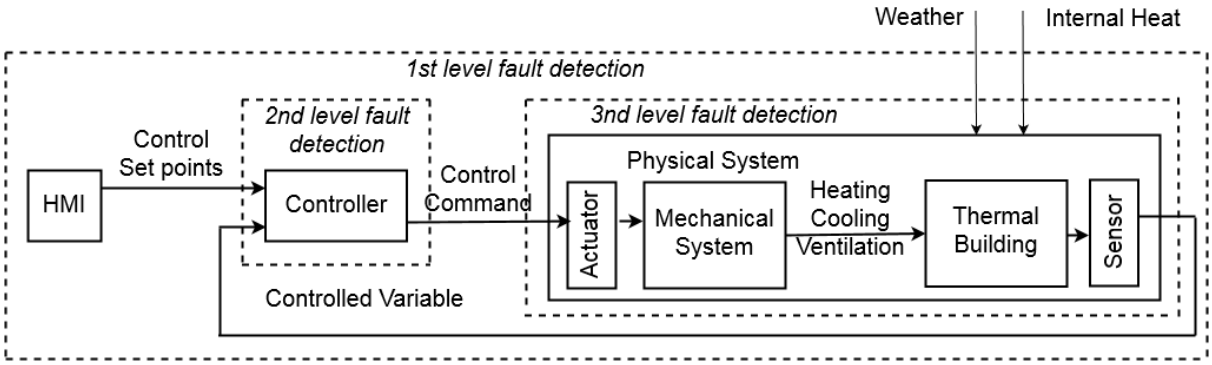


Figure 16. Components of building energy systems and three levels of fault detection

Figure 16 illustrates all the above mentioned components of building energy systems. In the best scenario, if we know all the variables going from one component to another (shown as arrows in the figure), we would be able to detect faults and also diagnose exactly which component is at fault. Different fault types require different detection method:

- HMI faults: compare real set point and schedule to expected value.
- Controller faults: detect large control error and oscillation caused by unreasonable control commands.
- Physical systems faults: modelling and prediction method. For each component, the output is a function of the inputs. Using this function to predict the output and compare the prediction with real value, analyse the residuals to detect faults. It is possible to test any specific part of the physical system depending on the chosen inputs and outputs.

Very often we do not have measurement of all the variables. In this case it is very important to know the system boundary and which faults are included. We defined three levels of fault detection (as shown in Figure 16), which cover faults of different components in the building energy system.

1st level fault detection:

Take weather and internal heat gains as inputs to predict heating / cooling power as Figure 17 shows. The model includes the whole building energy system. Therefore, 1st level covers all possible faults except weather and internal heat gain input errors. Change of any component from normal behaviour is detected, including HMI. However, it is not possible to diagnose the error source. Be aware that sometimes a change of set points are on purpose, and doesn't belong to faults.

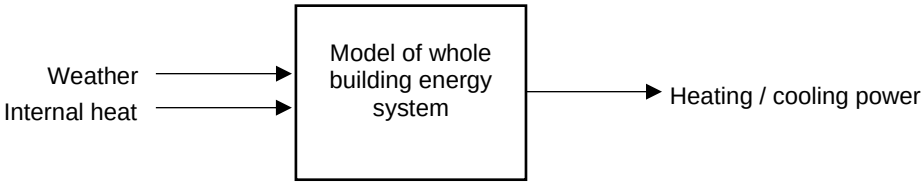


Figure 17. 1st level fault detection detecting any fault in the system, however no diagnostics possible

2nd level fault detection:

Investigate control loops (control set points, controlled variable, and control command) to detect controller faults as Figure 18 shows. For example, if the room temperature is lower than set point for a very long time, but the heating valve command stays at zero, it is detected as a control fault.

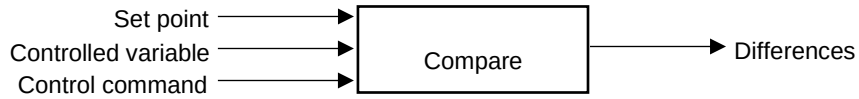


Figure 18. 2nd level fault detection detecting controller faults

3rd level fault detection:

Take weather, internal heat, and control commands as inputs to predict room temperature as Figure 19 shows. The model focusses on physical systems. Faults of actuators, mechanical systems, thermal building, and sensors will be detected. However, these are lumped together and not clearly distinguishable for diagnostics. The 3rd level fault detection can have many variations. If the output is heating/cooling power instead of room temperature, then it covers only the actuator and mechanical system. If the input is heating/cooling power instead of control commands, then it covers only the thermal building and sensors.

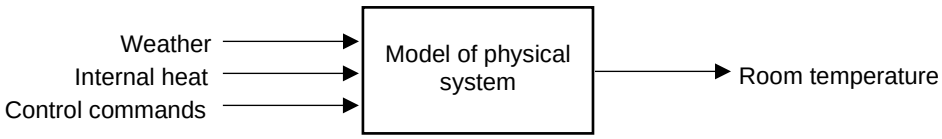


Figure 19. 3rd level fault detection detecting physical system faults

3.2.3. Linking Twin House potential faults to modelling and detection approaches

Different types of faults require different types of modelling and/or detection approaches. This section classifies Twin House potential faults and introduces different modelling and detection techniques that can be used in order to detect each type of fault.

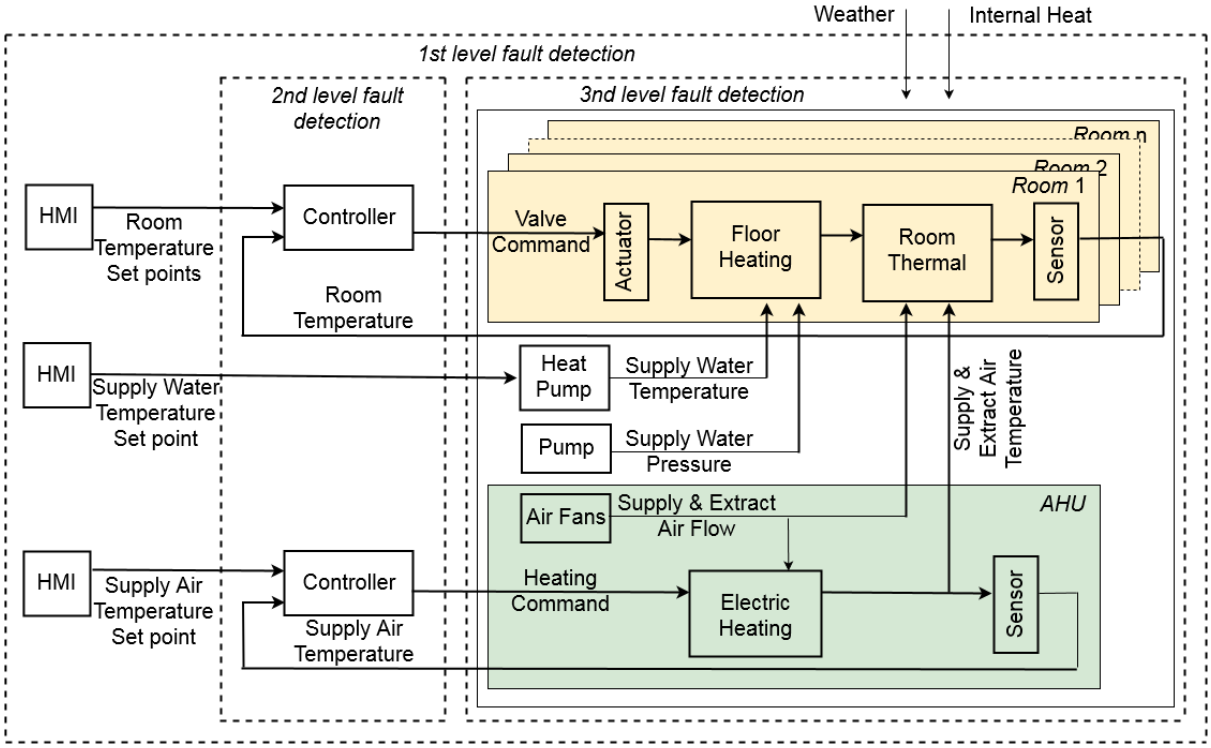


Figure 20. Three level fault detection for the Twin Houses

The system of the twin house is composed by floor heating, heat pump, pump, and Air Handling Unit (AHU) as shown in Figure 20. The potential faults include the following.

HMI faults:

- abnormal room temperature set point
- abnormal AHU supply air temperature set point
- abnormal heat pump supply water temperature set point

Controller faults:

- Room temperature control disabled or overridden
- AHU supply air temperature control disabled or overridden

Physical faults:

- Floor heating valve actuator stuck closed or stuck open
- Floor heating leakage, or fouling
- Heat pump failure or reduce capacity
- Pump failure
- AHU air fan failure
- AHU electric heater failure
- Room temperature sensor drift, out of range, or frozen
- AHU supply air temperature sensor drift, out of range, or frozen
- Damage of building enclosure or open window.

All the fault types could be detected with component level tests if enough data was available. For HMI component level fault detection the ranges or rulesets or intention for suitable set points should be defined and compared to the actual set points. Detecting controller faults, electric signal given to actuators should be available which are not in the Twin Houses. Most of the physical faults would not be detectable separately as for that all inputs and outputs for the component would have to be available.

The physical faults are detectable with 3rd level fault detection. For controllers, 2nd level fault detection is intrinsically component level fault detection and cannot be applied here. However, all the faults are detectable with 1st level fault detection. This does not enable to diagnose the source of the error but finds the existence of fault and its effect. Same applies to the 3rd level fault detection.

All in all, most potential faults are not diagnosable separately but only detectable within 1st or 3rd level fault detection. On both of these levels, faults can be detected with modelling techniques. This often applies to component level fault detection as well. Therefore, in the following, only fault detection with modelling approaches will be discussed.

3.2.4. Statistical tests to determine faulty behaviour

While it is often possible to detect faulty behaviour from a dataset using expert judgement, automating this detection process by formalizing the detection criteria is a key challenge. In this section, statistical methods are introduced that allow to formalize this detection process.

One possibility of fault detection is the use of a residual analysis. Residuals are the difference between the observed values and the predicted values from a statistical model or more general of a prediction model based on methods of machine/statistical learning. In a well-adapted model on data without faults, the residuals should possess certain stochastic properties. The exact properties depend on different factors such as the used modelling technique, the training data and the specific data structure. For simplification, the following properties for the residuals which usually are the result of an adequate model adjustment to the data are assumed to be valid:

· The residuals have a median of zero.

· The residuals are independent.

· The residuals are uncorrelated.

· The residuals behave randomly.

These properties are typical assumptions for the theoretical random error term in regression-based models. The residuals can be considered as an approximation of the unknown and inestimable random error terms. A median of zero indicates at this point that there is no systematic over- or underestimation (model bias). Whereas independence and uncorrelatedness assume that residuals at one point in time do not depend on the residuals at another point in time. This means that structures like positive and negative residuals alternating or similar are excluded. In addition, the error term is often assumed to be normally distributed. Here, instead, it is assumed that the residuals behave randomly, which is basically a generalization of the normal distribution assumption. Note, just because these properties are assumed for the random error term does not mean they automatically apply to the residuals. For example, in linear regression models, the residuals are not uncorrelated, see S. N. Wood (2006) page 16. Since in many prediction models it is not clear in advance which properties apply to the residuals. The residuals at the fault-free period are used to find out which properties apply as long as there is no fault. Subsequently, the residuals for the period under investigation are used to check which of these properties are still valid. If a noticeable number of these properties are fulfilled in the fault-free period and deviate significantly from it in the period to be examined, then a fault is assumed at these points in time. In the following, suitable tests are listed for each propagated property of the residuals with links to the corresponding implementations in the statistics software R.

Tests for the median is zero:

- Sign test [R: BSDA]
- Wilcoxon signed-rank test [R: stats]

Independence test:

- Turning point test [R: spgs]

Correlation tests:

- Box-Pierce test [R: stats]
- Ljung-Box test [R: stats]

Tests for randomness:

- Bartels-rank test [R: randtest]
- Cox-Stuart trend test [R: randtest]
- Difference-sign test [R: randtest]
- Mann-Kendall rank test [R: randtest]

For the application, it is not useful to test all residuals at once. Instead, a time-window is defined that is shifted one after the other over all residuals and only tests within these residuals windows. For this purpose, the sample size within the window must be determined beforehand. For example, if the sample size is set to 10, the first ten residuals of the observed time series are tested first, then the window is shifted and the residuals two through eleven are tested and so on until the window cannot be shifted any further. For each shift a p-value is calculated for each test, this value is compared with the significance level α that was selected before. If the p-value is smaller than the significance level, then the behaviour of the residuals differs significantly from the tested property. Over all, three parameters are required for the procedure; these are the sample size for the residual window (l), the significance level (α) and the minimum number of tests (h) whose p-value should be smaller than the significance level to assume faulty behaviour. Each combination of l, α and h will be called CDR in the following, which is the abbreviation for *combined decision rule*. The goal is to find the optimal CDR to detect the faults.

Without data of already observed faults, it is very difficult to find the best CDR. For this case (Parzinger et al., 2020) have proposed a procedure in which many CDRs are used to calculate the rate of estimated faults. This rate of estimated faults is then used to detect the faults. The basic idea runs in several steps. First, as many different CDRs as possible are applied to the residuals from the fault-free month data. The CDR that assume any faults within this period are initially sorted out. Assuming that there are m CRDs left after sorting out, all m CDRs are used simultaneously to calculate m predictions for fault or no fault at each point in time. Finally, the rate is calculated at each point in time, i.e. the number of CDRs that assume a fault divided by the total number of CDRs m. If this the value of this rate is

greater than a previously defined threshold value between 0 and 1, then a fault is assumed. In summary, for the fault prediction technique the following steps are applied:

1. Calculate the residuals.
2. Use the residuals for the calculation of the p-values of each used statistical Test.
3. Use many CDRs one after another during the fault free time period.
4. Remove all CDRs that indicate any fault during the fault free time period.
5. Apply the remaining CDRs during the time period that is to be checked for faults.
6. Use these results to calculate the rate of estimated faults.
7. Use the rate of estimated faults to detect the faults.

The details for all steps are shown in (Parzinger et al., 2020).

3.3. Application to common exercise

The theoretic framework for FDD - as introduced in the previous sections - has been tested and applied by different modelling teams (7 teams from different universities and research institutes) as part of a second common exercise of this subtask. The goal of that common exercise was twofold. First and foremost, the goal of this exercise was to explore different modelling techniques in their ability and limitations related to the detection of different types of faults using on-board monitoring. Secondly, the exercise was set up in two rounds, starting with data generated using a detailed building energy simulation model and finally progressing towards actual measured data. This two-step approach allowed an early start of the common exercises, but more importantly provided an interesting research opportunity that allowed to first test the model identification and FDD techniques on well-conditioned data, allowing participants to compare different lengths and complexities of training data.

Moreover, this approach allowed us to analyse an underlying research question: “Can detailed building energy simulations models be used to identify (preliminary) models that can be later used in actual application?”

The test case for this common exercise were the Twin House experiments carried out at Fraunhofer in Holzkirchen (Germany) as part of this Annex project[1]. The next subsection briefly presents the test case and the faults that were introduced. Subsequently, the modelling approaches and most important findings from the common exercises are discussed.

3.3.1. Case study: Twin House (Holzkirchen, Germany)

The central case study for the common exercises on FDD was the side-by-side experiment carried out during this Annex 71 project in the Twin House buildings at Fraunhofer in Holzkirchen (Germany). The Twin Houses have already been the focus of model identification and validation exercises during the earlier IEA EBC Annex 58 project (https://www.iea-ebc.org/projects/project?AnnexID=58). The houses are situated in a flat location at Holzkirchen, Germany (near Munich). The latitude of the buildings is 47.874 °N, the longitude is 11.728 °E. The elevation above mean sea level (MSL) is 680 m. Time of all data provided is in Central European Winter Time i.e. (UTC/GMT +1).

The BES model validation study, conducted during Annex 58 (Strachan, Svehla, Kersken, & Heusler, 2014) was designed to focus on the fabric-related functionality of Building Energy Simulation (BES) programs including transmission heat losses, thermal bridges, solar gains, internal heat gains, window / blind models and internal and external air exchange. It did not consider occupancy user behaviour or typical heating and cooling systems. The following were deliberately not included in order to reduce complexity:

- No internal gains representative of occupants (heat, moisture and CO2)
- Constant set temperatures in constant temperature phases, no temperature profile or night setback
- Constant operation of a simple mechanical ventilation

[1] https://doi.org/10.1051/e3sconf/202017222003

- No windows opening
- No operation of internal doors
- No building service equipment, just electrical heating

The new experiments carried out in this Annex increase the realism and complexity. Key aspects of the changes are as follows.

- Including building services equipment: one of the Twin Houses (House O5: the test house) will have an underfloor heating system supplied by an air source heat pump. The other Twin House (House N2, the reference house), for comparison, will have electrical heating as for the Annex 58 experiments.
- Inclusion of attic space in the experimental configuration in addition to the ground floor rooms that were the focus in Annex 58. The construction properties of the walls of the buildings have also changed, although these changes are small.
- Including synthetic occupancy profiles: it was considered too complex to monitor real occupants, so a realistic synthetic occupancy profile was developed for the various rooms in the house, including window and door opening in part of the experiment.
- Including moisture injections for assessing moisture buffering effects (not during FDD experiment).
- A multi-stage operational schedule comprising a constant temperature phase (for Co-heating test assessment), a simple User-1 period with a temperature profile consistent across all rooms, a User-2 phase with a more complex user profile which varies from room to room and includes window and door opening, and a phase which includes moisture injection schedules.

Figure 21 and Figure 22 show an overview of the Twin Houses' geometry including the ventilation and door elements. The connection between both floors is a stair that is open in the living room in the ground floor and ends into a staircase in the attic from where doors lead to the children's rooms. This door can be sealed by a double trap door to create two separate air spaces of the ground floor and the attic. Sealed door, operable internal door supply air point extract, air point trap door (sealed/open according to schedule) are shown in the figures below.

