International Energy Agency, EBC Annex 58

Reliable building energy performance characterisation based on full scale dynamic measurements

Report of Subtask 4a: Empirical validation of common building energy simulation models based on in situ dynamic data

Paul Strachan, Katalin Svehla, Matthias Kersken, Ingo Heusler

EBC is a programme of the International Energy Agency (IEA)

International Energy Agency, EBC Annex 58

Reliable building energy performance characterisation based on full scale dynamic measurements

Report of Subtask 4a: Empirical validation of common building energy simulation models based on in situ dynamic data

Authors

ESRU, University of Strathclyde, Glasgow, UK (www.strath.ac.uk/esru)

Paul Strachan, Katalin Svehla

Fraunhofer Institute for Building Physics IBP, Holzkirchen, Germany (www.ibp.fraunhofer.de)

Matthias Kersken, Ingo Heusler

With contributions from:

25 modellers/modelling teams (listed in Section 4.1)

Reviewed by:

Internal reviewers: Søren Østergaard Jensen (Danish Technological Institute), Gabriel Rojas-Kopeinig (University of Innsbruck), Gabrielle Masy (University of Liege), Ricardo Enriquez Miranda (CIEMAT), Jos van Schijndel (University of Eindhoven)

External reviewers: Michele Zinzi (ENEA), Ron Judkoff (NREL)

Published by KU Leuven, Belgium

ISBN: 9789460189852

Participating countries in EBC:

Australia, Austria, Belgium, Canada, P.R. China, Czech Republic, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Republic of Korea, the Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States of America.

Additional copies of this report may be obtained from:

www.iea-ebc.org

essu@iea-ebc.org

cover picture: Tailored heating experiment performed for one of the industrial partners of the Annex 58-project to determine the overall heat loss coefficient of a newly built dwelling based on on-site measured data. (source: Geert Bauwens, KU Leuven)

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 co-operation among the 29 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. The mission of the Energy in Buildings and Communities (EBC) Programme is to develop and facilitate the integration of technologies and processes for energy efficiency and conservation into healthy, low emission, and sustainable buildings and communities, through innovation and research. (Until March 2013, the IEA-EBC Programme was known as the Energy in Buildings and Community Systems Programme, ECBCS.)

The research and development strategies of the IEA-EBC Programme are derived from research drivers, national programmes within IEA countries, and the IEA Future Buildings Forum Think Tank Workshops. The research and development (R&D) strategies of IEA-EBC aim to exploit technological opportunities to save energy in the buildings sector, and to remove technical obstacles to market penetration of new energy efficient technologies. The R&D strategies apply to residential, commercial, office buildings and community systems, and will impact the building industry in five focus areas for R&D activities:

- Integrated planning and building design
- Building energy systems
- Building envelope
- Community scale methods
- Real building energy use

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 (*):

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 & Evaluation Methods (*)
Annex 54: Integration of Micro-Generation & Related Energy Technologies in Buildings (*)
Annex 55: Reliability of Energy Efficient Building Retrofitting - Probability Assessment of Performance & Cost (RAP-RETRO) (*)
Annex 56: Cost Effective Energy & CO2 Emissions Optimization in Building Renovation
Annex 57: Evaluation of Embodied Energy & CO2 Equivalent Emissions for Building Construction
Annex 58: Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements
Annex 59: High Temperature Cooling & Low Temperature Heating in Buildings
Annex 60: New Generation Computational Tools for Building & 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 Simulation
Annex 67: Energy Flexible Buildings
Annex 68: Design and Operational Strategies for High IAQ in Low Energy 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

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 (*)

IEA EBC Annex 58: Reliable Building energy performance characterisation based on full scale dynamic measurements

Annex 58 in general

To reduce the energy use of buildings and communities, many industrialised countries have imposed more and more stringent requirements in the last decades. In most cases, evaluation and labelling of the energy performance of buildings are carried out during the design phase. Several studies have shown, however, that the actual performance after construction may deviate significantly from this theoretically designed performance. As a result, there is growing interest in full scale testing of components and whole buildings to characterise their actual thermal performance and energy efficiency. This full scale testing approach is not only of interest to study building (component) performance under actual conditions, but is also a valuable and necessary tool to deduce simplified models for advanced components and systems to integrate them into building energy simulation models. The same is true to identify suitable models to describe the thermal dynamics of whole buildings including their energy systems, for example when optimising energy grids for building and communities.

It is clear that quantifying the actual performance of buildings, verifying calculation models and integrating new advanced energy solutions for nearly zero or positive energy buildings can only be effectively realised by in situ testing and dynamic data analysis. But, practice shows that the outcome of many on site activities can be questioned in terms of accuracy and reliability. Full scale testing requires a high quality approach during all stages of research, starting with the test environment, such as test cells or real buildings, accuracy of sensors and correct installation, data acquisition software, and so on. It is crucial that the experimental setup (for example the test layout or boundary conditions imposed during testing) is correctly designed, and produces reliable data. These outputs can then be used in dynamic data analysis based on advanced statistical methods to provide accurate characteristics for reliable final application. If the required quality is not achieved at any of the stages, the results become inconclusive or possibly even useless. The IEA EBC Annex 58-project arose from the need to develop the necessary knowledge, tools and networks to achieve reliable in situ dynamic testing and data analysis methods that can be used to characterise the actual energy performance of building components and whole buildings. As such, the outcome of the project is not only of interest for the building community, but is also valuable for policy and decision makers, as it provides opportunities to make the step from (stringent) requirements on paper towards actual energy performance assessment and quality checking. Furthermore, with the developed methodology it is possible to characterise the dynamic behaviour of buildings, which is a prerequisite for optimising smart energy and thermal grids. Finally, the project developed a dataset to validate numerical Building Energy Simulation programs.

Structure of the project

Successful full scale dynamic testing requires quality over the whole process chain of full scale testing and dynamic data analysis: a good test infrastructure, a good experimental set-up, a reliable dynamic data analysis and appropriate use of the results. Therefore, the annex-project was organised around this process chain, and the following subtasks were defined:

Subtask 1 made an inventory of full scale test facilities available all over the world and described the common methods with their advantages and drawbacks for analysing the obtained dynamic data. This subtask produced an overview of the current state of the art on full scale testing and dynamic data analysis and highlighted the necessary skills.

Subtask 2 developed a roadmap on how to realise a good test environment and test set-up to measure the actual thermal performance of building components and whole buildings in situ. Since there are many different objectives when measuring the thermal performance of buildings or building components, the best way to treat this variety has been identified as constructing a decision tree. With a clear idea of the test objective, the decision tree will give the information of a test procedure or a standard where this type of test is explained in detail.

Subtask 3 focused on quality procedures for full scale dynamic data analysis and on how to characterise building components and whole buildings starting from full scale dynamic data sets. The report of subtask 3 provides a methodology for dynamic data analysis, taking into account the purpose of the in situ testing, the existence of prior physical knowledge, the available data and statistical tools,… The methodologies have been tested and validated within different common exercises, in a way that quality procedures and guidelines could be developed.

Subtask 4 produced examples of the application of the developed concepts and showed the applicability and importance of full scale dynamic testing for different issues with respect to energy conservation in buildings and community systems, such as the verification of common BES-models, the characterisation of buildings based on in situ testing and smart meter readings and the application of dynamic building characterisation for optimising smart grids.

Subtask 5 established a network of excellence on 'in situ testing and dynamic data analysis' for dissemination, knowledge exchange and guidelines on testing.

Overview of the working meetings

The preparation and working phase of the project encompassed 8 working meetings:

Meeting	Place, date	Attended by
Kick off meeting	Leuven (BE), September 2011	45 participants
Second preparation meeting	Bilbao (SP), April 2012	46 participants
First working meeting	Leeds (UK), September 2012	44 participants
Second working meeting	Munich (GE), April 2013	53 participants
Third working meeting	Hong-Kong (CH), September 2013	26 participants
Fourth working meeting	Gent (BE), April 2014	49 participants
Fifth working meeting	Berkeley (USA), September 2014	37 participants
Sixth working meeting	Prague (CZ), April 2015	39 participants

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, a Round Robin experiment on characterising a test box was undertaken, and several common exercises on data analysis methods were introduced and solved.

Outcome of the project

The IEA EBC Annex 58-project worked closely together with the Dynastee-network (www.dynastee.info). Enhancing this network and promoting actual building performance characterization based on full scale measurements and the appropriate data analysis techniques via this network is one of the deliverables of the Annex-project. This network of excellence on full scale testing and dynamic data analysis organizes on a regular basis events such as international workshops, annual training,... and will be of help for organisations interested in full scale testing campaigns.

In addition to the network of excellence, the outcome of the Annex 58-project has been described in a set of reports, including:
Report of Subtask 1A: Inventory of full scale test facilities for evaluation of building energy performances.
Report of Subtask 1B: Overview of methods to analyse dynamic data
Report of Subtask 2: Logic and use of the decision tree for optimizing full scale dynamic testing.
Report of Subtask 3 part 1: Thermal performance characterization based on full scale testing: physical guidelines and description of the common exercises
Report of Subtask 3 part 2: Thermal performance characterization using time series data – statistical guidelines.
Report of Subtask 4A: Empirical validation of common building energy simulation models based on in situ dynamic data.
Report of Subtask 4B: Towards a characterization of buildings based on in situ testing and smart meter readings and potential for applications in smart grids
IEA EBC Annex 58 project summary report

Participants

In total 49 institutes from 17 countries participated in Annex 58. The different participants are listed below:

Austria	Gabriel Rojas-Kopeinig, Universität Innsbruck
	Susanne Metzger, Vienna University of Technology
Belgium	Gilles Flamant, Belgian Building Research Institute
	Guillaume Lethé, Belgian Building Research Institute
	Luk Vandaele, Belgian Building Research Institute (subtask 5 co-leader)
	Paul Steskens, Belgian Building Research Institute
	Gabrielle Masy, Haute Ecole de la Province de Liège
	An-Heleen Deconinck, Katholieke Universiteit Leuven
	Dirk Saelens, KU Leuven (subtask 4 co-leader)
	Geert Bauwens, KU Leuven (secretary)
	Glenn Reynders, KU Leuven
	Ruben Baetens, KU Leuven
	Roel De Coninck, KU Leuven
	Staf Roels, KU Leuven (operating agent)
	Frédéric Delcuve, Knauf Insulation
	Philippe André, Université de Liège
	Arnold Janssens, Universiteit Gent (subtask 1 leader)
	Eline Himpe, Universiteit Gent
China	Gongshen Huang, City University of Hong Kong
	Tin-Tai Chow, City University of Hong Kong
	Linda Xiao Fu, The Hong Kong Polytechnic University
	Shengwei Wang, The Hong Kong Polytechnic University (subtask 4 co-leader)
	Xue Xue, The Hong Kong Polytechnic University
Czech Republic	Kamil Stanek, Czech Technical University Prague
	Pavel Kopecký, Czech Technical University Prague
Denmark	Christian Holm Christiansen, Danish Technological Institute
	Søren Østergaard Jensen, Danish Technological Institute
	Henrik Madsen, Technical University of Denmark (subtask 3 co-leader)
	Kyung Hun (Peter) Woo, Technical University of Denmark
	Peder Bacher, Technical University of Denmark
France	Bouchie Remi, Centre Scientifique et Technique du Bâtiment
	Pierre Boisson, Centre Scientifique et Technique du Bâtiment
	Mohamed El Mankibi, Ecole Nationale des Travaux Publics de l'Etat
	Christian Ghiaus, INSA de Lyon
	Ibán Naveros, INSA de Lyon
	Guillaume Pandraud, Isover Saint-Gobain
	Simon Rouchier, Université de Savoie
Germany	Franz Feldmeier, Fachhochschule Rosenheim
	Lucia Bauer, Fachhochschule Rosenheim
	Herbert Sinnesbichler, Fraunhofer-Institut für Bauphysik
	Ingo Heusler, Fraunhofer-Institut für Bauphysik
	Matthias Kersken, Fraunhofer-Institut für Bauphysik
	Soeren Peper, Passive House Institute
Italy	Fabio Moretti, ENEA
	Hans Bloem, European Commission - DG JRC (subtask 5 co-leader)
	Lorenzo Pagliano, Politecnico di Milano
	Giuseppina Alcamo, Università degli Studi di Firenze
The Netherlands	A.W.M. van Schijndel, Technische Universiteit Eindhoven
	Rick Kramer, Technische Universiteit Eindhoven
Norway	Nathalie Labonnote, Norges teknisk-naturvitenskapelige universitet
Spain	Gerard Mor-Lleida, Centro Internacional de Métodos Numéricos en Ingeniería
	Xavi Cipriano, Centro Internacional de Métodos Numéricos en Ingeniería
	Aitor Erkoreka, Escuela Técnica Superior de Ingenieria Bilbao (subtask 2 co-leader)
	Koldo Martin Escudero, Escuela Técnica Superior de Ingenieria Bilbao

	Roberto Garay Martinez, Tecnalia Research & Innovation
	Luis Castillo López, CIEMAT
	Maria José Jiménez Taboada, CIEMAT (subtask 3 co-leader)
	Ricardo Enríquez Miranda, CIEMAT
United Kingdom	Richard Fritton, Salford University
	Chris Gorse, Leeds Beckett University (subtask 2 co-leader)
	Martin Fletcher, Leeds Beckett University
	Samuel Stamp, University College London
	Filippo Monari, University of Strathclyde
	Paul A. Strachan, University of Strathclyde (subtask 4 co-leader)
United States	Stephen Selkowitz, Lawrence Berkeley National Laboratory

IEA, EBC Annex 58, Report of Subtask 4a

Empirical validation of common building energy simulation models based on in situ dynamic data

Paul Strachan, Katalin Svehla, Matthias Kersken, Ingo Heusler

March 2016

TABLE OF CONTENTS

1 Introduction

There is a marked lack of high quality datasets from real buildings (as opposed to test cells) suitable for validating the dynamic thermal simulation programs that are commonly used in predicting the energy and environmental performance of buildings. This report describes two detailed validation experiments that were undertaken on the Twin Houses, a Fraunhofer Institute for Building Physics (IBP) experimental facility at Holzkirchen, Germany. The first experiment was conducted in summer 2013, the second in spring 2014. A comprehensive validation methodology was adopted, and over 20 modelling teams were involved in following the experimental specification and submitting predictions. This report describes the background to the study, details of the experiments, the main results obtained, and information on the benefits to the wider modelling community of these experimental datasets and detailed specification, which have been made available under Open Access arrangements (see Section 6.4).