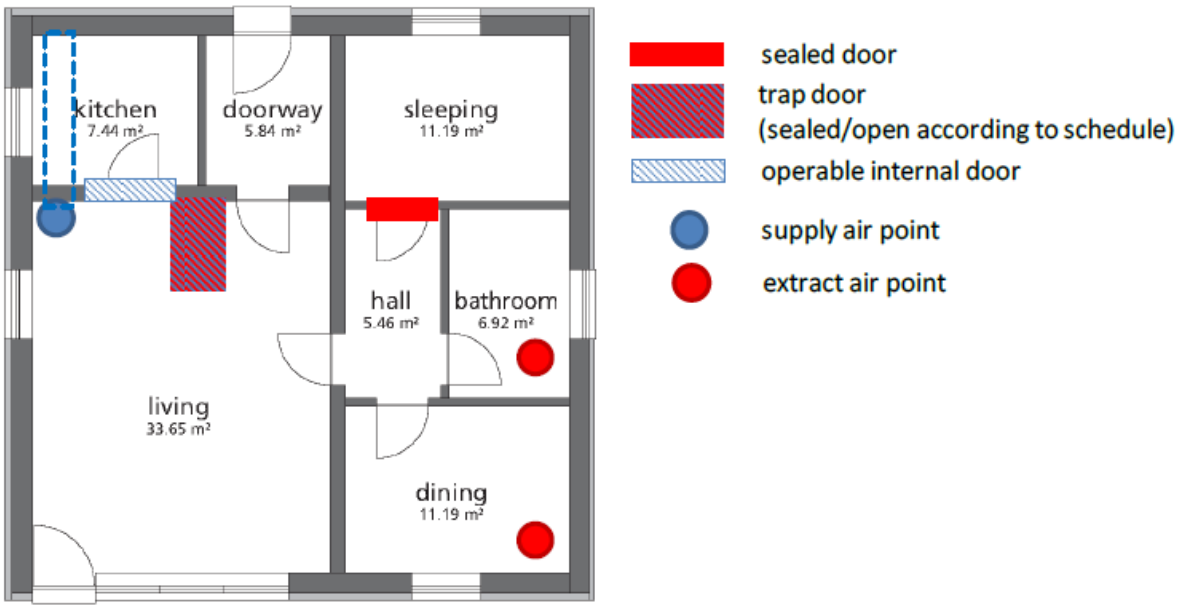


Figure 21. Floor plan ground floor.

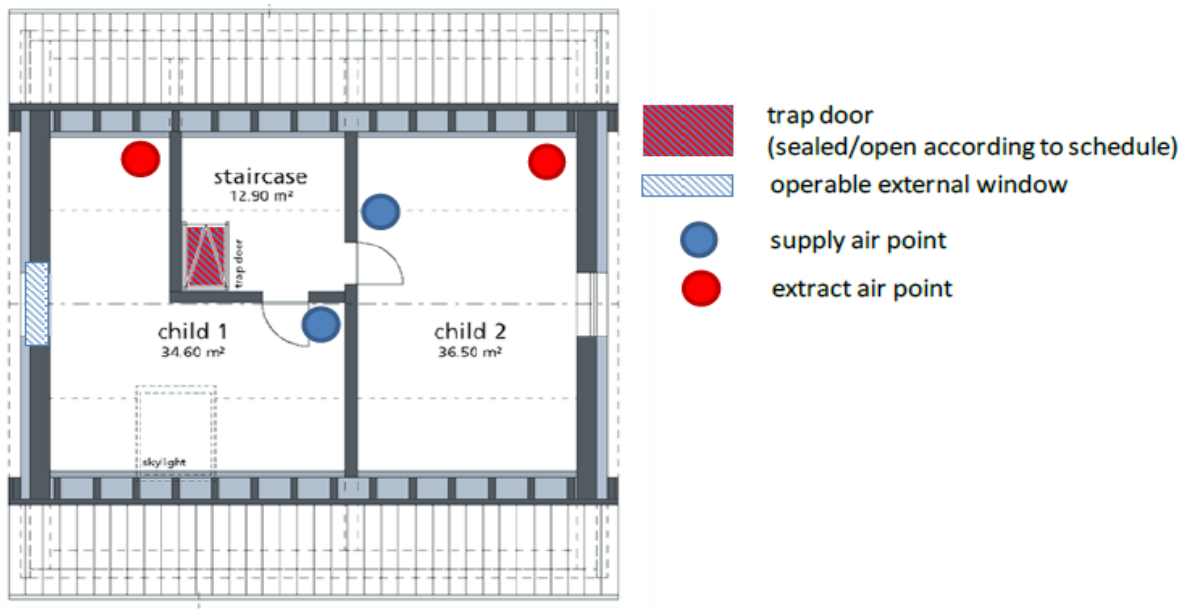


Figure 22. Floor plan attic

The main experiment was designed in 4 blocks. After an initialization period, first a Co-heating test ('coheat') was performed especially as input for the estimation of the overall heat loss coefficient in Subtask 3 of this Annex. Next, two 'in-use' experiment are carried out during which the building is emulated to represent normal occupation. Two different user profiles were imposed influencing the internal heat gains and temperature setpoint schedules for the heating system. Lastly, a dedicated FDD experiment is carried out. In this experiment – continuing on the User-2 operation – two faults have been introduced to the building systems. The common exercises on FDD are then specifically focused on exploring and testing different modelling techniques – trained on the 'in-use' period – for their ability to identify the faults and diagnose their causes.

Before going into the results, it is important to note that in the design phase of these experiments, a detailed building energy simulation model was implemented in EnergyPlus to aid the experiment designs (Mantesi et al., 2019). In this subtask, that model was used to generate a synthetic dataset for the 'User 2' and 'FDD' periods. In that FDD period, similar 'faults' were introduced in the simulation model as in the actual experiments. This allowed the modelling teams to go ahead with their modelling exercises without waiting for the actual measurements to be conducted. On top, it allows to compare the performance of modelling techniques between the synthetic data and the actual measurement data.

Data provided

- Indoor air and operative temperatures for each room
- Total heating power: total heat supplied to the floor heating system
- Plug loads
- Outdoor temperature
- Solar irradiation: global and direct irradiation on horizontal surface
- Wind: speed and direction

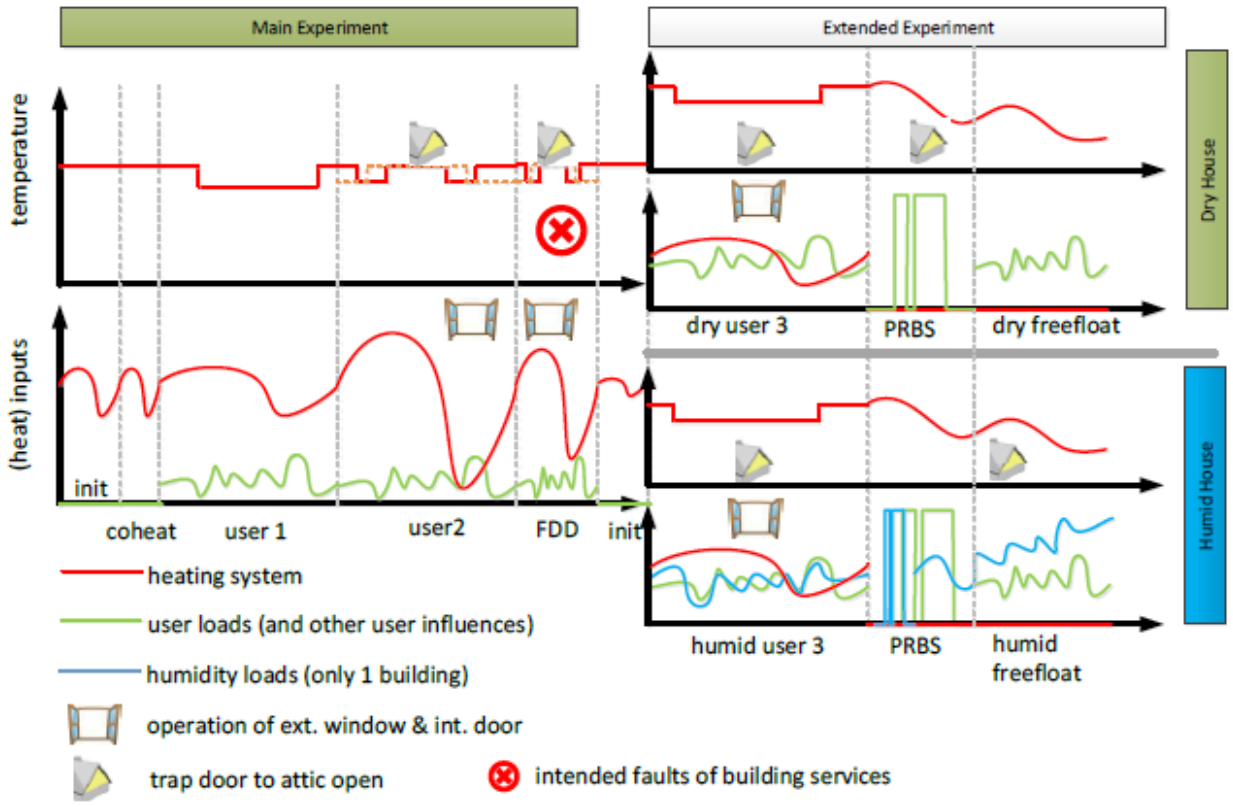


Figure 23. Overview of the experiment design indictating the sequence of tests.

3.3.2. Application to simulated data

A first common exercise to evaluate modelling techniques for FDD was launched on the results of a detailed building energy simulation model. Figure 24 (top) shows the indoor (red) and outdoor (blue) temperature for living room in the simulated Twin House experiment. Figure 24 (bottom) shows the corresponding heating power for the whole building. Two months of data was provided to the modelling teams, consisting of a training period (1 Feb 2019 – 28 Feb 2019) and an FDD period (1 March – 22 March). During the latter period, 3 system faults[2] were introduced to the model (Figure 25):

A. Under floor heating (02.03.2019 04:30 – 04.03.2019 10:30). During this period the underfloor heating in the living room was set to maximum power. Simulating a valve malfunction (can be mechanical or a sensor malfunction)
B. Ventilation ground floor (12.03.2019 16:20 – 22.03.2019 06:10) during this fault the supply and exhaust rates for the ground floor area were set to 0 m²/h, simulating a first fan issue.
C. Ventilation whole building (16.03.2019 06:00 – 22.03.2019 06:10) during this fault the supply and exhaust rates for the whole buildings were set to 0 m²/h, simulating a second fan issue.

Note that the models were not known to the modelling teams during the execution of the exercise. Instead, they were asked to identify start and stopping times of the faults, as well as to diagnose the source of the fault.

[2] An interesting anecdote in this exercise is that during the exercise design and simulation setup only faults 1 and 2 were planned. Nevertheless, due to a modelling error the 3rd fault was introduced. This error was discovered by the modelling teams who found not 2 but 3 faults. A finding that was confirmed when reviewing the simulation setup.

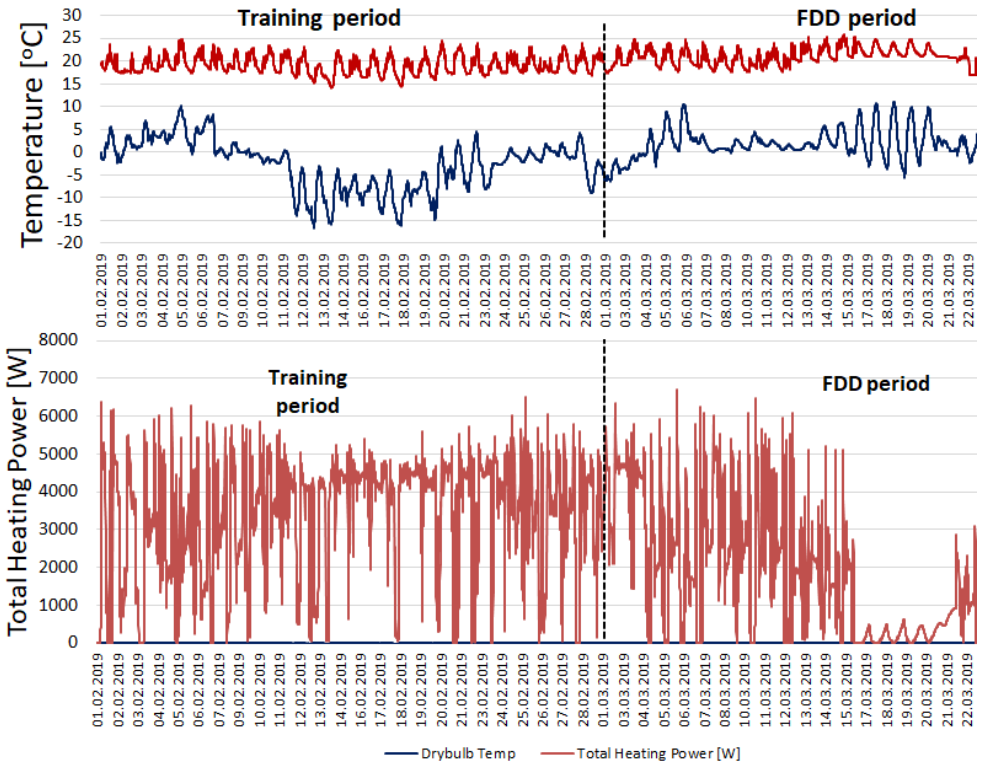


Figure 24. Indoor temperature and heating power of the Twin House for FDD common exercise based on simulated data

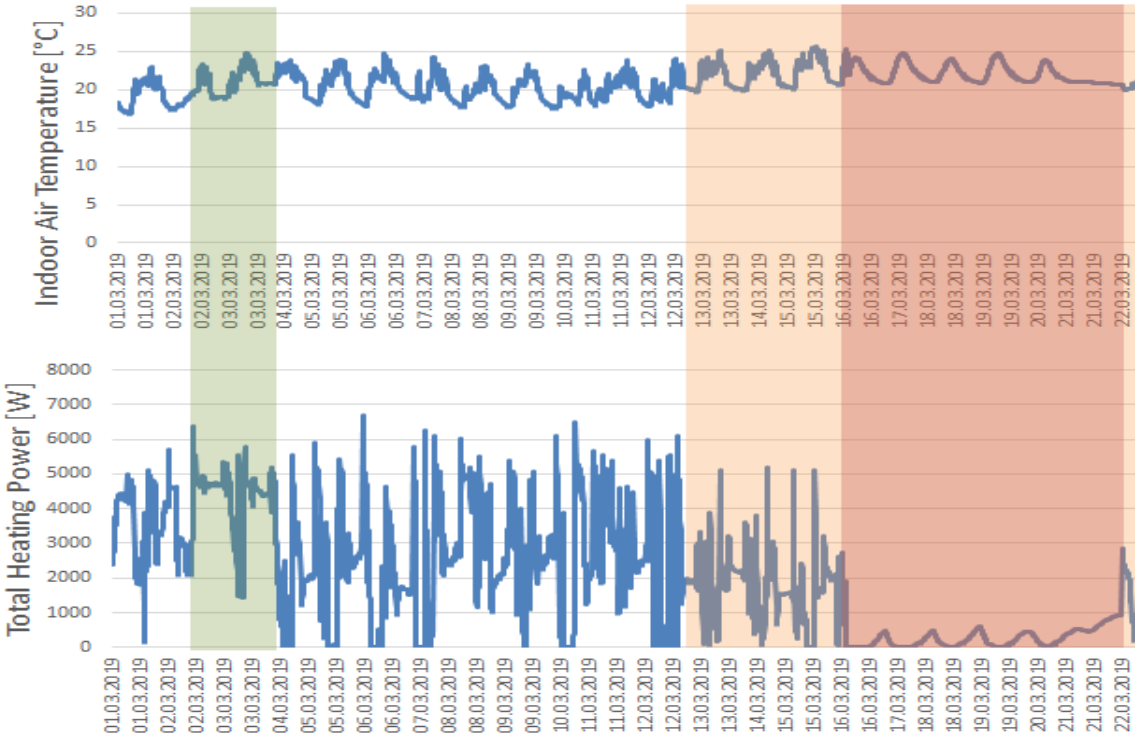


Figure 25. Details of the living room temperature and heating power, highlighting the faults. Green = fault A, Orange = fault B and Red = fault C.

In total five modelling teams performed the fault-detection exercise using 3 distinct modelling techniques. Two of the teams trained grey box models to predict the indoor temperature based on the climate data, plug loads and heating power. Two teams investigated regression-based techniques identifying respectively ARX and ARIMAX models. Lastly, one modelling team used a machine learning technique (random forest) to predict the heating power and from it detect the faults. Note that in all three cases, the modelling teams used an input-output modelling technique that predicts a physical parameter (indoor temperature or heating power in this case) by means of physical input parameters (e.g. weather data). While each team used – one or more – different indicator(s) to characterize the starting and stopping times of the faults, in each case the faults were identified from the model by looking for a sudden change in the prediction performance of the model. Thereby, as the models represent the input-output behaviour of a physical system, it is assumed that a sudden change in the prediction performance of the model coincides with a change in the physical system. In the terminology of section 3.2, these techniques hence are 3rd level FDD techniques focussing on the physical system.

Table 5. Overview of fault detection on the simulated data

Method	FA - timing	FA - cause	FB - timing	FB - cause	FC - timing	FC - cause
Greybox	/	/	11/3 17:19-15/3 12:00	**living**	**16/3 5:16 – 21/3 11:10**	**child 1**
Random forest	tbc	tbc	/	/	16/3 – 19/3	/
ARX	/	/	**12/3 – 22/3**	Local heating system issue (list given in report)	/	/
ARIMA	**3/3 7PM – 4/3 8 AM**	Added heat gain	23/3 - end	Added heat loss		
Greybox	/	/	**12/3 16:30 – end**	Reduction ventilation (28W/K reduction of HLC)	*16/3 10:40 –* ***22/3 6:20***	Second: undetected heat source (25W/K drop HTC)

Table 5 shows the results submitted after the first 'blind' phase during which the faults were unknown to the modelling teams. While these results only represent a first exploration of the techniques and further optimization of both the models as the identification techniques is still possible, some first interesting conclusions were drawn from this table. Firstly, only 1 team was able to detect an issue with the heating power, even though the starting and stopping time did not perfectly coincide with the actual fault A. This is because in the other modelling techniques the heating power was used as in input. A sudden change in heating power was hence not detected by most models as the response (temperature increase) to that change in heating power follows the same physical relation as to which the model is trained. The ARIMA model was still able to detect the change though, most likely as a result of the moving average part in the model structure. This hypothesis should be explored further.

As the ventilation flowrate was not used as an input signal by the modelling teams, faults 2 and/or 3 were identified by most teams. For these faults, the reduction in flow rates resulted in a different relation between model inputs and output (the physics of the underlying system changed). Although most teams were able to identify faults B and C, the spread on the starting and stopping times was significant. These changes can be attributed to both the accuracy of the model as well as the accuracy of the detection method.

Important differences in the methodology to identify start and stop times of the faults were shown between the modelling teams. A first group used a manual and visual inspection of the predicted temperature profile and or the residuals between measured and predicted temperature to decide upon the starting and stopping time of the faults. While being still labour-intensive, the use of a model and the comparison of the predicted temperature against the measured data, as exemplified in Figure 26, simplified the visual detection of the faults.

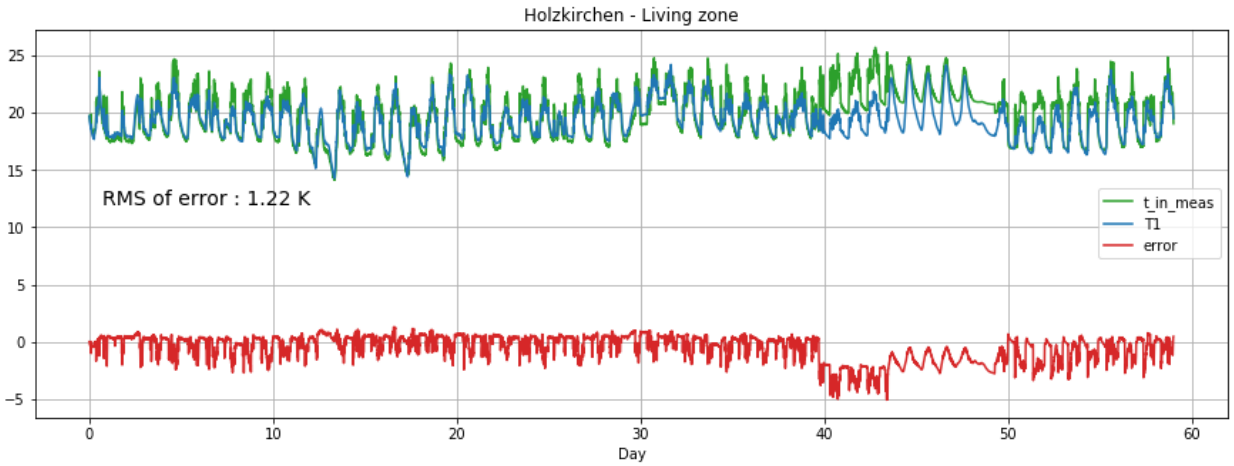


Figure 26. Comparison of measured temperature (green) compared to predicted (blue) and the residuals (red) using one of the Grey-box models

In addition to a visual inspection, few teams used a quantitative analysis of the prediction error at time k against the RMSE of the residuals on a cross validation test during the training period. For example, errors were defined to start when the prediction error exceed a threshold of 5 times the RMSE. All teams thereby distinguished between positive and negative errors, indicating the predicted indoor temperature (or heating power) to be too high or too low compared to the measurements. Alternatively, the confidence intervals around the predicted temperature were used and errors were defined when the measured temperature would fall outside the confidence intervals of the predicted temperature (Figure 27). In this approach, it was argued that the choice between e.g. 90% confidence intervals or 95% confidence intervals can significantly influence the detection process. Choosing larger confidence interval (e.g. 95%) may already lead to large differences between measured and predicted temperature leaving small errors to go unnoticed (false negatives), while narrow confidence bands (e.g. 90% or 68%) may lead to many false positives. Based on these findings, a search for more robust statistical methods was started leading to the methods discussed on section 3.2.

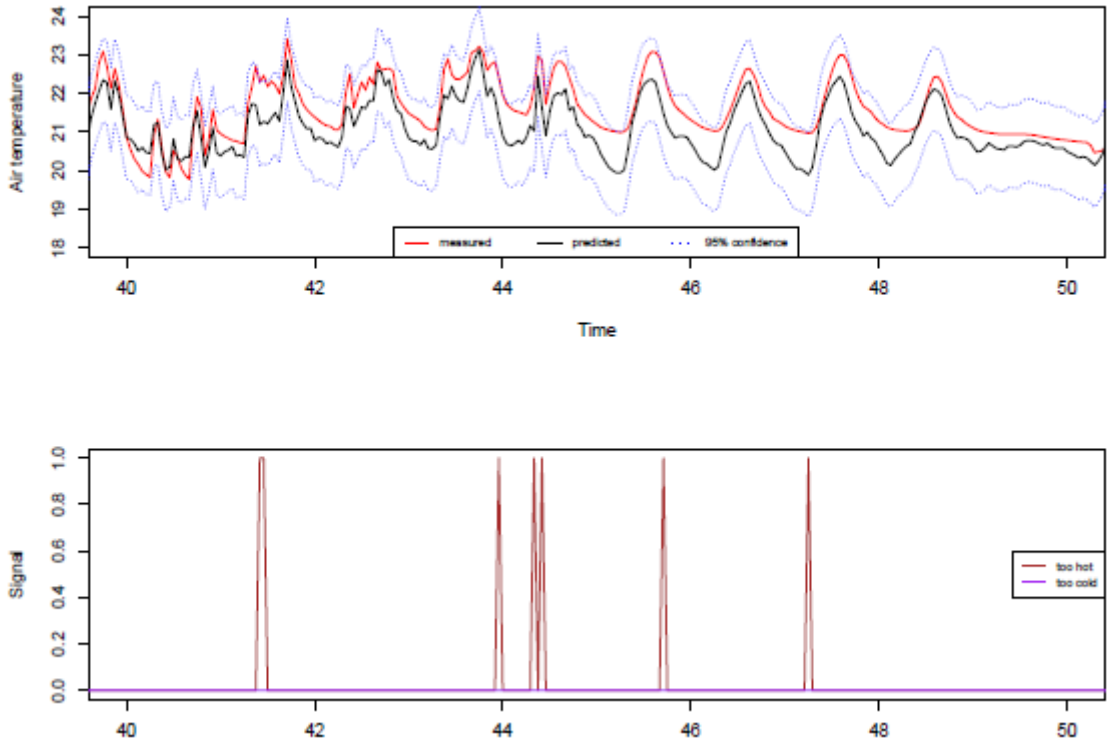


Figure 27. Example of a fault detection process based on confidence intervals

In addition to the detection of the starting and stopping times, modellers were asked to explore if their methodology allowed to diagnose the source of the faults. By separating positive from negative residuals, most teams were able to identify if the fault was due to an increase in heat loss (or equivalent reduction of heat gains) or due to a decreasing heat loss (or equivalent increase in heat gains). In the case of the grey-box model, one team explored to possibility to use the physical interpretability of the model parameters to diagnose the faults. By re-estimating the model parameters on a daily basis and analysing the resulting estimate for the buildings heat transfer coefficient HTC (W/K) they noticed that during the period of fault 2 the heat transfer coefficient reduces by 61 ± 27.7 W/K (from the black dots to the blue dots in Figure 28). The second reduction of HTC was deemed too high to be attributed to an envelope or ventilation change by the modelling team. Assuming this second fault was due to an unmodeled heat source, the argued that a heat source of 2041.09 W/K was needed to explain the differences. While no solid proof of the fault could be delivered, it was interesting to see that using this approach physically viable hypotheses could be formulated. Further research is needed to explore to what extend more detailed grey-box models – using additional measurement data – could further improve the diagnosis.

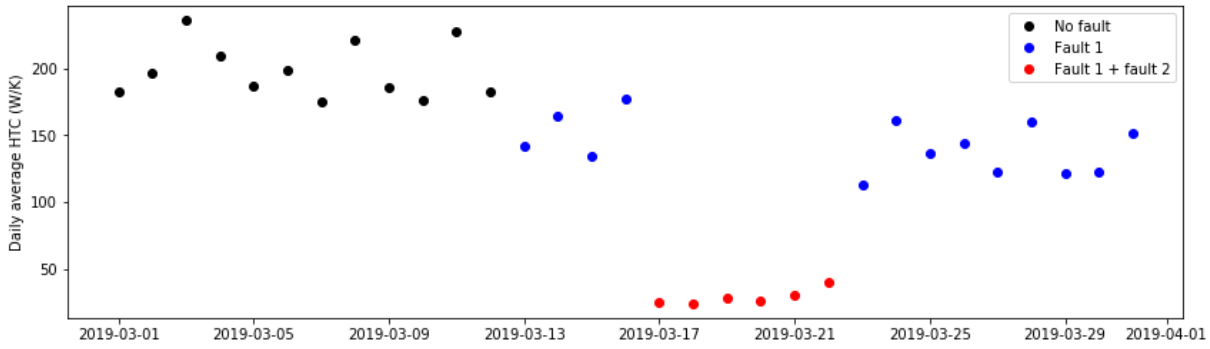


Figure 28. Diagnosis of the heat transfer coefficient based on the grey-box modelling approach for FDD

For the detection based on the random forest an update of the feature selection was explored to identify possible sources of the faults. By comparing the most influential features between the training period and the period during which the faults were detected, the modelling team tried to identify to which variables the change was correlated. While Figure 29 shows that the load and temperature in the child rooms have increased in importance, suggesting that something is happening in these rooms, further research is needed to verify the generalisation of these findings.

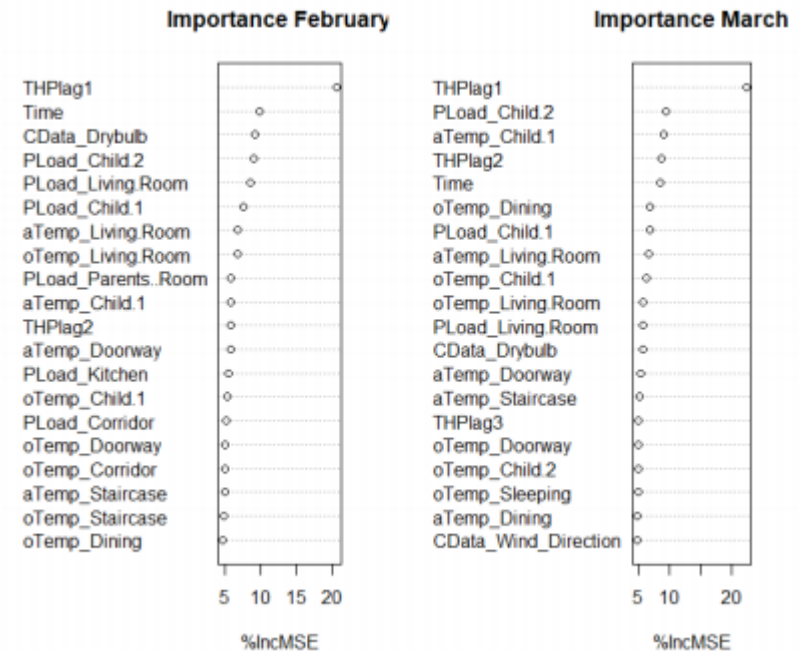


Figure 29. Comparison of importance of features for random forest between training period and faulty period

3.3.3. Application to real data

Following the FDD exercise on the simulated data, another exercise was launched using the data of the actual Twin House experiments. In the common exercise, the focus was put on the O5-building equipped with floor heating. During the FDD phase of the experiment, two faults were introduced:

A. The ventilation on the ground floor (04.03.2019 17:30 – 15.03.2019 09:30): the ventilation was shut down during that period by stopping the supply in the living room and the exhaust in bathroom and dining.
B. Heating in living room to 100% (08.03.2019 17:50 – 11.03.2019 13:15). Underfloor heating is set to maximum power by increasing the set point temperature for the living room to 40°C.

As for the common exercise on simulated data a significant spread on modelling approaches was followed ranging from 'white-box' building energy simulation models (TRNSYS & Contam), grey-box models and model-free approaches using principle component analysis and machine learning techniques (random forests) to detect the starting and stopping times of the two faults. Thereby it was noted that, following on the difficulties encountered and lessons learned on the simulated data, all times restricted themselves to the detection state and did not go into detail on the diagnosis. In this detection, modelling teams have detailed further a quantitative approach for the definition of starting and stopping times. A method based on daily cumulative residuals was used in combination with a grey-box model approach. Thereby a 3rd order grey-box model was fitted for each room using detailed measurements of the zonal indoor temperatures, surface temperatures and the heating power. Based on the comparison of the predicted indoor temperate and the measured room temperature, the maximum positive and negative cumulative daily residuals were calculated during the training period. A fault (positive or negative) was then defined when the daily cumulative residual was higher than (or lower than) 1.05 times the maximum positive (or negative) daily cumulative error during the training period. Based on this method, a fault in the living room was detected start at March 9th at 7 AM and to stop at March 12 at 5 AM. On top of the detection in the living room, the error was also identified in the other rooms on the ground floor, which is likely to follow from the secondary effect that because of the increasing temperature in the living room, the temperature in the adjacent rooms will also increase. Moreover, it was argued that the delay in the fault detection might be attributed to the use of daily cumulative values.

For the Random Forest approach, an elaborate statistical analysis was carried out comparing the p-values of a set of statistical validation techniques (described in section 3.2). The setup of this analysis was design to optimize the fault detection procedure in order to minimize the occurrence of "false positives" and "false negatives". Further analysis is however needed to generalize the findings of this exercise.

3.4. Lessons learned

The goal of this activity was to evaluate and test the contribution of prediction models - obtained through building behaviour characterization – to automated fault detection and diagnosis. It was identified based on a literature review that FDD in commercial applications still relies on the individual component or system level, with individual processes being monitored by dedicated sensors. While building energy management systems are finding their way into the market, only recently methodologies that exploit overarching data are being developed.

The work carried out in this activity contributes to that development in three significant areas. First, a conceptual framework has been described that defines and organizes different types of faults. For each type of faults, different detection methods are proposed. The development and layout of this framework is inspired by the observations in the common exercises that demonstrated that despite detailed and accurate modelling techniques some type of faults could not be detected. For example, when using input-output models for anomaly detection that use the heating power as an input, it becomes impossible to detect errors in the control of that heating power (e.g. a thermostat malfunction). Consequently, as discussed in the framework, the type of errors to be detected should be properly matched to the modelling approach, taking into account the physical behaviour of the building.

Second, an overview was made on statistical methods used to detect the actual faults. The methods discussed focus on the detection of faults by comparing the predicted behaviour against the actual measurements. With this overview, guidelines are provided to move beyond the need for modeller/operator interpretation and move towards an automated detection process.

Third, by demonstrating the application of the identification of prediction models and the fault detection process for both simulation and actual measurement data for the same case study, this activity indicates that detailed building energy simulation models can play a significant role in the further development and research on automated fault detection and diagnosis methods. Further research is however needed to generalize these findings are a wider set of building and addressing a broader range of anomalies.

4. Model predictive control

4.1. Introduction

After assessing the performance of building behaviour models in FDD applications, it was decided to assess the performance of those models in another operational application. The application chosen here is called Model Predictive Controller (MPC) which is briefly described below.

In order to increase the energy efficiency of buildings many solutions exist. Reducing the energy demand by increasing the quality of the building envelope (such as using better insulation, lowering the solar gains with shading devices, and increasing the air-tightness) remains the most important measure. It not only reduces the need for energy but also facilitates the use of Renewable Energy Sources (RES) and efficient building energy systems. The integration of RES and the deployment of smart technologies in buildings to facilitate incorporation of buildings in demand-side management schemes requires advanced, smart control technologies. Conventional controllers are not able to handle the complexity introduced to buildings energy structure by RES and smart technologies. One of the most promising methods for integrating RES and smart technologies in buildings is Model Predictive Control (MPC) which has proven to have superior performance compared to traditional control techniques applied to buildings (Afram & Janabi-Sharifi, 2014). MPC is an active control strategy which optimizes performance of a system over a given time horizon (Maciejowski, 2002). It has shown considerable potential in optimizing performance of HVAC systems along with facilitating the integration of RES in buildings (Atam & Helsen, 2016). An important feature of MPC, which boosts its performance in buildings, is its ability to handle slow moving dynamics, which suits the requirement of a good optimization strategy for buildings. To achieve the latter, MPC uses a model of the building to predict its hygrothermal behaviour in the future. Deploying this model allows MPC to calculate the optimal input trajectory of the system in a time horizon called the "*control horizon".* This prediction feature gives MPC a crucial edge compared to other controllers (Atam & Helsen, 2016). Essentially, MPC is composed of two main parts; a predictive model and a solver. Developing a predictive model has been identified as one of the costliest and most challenging steps in designing an MPC for buildings. Hence, given the ever-increasing interest in the application of Model Predictive Control in buildings and the crucial role of predictive models deployed in it, we have defined a common exercise to assess the performance of different building behaviour models in the context of an MPC.

In this section, we first describe the concept of model predictive control and the performance indicators used to assess the predictive models and the controller performance. Section 4.2 is devoted to the description of the framework of the MPC applied to the building under study. Some considerations about the common exercise along with the formulation of Optimal Control Problem (OCP) are presented in section 4.3. Section 4.4 describes the development of the predictive models used as the predictor in this exercise inside the defined framework. Afterwards, validation of these models is reported in section 4.5. In section 4.6 the results of the models applied in MPC are shown and discussed.

4.1.1. Brief Introduction of Model Predictive Control

Model-based Predictive Control (MPC) is categorized as an advanced control technique which was first introduced in 1969 but it wasn't given much attention up until late 70's where it was used by chemical engineers to a fluid catalytic cracker process which marks the first physical implementation of MPC (Maciejowski, 2002; Wishart, Lee, & Markus, 1969). But this wasn't enough to make this new technique popular in other fields, application of MPC remained confined to the chemical process until the very beginning of 21th century when theoretical foundations of MPC regarding stability, solution feasibility and other theoretical backbones were defined and established which made MPC quite popular in multiple fields (Mayne, Rawlings, Rao, & Scokaert, 2000).

In model predictive control, an optimization problem is formulated and solved to find the optimal control input sequence over a time horizon. MPC essentially aims at minimizing a user-defined objective function in a given time horizon subject to some constraints. In general, the Optimal Control Problem (OCP) could be formulated as Equation (4.1).