1.1 Need for Empirical Validation

Energy regulations for new and retrofitted buildings require increasing attention to energy efficiency as part of the worldwide drive to reduce carbon emissions and reliance on fossil fuels. However, a so-called "performance gap" has been observed between the as-designed energy consumption and that experienced in practice, for which there are many reasons, including poor controls, poor workmanship, poor commissioning, different weather conditions, different operating conditions and user behaviour. Nevertheless, there is still a need for design tools to predict energy consumption for compliance purposes, for identifying energy and carbon efficient designs, and to ensure good indoor environmental quality. This typically involves the use of simulation programs to represent dynamic response and to predict overheating risk, indoor air quality, lighting comfort etc. There needs to be documented evidence that the programs used to make these predictions are both adequate and appropriate. This study focuses on the predictive capability of programs with regard to thermal performance.

In spite of significant international effort, notably within previous IEA projects, it is perhaps surprising that there are still questions regarding the reliability of commercial and research programs to predict energy and environmental performance accurately. However, some of these previous studies have yielded inconclusive results because of too much uncertainty in inputs and/or experimental measurements to be useful for diagnosing sources of disagreement between simulation results and empirical data. Other studies have focused on experimental datasets from small test cells, and questions have been raised whether these are properly representative of full-scale buildings. There have also been many papers published that claim, following a limited monitoring exercise, that the program being used has been "validated", but in reality this may only be true for a particular building type, climate, operational regime, construction type etc. There are also questions regarding how extensive and accurate these monitored datasets are, whether all the relevant influencing parameters have been measured, and whether in fact the level of agreement can be classified as "good". Useful validation work has been carried out using inter-program comparisons against a range of benchmark models (e.g. BESTEST which was developed within IEA Annex 21 (1995)), but these tests are not based on measured data from real buildings and there is no "truth" standard in such tests (Judkoff and Neymark 2006).

2 Validation Methodology

The overall empirical validation methodology applied in this study was similar to that employed in previous IEA validation studies (e.g. Lomas et al 1997, Loutzenhiser et al 2007, Kalyanova et al 2009). The steps were as follows:

1. Experimental design. Model the selected building using a representative local climate dataset. The first objective of this phase is to design the overall experiment by determining building time constants, suitable test sequences, magnitudes of heat inputs and variation in internal temperatures. The second objective is to design the monitoring scheme. This is achieved primarily with sensitivity tests to identify important simulation parameters that need to be measured.
2. Experimental set-up. Calibrate and install all required sensors, install and check the data acquisition system and program the heating and/or cooling as required.
3. Experimental specification. Develop the specification to describe all parameters of the buildings required for modelling.
4. Experiment. Undertake the experiment and process the experimental data.
5. Blind validation (Phase 1). Modellers predict internal conditions using the experimental specification, measured climate data and operational schedules but without knowledge of internal conditions. At this stage, there are usually additional questions regarding the experimental details – these questions and answers are distributed to all modelling teams. Modelling teams submit modeller reports with details of the programs used, and assumptions made.
6. First stage analysis. This compares predictions against experimental data for internal temperatures and heat fluxes. Inevitably at this stage, differences are due to a combination of user and modelling error (and potentially measurement errors).
7. Re-modelling (Phase 2). The measured internal temperature and heat flux data is disseminated, so the modelling teams now have all the information describing the experiment and the measurements. Modelling teams are encouraged to investigate differences between measurements and predictions and resubmit predictions and updated reports. Only changes that correct user modelling errors or alter a modelling assumption (with documented rationale) are allowed. It is important to ensure that model input parameters are not simply tuned to improve agreement with measurement. In principle, this step separates the modelling errors from the user errors by eliminating the user errors.
8. Final analysis and reporting. This should provide definitive documentation of the analysis and outcomes.
9. Archiving of high quality data sets. The intention is that the resulting specification and datasets will be useful for developers of new programs, for those improving modelling algorithms, for teaching purposes, and for those intending to design new empirical validation data sets.

The datasets were also designed to be of use to the modelling teams involved in system identification (Subtask 3) for the determination of performance characteristics such as heat loss coefficients, solar aperture and effective capacity. This was a key reason for including a pseudo-random heat injection sequence as part of the experimental schedule.

3 Validation Experiments

3.1 Process for Selection of Test Building

At the start of IEA Annex 21, a comprehensive worldwide review of existing datasets suitable for empirical validation was reported (Lomas et al 1997). The majority of the datasets investigated were found to be of limited use for program validation, primarily because of missing monitored data of key parameters. There have subsequently been some successful large-scale international projects for empirical validation purposes, but these have been at component level, e.g. for testing micro-cogeneration models in IEA Annex 42 (2007), or on outdoor test cells, e.g. IEA Annex 43 (2007), or focused studies such as moisture buffering, e.g. IEA Annex 41 (2007). The reason why few large-scale whole building empirical validation tests have been undertaken is a combination of the lack of suitable test facilities, cost and time.

The criteria for suitable validation-quality experimental datasets are exacting. Building on previous work by Lomas et al (1997) and Judkoff et al (2008), the following are important requirements:

- The test building should be unoccupied because of the difficulty of separating uncertainty in calculating building performance from uncertainty in occupant behaviour (although it is possible to simulate occupancy under controlled conditions with injections of heat, moisture and carbon dioxide).
- There should be options for heating/cooling and an accurate/flexible control system.
- All program data inputs should be measured (building parameters, schedules etc).
- Detailed weather data must be measured on-site to record all the weather inputs required by the simulation programs.
- The instrumentation system should be comprehensive and reliable with traceability for all sensor calibrations. Data recording should be sub-hourly – in the order of 10-minutely data should be adequate for most thermal response unless detailed HVAC plant and control response is studied. The experimental team should be able to respond to modeller requests for additional sensors to be added.
- Data should be recorded for the overall building to enable the calculation of the overall building heat loss and effective capacity.
- As far as possible, data should be recorded for individual heat transfer paths. It can help to identify causes of differences between measurements and predictions, but it becomes more difficult when moving from test cells to larger scale buildings. Alternatively, if two essentially identical buildings are available, side-by-side experiments can be devised to focus on individual heat transfer processes.
- Infiltration and ventilation should be controlled and measured.
- Measured data should include uncertainty estimates.
- The experimental team should be experienced in detailed high quality dataset collection and the test facility must be well documented.
- The test facility should be available for extended test periods to cover a range of weather conditions.

A detailed checklist was constructed of the requirements. This was circulated to potential experimental teams within IEA EBC Annex 58. Table 1 shows the list of information requirements and the availability (at the time) for four potential test sites.

Table 1: Information requirements and evaluation for full-scale building validation study

Code:	*+ : facility available;*	*(+): partly available or could be made available*	*- : facility not available*	*0 : not applicable or neutral*

Building description and location:	**Energy House University of Salford, UK**	**Fraunhofer Twin Houses Holzkirchen Germany**	**Almeria PSE-ARFRISOL Solar Building Spain**	**EnergyFlex Lab Denmark**
References/reports available with building and instrumentation description	+	+	+	+
Availability of the building for testing for an extended period	+	+	+	+
Building fabric				
Dimensional details and orientation	+	+	+	+
Construction materials and layer thicknesses data	+	+	+	+
Measured thermophysical properties (particularly conductivity)	(+)	(+)	-	-
Measured surface properties – emissivity and absorptivity	-	(+)	(+)	-
Glazing system – optical transmittance/absorptance/reflectance data	-	+	(+)	(+)
Information on shading by surrounding buildings, shading devices	0	0	0	+
Information on thermal bridges (constructional details)	(+)	(+)	(+)	(+)
Internal heat gains (assume unoccupied)				
Measured lighting loads	+	+	+	+
Measured equipment loads	+	+	+	+
Ventilation				
Pressurization test data	+	+	(+)	+
Ventilation system: natural, mechanical or mechanical with heat recovery	0	0	0	0
Possibility for tracer gas measurements during experiments	(+)	(+)	+	-
Measurement of air movement between spaces (or air movement prevented by sealing)	(+)	(+)	+	-

Control				
Possibilities for scheduling heating/cooling inputs and measuring resulting temperature	+	+	+	+
Possibilities to select temperature setpoint and measure heat/cooling required	+	+	+	(+)
Type of temperature control possible (on/off, PID etc)	+	+	+	+
Heating and cooling system options				
Range of options available – conventional and/or renewable	0	(+)	+	+
Manufacturer's data available for the heating/cooling system	+	+	+	+
Measured performance data available for the heating/cooling system?	+	+	-	+
Performance data for renewable technologies available	0	+	+	+
Instrumentation				
Air temperatures in spaces: number of sensors and location and whether sensors are shielded	+	+	+	+
Surface temperatures: number and location	+	(+)	+	+
Electrical power consumption	+	+	+	+
Delivered heating	+	+	+	+
Delivered cooling	0	0	+	0
Instrumentation for heating/cooling plant (flow rates, return/supply temperatures etc)	+	+	+	+
Ventilation	0	+	-	+
Other instrumentation				+
Climate and other boundary conditions				
Air temperature	+	+	+	+
Solar radiation – global horizontal, diffuse horizontal, total vertical	0	+	+	+
Wind speed	0	+	+	+
Wind direction	0	+	+	+
Relative humidity	0	+	+	+
Long wave radiation	0	+	+	+
Ground reflectivity	0	-	+	-
Ground temperature	0	+	+	+

Based on the analysis of the evaluation in Table 1, the Twin Houses at the Fraunhofer Institute for Building Physics (IBP) at Holzkirchen, Germany) were selected. The houses are representative in terms of construction and thermal performance of recent housing (at least in some European countries).

3.2 Twin Houses

External and internal views of the Twin Houses are shown in Figures 1 and 2 respectively.

Figure 1: External views of Twin Houses in Holzkirchen, Germany

Figure 3 shows the site layout – the two houses and the meteorological station. The main glazed areas are on the south façade and are unshaded. There is a minimal amount of shading on other facades from adjacent low-rise buildings. External walls are externally insulated with U-values in the range 0.20 to 0.28 W/m²K. Windows are double-glazed with a glazing U-value of 1.2 W/m²K and with electric external roller blinds. Full details of the glazing, frame and blinds are included in the full specification.

These two houses had the added advantage that they were essentially identical (see following text), so could be used for side-by-side testing. Two separate pressurization tests were performed on each building. The first one examined all the ground floor rooms as used in the validation exercise. The air change rates at the standard 50 Pa pressure difference were measured as 1.62 ac/h and 1.54 ac/h (in the order of 0.113 ac/h and 0.108 ac/h for typical averaged pressure differences according to the European Standard EN 832). This indicates a difference of approximately 5% between the buildings, confirming that they are quite similar in terms of air leakage. Since the accuracy of the pressurization test is 14 % according to EN 13829 no real difference in air tightness can be determined. Also, since 1.5 ac/h at 50 Pa represents a typical value for buildings of this thermal quality and age, this aspect of the validation scenario can be considered realistic.