In this formulation, *z(.)* is the objective function, *f(.)* and *g(.)* describe the dynamics of the system, *k* represents the future time steps and *N* is the control horizon. The inequalities represent the constraints imposed on system inputs and states. At each time step, the optimal control problem is solved and the optimal control policy $(\vec{u}_k, \vec{u}_{k+1}, \ldots \vec{u}_{k+N-1})$ is calculated by solving the OCP presented in Equation (4.1)**.** Only the first element of the control policy namely $\vec{u}_k$ is applied to the system. In the next time step, OCP is solved again and the first element of the optimal control policy is applied to the system. This process is repeated at each time step and is called "*receding horizon method"* (Maciejowski, 2002). This scheme is shown in Figure 30 Given the current states of the system and a model of the plant, an optimal

control policy is calculated at time *t* but only the first element of this optimal policy (u_0) is applied to the system and then the whole process is repeated at time step *t+1.*

$$
\begin{aligned}
&\min_{\bar{u}_0,\bar{u}_1,\ldots\bar{u}_{N-1}} \sum_{k=0}^{N-1} z(x_k,u_k) \\
&s.t: \\
&x_0 = x_t \\
&x_{k+1} = f(x_k,u_k) \\
&y_k = g(x_k,u_k) \\
&x_{k,min} \le x_k \le x_{k,max} \\
&u_{k,min} \le u_k \le u_{k,max} \\
&k = 0,1,\ldots N-1
\end{aligned}
\qquad (4.1)
$$

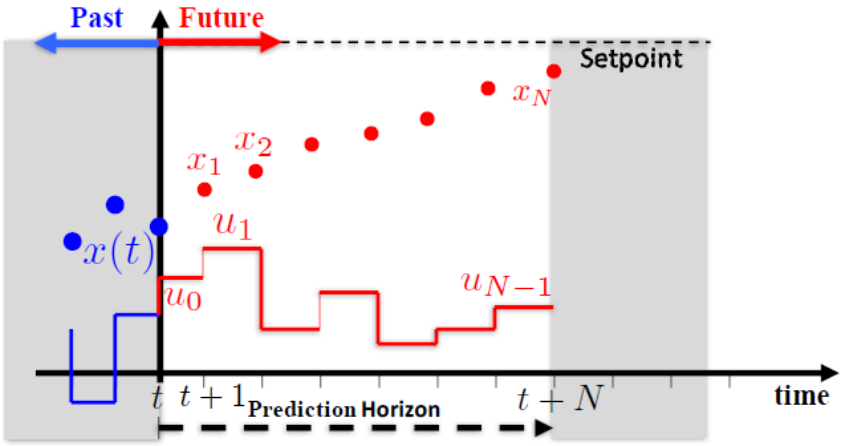


Figure 30. General Scheme of MPC (Sarimveis et al., 2008)

Given these descriptions, one can distinguish two main components in an MPC: the solver and the predictive model. The solver is used to solve the OCP formulated as Equation (4.1)**.** The Predictive model is used to predict the dynamics of the plant in the optimization problem and it is an integral part of the MPC formulation. Other components might be a part of an MPC such as state observer which based on the type of model might be needed in the control structure (Arroyo, Spiessens, & Helsen, 2020). Compared to other control methods MPC has many advantages:

- By deploying a predictive system model, MPC is a proactive controller rather than a reactive one
- A disturbance model can be easily incorporated in MPC for minimizing the effect of disturbances on the system
- MPC inherently handles constraints including physical, performance and safety constraints
- MPC can handle a wide range of operating conditions
- MPC is able to tackle multiple objectives concurrently

4.1.2. Evaluation indicators for model and controller

MPC is not exactly an algorithm but rather a strategy. Hence, many types of models could serve in MPC as the predictive model such as the models described in chapter 2 of this report. To quantify the performance of these models in the MPC Key Performance Indicators (KPIs) for both model and the controller have to be identified. We need KPIs for scoring the models used in the MPC as well in order to quantify their quality. Some of the most important KPIs used to evaluate the performance of MPC in buildings are energy use, operational cost, discomfort level, demand charge,

real-time response, CO_2 emissions, capital cost and robustness (D. Blum, F.Jorissen, 2010). In this report, we narrow down our attention to two of these KPIs, which are energy cost and thermal comfort.

KPIs used for assessing performance of predictive models could be divided into three main categories:

1. KPIs focusing on the model accuracy

   These KPIs evaluate models ability to predict the future states of the system. Most popular ones are listed in Table 6.

Table 6. List of model's accuracy KPIs

Measure	Formula
Prediction Error (Verhelst, 2012) (PE)	$\sum_k [\hat{y}(k) - y_m(k)]$
Mean Bias error (Zhan & Chong, 2021) (MBE)	$\frac{\sum_k [\hat{y}(k) - y_m(k)]}{\bar{y}_m}$
Maximum Absolute Error (Bourdeau, Zhai, Nefzaoui, Guo, & Chatellier, 2019) (MAE)	$\max \lvert \hat{y}(k) - y_m(k) \rvert$
Mean Absolute Error (Afram & Janabi-Sharifi, 2014; Zhan & Chong, 2021) (MAE)	$\frac{\sum_k \lvert \hat{y}(k) - y_m(k) \rvert}{\bar{y}_m}$
Mean absolute percentage Error (Afram & Janabi-Sharifi, 2014; Zhan & Chong, 2021) (MAPE)	$\frac{\sum_k \lvert \hat{y}(k) - y_m(k) \rvert}{\bar{y}_m} * \frac{100}{n}$
Mean Squared Error (Afram & Janabi-Sharifi, 2014; Bourdeau et al., 2019; Zhan & Chong, 2021) (MSE)	$\frac{\sum_k [\hat{y}(k) - y_m(k)]^2}{n}$
Root Mean Square Error (Ferracuti et al., 2017; Hong & Kim, 2018) (RMSE)	$\sqrt{\frac{\sum_k [\hat{y}(k) - y_m(k)]^2}{n}}$
Coefficient of Variation of RMSE (CVRMSE)(Aftab, Chen, Chau, & Rahwan, 2017; Bourdeau et al., 2019; Coakley, Raftery, & Keane, 2014; Feng, Chuang, Borrelli, & Bauman, 2015; Reynolds, Rezgui, Kwan, & Piriou, 2018)	$\frac{\sqrt{\frac{\sum_k [\hat{y}(k) - y_m(k)]^2}{n}}}{\bar{y}_m}$
Normalized Mean Square Error (NRMSE)(Bourdeau et al., 2019; Hedegaard, Pedersen, Knudsen, & Petersen, 2018; Huang, Chen, & Hu, 2015)	$\frac{\sqrt{\frac{\sum_k [\hat{y}(k) - y_m(k)]^2}{n}}}{y_{m,max} - y_{m,min}}$
Coefficient of determination (Bourdeau et al., 2019; Zhan & Chong, 2021) (R^2)	$1 - \frac{\sum_k [\hat{y}(k) - y_m(k)]^2}{\sum_k [y_m(k) - \bar{y}_m]^2}$

2. KPIs focusing on the model complexity

   These KPIs present the level of complexity of the predictive models.

   The most important KPIs used in this category the order of the model which is the number of states for state space models, the number of output lags for ARX models or the number of neurons in an ANN model (Abu-Mostafa, 1992; Atthajariyakul & Leephakpreeda, 2005; Erfani, Rajabi-Ghahnaviyeh, & Boroushaki, 2018).

   The other important KPI in this category is the time required for developing a model by the modeller. Although it is quite difficult to measure but it gives an impression on the level of skills that are required to implement a modelling technique.

3. KPIs focusing on the model usability

   One of the most important of these KPIs is the generalizability of the model, which is defined as the possibility of applying the model to different types of buildings.

   The other important KPI in this class is the model's adaptability which is defined as the model's capability to cope with the changes in the real-life plant (which is building in our case) (Atam & Helsen, 2016).

   As for the previous category, these KPIs are rather difficult to quantify but are of great importance to evaluate the ease of implementation and hence a more widespread adaptation of MPC.

Although in this work, the focus is on the first category of KPIs, we will reflect on the other categories where appropriate. It is important to note that the KPIs defined in the first category evaluate the one-step-ahead prediction error (OSPE) of the predictive model. As can be seen from the OCP formulation in Equation (4.1) the MPC solves the optimization problem in a time horizon not just for one time-step. Hence, MPC needs a model, which has good predictive abilities throughout the control horizon not just one-step ahead. For instance, if a model is over-fit, it might yield very good results in term of one-step ahead prediction error but poor predictions throughout the rest of the horizon. The importance of this issue has led to the emergence of a new concept in the field of MPC, which is mainly known as *Model Predictive Control Relevant Identification* (MRI). Modelling techniques, which work based on this concept aim at minimizing the multi-step ahead prediction error (MSPE) of the predictive model not just the OSPE (Pčolka, Žáčeková, Robinett, Čelikovský, & Šebek, 2016; Prívara, Cigler, Váňa, Oldewurtel, & Žáčeková, 2013; Žáčeková, Vana, & Cigler, 2014; Žáčeková & Prívara, 2012). We will also use the MSPE to score the predictive models serving in the MPC later on in this work.

4.2. Description of the framework

This section explains the general framework for the common exercises that was designed for assessing the performance of predictive models in an MPC. Before going into the details of the exercise, we need to pinpoint the aim of doing this exercise. The aim of this exercise is two-fold:

1. Assess the quality of predictive models in an MPC:

   This is the main objective of this common exercise. To achieve this goal, first, we need to find a suitable KPI to score the performance of predictive models serving in the MPC. Afterwards, we assess the effect of selecting different modelling approaches on the performance of MPC with the help of the suitable KPI. To conduct this evaluation, models' quality and controllers' quality need to be qualified using relevant KPIs. This qualification facilitates a systematic comparison between the models and paves the way for a thorough analysis on how predictive models would affect the whole control scheme.

2. Demonstrate opportunities of the MPC:

   To this end, we first compare the performance of MPC with a traditional controller commonly used in buildings and then we investigate how MPC manages to integrate Renewable Energy Sources (RES) in buildings' energy system. To achieve this, a Time of Use (ToU) electricity pricing has been included in the exercise, which reflects the integration or RES.

4.2.1. Test case

Now that we have defined the goal of the exercise, we need to select a suitable test case for this common exercise. To find a suitable test case, we need to go back to the objective of the controller that we are going to design. Goal of our

MPC is to minimize energy costs of the heating system while minimizing thermal discomfort level simultaneously. As in subtask 2 we intend to assess predictive models developed based on measurements from a building, we wanted to choose a building, which has extensively available measured data. The Twin Houses –which have been considered many times in this project-, fulfil this criterion. However, to choose one of the twin houses as the use case in this study we need to look at the differences with regard to the aim of the exercise, which is assessing different predictive models in MPC. The main difference between O5 and N2 buildings is the fact that O5 is equipped with an underfloor heating system while N2 is equipped with electric heaters. This variance gives us enough material to choose the suitable test case.

An underfloor heating system deploys the passive thermal storage of a building's thermal mass, which increases the time constant of the system. Hence, controlling such systems is more challenging than controlling conventional heaters. However, MPC has shown superior performance for such slow-moving systems with time delays. Therefore, we selected the O5 house as the test case in our study (Cigler, Gyalistras, Siroky, Tiet, & Ferkl, 2013; Pichler, Goertler, & Schranzhofer, 2017; Schmelas, Feldmann, Wellnitz, & Bollin, 2016; Viot, Sempey, Mora, Batsale, & Malvestio, 2018; Zakula, Armstrong, & Norford, 2015).

For the purpose of this exercise, the User-1 phase of the experiment is considered with the duration of 44 days. In User-1 phase, occupants are introduced to buildings as internal heat gains with a pre-defined schedule by means of electrical heaters (see section 3.3.1 for more details). The building simulation model of the O5 building was developed using OpenIDEAS which is an open-source library using the Modelica language through Dymola (a commercial tool for Modelica language). Further details on the test case can be found in section 3.3.1 of this report. It should be mentioned that the only difference between the model described there and the one we are using here for this common exercise is adding the model of the heat pump to be able to control the performance of the heat pump. The heat pump is an air-to-water heat pump, which supplies the water for the underfloor heating system.

Another important note is that for simplicity in this common exercise, different zones of the building are lumped together. In this exercise, the volume-averaged temperature of all 10 thermal zones of the building is to be controlled. The OpenIDEAS simulation model serves as the emulator in this exercise. Henceforth, we call this simulation model *"the emulator"* and the models, which are delineated in chapter 4 of this report as *"predictive models."*

Now with clear goals in mind and a general description of the test case we can move one-step further and translate those goals into the objectives, which are to be achieved by the controller. The controller aims at optimizing the electricity consumption of the heating system of the O5 building while maintaining indoor thermal comfort. In other words, two objectives are to be concurrently satisfied by the controller: minimization of heat pump's electricity consumption and minimization of thermal discomfort.

4.2.2. MPC Framework

Defining the aim of the exercise (assessing performance of predictive models in the context of MPC) and the controller objectives (minimizing electricity cost and thermal discomfort), we are now able to design a framework for MPC development and assessments. To evaluate the models in an MPC we need to design a complete and detailed framework in which the performance of the predictive model could be evaluated. This framework is illustrated in Figure 31.

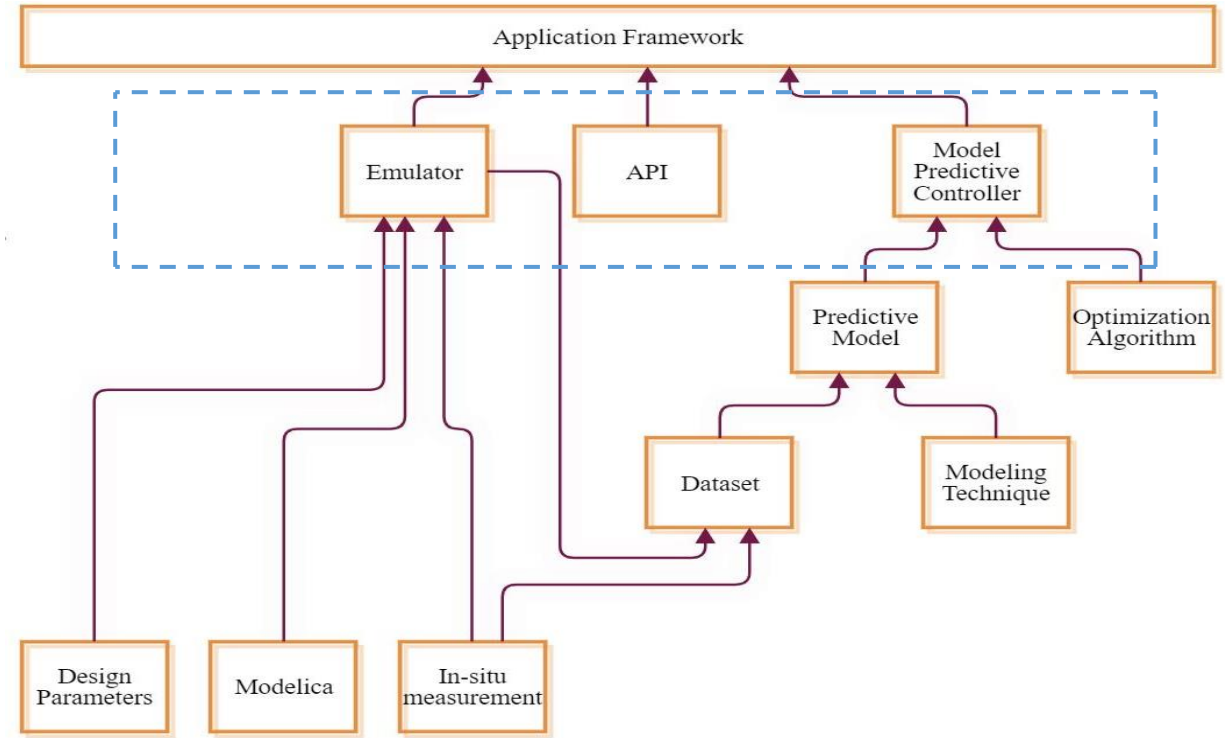


Figure 31. General Framework of an MPC with a data-driven model

Henceforth, we are going to describe this flowchart from top to bottom. First step of this framework is collecting building's physical and geometrical inputs (design parameters) which are in turn used in a building simulation software such as Modelica to develop an accurate simulation model of the building, which we call emulator in this exercise. This model is then validated based on in-situ measurements of the building. In-situ measurements or proxy data generated from the emulator are then fed to different modelling techniques to train data-driven predictive models. These predictive models are a part of OCP (see equation 4.1) which forms a full MPC design along with the solver and constraints.

In this section, we focus on the top three components of this flowchart indicated by the blue dotted rectangular. As it could be seen in Figure 31 to have a complete framework, we need three main components: an Emulator, a model predictive controller and an Application Programming Interface (API), which are going to be elaborated further on.

Emulator

To be able to evaluate the performance of different models in MPC, we need to develop an emulator, which simulates the performance of the building under study. This emulator will replace the real building in a simulation environment. As explained earlier, the simulation model of the O5 building used has been developed using Modelica language in Dymola and is used as the Emulator model here.

Complete Model Predictive Controller

As could be inferred from Figure 31 MPC is composed of two main components; a solver (optimization algorithm) and a predictive model.

Optimization Algorithm:

To select a suitable solver we need to keep the features of our OCP in mind. Our optimization problem includes both integer and continuous decision variables. Therefore, the solver must be capable of handling mixed integer programming problems. Another feature of the solver is regarding the convexity of the optimization. Assuming that the constraints on states and inputs in Equation (4.1) are all linear and the objective function is convex then the predictive model determines the convexity of the optimal control problem. Using a non-linear model would most probably lead to a non-convex optimization while using a linear model would yield a convex

optimization problem. Convex OCPs could be solved much faster and reaching the global minimum in the solution space is guaranteed by most of the methods for such programming problems, while the same doesn't hold for non-convex OCPs; they could be computationally demanding to solve and reaching the global optimum couldn't be guaranteed in most of the cases (Atam & Helsen, 2015). In addition, since we want to minimize the effect of solver on the results obtained from implementing the controller, we will use the same solver for all models. Given the former explanation, the solver has to be able to cope with non-convex optimization and Mixed Integer programming as well. In this study, we used Genetic Algorithm (GA) as the solver since it has proven to be able to handle both of these complexities.

Predictive Model:

This part of the framework is the one that we pay most of our attention to. As it could be deduced from Figure 31 predictive model is developed by deploying a modelling algorithm and applying it to a dataset. This data could be collected from the building itself or from the emulator. In this exercise as the main aim was to assess the impact of different modelling techniques on the performance of MPC, we resolved to use proxy data from the emulator to avoid the undesired impact of emulator-building mismatch on the performance of MPC. This data is then used to identify the parameters of different modelling techniques or in other words "training" them. Each model based on its characteristics might use different types of data to use for training. Another dataset is then used to test the performance of these models (section 4.3.1). Development of predictive models are exhaustively explained in section 4.4 of this report.

An Application Programming Interface

Since the emulator is developed in Dymola and the controller is developed in Matlab, an interface is needed for the communication of the two softwares with each other. To handle this issue, we use a co-simulation platform, which facilitates the connection between Modelica and Matlab. This co-simulation is carried out by means of Functional Mock-up Interface (FMI). We are going to briefly describe the FMI here. Many of the simulation languages such as Modelica essentially translate the models into equations and solve those equations afterwards. In case of Modelica, these equations are derived from physical laws governing the building or the physical plant. These equations are then solved using various existing solvers. Keeping this in mind, what the FMI does is that it translates those equations into a universal binary format, which is compatible by many simulation tools. Hence, other tools could load and run the model. A straightforward method was implemented: after developing the emulator model of the O5 building, it is compiled as an FMU compatible with the FMI 2.0 standard. Then the emulator is imported into Matlab Simulink using the FMU block in Matlab, which is used for reading FMUs. From there on, the FMU serves as the emulator model in our MPC framework and communicates with the controller in the Simulink environment.

4.2.3. Modelling options

Two options were provided to implement models in the MPC framework:

Option 1: a discrete-time model of the Twin House is to be provided. The solver along with the whole MPC setting is to be carried out by the team who designed MPC.

Option 2: Full implementation of MPC is to be carried out by modelling teams who choose this option. In another word, they are required to develop predictive model and integrate them with a functioning MPC design.

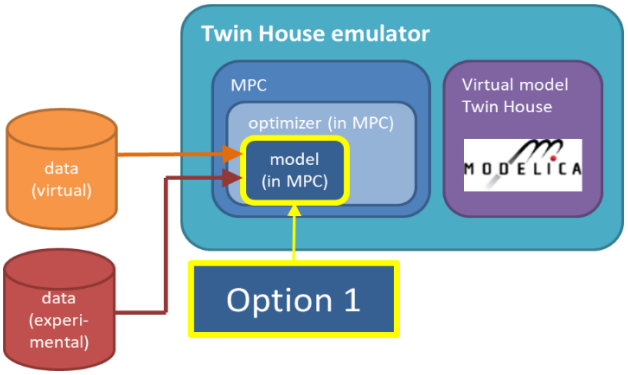


Figure 32. Framework of option 1 for the participants (designing only predictive model)

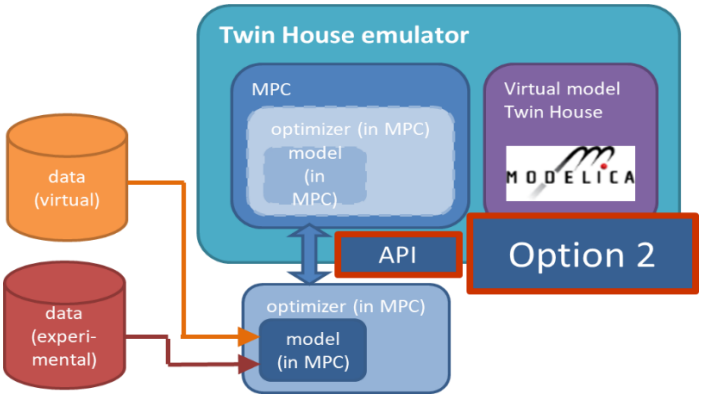


Figure 33. Framework of option 2 provided for the participants (design of whole MPC)

The latter option was foreseen for the participants who wished to design their own MPC. The modelling teams who chose this option, were supposed to develop not only the predictive model but also the whole optimizer along with it. This option is called *"option 2 "throught* the text.

As mentioned, one of the modelling teams decided to go with option 2. This team was provided with the building's FMU file as the emulator. In this approach, GA was chosen as the optimizer to minimize impact of factors other than the predictive model on the controller's KPI. The only difference in the OCP formulation is the fact that thermal discomfort is penalized with an exponential function while in the main OCP formulation it is penalized in terms of Kelvin hours outside thermal comfort bands. They developed their MPC in R, which communicates, with the building's FMU through Python while the rest of the models were tested in MATLAB SIMULINK.

4.3. MPC Pre-requisites

In this section, we thoroughly describe the procedure used for developing the data-driven models. As seen in Figure 33 for developing a data-driven model, a dataset and a modelling technique are acquired. An overview of the modelling techniques used in this study is provided in section 4.4 of this report. Here, we first describe the process of dataset generation applied in this common exercise. Then the general structure of the MPC used in this study is explained along with the constraints and the optimal control problem (OCP) formulation used for this exercise.

4.3.1. Dataset

In this exercise as the main aim was to assess the impact of different modelling techniques on the performance of MPC, we resolved to use proxy data from the emulator. By doing so we avoid the undesired impact of emulator-building mismatch on the performance of MPC, which in turn complicates the assessment of predictive models, impact on the MPC. To explain this further we need to remember that the emulator no matter how accurate is still a model and models do not completely represent reality. There is always a mismatch between model and the real system no matter how small. Therefore, if we use in-situ measurements to train the models, these predictive models would be representing the real building, which is not exactly the emulator. Consequently, accuracy of predictive models in the MPC framework would not only be determined by the predictive model's quality itself but also by the accuracy of the emulator. Therefore, we resolved to use proxy data from the emulator.

Data generated from the emulator model throughout the User-1 period (see section 3.3.1) was used for training and validation of the data-driven models in this exercise. Data used for training data-driven models should include a wide-enough range and it should be reach in frequency. The latter mandates that the system is excited with a wide range of inputs which includes a wide range of frequency (from steady-state to highly transient). To fulfil these criterions, two random sequences for the heat pump's status (S_{hp}) and heat pump's supply water temperature (T_s) (heat pump's control inputs) were generated and then applied as inputs to the emulator model. These sequences have different holds, which range from one hour to a day to include different levels of frequency in the dataset. The resulting temperatures profile is shown in Figure 34, illustrating that the indoor temperature varies between 19.5 °C and 24.5 °C. As it would be described later on, the comfort band considered for the building for this exercise is [21:25] during the day with a night setback to [18:22]. Therefore, the range of the output does not fully cover the temperature range that our MPC is most likely to operate within. This issue caused problems in development of different models, which are explained in section 4.4 . Henceforth, we call this dataset, the *"dataset 0".*

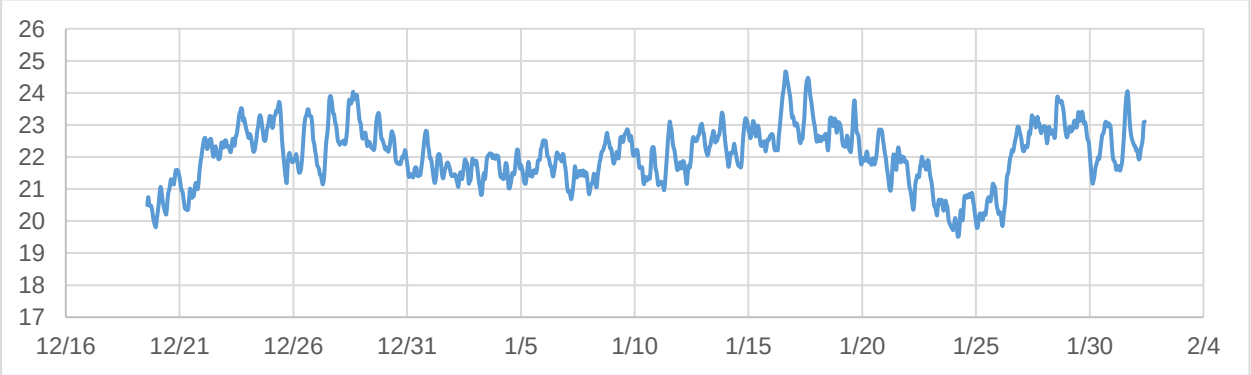


Figure 34. Temperature profile of the emulator in dataset 0

To obviate the aforementioned issues, we later generated another dataset, which covers a wider operation range of the twin house. In doing so, we revised the random sequence of commands applied to the heat pump. To have a wider span in the temperature output we divided the command signal for heat pump status into two parts with different probabilities for being on or off. In one half, the probability of generating an *on* command for the heat pump is 70% and in the other half the probability of generating an *off* command for the heat pump is 70%. The resulting temperature profile of the twin house is presented in Figure 35.

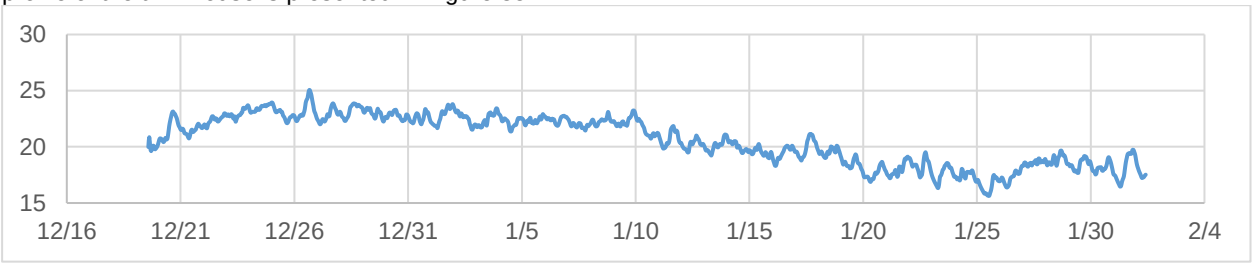


Figure 35. Temperature profile of twin house from dataset 1

Figure 35 illustrates that this dataset covers the whole range of thermal comfort and is referred to as *dataset 1* throughout this report. It is used to train and validate all the data-driven models applied in this study. The boundary condition used to carry out the simulations are derived from the in-situ measurements of the Twin House s.

4.3.2. OCP formulation

In this section, we define the OCP formulation used in this exercise. As explained in section 4.1.1, Model Predictive Control essentially solves an optimization problem at each time step. Like other optimization problems, we need to define the objective function and the constraints, to which the optimization is subjected.

As stated in section 4.2, the goal of the MPC in this exercise is to optimize the performance of the heating system of the O5 building while maintaining thermal comfort. This information is the starting point for setting up a suitable objective function, which adheres to these criteria. In this study, soft constraints are applied on thermal comfort bands to avoid convergence issues when solving OCP. To allow indoor temperature to go outside the comfort band, we have turned the constraints on indoor temperature to soft constraints using slack variables (v_k). The value of these slack variables are then penalized in the objective function to minimize thermal discomfort:

$$\min_{\bar{u}_k, \bar{u}_{k+1}, \ldots \bar{u}_{k+N-1}} \sum_{i=0}^{N-1} Lv_{k+i+1} + C_{el,k+i+1} P_{el,k+i+1} \qquad (4.2)$$

$$T_{in,k+i+1} + v_{k+i+1} \geq T_{low,k+i+1} \qquad (4.3)$$

$$T_{in,k+i+1} - v_{k+i+1} \leq T_{up,k+i+1} \qquad (4.4)$$

$$\vec{u} = [u_1, u_2]$$

$$v_{k+i+1} \geq 0 \ , \ i = 0,1,...N-1$$

In Equation (4.2), *k* represents the current time step, *v* represents the slack variable for softening the discomfort bands, T_{up} and T_{low} represent the upper and lower level of comfort band at each time step. C_{el} is the electricity price and P_{el} represents the electricity consumption of the heat pump while u_1 and u_2 represent heat pump status and heat pump's supply water temperature respectively. As for C_{el}, a Time of Use (ToU) electricity pricing has been included in the exercise, which reflects the integration or RES. N is the control horizon which has been set to 12 hours (12 time steps) in this work.

As for the other constraints, the most important ones are the constraints on the heat pump performance. Performance of the heat pump is constrained mainly by its compressor. If the required output by the heat pump goes below a certain level, then the compressor would have to modulate at a low frequency with a low Coefficient of Performance (COP) (Verhelst, Logist, Van Impe, & Helsen, 2012). This situation must be avoided as much possible. On the other hand, determining the minimum level of modulation is not easy and depends on many variables. For the sake of simplicity, in this study we have considered a minimum level for the heat pump's supply water temperature. On the other end of the spectrum, there is a limitation on the upper level of the heat pump's supply water temperature (T_s), which is chosen in accordance with specification of the real heat pump to comply with the real physical system as much as possible. These constraints are included in the OCP formulation as in Equation (4.5) in which T_s stands for heat pump's supply water temperature.

$$28 \leq T_s \leq 45 \qquad (4.5)$$

Moreover, to estimate the electricity consumption of the heat pump through the control horizon, a polynomial regression has been applied with the inputs of heat pump supply water temperature, ambient temperature, heat pump status and indoor air temperature. Selection of these variables for the heat pump's electricity consumption is based on the heat transfer laws governing the heat pump. To make the exercise simpler, the return water temperature is not used to estimate the heat pump's electricity consumption since using this variable requires another regression to forecast the return temperature of the heat pump during the control horizon. The simplification made here was that instead of the heat pump's return water temperature we used indoor temperature of the building. The reason behind this choice is two-fold; first, return water temperature varies based on the supply water temperature and the building's temperature as well so the indoor temperature would change the temperature of the return water. The other reason is that since the predictive model already provides us with an estimation of the indoor temperature, we could use the prediction of those models for estimating heat pump's electricity consumption. Therefore, no further modelling effort is required. In short, P_{el} is calculated as follows:

$$P_{el} = S_{hp} * f(T_s, T_e, T_{in}) \qquad (4.6)$$

In Equation (4.6)**,** *f()* represents a polynomial function. S_{hp} represents heat pump's status whether it is on or off. T_e and T_{in} indicate ambient temperature and indoor temperature respectively.

To avoid complexity in the simulation, the flow rate of the hot water provided by the heat pump is also considered constant throughout this exercise (naturally this is valid only in case that the heat pump is working). The water flow rate share of each of the rooms equipped with underfloor heating are kept constant throughout the simulations. The other part, in which the exercise differs from the real system, is that in this exercise, the three underfloor circuits in the building are aggregated and only one heating circuit is considered for the whole building in this exercise. By defining Equations (4.2)-(4.6) we have the full formulation of the OCP available and now we can move on to describing different predictive models at length.

4.4. Modelling techniques

Participants were asked to provide their predictive models for the test case using any technique and modelling type that they see fit for this exercise. In this section, the different modelling techniques, which have been applied for the purpose of this study, are explained. All these models have been developed with a time-step of 1 hour.

Table 7 shows the general structure of these models.

Table 7. Description of models

Model name	Model order	Inputs	Modelling team
Grey-box 1	1	T_e, GHI, H_{in}	Norwegian University of Science and Technology
Grey-box 2	2	T_e, GHI, T_s, S_{hp}	Technical University of Denmark
State Space 1	1	T_e, GHI, T_s, S_{hp}, IHG,	Catholic University of Leuven
State Space 2	7	T_e, GHI, T_s, S_{hp}, IHG, VFR_{av}	Catholic University of Leuven
ARX 1	3	T_e, GHI, T_s*S_{hp}, DHI, VFR_{living}	Tallinn University of Technology
ARX 2	3	T_e, IHG, T_s, S_{hp}, DNI	CIMNE- Polytechnic University of Catalonia
ANN	7	T_e, GHI, T_s, S_{hp}, IHG, VFR_{av}	Catholic University of Leuven

Input variables in the table are as follows:

T_e: Ambient temperature (^oC)

GHI: Global horizontal irradiance (W/m^2)

T_s: Supply water temperature (u_2) (^oC)

S_{hp}: Heat pump status (u_1)

H_{in}: Heat injected to building by underfloor heating (W)

IHG: Internal heat gains (W)

VFR: Volumetric flow rate of ventilation system (m^3/h)

DHI: Diffuse horizontal irradiance (W/m^2)

DNI: Direct Normal irradiance (W/m^2)

4.4.1. Grey-box Model 1

As a popular building identification method, grey-box models were used by the participants in this study. They started with a simple structure for grey-box models and then built up complexity onward. For each of the grey box models, first, the structure of the model is determined based on the law of energy conservation, then the parameters of the model are identified using the identification dataset. Interested readers are referred to chapter 2 of this report, Reynders et al., (2014), report of Annex 58 (Staf Roels, 2016), as well as the report on subtask 3 of Annex 71 itself for more details on model development. This grey-box model only has one state, which represents the average temperature of the indoor air and its structure is shown in the form of a Resistance Capacitance (RC) electrical circuit in which Φ_h represents the heat input injected to the building, which is calculated using Equation (4.6).

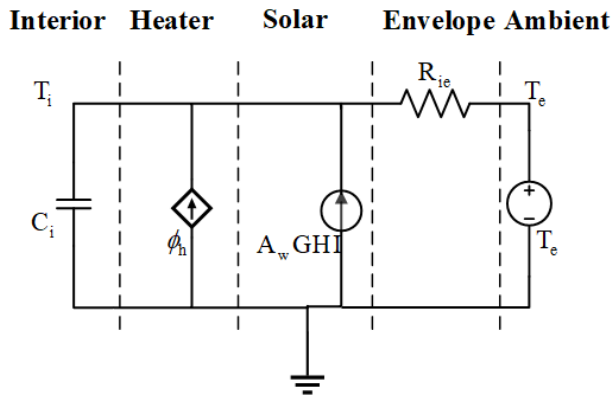


Figure 36. Grey-box model with one state

4.4.2. Grey-box Model 2

Another model developed by the participants for this exercise is a grey-box model which has two states, one for the air temperature, while the other one represents the floor temperature (as the heating medium) of the building (Figure 37). The inputs are listed in

Table 7. Φ_h in this model is calculated based on the temperature difference between the heating medium (T_h) e.g. floor in this case and the supply water in the heat pumps (T_s) with a constant specific heat capacity. Interested readers are referred to (Bacher & Madsen, 2011) for more details on this model.