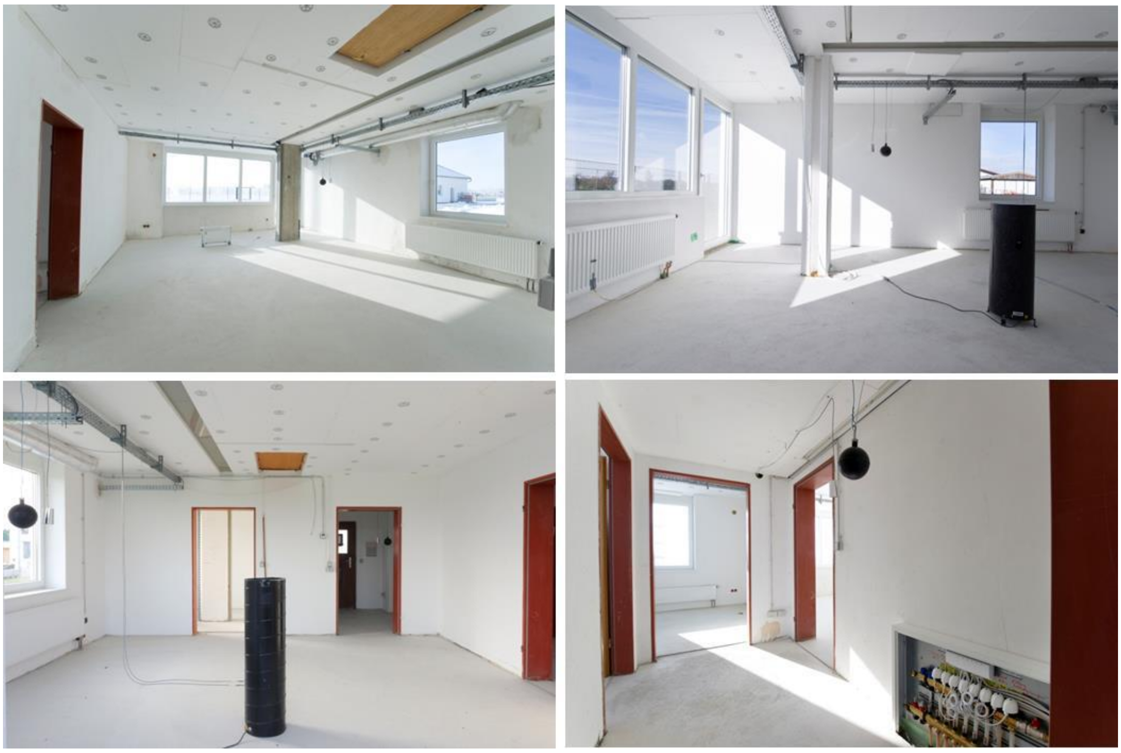

Figure 2: Internal views of the Twin Houses

Figure 3: Site layout

A second pressurization test was performed for the four southern, ventilated rooms (see Figure 6). The doors to the north bedroom, lobby and the kitchen were already sealed and the blower door fan was installed in the patio door. In this case, values of 2.2 ac/h and 2.3 ac/h were obtained at the 50 Pa pressure difference, again indicating close correspondence between the buildings. Note that the higher value compared to the first pressurization test is due to the reduced room volume.

Before the validation experiments, the in situ heating power requirements of the two houses to maintain constant internal temperatures was measured, ensuring the same temperature set points and shading configuration. For this test there was no mechanical ventilation system operating. This was to check that there were no undetected hidden flaws in the construction. Figure 4 shows the performance of the Twin Houses during this baseline measurement. The black line indicates the deviation between the cumulative heating energy consumption of both buildings; it shows that the deviation was within 0.5% at the end of the measurement period, and never exceeded 2 %.

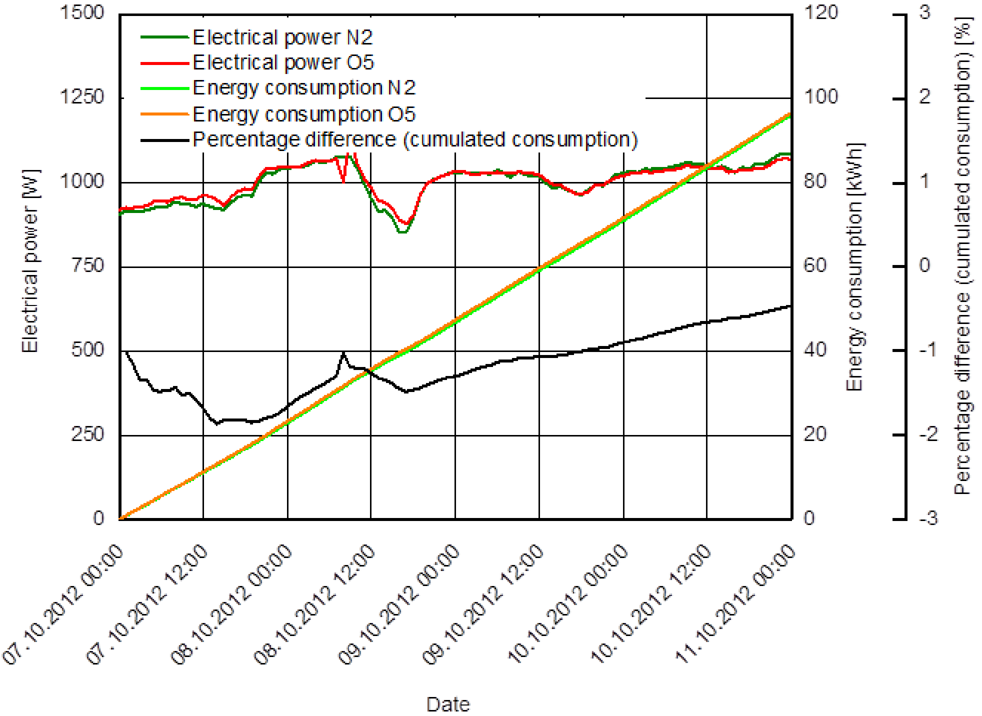

Figure 4: Base line measurement of the Twin Houses

Thermographic images were also taken (an example is shown in Figure 5). There were no obvious differences or anomalies in the fabric heat loss of the ground floor rooms.

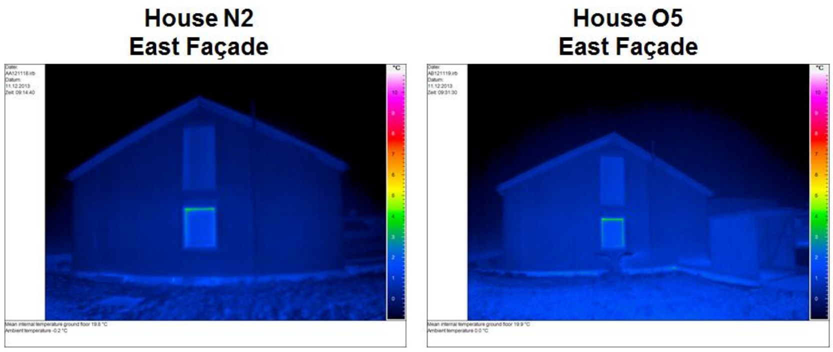

Figure 5 Thermal images of the East Facades of the two houses.

3.3 EXPERIMENTAL DESIGN

3.3.1 *EXPERIMENT 1*

For a validation study, it is necessary to develop a suitable dynamic test that ensures that there are significant heat flows for each of the main heat flow paths. It was decided to have a multi-stage test sequence with three main components – steady state internal temperatures, a sequence of pseudo-random heat injections, and a free-float period. For the first experiment, there was one significant constraint – the houses were only available in the summer period for testing. Because heating energy consumption usually dominates in Europe over cooling energy and also for accuracy reasons, it was decided to only use heat inputs, and to keep the heating system simple by using fast responding electrical heaters. Experimental design was undertaken by modelling the houses using representative climate for Munich (Munich IWEC 2014) with the following aims:

1. To ensure the mechanical ventilation rate was sufficient to prevent significant overheating above the heating set point.
2. To size the heaters required to maintain a suitable setpoint.
3. To decide on the magnitude and schedule for a pseudo-random series of heat injections that would not exceed temperature limits and which would test the building over its inherent time constants.
4. Through the use of sensitivity studies, to identify additional measurements needed to ensure that experimental uncertainty was small and that all significant parameters for model inputs were available. Based on this knowledge the most critical parameters were investigated in more detail during the experiment.

To make use of the two houses in this summer test, it was decided to have the automated external roller blinds down on the south facing windows of one building and fully up on the other – the difference between the two houses would then largely depend on the solar gains. This was implemented in the following way: all blinds were up all the time, except for the south windows; the blinds on the south façade were closed on one house permanently (house N2) and were closed only for the initialisation and the constant temperature scenarios on the other house (house O5).

Although the existing instrumentation on-site was extensive, additional measurements were made as a result of the sensitivity analysis – in particular the solar absorptivity of the external surfaces and the ground reflectivity. Thermal bridges were identified as significant and a 2-D analysis of thermal bridges at the external wall/floor junction, the external wall/ceiling junction and the wall/wall junction with THERM (2014) was carried out by several Annex 58 participants, with linear thermal transmittances included in the specification. Further analysis was later carried out into internal thermal bridges (Section 4.3).

The experimental configuration is shown in Figure 6.

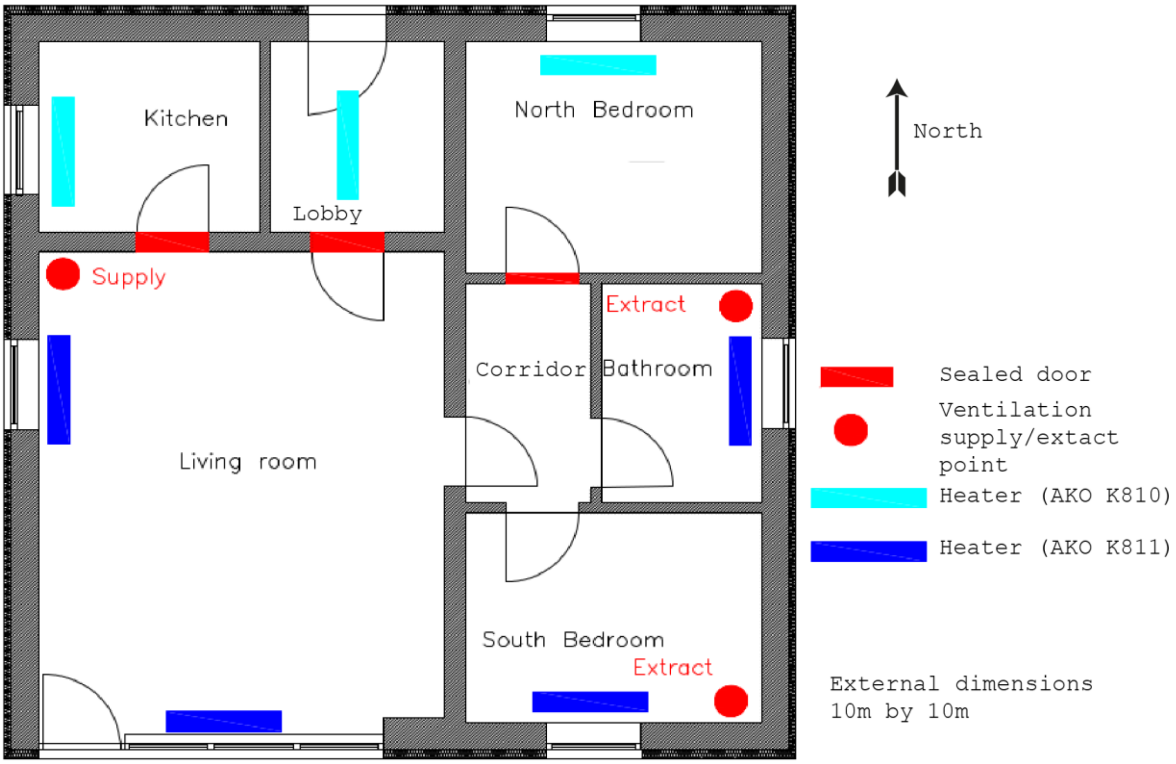

Figure 6: Experimental layout

To reduce complexity, the temperatures in the cellar and attic spaces were measured and treated as boundary conditions (although modelling those spaces was also possible).

The first experiment was undertaken in summer 2013. The planned schedule is shown schematically in Figure 7. It was divided into five different periods. The control in these periods was chosen to reflect common conditions in buildings as well as ensuring the dynamic response was tested.

Period 1: Initialization phase (7 days) in which both buildings were heated to a constant temperature of 30 °C to obtain identical and well-defined start conditions.

Period 2: Room air temperatures were kept constant at 30 °C for 7 days with a required heating power controlled by the building management system.

Period 3: A Randomly Ordered Logarithmic Binary Sequence (ROLBS) for heat inputs into the living room was implemented, with heat injections of 0 and 500 W (with a nominal radiative:convective split of 30 % : 70 %). The use of a pseudo-random sequence of heat injections ensures that the solar and heat inputs are uncorrelated, which should help to disaggregate the fabric heat transfer and solar gains in the analysis. This test sequence lasted for 2 weeks – the sequence has heat pulses ranging from 1 hour to 90 hours in duration to cover the expected range of time constants in the building as determined in the experimental design simulations. These sequences were developed in the EC COMPASS project (van Dijk and Tellez 1995) and customized in this case to cover the maximum expected time constant of the Twin Houses – large in this case as the houses contain a significant amount of thermal mass. All other rooms were without heating power in this period to increase the interaction between the rooms. One reason for including the pseudo-random sequence was to make the dataset more useful for testing system identification techniques, the role of Subtask 3 of Annex 58. For this purpose, it is particularly important to avoid correlation between inputs as much as possible.

Period 4: A constant temperature period of 7 days was to re-initialize the two houses to the same state. The controlled temperature level was set at 25 °C (lower temperatures as the external temperatures were expected to decrease in late summer).

Period 5: In this 7-day period, there were no artificial heat injections.

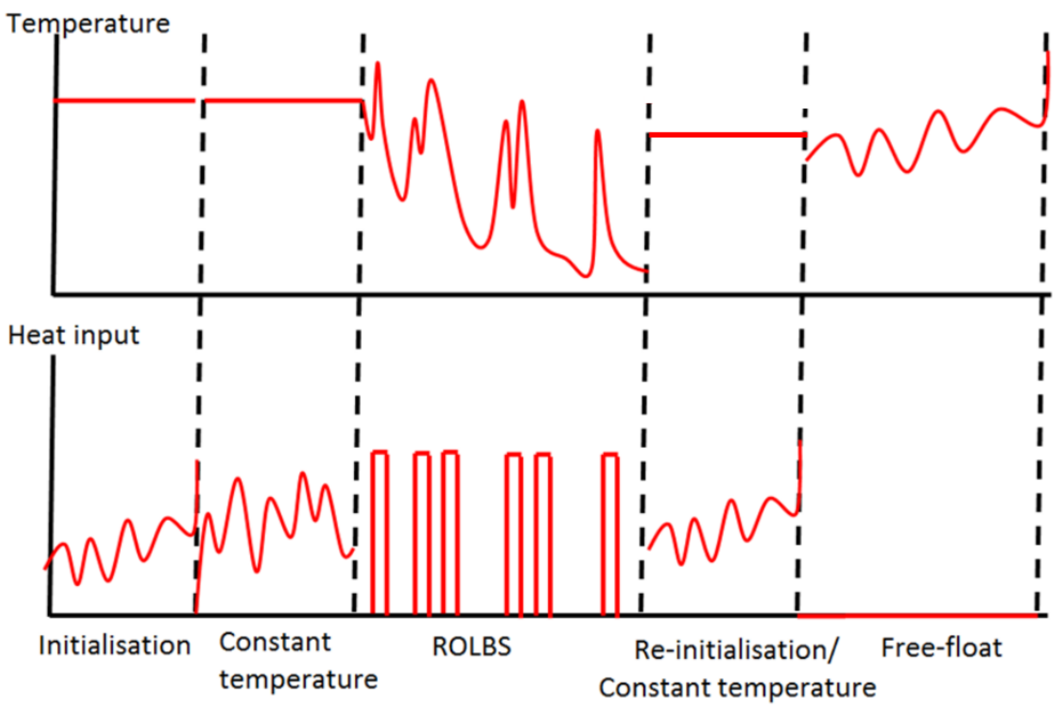

Figure 7: Schematic of test schedule

Heating and Ventilation Systems

The heating power was provided to the rooms through fast responding 2 kW electric convectors driven by a phase-controlled modulator during the constant temperature set point periods.

The southern rooms of the ground floor were ventilated as can be seen in Figure 6. A balanced ventilation system was implemented, with supply air entering the living room with a volume flow rate of 120 m³/h and extracted through the bathroom and the south bedroom with a flow of 60 m³/h each. Because the mechanical ventilation system is a major component of the energy balance, high accuracy is required when controlling and recording the ventilation air temperatures and air volumes during the measurements. To guarantee identical volume flow rates in this experiment both the supply and the extract air ducts were equipped with thermo-anemometers for measuring the air velocities in the ducts. Using profile factors these velocities can be converted to volume flow rates. Since the ventilation system is mass balanced, a volume difference can occur depending on the temperature difference between supply and exhaust air. By phase modulation the fan power was controlled to keep the desired flow rate of 120 m³/h, which was achieved with a standard deviation of only ±0.2 m³/h, less than the uncertainty in the anemometer measurement. To ensure that the exhaust air amount is equal from the two outlets of the bathroom and south bedroom, during the experimental setup the disc valves in both rooms were adjusted using a second, temporary flow meter. All duct joints were sealed carefully using tape to minimize unwanted air flows throughout the ducts' length. The supply air temperature was measured in the cellar after the fan, so the fan's waste heat was included in this temperature. This is discussed further in Section 4.3. The exhaust air temperature was measured before the fan so its heat was not included, as required.

The supply air temperatures and flow rates to the ground floor living room were provided as inputs to the simulation programs.

3.3.2 *Experiment 2*

Experiment 1 provided a useful validation dataset, with an extensive set of modelling undertaken as described later in the report. However, it was considered not ideal for identification exercise as part of Subtask 3 (IEA EBC Annex 58 Final Report: Subtask 3, 2015) because the boundary conditions were allowed to free-float, making it difficult to separate the heat loss coefficient to the outside and the heat loss coefficient to internal boundary spaces. A second experiment was therefore undertaken when one of the Twin Houses became available for testing in April and May 2014. The objectives were as follows.

1. To provide a further validation dataset in a cooler time of year, so that internal heat injections could be increased and mechanical ventilation rate decreased. This should result in more emphasis on fabric losses and less on solar gains compared to the first experiment.
2. To provide an experimental dataset suitable for system identification - this should be useful as a common exercise to compare different analysis approaches as well as provide useful information to explain differences between measurements and predictions from detailed modelling tools.
3. To keep the experiment simple and similar to the first experiment in summer 2013, in order to minimise set up time for the control and instrumentation.

The experimental configuration and test sequence was the same as Experiment 1 except for the following.

- Additional sensors were installed to measure internal surface temperature and heat flux in the middle of each wall and window of the south rooms. These were on the south wall of the south bedroom (to the left of the window, facing outwards), on the east wall of the south bedroom, and on the west wall of the living room. A black globe temperature sensor was added in the living room. Additional air temperature sensors were added to the south bedroom and bathroom, to allow a better analysis of the internal temperature for identification and the magnitude of any stratification: this had already been shown to be important from measurements in the living room in Experimental 1. Near the start of the experiment, an additional temperature sensor was placed at the outlet of the supply air duct to measure the air temperature entering the living room.
- The ROLBS sequence of heat injections was applied in the bathroom and south bedroom as well as the living room as was done in Experiment 1.
- Boundary spaces in the north rooms (kitchen, lobby and north bedroom), attic and cellar were maintained at fixed temperatures of 22°C throughout the test period to increase the identifiability of heat loss coefficients through different flow paths. According to the experimental design, the cellar was expected to require heat inputs up to 3.5 kW, although this was a bit uncertain as the cellar wall and floor constructions are not well known. Blinds were kept closed in the attic to avoid overheating and solar absorption on the floor. Heat inputs were measured for maintaining the fixed temperatures in the north rooms: heat requirements were expected to be small – estimated to be 150 W maximum in the individual spaces. Blinds were closed in the kitchen and north bedroom to reduce the chance of overheating.

Further simulations were undertaken to determine suitable magnitudes for the ROLBS heat injections and mechanical ventilation rate: the objective was to make the heat injections large

enough to lead to significant temperature rises but without excessive overheating. The heat injections were increased to 1800W in the living room, synchronised with 500 W inputs in the south bedroom and 500 W in the bathroom. The supply flow rate was set at a nominal 60 m^3/h (half that of Experiment 1) into the living room, with extract at 30 m^3/h in the bathroom and 30 m^3/h in the south bedroom. The constant temperature phase was set at 30°C in all the south rooms (estimated to require up to 1500W in the living room and up to 300 W in the bathroom and 300 W in the south bedroom).
It was not possible to undertake a blind validation in Experiment 2, because at that stage measured data from Experiment 1 had been made available to modelling teams. Nevertheless, the measured temperature data of the free-float period of Experiment 2 was initially withheld so an element of blind validation was maintained.

3.4 EXPERIMENTS

3.4.1 EXPERIMENT 1

The data sequence for the side-by-side Experiment 1 was from 23rd August to 29th September 2013. As can be seen in Figure 8, the ambient temperature varied from approximately 4°C to about 25°C, with both sunny and cloudy periods. The weather data was gathered continuously. The data from the Twin Houses (O5 and N2) was also continuous apart from a 4-hour gap due to a logging failure (although the ventilation and heating was still operating) on the 3rd September. The 1-minutely logged data was processed to both 10-minutely and hourly averages.

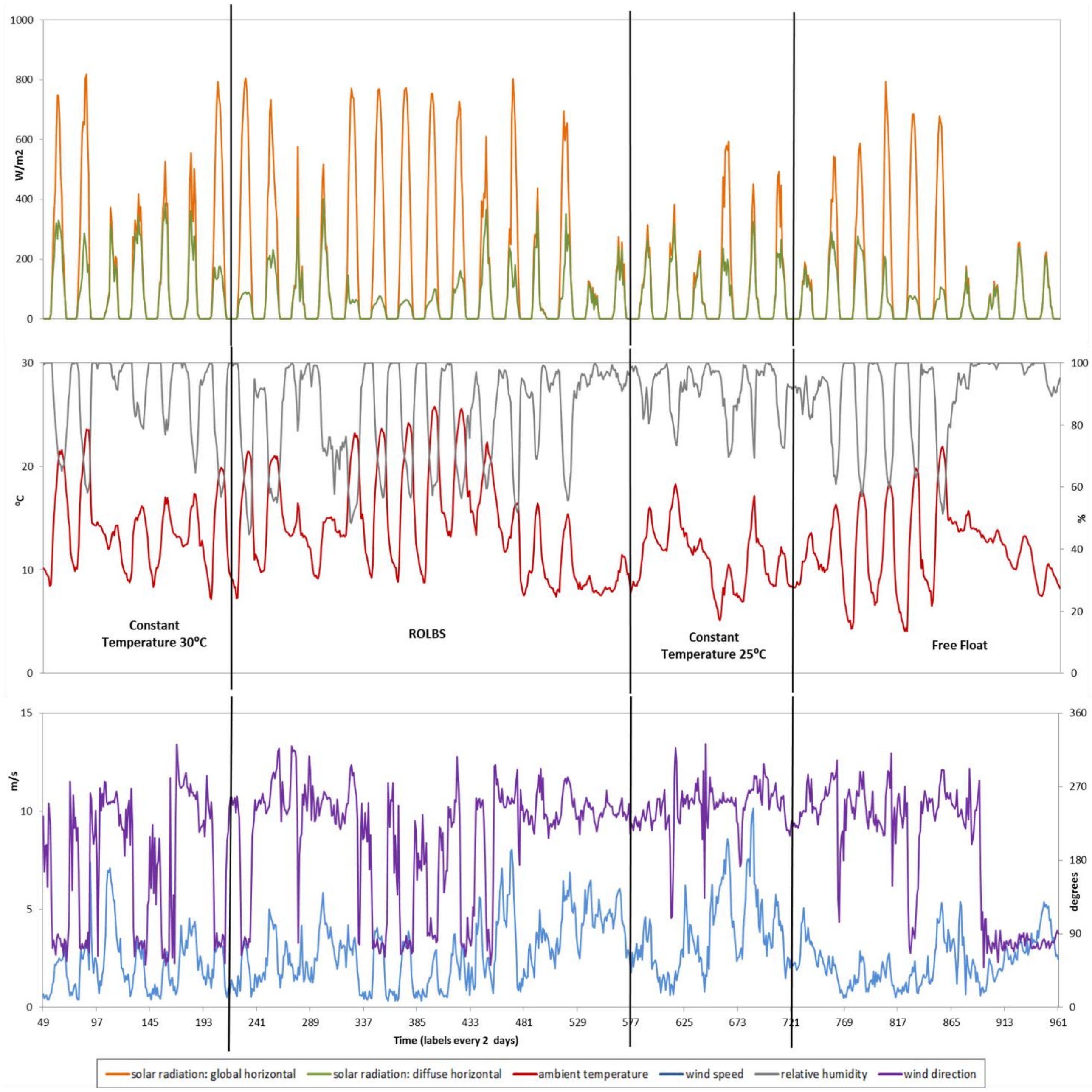

Figure 8: Weather data for the 1st Experiment (23rd August to 29th September 2013)

3.4.2 EXPERIMENT 2

This experiment was conducted on House 05 from 9th April 2014 to 2nd June 2014. Weather data starting on 14th April is shown in Figure 9. Ambient temperature varied from a minimum of -4°C to a maximum of 29°C, with significant variations in solar radiation. The weather data was gathered continuously. During the initialisation period the heating control failed so the internal temperatures dropped below the set-point (although the data acquisition system was still running), so the analysis for the constant temperature period was undertaken with data from 24th April onwards. (Note that the data before this date could be used by researchers as an additional free-float period, but it wasn't used in the analysis because the rooms were not maintained at a constant temperature as intended.). There was also a failure of the data logging in the free float period from 23rd to 26th May. However, as it was in the free-float period, comparisons between measurements and predictions could easily be made by excluding this period.