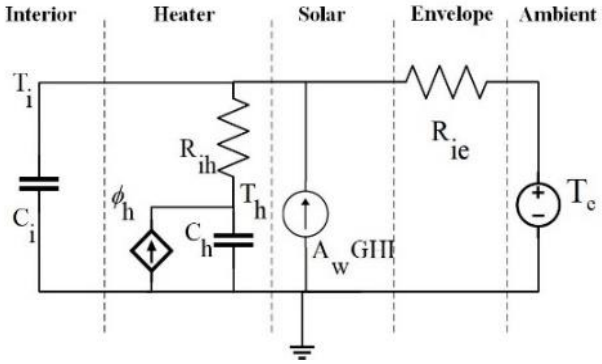


Figure 37. Grey-box model with two states

4.4.3. Autoregressive with exogenous inputs 1

One of the most common black box methods applied for building behaviour identification is Auto-Regressive with eXogenous input (ARX) models (Bourdeau et al., 2019). To develop this model, a Principle Component Analysis (PCA) has been carried out which led to the selection of optimal set of inputs as well as the output lags used for the building behaviour identification, which is three (see Table 7). The general structure of ARX models for identifying a multi input single output system is given in Equation 4.7.

$$A(z)y(z) = B_1(z)u_1(z) + ...B_{n_u}(z)u_{n_n}(z) + C(z)e(z) \quad (4.7)$$

In this equation, n_u stands for the number of input signals, which is five in this case. A(), B() and C() are polynomials representing the parameter of the ARX model which are estimated using the training dataset described earlier.

4.4.4. Autoregressive with exogenous inputs 2

Another ARX model approach is developed by another group of participants, their strategy consists in develop 4 different models to capture the dynamics of the building. To be able to predict the building behaviour and execute the MPC is modelled the supply temperature, the heat pump consumption, the COP and the internal temperature.

The supply temperature is used when the heat pump is OFF; to capture the decay of the supply temperature otherwise when the heat pump is ON the temperature is determined by the set point. This model depends on the lagged terms of T_s, T_e, S_{hp} and T_e.

The heat pump consumption is required to evaluate the economical MPC and used with the COP to compute the heat injected into the building by radiant floor. Similar to the supply temperature model, the set point determinate if it is required to predict the consumption or not. When the heat pump is set to OFF the heat pump consumption is zero, otherwise is used the heat pump consumption model which depends on the lagged terms of T_s, T_e and T_i.

The COP is required by the internal temperature model because is used to calculate the heat delivered by the radiant floor. Equal to the heat pump consumption if the heat pump is set to OFF the COP will be zero and otherwise is used the COP model, which depends on the lagged terms of T_s and T_e. Differently from the other data used, the COP is not commonly an available operation data. In the studied case, it was available, so the model was trained with the historical data. On the other hand, when the measured historical data is not available, it can be obtained by the datasheet of the heat pump.

The thermal contribution of the heated floor (H_{in}) is computed using the COP and the heat pump consumption. With that, predicted data it can be forecasted the internal temperature of the building depending on the lagged terms of T_i, T_e, H_{in}, IHG and DNI.

4.4.5. State Space

Another popular modelling technique in the category of black-box models is state space identification. One of the advantages of linear state space models is that most linear systems could be described using this formulation and most of the notations and theorems developed regarding MPC and OCP are based on state space representation of the systems (Rawlings, Mayne, & Diehl, 2019). In this study, the focus was on the Linear Time-Invariant (LTI) state space models since they are easy to develop and have been successfully applied in the context of MPC design for buildings (Bourdeau et al., 2019). General form of these models is formulated below:

$$\dot{x} = Ax(t) + Bu(t) \qquad (4.8)$$

$$y = Cx(t) + Du(t) \qquad (4.9)$$

In this form, Matrixes A, B, C and D describe the dynamics of the system states and they are called system matrix, Input matrix, output matrix, and feedthrough matrix. In this research, two different state space models are deployed. One only has one state, which is the simplest state space possible; as for the other model, the number of states has been determined based on Singular Value Decomposition (SVD) of the Henkel Matrix for which 7 states is selected as the optimum number of states (Drgoňa, Picard, Kvasnica, & Helsen, 2018). Identification dataset is used for estimating elements of the four matrixes in Equation (4.8) and (4.9).

4.4.6. Artificial Neural Network

Artificial Neural networks (ANNs) are known as a powerful tool in machine learning. They are inspired by the structure of the brain (Abu-Mostafa, 1992). There is an ever-increasing interest in applying ANNs for HVAC system optimization applications. There are various architectures of ANNs available. One of the architectures deemed suitable for the application of building characterization is Non-linear Auto-Regressive with eXogenous inputs (NARX) which has proven successful in capturing dynamics of HVAC systems and it has been used by participants here as well (Bourdeau et al., 2019). These models have essentially the same input-output structure as ARX models with the main difference being using neurons for capturing system's dynamics instead of linear mapping in the ARX case. Interested readers can refer to (Erfani et al., 2018) for further details on the NARX model application for building behaviour identification.

4.5. Model validation

Since different combinations of the training dataset were used to train and validate the models, a second dataset was generated solely with the purpose of testing the models. As stated earlier, MPC solves an optimization problem over a

given time horizon. Hence, predictive model used in the MPC should be able to provide acceptable predictions not only for one-step ahead in time but also throughout the whole control horizon. Therefore, we investigate whether one-step ahead prediction accuracy is a good enough indicator to reflect the quality of predictive models or if it necessary to look into multi-step ahead prediction accuracy. In other words, we analyse whether model's multi-step ahead prediction accuracy affects the KPIs of the controller or not. If the answer to this question is positive, it means that predictive models should be trained by the objective of minimizing multi-step ahead prediction error not only one-step ahead. To this end, prediction results obtained by running the models against this test dataset are presented for the whole prediction horizon in Figure 38. In the results section these models are implemented in our MPC framework and the performance of the resulting MPC is assessed against each model's one-step and multi-step ahead accuracy.

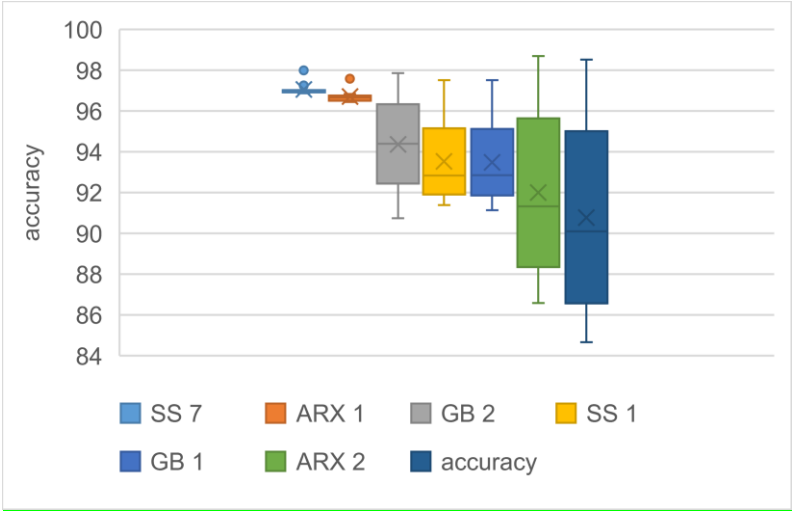


Figure 38. Accuracy of different models for the whole control horizon

This figure provides the boxplot accuracy of different modelling techniques in terms of their coefficient of determination (R^2). The maximum in each box corresponds to the OSPE while the minimum corresponds to N (Control horizon) step ahead prediction accuracy. As stated earlier N is 12 in this work. As could be seen in Figure 38, ANN model is the best performing models in terms of one-step ahead prediction accuracy but they are not the best models when looking into the multi-step ahead prediction. Such a model with such prediction performance is an interesting candidate that might help us decide whether one-step ahead prediction is a better KPI or multi-step ahead accuracy better reflects the quality of a predictive model in MPC.

4.6. Results and Discussion

In this section, the results of applying different predictive models in the MPC are presented and discussed. The simulations have been carried out for a total duration of two weeks from 19th of December to 2nd of January. The weather data used for this study is from the in-situ measurements of the building. Perfect forecast is considered for the climatic conditions.

The KPIs considered in this study for comparing the MPCs based on different models, are the total thermal discomfort level and the total electricity cost of the heating system. Total discomfort is calculated as the Kelvin hours outside the thermal comfort band. Electricity usage of the heat pump is directly obtained from energy demand simulation in the emulator model (FMU).

First, results obtained by applying the base-line controller are presented and shortly explained. Afterwards, we address the results obtained by deploying different predictive models in the MPC. To compare the performance of MPC against traditional controllers (the second objective of this common exercise) a Rule-Based Controller (RBC) (See Figure 39) has been designed and applied to the building. In this RBC, heat pump's status (S_{hp}) and heat pump's supply temperature (T_s) are the control inputs, which are called u_1 and u_2 respectively. These control inputs are to be computed at each time step. For the heat pump's status (u_1), a hysteresis logic is applied in which the hysteresis band and its set point are to be determined. As for heat pump's supply water temperature (u_2), a proportional controller is applied. Since thermal behaviour of a building is non-linear and complex, traditional tuning methods are not applicable in this case.

Therefore, a suitable value has been found using trial and error. To compensate for the high inertia of the underfloor heating, RBC takes the comfort band of 2 hours ahead for tracking task. Consequently, heat pump is turned on 2 hours before the night set-back period is finished. Results shown in Figures 39-44 are all from 28th of December until 31th of December midnight just like the base-line controller. It could be seen in these figures that MPCs are able to maintain the temperature within the thermal comfort band although there are some minor violations (these violation still lead to a feasible solution by formulating thermal comfort as a soft constraint as explained in section 4.3.2). Fluctuations outside thermal comfort band could have two main causes. First, the magnitude of weight (L) scalar in the objective function, which allows thermal discomfort to some extent especially when the electricity cost is relatively high. The second reason behind the minor thermal discomfort could be the mismatch between the predictive model and the emulator. As seen in electricity consumption profile of each MPC, the load profile does not completely correspond with the time-of-use price. This observation is expected since the MPC does not optimize the building's behaviour only for one time-step but for the whole control horizon.

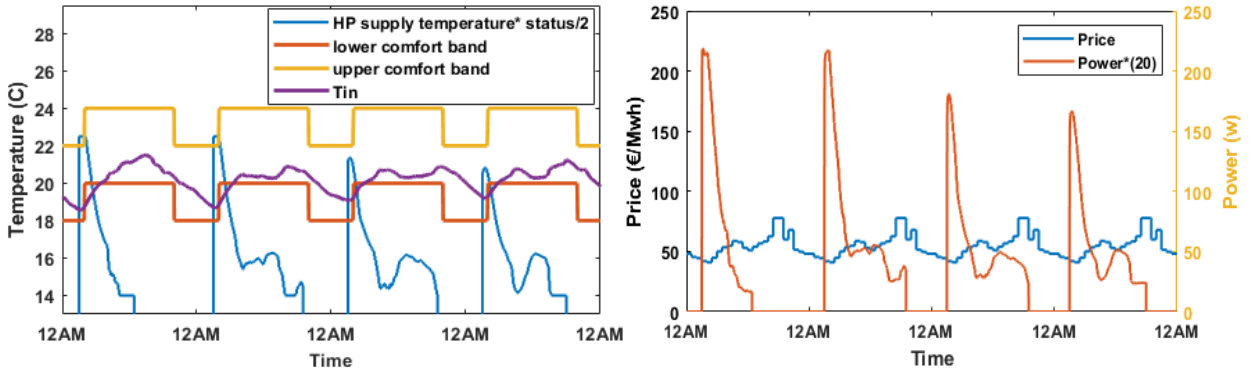


Figure 39. Results of applying Rule-Based Controller (RBC)

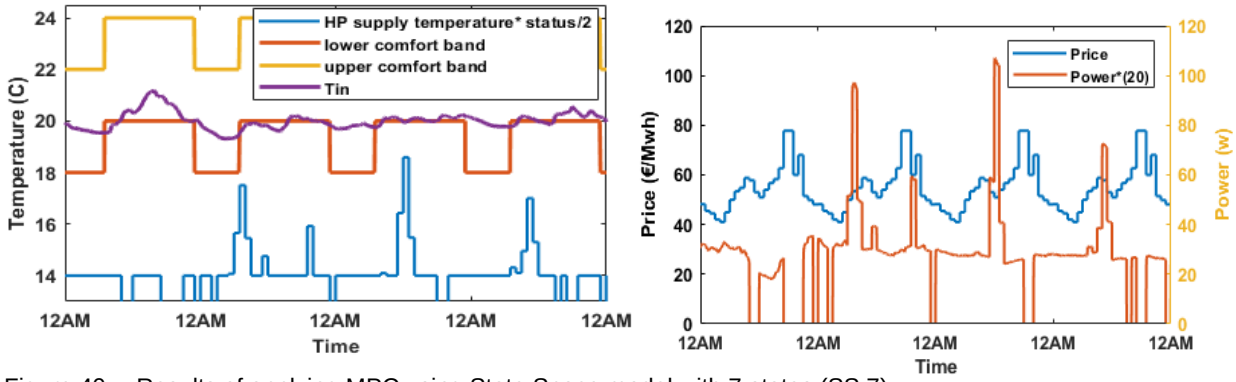


Figure 40. Results of applying MPC using State Space model with 7 states (SS 7)

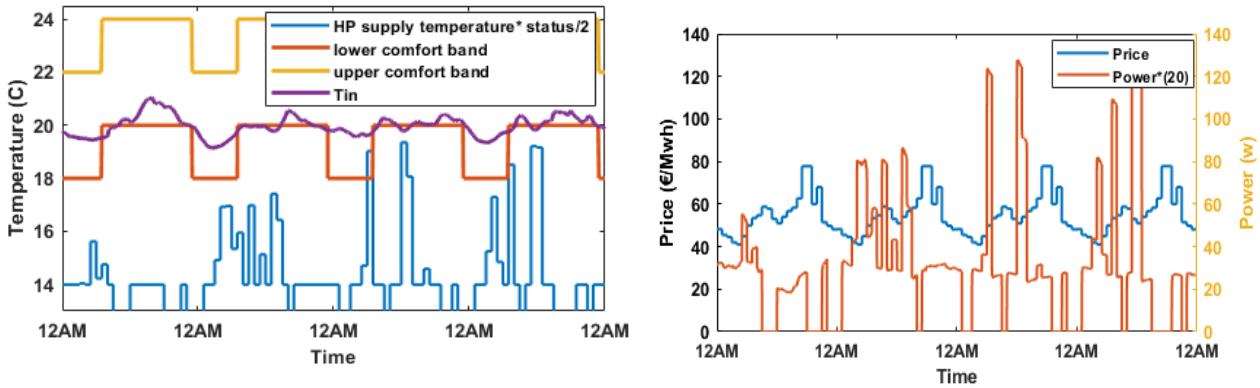


Figure 41. Results of applying MPC using ARX model with the order of 3 (ARX 1)

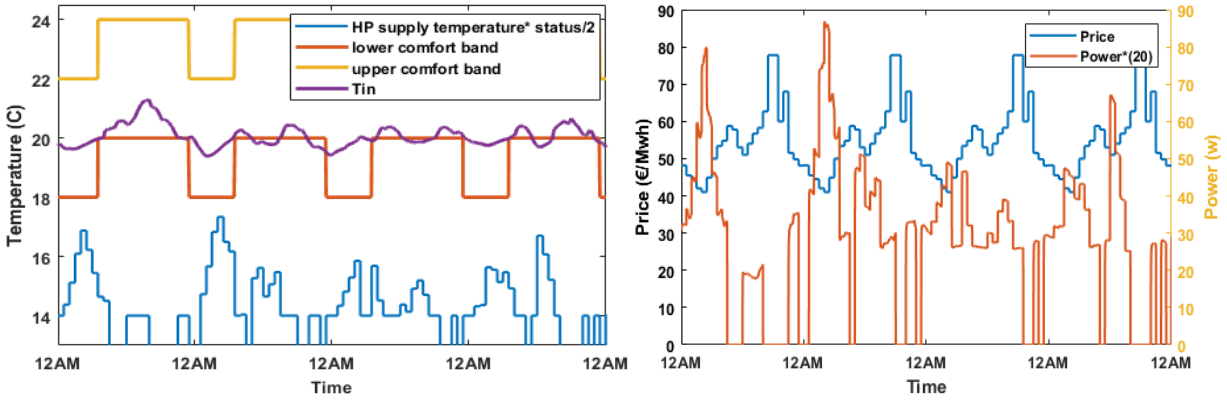


Figure 42. Results of applying MPC using Grey-box model with 2 states (GB 2)

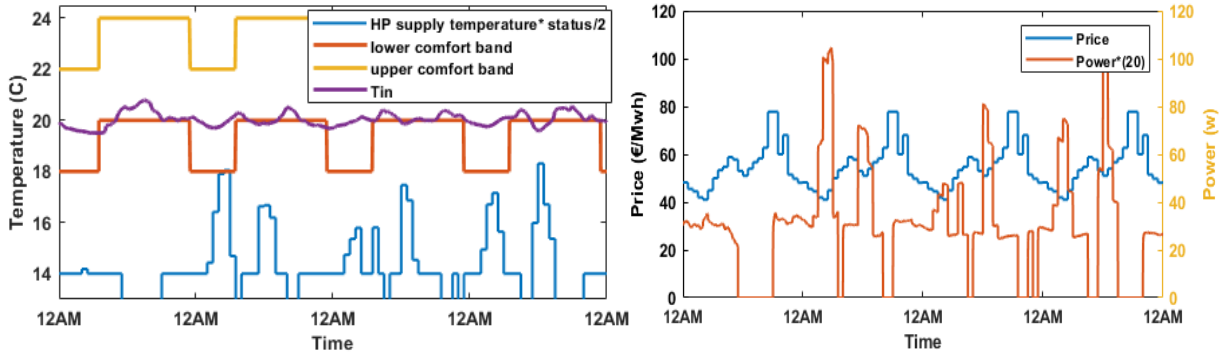


Figure 43. Results of applying MPC using State Space model with 1 state (SS 1)

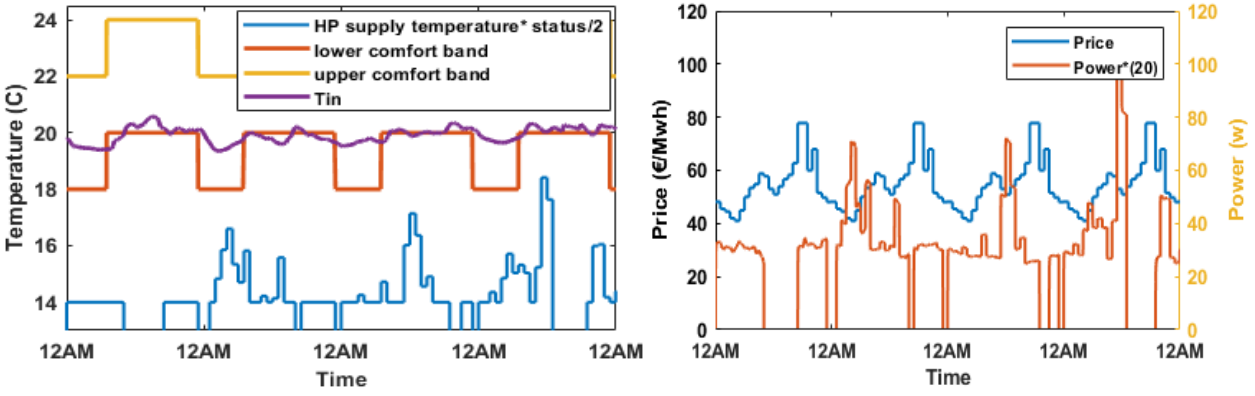


Figure 44. Results of applying MPC using ANN model with 7 neurons (ANN)

It could be seen that the MPCs resulting from each of the predictive models are capable of maintaining thermal comfort of the occupants in respect to the given constraints and electricity cost. In all of the MPCs shown in Figure 40-44, indoor temperature is kept inside the comfort band and only small values of thermal discomfort are allowed. An important note for evaluating these results is that since the MPC solves the optimization problem for a given time-horizon, electricity use does not exactly correspond to the electricity price profile. In order to have an overall view of the performance of the models in the MPC, we aggregate results from all models in Figure 45 . Attempting to compare the results of different MPCs we come across an impediment, which blocks the way of a straightforward comparison of the controllers. This barrier arises from the fact that the MPC aims at optimizing two objectives (thermal discomfort and electricity cost) which are not physically related to each other. Therefore, by changing the weight (L in Equation (4.2)) optimal performance of the controllers are obtained in a way that they yield similar discomfort levels as could be seen in Figure

45. By employing this method, we ensure that all controllers have a similar thermal discomfort level so that we can compare the controllers only based on electricity cost. As could be seen in Figure 45, all the MPCs have a total thermal discomfort of between 10 and 11 Kh. Therefore, we evaluate them based on their total electricity cost throughout the simulation. From this figure, it could be realized that deploying the state space model with 7 states (SS 7) leads to the best KPIs of the controller compared to other models.