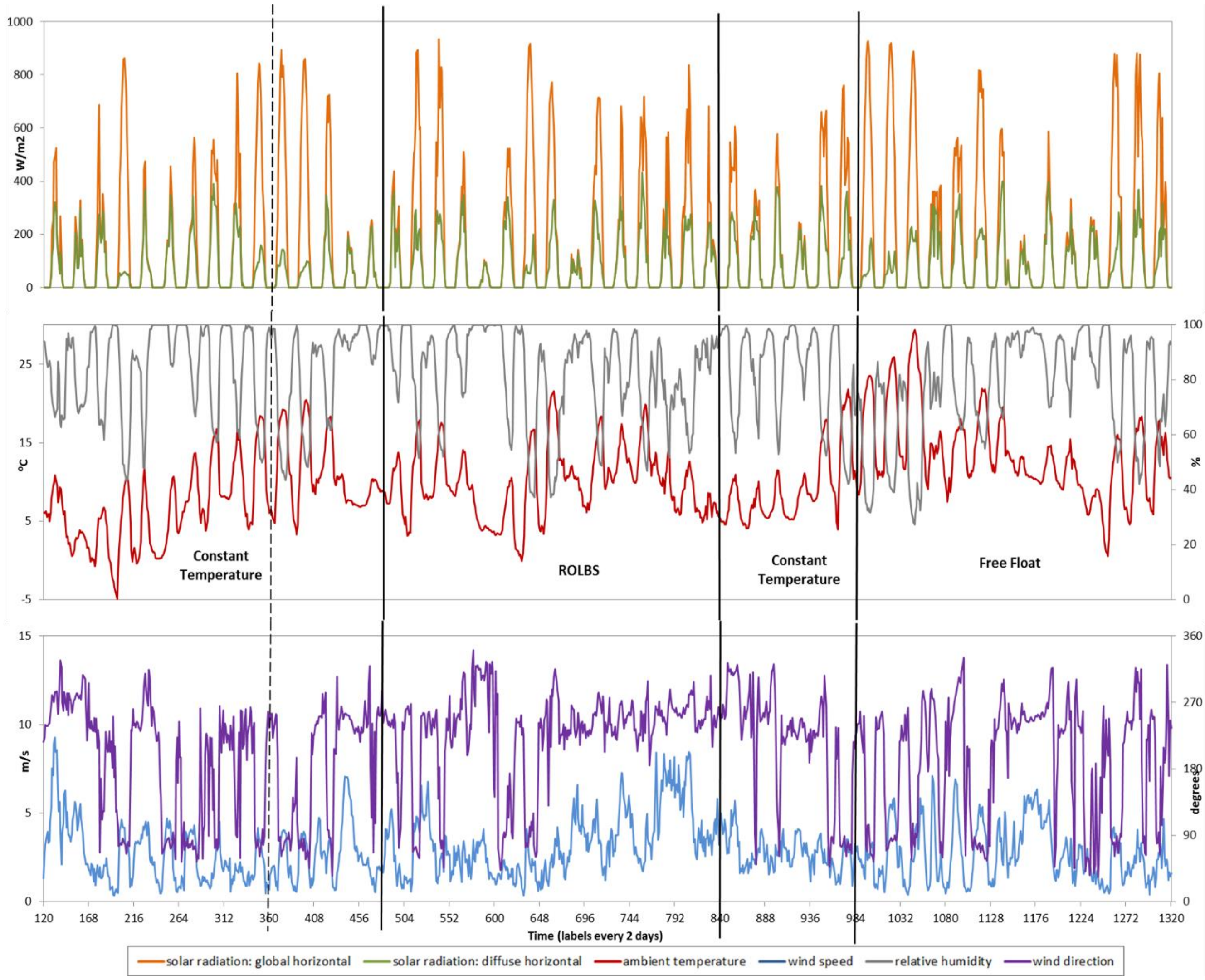

Figure 9: Weather data for the 2nd Experiment (14th April to 2nd June 2014)

3.5 INSTRUMENTATION

Table 2 lists the sensors in each of the Twin Houses in Experiment 1 for monitoring the thermal conditions in the buildings, together with an estimate of their total accuracy including the entire measuring chain following calibration before the experiment. Some of these sensors can be seen in the internal views of the Twin Houses in Figure 2. The climate data from the Fraunhofer IBP's weather station were provided as boundary conditions. These sensors, listed in Table 2, are calibrated regularly as recommended by the manufacturer. Full details of the sensors and calibration results are included in the specification made available to modellers.

Table 2: Instrumentation of the Twin Houses

Each Twin House		**Meteorological**	
Sensor	**Accuracy**	**Sensor**	**Accuracy**
Air temperature in all 7 rooms at a height of 125 cm (radiation shielded)	±0.12 K	Ambient air temperature (ventilated)	±0.10 K
Additional air temperatures in the living room at a height of 67 cm and 187 cm (radiation shielded)	±0.14 K	Ambient relative humidity	±2.0 %
Air temperatures in the cellar and attic spaces	±0.14 K	Ground temperatures, depth of 0, 50, 100 and 200 cm	
Relative humidity living room	±2.3 %	Wind speed (@ 10 m height)	±0.1 m/s
Fresh, supply and exhaust air temperatures measured in the cellar	±0.04 K	Wind direction (@ 10 m height)	±1.0 °
Heating power of the 6 heated rooms	±1.5 %	Solar radiations: global, diffuse and vertical (north, east, south, west)	±2.0 %
Supply and exhaust fan power	±1.5 %	Long wave radiation (horizontal, west)	<34 W/m²
Ventilation flow rates	±3.5 m³/h		
Heat flux at the west facade	±0.65 W/m²		
West wall temperatures: Internal, external and between layers	±0.14 K		

As mentioned above, additional sensors were installed for Experiment 2. These measured internal surface temperature and heat flux in the middle of each wall and window of the south rooms, globe temperature sensor in the living room, additional air temperature sensors locations in south bedroom and bathroom, and in the supply air duct at the living room outlet.

Some images of these sensors are shown in Figures 10 to 13. Figure 14 shows sensors at the on-site weather station.

For the measured and predicted results presented later in this report, the comparisons were made with internal air temperature for the pseudo-random heat input and free-float periods, and with heat inputs for the constant temperature setpoint periods. As shown in Table 2, the heating power accuracy is ±1.5 %; individual calibrated shielded temperature sensors have an accuracy of better than 0.15 °C. However, stratification was observed in rooms as discussed later, so modelling choices made in selection of appropriate room average temperatures may account for offsets in the order of 1 °C, especially during higher heating power inputs.

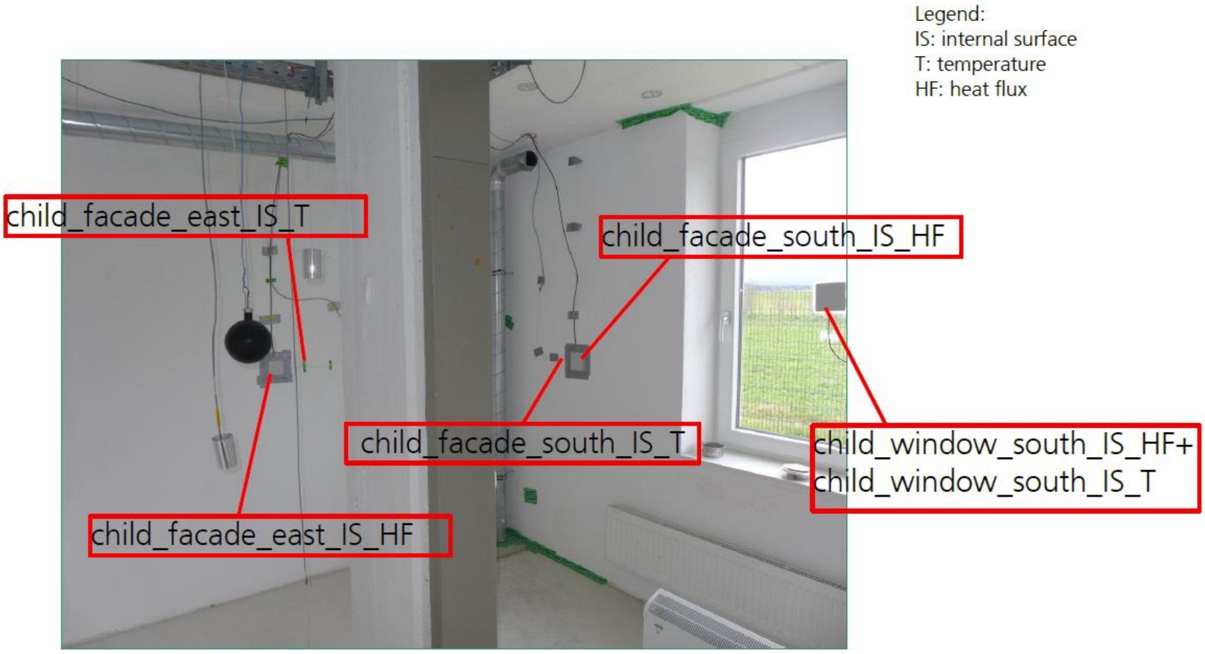

Figure 10: Heat flux and surface temperature sensors in the south bedroom (also called children's room)

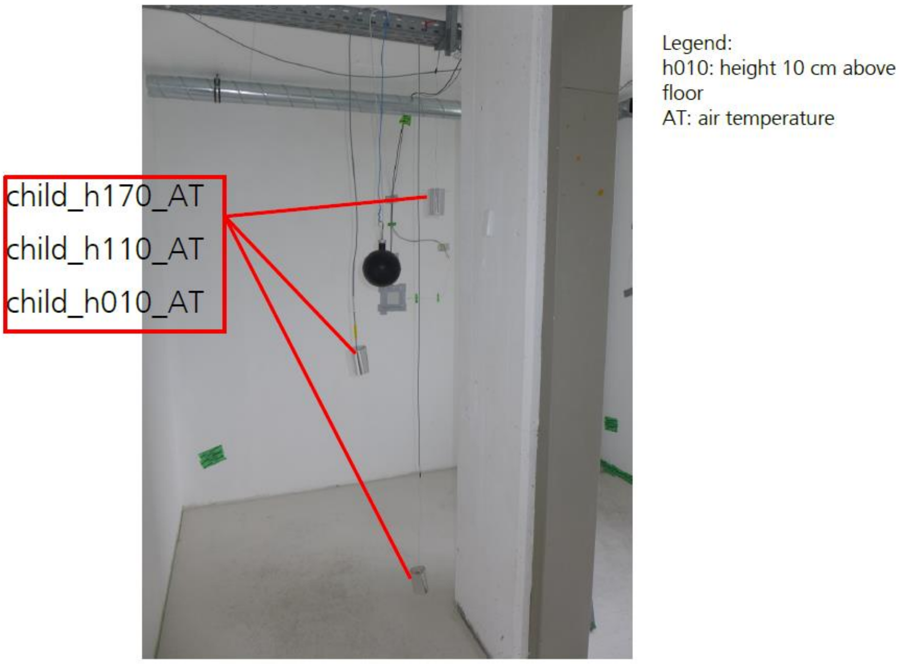

Figure 11: Air temperature sensors in the south bedroom (also called children's room)

living room

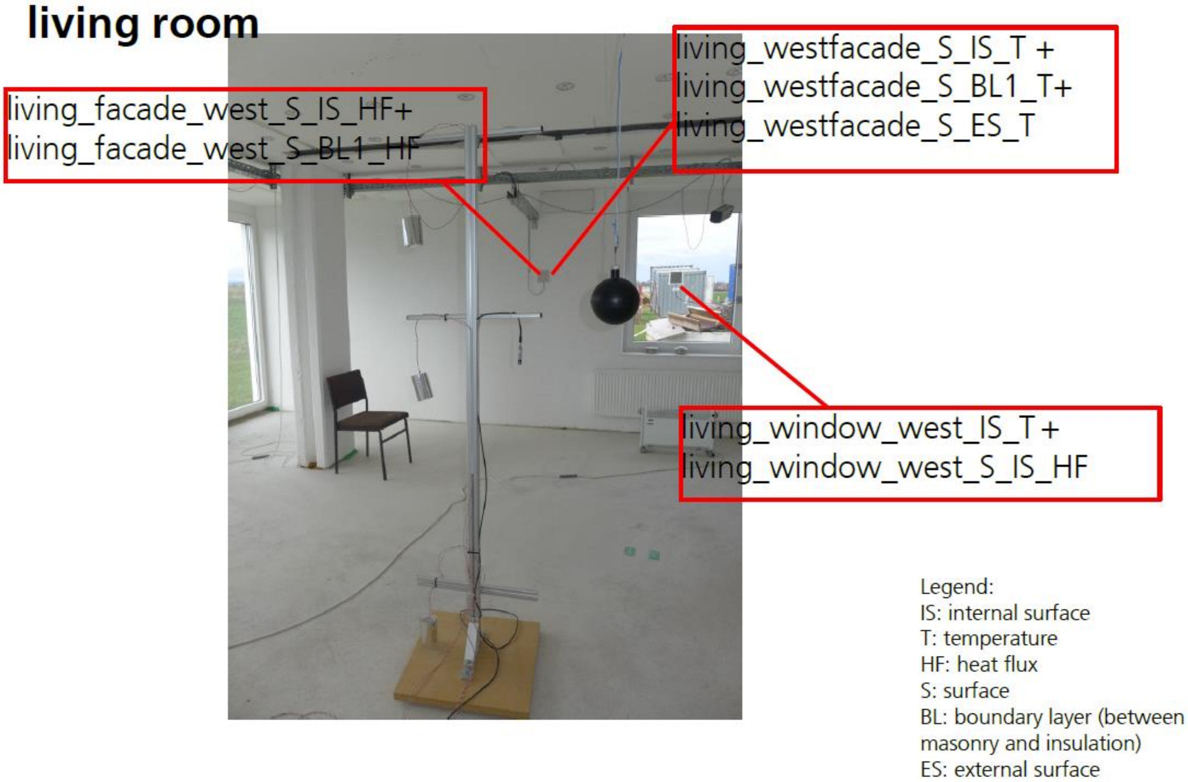

Figure 12: Heat flux and temperature sensors in the living room: view to west

Figure 13: Heat flux and temperature sensors in the living room: view to south

Figure 14: On-site weather station

3.6 SPECIFICATION

Modelling teams were given a comprehensive specification covering:

- Geometrical details (including location and size of surrounding buildings)
- Constructional details
- Roller blind details
- Thermal bridge details
- Glazing and frame properties – optical and thermal
- Internal contents (thermal mass)
- Pressurization test data
- Ventilation system details