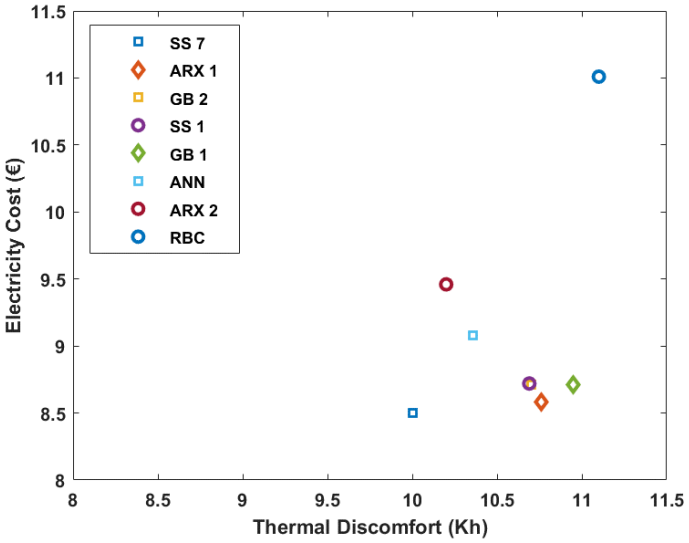


Figure 45. KPIs of MPCs deploying different predictive models

To assess the impact of model quality and its performance in the MPC, we should consider both Figure 38 & Figure 45 together. As could be seen in Figure 38 ANN model is the best performing model in terms of one-step ahead prediction accuracy (the maximum in its corresponding box plot is the highest). Nevertheless, this model is not the best performing model in this exercise but the worst one. In spite of the highest one-step ahead prediction accuracy, the ARX 2 and ANN model lead to the highest electricity cost compared to other models. In order to figure out what is the cause behind this finding, we looked into the Multi-Step ahead Prediction Error (MSPE) of the models. Now by evaluating the predictive models based on their multi-step ahead performance, one can easily see that although ARX 2 and ANN models have the highest one-step ahead R^2 value (see Figure 38), but their multi-step ahead prediction performance is the poorest amongst all the models. Hence, ANN and ARX 2 are not the best performing predictive model but the worst between our models. The reason for this observation is explained by one of the important drawbacks of ANN-based models. ANNs easily become over-fit to training data if no regularization of some sort is used (Afroz, Shafiullah, Urmee, & Higgins, 2018). This issue should be tackled when using ANNs as predictive models in the context of MPC otherwise one might end up with an ANN model, which is highly accurate for one-step ahead prediction but provides poor forecasts for multi-step ahead prediction. As for ARX models, the coefficients should be chosen to minimize MSPE rather than OSPE. Consequently, MSPE reflects the quality of the predictive model in the MPC better than the OSPE and it should be used to quantify the models instead of the one-step ahead accuracy.

As it is shown in Figure 38, the ARX model developed by the participants who chose the 2nd option has a close quality to the one of ANN model but looking at Figure 45 we can see that the performance of the MPC designed by this team yielded similar results compared to the ANN although a difference still exists. This could be explained by the reason that the other team used an exponential function to penalize thermal discomfort while the main OCP formulation uses a linear function to penalize thermal discomfort. Another possible contributing factor could be the use of different platforms for implementation of the MPC since this team used R to develop their model and Python to implement their MPC.

Analysing the results as illustrated in Figure 45, it could be concluded that the best performing MPC (deploying state space model with 7 states) compared to the RBC, reduces electricity cost from 11€ to 8.5 € which corresponds to 22.7%. Comparing different MPCs using Figure 45 one can deduce that the difference between electricity cost resulted from using different predictive models in the MPC is 7% (Electricity cost of 8.5 € in the SS7 model compared to 9.1 € achieved by using the NARX model). It could be deduced that the models used here vary by 24% in terms of activating the potential energy savings achieved by MPC. Therefore, we can conclude that the quality of the predictive model has a non-negligible impact on the performance of the controller for which it provides predictions. It could be seen from the results that designing the whole MPC with another team yielded KPIs, which are a bit different from the other MPC in

this exercise despite a lot of communication, which took place between the two designing teams to eliminate any discrepancy between their designs as much as possible.

The different performance of MPC from using different predictive models is easily explained by comparing state space-based MPC with the ANN-based MPC. In spite of higher one-step ahead prediction accuracy of the ANN, it leads to higher electricity cost compared to the state space model, which is explained by ANNs poorer multi-step ahead prediction performance. Hence, we concluded that Multi-step ahead prediction performance of a model is more suitable for assessing its quality in an MPC framework and predictive models used in the context of MPC should be trained and evaluated based on MSPE rather than OSPE.

4.7. Lessons Learned

The aim of this exercise was to investigate the impact of different modelling techniques on the performance of MPC for optimization of thermal response on an experimental building. In the literature, it has been shown that predictive model has a notable impact on the quality of the MPC. To identify the mechanism of this impact, first we developed a conceptual framework in which different models could be integrated in the designed MPC to see how they alter the performance of the controller. Second, various predictive models were developed and their quality were reported based on their short term and mid-term (throughout the control horizon) prediction performance. In the next step, these models were utilized to optimize thermal behaviour of the test case in the context of the MPC framework.

Keeping in mind the goal of this exercise, we looked into suitable indicators for identifying the suitability of predictive models for the controller. Analysing results obtained from applying different models in the MPC, showed that the mid-term prediction performance of the models better reflects their quality in the MPC. There are models that yield powerful prediction capabilities for one-step ahead prediction yet provide poor predictions throughout the rest of the control horizon. Such models lead to worse performing MPCs compared to the ones with better mid-term predictions. Therefore, it was shown that contrary to common practice in modelling, special attention should be given to mid-term performance of predictive models which are to be deployed in MPC schemes.
Another point of interest might be that the configuration of the controller such as objective function formulation also plays a nontrivial role in the performance of the controller. However, this common exercises showed that developing a platform in which different modellers could develop and test their models in an MPC scheme is feasible through close collaboration and extensive documentation.

5. Conclusions

IEA ECB Annex 71 focusses on an accurate characterization of the as-built energy performance of buildings based on in-situ measurements. This characterization may serve to identify the performance gap with focus on physical parameters or looks into the energy behaviour of the building. Subtask 2 focuses on the latter and assesses data analysis methods suitable for describing and predicting the energy dynamics of buildings. Knowledge on the energy behaviour of buildings is important in the ongoing energy transition. The shift towards renewable energy sources introduces a new paradigm where not only the amount of energy use but also the time of usage becomes significant and requires matching the energy demand with the intermittent production of renewable energy sources. During the operational phase, knowledge on the expected energy behaviour can be used to determine correct functioning of the building and its installations.

The activities in Subtask 2 were organized through setting up common exercises in which participants could contribute on a particular topic. The common exercises first explored the existing modelling techniques for building behaviour identification. The results of this analysis are reported in Chapter 2. Subsequently, two applications in which building behaviour identification plays an important role were identified. Fault Detection and Diagnosis (FDD) was chosen as a first application in which the behavioural models are used to detect and diagnose errors in the operation of the building and its systems. The results of this analysis are reported in Chapter. Chapter 4 focused on identifying models for Model Predictive Control (MPC) applications. The main findings on these two applications are summarized in the next 2 sections.

5.1. Conclusions on FDD

The goal of this activity was to evaluate and test the contribution of prediction models - obtained through building behaviour characterization – to automated fault detection and diagnosis. It was identified based on a literature review that FDD in commercial applications still recides on the individual component or system level, with individual processes being monitored by dedicated sensors. While building energy management systems are finding their way into the market, only recently methodologies that exploit overarching data are being developed.

The work carried out in this activity contributes to that development in three significant areas. First, a conceptual framework has been described that defines and organizes different types of faults. For each type of faults, different detection methods are proposed. The development and layout of this framework is inspired by the observations in the common exercises that demonstrated that despite detailed and accurate modelling techniques some type of faults could not be detected. For example, when using input-output models for anomaly detection that use the heating power as an input, it becomes impossible to detect errors in the control of that heating power (e.g. a thermostat malfunction). Consequently, as discussed in the framework, the type of errors to be detected should be properly matched to the modelling approach, taking into account the physical behaviour of the building.

Second, an overview was made on statistical methods used to detect the actual faults. The methods discussed focus on the detection of faults by comparing the predicted behaviour against the actual measurements. With this overview, guidelines are provided to move beyond the need for modeller/operator interpretation and move towards an automated detection process.

Third, by demonstrating the application of the identification of prediction models and the fault detection process for both simulation and actual measurement data for the same case study, this activity indicates that detailed building energy simulation models can play a significant role in the further development and research on automated fault detection and diagnosis methods. Further research is however needed to generalize these findings are a wider set of building and addressing a broader range of anomalies.

5.2. Conclusions on MPC

As a second application, it was chosen to determine how models used in Model Predictive Controllers (MPC) could be identified and what the impact of their predictive power was on achieving the goals set for the MPC. MPC was chosen as an application as it is a promising method for integrating Renewable Energy Sources (RES) and smart technologies in buildings. The developed common exercise aims at evaluating the performance of building behaviour models in MPC for the Holzkirchen Twin House O5 building and to demonstrate the opportunities of MPC. This MPC aimed at minimizing the heating system’s energy cost while maintaining indoor thermal comfort.

An important question while developing a model for building energy assessment is: which model is better suited for the application at hand? Therefore, in this study we searched for a suitable Key Performance Indicator (KPI) to score the performance of predictive models for a popular building optimization application. This application is called Model Predictive Control (MPC), which has been proven successful in optimizing building's energy use while maintaining thermal comfort. The building is heated by an underfloor heating system, which is coupled to an air-water heat pump. A varying electricity price of the occupants in many cases. MPC employs a predictive model of the building to optimize its load profile was applied to mimic the dynamic behaviour of RES generation over a time horizon (Drgoňa et al., 2020).

In the exercise, an OpenIDEAS simulation model developed in Modelica serves as the emulator in which the MPC was implemented. Different modelling teams were asked to identify behavioural models based on data that were generated by the emulator. Subsequently, the provided models were implemented in the emulator. To do so, two options were provided: in option 1 partners provided their predictive models that were implemented in the MPC of the developed framework while in option 2 the participants could develop their own MPC and communicate to the building part in the emulator through an API. The exertion of option 2 revealed the need for an in-depth report of the emulator along with its inputs and outputs. Detailed report of the emulator becomes more relevant when one wants to compare controllers that have been implemented in different environments. Models which were studied for this exercise are amongst the most common modelling techniques used for developing a predictive model in the context of MPC; namely Grey-box RC models, AutoRegressive models with eXogenous inputs (ARX), State Space models (SS) and Artificial Neural Networks (ANN). At first, models' quality were reported based on their one-step ahead forecast but these reported accuracies did not reflect the performance of the resulting controller very well. As the predictive model is required to provide the controller with forecast of building's thermal behaviour throughout a time horizon, we opted to look into Multi-Step ahead Prediction Error (MSPE) as a performance indicator.). Evaluating MSPE of different models shed some light on the poor performance of controllers, which incorporated these predictive models. For example, ANN has the best prediction performance in terms of one-step ahead accuracy but the MPC, which uses ANN as its predictor, yielded the highest energy cost compared to other controllers, which deployed other modelling techniques. This could be explained by looking into MSPE of different models where we can see that ANN yielded the poorest prediction performance and therefore the MPC, which employed ANN model, led to the worst KPIs compared with other MPCs in this exercise. It was concluded that for scoring building energy assessment methods that are to be used in applications such as MPC, modellers should consider MSPE instead of one-step ahead prediction error. It has also been shown that MPC outperforms a well-tuned Rule-Based Controller (RBC) in this case by 22.7%.

6. References

Abu-Mostafa, Y. S. (1992). *Neural networks and learning*. *Institute of Physics Conference Series* (Vol. 127).

Abushakra, B., & Paulus, M. T. (2016). An hourly hybrid multi-variate change-point inverse model using short-term monitored data for annual prediction of building energy performance, part II: Methodology (1404-RP). *Science and Technology for the Built Environment*, *22*(7), 984–995. https://doi.org/10.1080/23744731.2016.1215199

Afram, A., & Janabi-Sharifi, F. (2014). Theory and applications of HVAC control systems – A review of model predictive control (MPC). *Building and Environment*, *72*, 343–355. https://doi.org/10.1016/j.buildenv.2013.11.016

Afroz, Z., Shafiullah, G. M., Urmee, T., & Higgins, G. (2018). Modeling techniques used in building HVAC control systems: A review. *Renewable and Sustainable Energy Reviews*, *83*(October 2017), 64–84. https://doi.org/10.1016/j.rser.2017.10.044

Aftab, M., Chen, C., Chau, C. K., & Rahwan, T. (2017). Automatic HVAC control with real-time occupancy recognition and simulation-guided model predictive control in low-cost embedded system. *Energy and Buildings*, *154*, 141–156. https://doi.org/10.1016/j.enbuild.2017.07.077

Ajib, B., Lefteriu, S., Caucheteux, A., Lecoeuche, S., & Gauvrit, J. (2018). Prediction of Standardized Energy Consumption of Existing Buildings Based on Hybrid Systems Modeling and Control. In *2018 IEEE Conference on Decision and Control (CDC)* (pp. 3880–3886). IEEE. https://doi.org/10.1109/CDC.2018.8619086

Araya, D. B., Grolinger, K., ElYamany, H. F., Capretz, M. A. M., & Bitsuamlak, G. (2017). An ensemble learning framework for anomaly detection in building energy consumption. *Energy and Buildings*, *144*, 191–206. https://doi.org/10.1016/j.enbuild.2017.02.058

Arroyo, J., Spiessens, F., & Helsen, L. (2020). Identification of multi-zone grey-box building models for use in model predictive control. *Journal of Building Performance Simulation*, *13*(4), 472–486. https://doi.org/10.1080/19401493.2020.1770861

Atam, E., & Helsen, L. (2015). A convex approach to a class of non-convex building HVAC control problems: Illustration by two case studies. *Energy and Buildings*, *93*, 269–281. https://doi.org/10.1016/j.enbuild.2015.02.026

Atam, E., & Helsen, L. (2016). Control-Oriented Thermal Modeling of Multizone Buildings: Methods and Issues. *IEEE Control Systems*, *36*(3), 86–111. https://doi.org/10.1109/MCS.2016.2535913

Atthajariyakul, S., & Leephakpreeda, T. (2005). Neural computing thermal comfort index for HVAC systems. *Energy Conversion and Management*, *46*(15–16), 2553–2565. https://doi.org/10.1016/j.enconman.2004.12.007

Bacher, P., & Madsen, H. (2011). Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings*, *43*(7), 1511–1522. https://doi.org/10.1016/j.enbuild.2011.02.005

Bako, L., Boukharouba, K., Duviella, E., & Lecoeuche, S. (2011). A recursive identification algorithm for switched linear/affine models. *Nonlinear Analysis: Hybrid Systems*, *5*(2), 242–253. https://doi.org/10.1016/j.nahs.2010.05.003

Baranski, M., & Voss, J. (2004). Detecting patterns of appliances from total load data using a dynamic programming approach. *Proceedings - Fourth IEEE International Conference on Data Mining, ICDM 2004*, 327–330. https://doi.org/10.1109/ICDM.2004.10003

Bourdeau, M., Zhai, X. qiang, Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*, *48*(April), 101533. https://doi.org/10.1016/j.scs.2019.101533

Chou, J.-S., Telaga, A. S., Chong, W. K., & Gibson, G. E. (2017). Early-warning application for real-time detection of energy consumption anomalies in buildings. *Journal of Cleaner Production*, *149*, 711–722. https://doi.org/10.1016/j.jclepro.2017.02.028