Full details of the specification are available for download (see Section 6.4); only a summary of key features is included in this report. Figure 15 lists the specification documents and the associated specification files that were made available to modellers (and which are available to other researchers).

```
Experiment_1_Empirical_Modelling_Specification_200514.pdf
Experiment_2_Empirical_Modelling_Specification_240614.pdf
Specification_files
├── Construction_details
│   ├── detail basement ceiling.jpg
│   ├── detail eaves.jpg
│   ├── detail entry door connection.jpg
│   ├── detail window connection.jpg
│   ├── door_dimensions.jpg
│   ├── ip_fl_4.ipe
│   ├── ip_ipl4E.ipe
│   ├── Window6.3_glazing_props_EN410.txt
│   ├── Window6.3_glazing_props_NFRC.txt
│   └── Window_types_201213.pdf
├── Experimental_details
│   ├── 2013_09_20_Measurement_uncertainty_of_sensors.xlsx
│   ├── 2013_10_30_Measurement_uncertainty_of_weather_data_sensors.xlsx
│   └── Expt_2_additional sensors.pdf
├── Images
│   ├── bathroom1.JPG
```

```
│   ├── bathroom2.JPG
│   ├── bed room1.JPG
│   ├── bedroom2.JPG
│   ├── childrens_room1.JPG
│   ├── childrens_room2.JPG
│   ├── corridor.JPG
│   ├── doorway1.JPG
│   ├── Ground in front of south facade Twin House N2.JPG
│   ├── Ground in front of south facade Twin House O5.JPG
│   ├── internal_door.jpg
│   ├── kitchen1.JPG
│   ├── kitchen2.JPG
│   ├── living1.JPG
│   ├── living2.JPG
│   ├── living room1.JPG
│   ├── living room3.JPG
│   ├── Livingroom_under_window_construction.pdf
│   └── Thermograms_Twin_houses.pptx
├── Layout_drawings
│   ├── Dachstuhl.dwg
│   ├── Internal and external door.jpg
│   ├── Plan EG_Experimentierhäuser_3D.dwg
│   ├── Plan EG_Experimentierhäuser.dwg
│   ├── SitePlanDimensions.pdf
│   ├── window sill.pdf
│   └── Zwillingshäuser_Plansatz.pdf
├── Measured_data
│   ├── Experiment_1
│   │   ├── Minute book test case TWIN_HOUSE_1.pdf
│   │   ├── Twin_house_exp1_house_N2_10min_ductwork_correction.xls
│   │   ├── Twin_house_exp1_house_N2_60min_ductwork_correction.xls
│   │   ├── Twin_house_exp1_house_O5_10min_ductwork_correction.xls
│   │   ├── Twin_house_exp1_house_O5_60min_ductwork_correction.xls
│   │   ├── Twin_house_exp1_weather_data_all_measurements_10min.xls
│   │   └── Twin_house_exp1_weather_data_all_measurements_60min.xls
│   ├── Experiment_2
│   │   ├── Minute_book_Test_Case_TWIN_House_2.pdf
│   │   ├── Twin_house_exp2_10min_wo_ff.xlsx
│   │   └── Twin_house_exp2_60min_wo_ff.xlsx
│   └── Raw_one-minutely_data
│       ├── Experiment_1
│       │   ├── Annex58_N2.csv
│       │   ├── Annex58_N2.xlsx
│       │   ├── Annex58_O5.csv
│       │   ├── Annex58_O5.xlsx
│       │   ├── Annex58_weather.csv
│       │   └── Annex58_weather.xlsx
│       ├── Experiment_2
│       │   ├── test_case_2.csv
│       │   └── weather_test_case_2.csv
│       └── Readme.txt
├── Other_docs
│   ├── heater_ako_k810-820_fd8206_gb-2009-07-20-1.pdf
│   ├── IEA58_ST4_CE_Twin_House_Influence Moisture.pdf
│   ├── Model_details.docx
│   ├── SupplyAirDuct-Heatloss_GRo-UIBK_MK_V2.xlsx
│   ├── SupplyAirDuct-Heatloss_GRo-UIBK_V1_10minute_data.xlsx
│   └── SupplyAirDuct-Heatloss_GRo-UIBK_V1.xlsx
└── Thermal_bridge_calcs
    ├── thermal_bridge_drawings.pdf
    ├── THERM_models
    │   ├── teHs_wallWall.THM
    │   ├── twHs_ceiling2.THM
    │   ├── twHs_ceiling.THM
    │   ├── twHs_extWall2.THM
    │   │
```

```
│    ├── twHs_extWall.THM
│    ├── twHs_intWallGFl_2.THM
│    ├── twHs_intWallGFl.THM
│    └── twHs_wallWall2.THM
├── THERM_PSI_tab.ods
├── TRISCO_ceiling140.png
├── TRISCO_ceilingW270.png
├── twinhouse_1_140_ceiling.doc
├── twinhouse_1_140_ceiling.trc
├── twinhouse_1_270_ceiling.doc
└── twinhouse_1_270_ceiling.trc
```

Figure 15: Structure of Specification Files Available to Modellers

Measured data with both 10-minutely and hourly averages were made available to modelling participants, who were asked to submit hourly-averaged predictions (irrespective of what simulation timestep they used).

4 Modelling

4.1 Modelling Teams

Table 3 shows the modelling teams who participated in the validation study. Most organisations were members of IEA EBC Annex 58 but others were encouraged to participate, particularly commercial program vendors and organisations involved in IEA EBC Annex 60 who were developing programs for building and community energy systems based on the Modelica and Functional Mockup Interface standards.

Table 3: Modelling participants and programs

Organisation	IEA58 participant	Program	Experiment 1 Blind	Experiment 1 Re-modelled	Experiment 2
CIEMAT	Yes	TRNSYS	✓	✓	
CTU Prague	Yes	Matlab	✓	✓	✓
CTU Prague	Yes	Matlab_Simulink	✓	✓	
DTU	Yes	ESP-r	✓	✓	✓
Equa	No	IDA-ICE	✓	✓	
Fraunhofer_1	Yes	TRNSYS	✓	✓	
Fraunhofer_2	Yes	Wufi	✓	✓	✓
HFT Stuttgart_1	No	INSEL		✓	✓
HFT Stuttgart_1	No	INSEL		✓	✓
Hong_Kong_City_1	Yes	eQuest	✓	✓	
Hong_Kong_City_2	Yes	EnergyPlus	✓		
IES	No	IESVE	✓	✓	✓
University Innsbruck	Yes	Dynbil	✓	✓	✓
Leuven_1	Yes	Modelica_model1: IDEAS library Build 01.12.2013	✓		✓
Leuven_2	Yes	Modelica_model2: IDEAS library Build 23.12.2013	✓		
Leuven_3	Yes	Modelica_model3: IDEAS library Build 23.12.2013	✓		
Leuven_4	Yes	TRNSYS	✓		
Liege_HEPL	Yes	EES	✓	✓	✓
Liege_JCG	Yes	EES	✓		
Liege_Ulg	Yes	Modelica: no library	✓		
Liege_Ulg_BEMS	Yes	TRNSYS 17		✓	✓
Politecnico di Milano	Yes	EnergyPlus	✓	✓	✓
RWTH Aachen	No	Modelica; version: 3.2; Dymola 2014		✓	
Univ Gent	Yes	TRNSYS	✓	✓	✓
Univ Strathclyde	Yes	ESP-r	✓	✓	✓

4.2 Experiment 1 Blind Validation

For Experiment 1 (blind validation), the measured data distributed to all modelling teams included all the climate data and the temperatures of the bounding cellar and attic spaces. In addition, the following house data was provided.

Period 1: Initialization phase: measured internal temperatures (but not the heating power).

Period 2: Constant temperature phase: measured temperatures were provided as inputs for the modelling, with heating power to achieve these measured temperatures to be predicted.

Period 3: ROLBS sequence of heat inputs into the living room: in this case, the heat inputs were provided for the modelling, with resulting temperatures to be predicted.

Period 4: Constant temperature phase: measured temperatures were provided as inputs for the modelling, with heating power to achieve these measured temperatures to be predicted.

Period 5: Free-float period (no heat injections): modelling teams were required to predict the resulting temperatures.

Figure 16 shows graphically what the modelling teams had to predict.

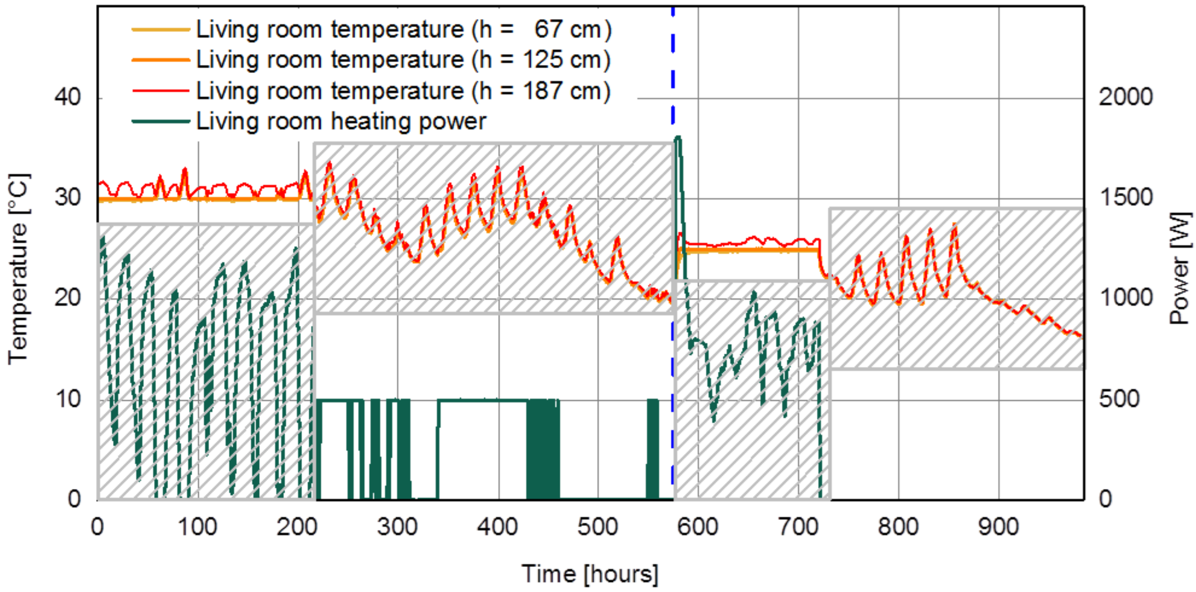

Figure 16: Profiles of the heating power and the air temperature at all three measurement heights of the living room of one of the houses. The grey shaded data are the validation targets that are not provided for the blind validation.

The specification described in Section 3.6 was sent to all participating teams. An important part of the methodology is to provide a response service to answer questions arising when modellers were setting up the models. This was done, with answers co-ordinated by ESRU and Fraunhofer IBP and sent to all teams. It often involved additional information being provided. A few examples of modelling questions are as follows.

- Can you explain in more details the timing convention and averaging of the data?
- Exactly where are the heaters located?
- Can you provide more accurate glazing and frame measurements?
- Are there measured ground reflectance coefficients? (additional measurements were made)
- Can you provide details of the roller blind materials?
- Can you provide photographs showing location of sensors?

At the end of the modelling work, all the additional information was added to the specification document and associated files. The fact that the specification has been implemented by over 20 modelling teams gives some confidence that all important experimental details are described in the document.

Modelling teams were requested to provide their predictions in a standard format together with a modelling report outlining the simulation program used and any assumptions made. A questionnaire was also issued asking teams to clarify key aspects of the heat and mass transfer modelling within their programs – for example, how windows and blinds were modelled, what internal convection coefficients were used, the solar diffuse sky model used.

4.3 EXPERIMENT 1 RE-MODELLING

After the blind validation results had been submitted, all the measured data was released to modelling teams. They were encouraged to check their predictions against the measurements, correct any obvious user errors, analyse reasons for differences, and then resubmit. In nearly all cases, modelling teams identified user input errors when they re-examined their models. Some of these were of minor consequence; others identified a major error, for example in interpretation of timing or of heater capacity.

As a result of comparing the blind validation results with measurements, a few specification and experimental errors were identified, so the teams were also issued with a slightly updated specification. The improvements were as follows:

- Internal thermal bridges between the partition walls and the floor and ceiling were identified as significant. 2-D and 3-D modelling was carried out by several of the modellers of these thermal bridges, as well as the thermal bridges associated with support pillars. Updated thermal bridge linear thermal transmittances (psi-values) were included in the specification. Psi-values associated with internal wall/ceiling junctions and chi-values associated with the pillars were added to the specification.
- The section of ventilation duct running through the kitchen was uninsulated, resulting in heat gain to the supply air and a heat loss to the kitchen air. An analysis was carried out with PHLuft (2014) to quantify the effect, with updated supply air temperatures and kitchen heat loss supplied as part of the modelling data.
- Internal walls solar absorptivity was measured (0.17).

At this stage it is possible that teams can adjust inputs (e.g. in the magnitude of thermal bridges or airgap resistance when roller blinds are down) to calibrate or “tune” their predictions. For this reason, all teams resubmitting result sets were asked to document changes that were made so that only user errors were corrected.

4.4 EXPERIMENT 2

For experiment 2, the modelling teams had access to all the data from experiment 1, so a blind validation phase was not possible. However, the measured temperatures in the final free-float period were withheld so predictions could be compared. Also, the dataset was made available to Subtask 3 for a common exercise: the intention was to compare the predictions for this free-float period using identified statistical models with predictions from the physically-based models.

5 Results

5.1 SELECTION OF COMPARATIVE METRICS

There are two categories of comparative metrics. The first is a timestep comparison; this is usually a time series display of all or part of the test sequence. It is largely based on a visual comparison and is useful for observing general trends and time shifts between measurements and predictions. The other category involves comparative statistics of the output variables (mean values, peak values, integrated heating energy input). These are useful for summarising the overall goodness of fit, but they do not provide information in cases where there is good fit over part of the period and poor fit elsewhere.

In the analysis carried out, comparative statistics were calculated for each simulation in each period in each room, covering all the standard comparisons made in Manz et al (2006): mean and standard deviation; maximum and minimum values; root mean square difference; average absolute difference and 95th percentile of absolute difference (to remove transients and outliers). For heat input, the total energy input to the room was calculated for each period. In addition, Spearman's rank correlation coefficient was calculated to give the level of correspondence between the shapes of the simulation and the experimental profiles (Kendall and Gibbons, 1990). Thiel's Inequality Coefficient was also calculated for each period: this is a composite metric that combines inequalities in the mean, the variance and the covariance of the two profiles (Williamson 1995). Table 4 is an example showing the comparative statistics for the living room temperature in Experiment 2.

Two metrics were selected as giving the most physically meaningful overall measures of how well the simulations matched the experiment:

1. the average absolute difference between measurement and prediction, to characterise the degree of fit in magnitude;
2. Spearman's rank correlation coefficient, to characterise the degree of fit in profile shapes.

Table 4: Comparative statistics by period – Living room temperature, Experiment 2

		Ex_O5	Sim 2	Sim 4	Sim 7	Sim 10	Sim 11	Sim 12	Sim 16	Sim 18	Sim 19	Sim 20	Sim 21	Sim 23	Sim 24
Mean	ROLBS	31.3	31.4	32.0	30.2	32.6	32.7	32.6	31.6	30.6	28.7	30.8	32.0	31.7	30.9
°C	Free float	24.4	23.4	23.6	23.3	22.9	23.6	23.7	24.3	23.3	22.0	24.0	23.9	22.0	22.6
StDev	ROLBS	4.5	4.7	5.2	4.2	6.1	5.6	5.7	4.5	4.6	5.7	4.0	5.4	5.3	5.4
°K	Free float	2.5	2.7	2.8	2.7	2.8	2.9	3.0	2.7	2.7	3.9	2.6	3.0	3.3	7.2
Xmax	ROLBS	39.5	39.9	42.0	37.6	43.4	43.1	43.0	40.3	39.2	39.7	38.9	42.4	41.2	40.9
°C	Free float	30.4	30.0	30.0	30.0	28.8	30.0	30.6	30.6	30.0	29.9	30.0	30.2	30.0	28.9
Xmin	ROLBS	23.4	23.0	23.3	22.7	22.6	23.2	22.9	23.5	22.5	18.8	23.7	22.7	22.7	21.8
°C	Free float	20.3	19.4	19.1	19.2	18.3	18.7	18.8	20.1	19.4	15.7	19.7	19.1	17.7	0.0
Diffrms	ROLBS		0.39	1.30	1.58	2.10	1.87	1.84	0.62	1.18	3.02	1.48	1.27	1.04	1.46
°K	Free float		1.06	0.91	1.41	1.53	0.95	0.92	0.34	1.09	2.75	0.59	0.74	2.61	7.27
AvAbsDiff	ROLBS		0.31	1.00	1.28	1.66	1.52	1.45	0.47	0.96	2.67	1.18	1.02	0.83	1.28
°K	Free float		1.00	0.81	1.29	1.46	0.82	0.84	0.27	1.03	2.52	0.45	0.66	2.38	1.88
AbsDiff95%	ROLBS		0.72	2.69	2.87	4.20	3.57	3.29	1.29	2.56	5.01	2.88	2.67	2.05	2.48
°K	Free float		1.71	1.44	2.11	2.09	1.57	1.32	0.68	1.56	3.91	1.14	1.14	3.74	2.74
Spearman's	ROLBS		0.997	0.988	0.964	0.993	0.994	0.992	0.992	0.977	0.985	0.952	0.992	0.993	0.974
coefficient	Free float		0.982	0.970	0.946	0.983	0.973	0.973	0.978	0.987	0.980	0.939	0.984	0.953	0.984
Thiel's IC	ROLBS		0.006	0.023	0.028	0.073	0.024	0.024	0.008	0.025	0.048	0.020	0.009	0.037	0.032
	Free float		0.022	0.019	0.029	0.032	0.020	0.019	0.007	0.023	0.059	0.012	0.015	0.056	0.157

In the experiments here, given the high levels of instrumentation, it is also possible to compare measurements with predictions of more focused areas of the building (e.g. surface temperatures and heat fluxes) and heat transfer processes (e.g. longwave and shortwave fluxes on different orientations). In addition, differences in predictions for the two houses can be compared with differences in measurements. Not all of this data has been fully analysed to date – this report gives an overview of the main results.

Regarding the assessment of goodness-of-fit, there are no definitions of “acceptable” bands within empirical validation, so these need to be guided by experimental uncertainty and subjective judgement. In the analysis undertaken to date, agreement of average absolute temperatures within 1°C is classed as “good”, as is the heating power within 100 W for the constant temperature periods. However, it is recognised that it would be useful to establish a more rigorous basis for categorising the level of agreement. This topic is worthy of further investigation in future validation studies in order to guide the experimental accuracy required and the evaluation of program predictions against measurements.

5.2 EXPERIMENT 1 BLIND VALIDATION

Some example graphs are presented. Figures 17 and 18 show the heat input predictions of the 21 submissions during the initial constant temperature phase in the living room of the house with blinds up (house O5). The x-axis shows the timeline in days; the y-axis shows the heat input predictions, with the thicker black line recording the measured data. As can be seen, 2 or 3 of the models had major discrepancies indicating a major user error or a mistake in the timestamp of the submitted predictions. On the other hand, many programs showed qualitatively good agreement with measurements.

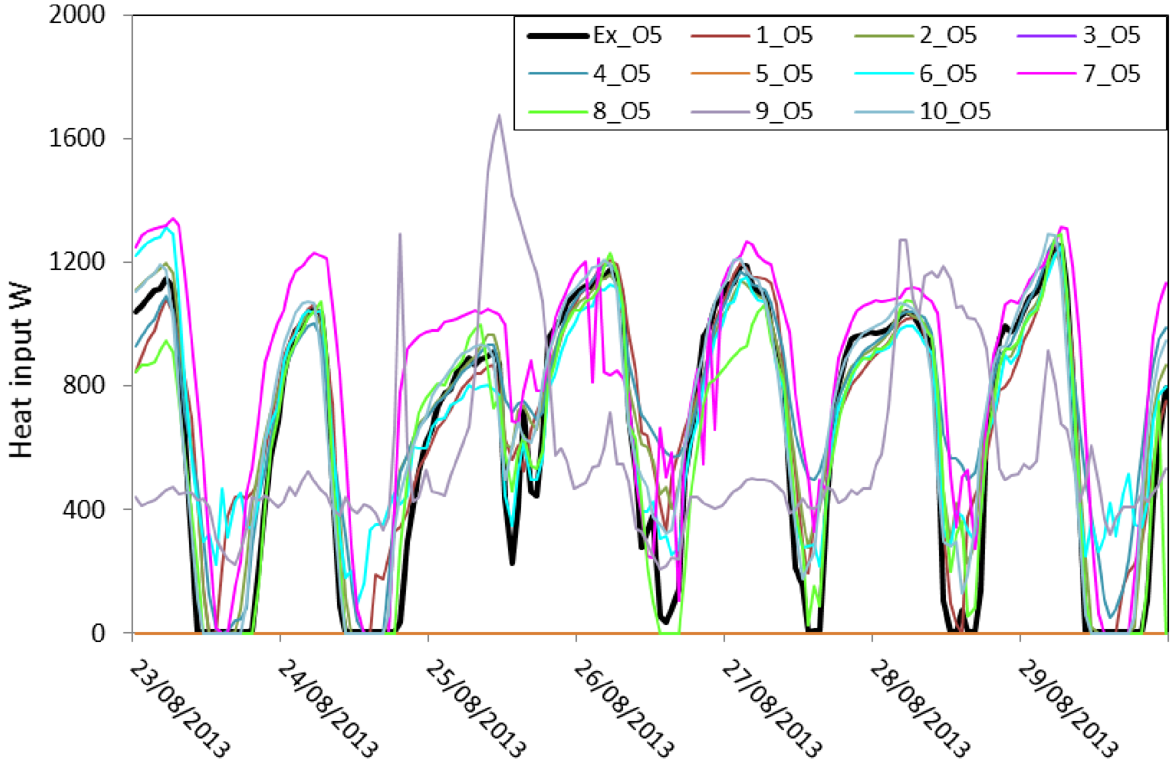

Figure 17: Experiment 1 Blind validation. Living room heat input: constant temperature phase (30°C): models 1-10 + experimental data: House O5

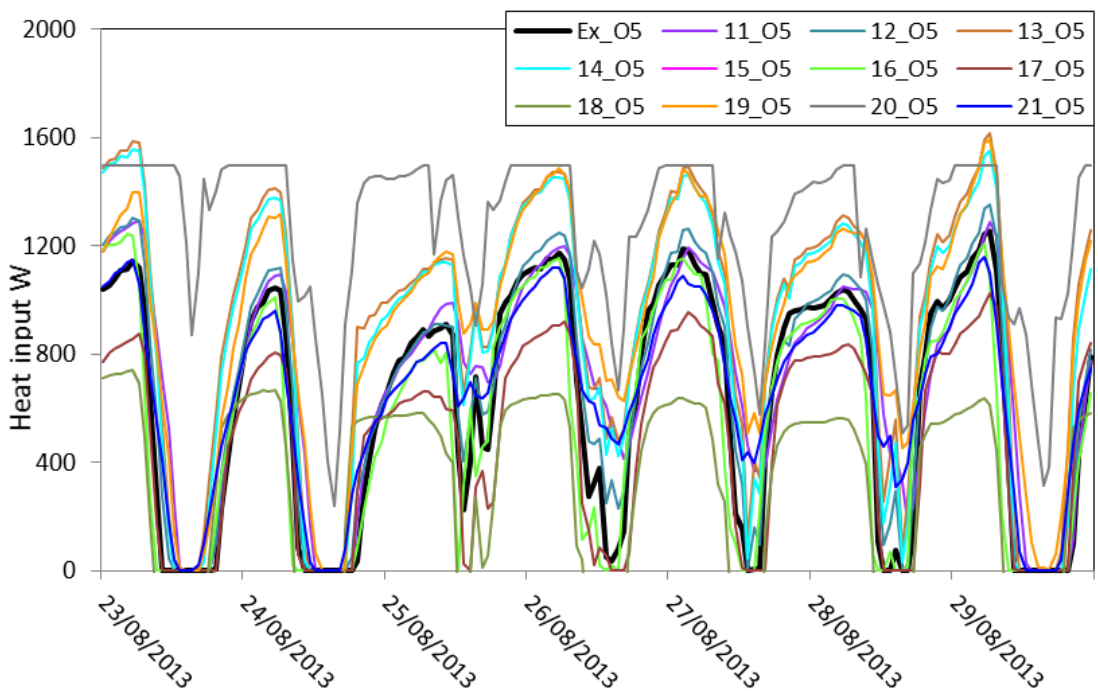

Figure 18: Experiment 1 Blind validation. Living room heat input: constant temperature phase (30°C): models 11-21 + experimental data: House O5

Figures 19 and 20 show the predicted and measured living room temperatures in the same house during the ROLBS input sequence. Again, a few models are clearly erroneous (e.g. models 1, 8, 9, 20), whereas others follow the trends well.

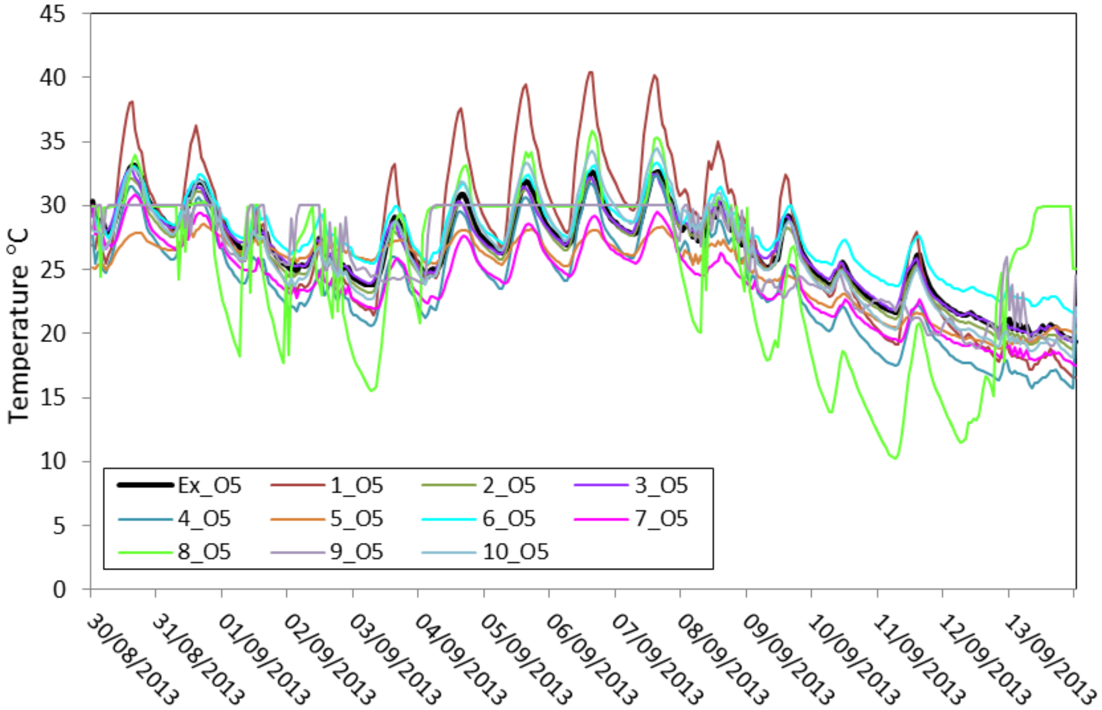

Figure 19: Experiment 1 Blind validation. Living room temperature: ROLBS sequence: models 1-10 + experimental data: House O5