Chou, J. S., & Telaga, A. S. (2014). Real-time detection of anomalous power consumption. *Renewable and Sustainable Energy Reviews*, *33*, 400–411. https://doi.org/10.1016/j.rser.2014.01.088

Cigler, J., Gyalistras, D., Siroky, J., Tiet, V.-N., & Ferkl, L. (2013). Beyond Theory: the Challenge of Implementing Model Predictive Control in Buildings. *11th REHVA World Congress and 8th International Conference Energy Efficient,*

Smart and Healthy Buildings, 1008–1018. https://doi.org/ISBN 978-80-260-4001-9

Coakley, D., Raftery, P., & Keane, M. (2014). A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews*, *37*, 123–141. https://doi.org/10.1016/j.rser.2014.05.007

D. Blum, F.Jorissen, S. H. et al. (2010). Prototyping the BOPTEST framework for Simulation-Based Testing of Advanced Control Strategies in Buildings, (July), 35–43.

de Wilde, P. (2014). The gap between predicted and measured energy performance of buildings: A framework for investigation. *Automation in Construction*, *41*, 40–49. https://doi.org/10.1016/j.autcon.2014.02.009

Dibley, M., Li, H., Rezgui, Y., & Miles, J. (2012). An ontology framework for intelligent sensor-based building monitoring. *Automation in Construction*, *28*, 1–14. https://doi.org/10.1016/j.autcon.2012.05.018

Drgoňa, J., Picard, D., Kvasnica, M., & Helsen, L. (2018). Approximate model predictive building control via machine learning. *Applied Energy*, *218*(March), 199–216. https://doi.org/10.1016/j.apenergy.2018.02.156

Erfani, A., Rajabi-Ghahnaviyeh, A., & Boroushaki, M. (2018). Design and construction of a non-linear model predictive controller for building's cooling system. *Building and Environment*, *133*(November 2017), 237–245. https://doi.org/10.1016/j.buildenv.2018.02.022

Feng, J., Chuang, F., Borrelli, F., & Bauman, F. (2015). Model predictive control of radiant slab systems with evaporative cooling sources. *Energy and Buildings*, *87*, 199–210. https://doi.org/10.1016/j.enbuild.2014.11.037

Ferracuti, F., Fonti, A., Ciabattoni, L., Pizzuti, S., Arteconi, A., Helsen, L., & Comodi, G. (2017). Data-driven models for short-term thermal behaviour prediction in real buildings. *Applied Energy*, *204*, 1375–1387. https://doi.org/10.1016/j.apenergy.2017.05.015

Galvin, R. (2014). Making the "rebound effect" more useful for performance evaluation of thermal retrofits of existing homes: Defining the "energy savings deficit" and the "energy performance gap." *Energy and Buildings*, *69*, 515–524. https://doi.org/10.1016/j.enbuild.2013.11.004

Gao, T. (2020). *Integrated building fault detection and diagnosis using data modelling and Bayesian networks*. Université de Lille, France.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hedegaard, R. E., Pedersen, T. H., Knudsen, M. D., & Petersen, S. (2018). Towards practical model predictive control of residential space heating: Eliminating the need for weather measurements. *Energy and Buildings*, *170*, 206–216. https://doi.org/10.1016/j.enbuild.2018.04.014

Hill, D. J., & Minsker, B. S. (2010). Anomaly detection in streaming environmental sensor data: A data-driven modeling approach. *Environmental Modelling & Software*, *25*(9), 1014–1022. https://doi.org/10.1016/j.envsoft.2009.08.010

Hong, G., & Kim, B. S. (2018). Development of a data-driven predictive model of supply air temperature in an air-handling unit for conserving energy. *Energies*, *11*(2). https://doi.org/10.3390/en11020407

Huang, H., Chen, L., & Hu, E. (2015). A new model predictive control scheme for energy and cost savings in commercial buildings: An airport terminal building case study. *Building and Environment*, *89*, 203–216. https://doi.org/10.1016/j.buildenv.2015.01.037

Jensen, S. Ø. (2019). *Energy in Buildings and Communities Programme: Annex 67 Energy Flexible Buildings*.

Jiménez, M. J., Madsen, H., Bloem, J. J., & Dammann, B. (2008). Estimation of non-linear continuous time models for the heat exchange dynamics of building integrated photovoltaic modules. *Energy and Buildings*, *40*(2), 157–167. https://doi.org/10.1016/j.enbuild.2007.02.026

Jimenez, M. jose. (2014). *Report of Subtask 3a: Thermal performance characterization based on full scale testing - description of the common exercises and physical guidelines*. *Build up Web Webinar*.

Junker, R. G. (2019). *Characterisation and Integration of Energy Flexibility through Stochastic Modelling and Control*.

Kim, S. H. (2013). An evaluation of robust controls for passive building thermal mass and mechanical thermal energy storage under uncertainty. *Applied Energy*, *111*, 602–623. https://doi.org/10.1016/j.apenergy.2013.05.030

Kristensen, N. R., Madsen, H., & Jørgensen, S. B. (2004). A method for systematic improvement of stochastic grey-

box models. *Computers and Chemical Engineering*, *28*, 1431–1449. https://doi.org/10.1016/j.compchemeng.2003.10.003

Labeodan, T., Zeiler, W., Boxem, G., & Zhao, Y. (2015). Occupancy measurement in commercial office buildings for demand-driven control applications—A survey and detection system evaluation. *Energy and Buildings*, *93*, 303–314. https://doi.org/10.1016/j.enbuild.2015.02.028

Lee, S. H., & Yang, C. S. (2017). An intelligent power monitoring and analysis system for distributed smart plugs sensor networks. *International Journal of Distributed Sensor Networks*, *13*(7). https://doi.org/10.1177/1550147717718462

Lodi, C., Bacher, P., Cipriano, J., & Madsen, H. (2012). Modelling the heat dynamics of a monitored Test Reference Environment for Building Integrated Photovoltaic systems using stochastic differential equations. *Energy and Buildings*, *50*, 273–281. https://doi.org/10.1016/j.enbuild.2012.03.046

Maciejowski, J. M. (2002). *Predictive Control: With Constraints* (2nd ed.). Pearson Education.

Madsen, H. (2015). Thermal Performance Characterization using Time Series Data ; IEA EBC Annex 58 Guidelines, (January 2016). https://doi.org/10.13140/RG.2.1.1564.4241

Mantesi, E., Mourkos, K., Hopfe, C., McLeod, R., Vatougiou, P., Kersken, M., & Strachan, P. (2019). DeployingBuilding Simulation to Enhance the Experimental Design of a Full-scale Empirical Validation Project. In *Proceedings of Building Simulation 2019: 16th Conference of IBPSA*. https://doi.org/9781775052012

Marinakis, V., Doukas, H., Karakosta, C., & Psarras, J. (2013). An integrated system for buildings' energy-efficient automation: Application in the tertiary sector. *Applied Energy*, *101*(2013), 6–14. https://doi.org/10.1016/j.apenergy.2012.05.032

Martani, C., Lee, D., Robinson, P., Britter, R., & Ratti, C. (2012). ENERNET: Studying the dynamic relationship between building occupancy and energy consumption. *Energy and Buildings*, *47*, 584–591. https://doi.org/10.1016/j.enbuild.2011.12.037

Masoso, O. T., & Grobler, L. J. (2010). The dark side of occupants' behaviour on building energy use. *Energy and Buildings*, *42*(2), 173–177. https://doi.org/10.1016/j.enbuild.2009.08.009

Mayne, D. Q., Rawlings, J. B., Rao, C. V., & Scokaert, P. O. M. (2000). Constrained model predictive control: Stability and optimality. *Automatica*, *36*(6), 789–814. https://doi.org/10.1016/S0005-1098(99)00214-9

Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, *22*(19), 3055–3071. https://doi.org/10.1002/sim.1545

O'Donnell, J. (2009). Specification of Optimum Holistic Building Environmental and Energy Performance Information to Support Informed Decision Making. *Doctorate, University College Cork, Ireland*, (April), 243.

O'Donnell, J., Keane, M., Morrissey, E., & Bazjanac, V. (2013). Scenario modelling: A holistic environmental and energy management method for building operation optimisation. *Energy and Buildings*, *62*, 146–157. https://doi.org/10.1016/j.enbuild.2012.10.060

Parzinger, M., Hanfstaengl, L., Sigg, F., Spindler, U., Wellisch, U., & Wirnsberger, M. (2020). Residual Analysis of Predictive Modelling Data for Automated Fault Detection in Building's Heating, Ventilation and Air Conditioning Systems. *Sustainability*, *12*(17), 6758. https://doi.org/10.3390/su12176758

Pathak, N., Ba, A., Ploennigs, J., & Roy, N. (2018). Forecasting Gas Usage for Big Buildings Using Generalized Additive Models and Deep Learning. In *2018 IEEE International Conference on Smart Computing (SMARTCOMP)* (pp. 203–210). IEEE. https://doi.org/10.1109/SMARTCOMP.2018.00092

Patteeuw, D., Reynders, G., Bruninx, K., Protopapadaki, C., Delarue, E., D'haeseleer, W., … Helsen, L. (2015). CO2-abatement cost of residential heat pumps with active demand response: demand- and supply-side effect. *Applied Energy*, *156*, 490–501. https://doi.org/10.1016/j.apenergy.2015.07.038

Paulo, P., Branco, F., de Brito, J., & Silva, A. (2016). BuildingsLife – The use of genetic algorithms for maintenance plan optimization. *Journal of Cleaner Production*, *121*, 84–98. https://doi.org/10.1016/j.jclepro.2016.02.041

Paulus, M. T., Claridge, D. E., & Culp, C. (2015). Algorithm for automating the selection of a temperature dependent change point model. *Energy and Buildings*, *87*, 95–104. https://doi.org/10.1016/j.enbuild.2014.11.033

Pčolka, M., Žáčeková, E., Robinett, R., Čelikovský, S., & Šebek, M. (2016). Bridging the gap between the linear and

nonlinear predictive control: Adaptations for efficient building climate control. *Control Engineering Practice*, *53*, 124–138. https://doi.org/10.1016/j.conengprac.2016.01.007

Peña, M., Biscarri, F., Guerrero, J. I., Monedero, I., & León, C. (2016). Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach. *Expert Systems with Applications*, *56*, 242–255. https://doi.org/10.1016/j.eswa.2016.03.002

Pichler, M. F., Goertler, G., & Schranzhofer, H. (2017). Test buildings with TABS for MPC-performance evaluation-Comparability and system identification. *2016 European Control Conference, ECC 2016*, 1177–1182. https://doi.org/10.1109/ECC.2016.7810449

Ploennigs, J., Chen, B., Schumann, A., & Brady, N. (2013). Exploiting Generalized Additive Models for Diagnosing Abnormal Energy Use in Buildings. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings* (pp. 1–8). New York, NY, USA: ACM. https://doi.org/10.1145/2528282.2528291

Prívara, S., Cigler, J., Váňa, Z. E., Oldewurtel, F., & Žáčeková, E. (2013). Use of partial least squares within the control relevant identification for buildings. *Control Engineering Practice*, *21*(1), 113–121. https://doi.org/10.1016/j.conengprac.2012.09.017

Rawlings, J. B., Mayne, D. Q., & Diehl, M. (2019). *Model predictive control: Theory, Computation, and Design*. *Nob Hill Publishing* (2nd ed.). Nob Hill Publishing. https://doi.org/10.1007/978-3-030-11869-3_4

Reynders, G., Diriken, J., & Saelens, D. (2014). Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals. *Energy and Buildings*, *82*, 263–274. https://doi.org/10.1016/j.enbuild.2014.07.025

Reynders, G., Nuytten, T., & Saelens, D. (2013). Robustness of reduced-order models for prediction and simulation of the thermal behavior of dwellings. In *Proceedings of Building Simulation 2013*. Chambery. Retrieved from http://www.ibpsa.org/proceedings/BS2013/p_1306.pdf

Reynolds, J., Rezgui, Y., Kwan, A., & Piriou, S. (2018). A zone-level, building energy optimisation combining an artificial neural network, a genetic algorithm, and model predictive control. *Energy*, *151*, 729–739. https://doi.org/10.1016/j.energy.2018.03.113

Sakurada, M., & Yairi, T. (2014). Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction. In *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis - MLSDA'14* (pp. 4–11). New York, New York, USA: ACM Press. https://doi.org/10.1145/2689746.2689747

Schmelas, M., Feldmann, T., Wellnitz, P., & Bollin, E. (2016). Adaptive predictive control of thermo-active building systems (TABS) based on a multiple regression algorithm: First practical test. *Energy and Buildings*, *129*, 367–377. https://doi.org/10.1016/j.enbuild.2016.08.013

Senave, M., Reynders, G., Sodagar, B., & Saelens, D. (2018). Uncertainty in Building Energy Performance Characterization: Impact of Gas Consumption Decomposition on Estimated Heat Loss Coefficient. In *7th International Building Physics Conference (IBPC): Healthy, Intelligent and Resilient Buildings and Urban Environments*. Syracuse, NY, USA. Retrieved from http://ibpc2018.org/

Shaikh, P. H., Nor, N. B. M., Nallagownden, P., Elamvazuthi, I., & Ibrahim, T. (2014). A review on optimized control systems for building energy and comfort management of smart sustainable buildings. *Renewable and Sustainable Energy Reviews*, *34*, 409–429. https://doi.org/10.1016/j.rser.2014.03.027

Sodagar, B., & Starkey, D. (2016). The monitored performance of four social houses certified to the Code for Sustainable Homes Level 5. *Energy and Buildings*, *110*, 245–256. https://doi.org/10.1016/j.enbuild.2015.11.016

Staf Roels. (2016). *Reliable building energy performance characterisation based on full scale dynamic measurements*. *Build up Web Webinar*.

Strachan, P., Svehla, K., Kersken, M., & Heusler, I. (2014). *Report of Subtask 4a: Empirical validation of common building energy simulation models basd on in situ dynamic data*.

Sumer, K. K., Goktas, O., & Hepsag, A. (2009). The application of seasonal latent variable in forecasting electricity demand as an alternative method. *Energy Policy*, *37*(4), 1317–1322. https://doi.org/10.1016/j.enpol.2008.11.014

Taal, A., Itard, L., & Zeiler, W. (2018). A reference architecture for the integration of automated energy performance fault diagnosis into HVAC systems. *Energy and Buildings*, *179*, 144–155. https://doi.org/10.1016/j.enbuild.2018.08.031

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4757-3264-1

Verbeke, S., Aerts, D., Reynders, G., & Waide, P. (2020). *Final report on the technical support to the development of a smart readiness indicator for buildings*. https://doi.org/10.2833/41100

Verhelst, C. (2012). *Model Predictive Control of Ground Coupled Heat Pump Systems for Office Buildings*. *Building Simulation* (Vol. 5).

Verhelst, C., Logist, F., Van Impe, J., & Helsen, L. (2012). Study of the optimal control problem formulation for modulating air-to-water heat pumps connected to a residential floor heating system. *Energy and Buildings*, *45*, 43–53. https://doi.org/10.1016/j.enbuild.2011.10.015

Viot, H., Sempey, A., Mora, L., Batsale, J. C., & Malvestio, J. (2018). Model predictive control of a thermally activated building system to improve energy management of an experimental building: Part II - Potential of predictive strategy. *Energy and Buildings*, *172*, 385–396. https://doi.org/10.1016/j.enbuild.2018.04.062

Volk, R., Stengel, J., & Schultmann, F. (2014). Building Information Modeling (BIM) for existing buildings — Literature review and future needs. *Automation in Construction*, *38*, 109–127. https://doi.org/10.1016/j.autcon.2013.10.023

Wang, S., & Xu, X. (2006). Simplified building model for transient thermal performance estimation using GA-based parameter identification. *International Journal of Thermal Sciences*, *45*(4), 419–432. https://doi.org/10.1016/j.ijthermalsci.2005.06.009

Wilson, C. (2014). Evaluating communication to optimise consumer-directed energy efficiency interventions. *Energy Policy*, *74*, 300–310. https://doi.org/10.1016/j.enpol.2014.08.025

Wishart, D. M. G., Lee, E. B., & Markus, L. (1969). Foundations of Optimal Control Theory. *Journal of the Royal Statistical Society. Series A (General)*, *132*(1), 110. https://doi.org/10.2307/2343766

Wrinch, M., El-Fouly, T. H. M., & Wong, S. (2012). Anomaly detection of building systems using energy demand frequency domain analysis. In *2012 IEEE Power and Energy Society General Meeting* (pp. 1–6). IEEE. https://doi.org/10.1109/PESGM.2012.6344790

Xue, X., Wang, S., Sun, Y., & Xiao, F. (2014). An interactive building power demand management strategy for facilitating smart grid optimization. *Applied Energy*, *116*, 297–310. https://doi.org/10.1016/j.apenergy.2013.11.064

Z̆a´c̆ekova´, E., Va´n̆a, Z., & Cigler, J. (2014). Towards the real-life implementation of MPC for an office building: Identification issues. *Applied Energy*, *135*, 53–62. https://doi.org/10.1016/j.apenergy.2014.08.004

Žáčeková, E., & Prívara, S. (2012). Control relevant identification and predictive control of a building. *Proceedings of the 2012 24th Chinese Control and Decision Conference, CCDC 2012*, 246–251. https://doi.org/10.1109/CCDC.2012.6244035

Zakula, T., Armstrong, P. R., & Norford, L. (2015). Advanced cooling technology with thermally activated building surfaces and model predictive control. *Energy and Buildings*, *86*, 640–650. https://doi.org/10.1016/j.enbuild.2014.10.054

Zhan, S., & Chong, A. (2021). Data requirements and performance evaluation of model predictive control in buildings : A modeling perspective. *Renewable and Sustainable Energy Reviews*, (January), 110835. https://doi.org/10.1016/j.rser.2021.110835

Zhao, L., Zhang, J. L., & Liang, R. B. (2013). Development of an energy monitoring system for large public buildings. *Energy and Buildings*, *66*, 41–48. https://doi.org/10.1016/j.enbuild.2013.07.007

Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, *56*, 215–225. https://doi.org/10.1016/j.rser.2015.11.050

Zhou, Z.-H. (2009). Ensemble Learning. *Encyclopedia of Biometrics*, 270–273. https://doi.org/10.1007/978-0-387-73003-5_293

ANNEX 71

EBC

Energy in Buildings and Communities Programme

www.iea-ebc.org