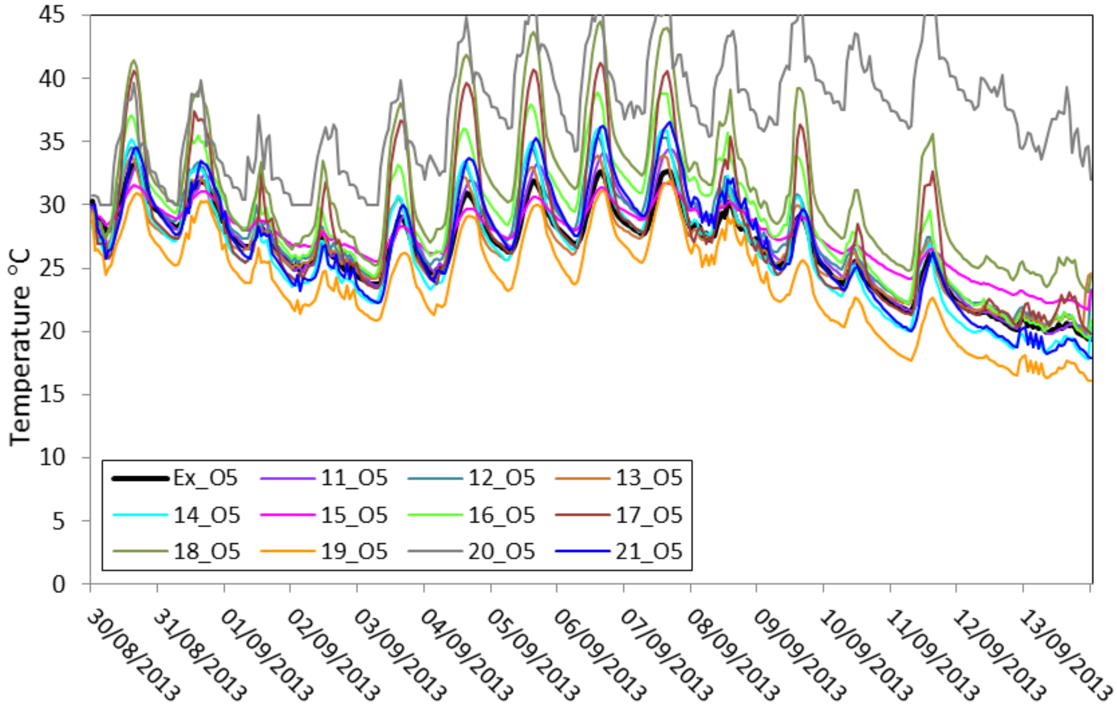

Figure 20: Experiment 1 Blind validation. Living room temperature: ROLBS sequence: models 11-21 + experimental data: House O5

Figure 21 and 22 show the predicted and measured living room temperatures in the same house during the free-float period. Again, some models are clearly erroneous, whereas others follow the trends well.

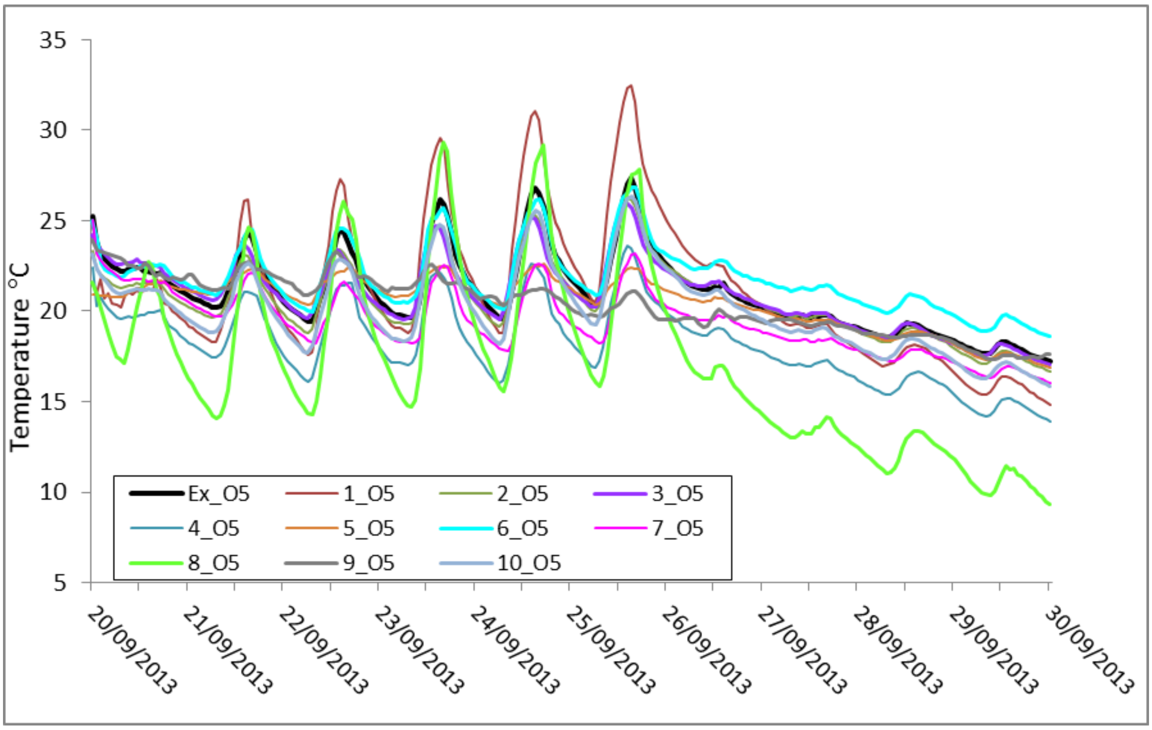

Figure 21: Experiment 1 Blind validation. Living room temp: Free-float period: models 1-10 + experimental data: House O5

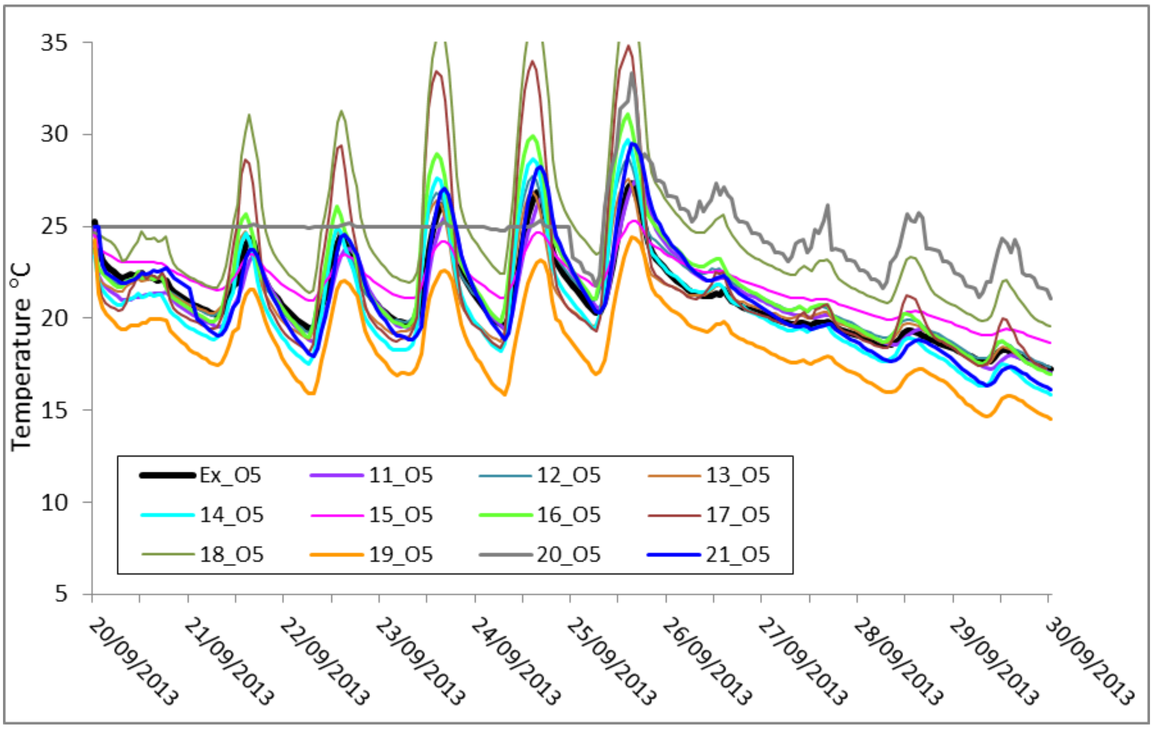

Figure 22: Experiment 1 Blind validation. Living room temperature: Free-float period: models 11-21 + experimental data: House O5

Table 5 compares the magnitude fit of temperature for all models, in the two periods with defined heat input: Period 3 (ROLBS) and Period 5 (free-float). Comparisons are given for the living room (LRT), south bedroom (SBDT), kitchen (KITT) and north bedroom (NBDT). Results are given for each room in both houses - House O5 with the blinds up, and House N2 with the blinds down. They are also included for the temperature difference between the two houses. For example, "N2–O5 LRT" is the difference in predictions of the living room temperature in the two houses: it is a good indicator of how well the models predict the difference in solar gains for the cases with blinds up and blinds down. The level of agreement is shown in bands, with green indicating average absolute differences between measurements and predictions of less than 1 °C; yellow in the range 1–2 °C; orange in the range 2–4 °C; red in the range 4-8 °C and purple showing outliers >8 °C.

As seen in the timeline comparisons, some submissions are clearly erroneous, but others show good levels of agreement overall. No program predicted temperature in every room and every period within 1^0C although two simulations came close. The bottom of the table (in the section labelled "fixed temperature periods") shows the temperatures submitted by modellers for the living room in the constant temperature periods (periods 2 and 4). The differences with measured data here should be close to zero because these were program inputs. The differences occur mainly because in the experiment there were a few times during the constant temperature periods when the setpoint was exceeded - especially in the living room with its large south facing windows, and in most cases modellers assumed the fixed setpoint rather than using the measured temperatures.

Table 6 shows the Spearman's rank correlation coefficient between the measurements and predicted temperatures for the same rooms for periods 3 and 5. In this case green represents a correlation of >0.9, yellow is 0.8 – 0.9, orange is 0.7 – 0.8, red is 0.35 - 0.7 and purple shows outliers <0.35. The significance associated with the bands was chosen to separate the performance of the submitted results. Most programs show good agreement for this metric, indicating that time offsets between measurements and predictions are generally small. Note that standard tests for significance are not helpful, as what they tell us is the likelihood that the two series do not describe the same phenomenon. In this case, with the very large number of data points, the standard test implies that with a coefficient over 0.35 there is less than 0.002% likelihood that the simulations are fundamentally incorrect. But all these simulation programs are already known to be capable of predicting building response to imposed climate boundary conditions, so the bands have been chosen to subjectively distinguish between the performance of the programs tested.

Table 7 shows the difference between the model predictions of heating to maintain the setpoint and the measurements in the constant temperature periods: period 2 at 30 °C, and period 4 at 25 °C. In this case, green represents agreement better than 100 W, yellow is 100-200 W, orange is 200-300 W, red is 300-500 W, and purple is >500 W. Comparisons are given for heat inputs into the living room (LRQ), south bedroom (SBDQ), kitchen (KITQ) and north bedroom (NBDQ). The data at the bottom of the table shows the heat inputs for the living room for the ROLBS and free-float periods. Again, these differences with measured data should be zero. A number of programs included the ROLBS heat inputs as internal (or "casual") heat gains rather than heater inputs, which accounts for those where the difference is around 240 W (so this is simply a function of modelling procedure rather than an error in predictions). Simulation results 5, 8, 9 and 20, however, show large errors that were caused by incorrect modelling of the simulation periods. Results are missing for a few models which combined the rooms and where the heat inputs to individual rooms could not be separated.

Table 5: Experiment 1. Blind validation results for the ROLBS sequence and free-floating periods: Temperature magnitude fit

Magnitude Fit — **Average absolute difference in temperature**

	Period	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6	Sim 7	Sim 8	Sim 9	Sim 10	Sim 11	Sim 12	Sim 13	Sim 14	Sim 15	Sim 16	Sim 17	Sim 18	Sim 19	Sim 20	Sim 21
Fixed heating periods																						
O5 LRT	ROLBS	2.2	0.5	0.4	2.5	1.9	1.2	2.4	4.3	1.7	0.8	0.7	1.2	0.7	1.2	1.3	2.4	2.0	4.8	2.6	10.8	1.2
O5 LRT	Free	1.5	0.5	0.4	2.9	0.9	0.8	1.7	4.1	1.2	1.0	0.6	0.6	0.6	1.0	1.1	0.9	1.4	3.6	2.5	3.8	0.8
N2 LRT	ROLBS	1.9	0.9	0.4	2.5	2.9	0.4	1.8	4.8	3.5	0.9	1.0	0.4	0.5	1.6	0.8	0.8	0.7	2.2	2.2	6.0	1.4
N2 LRT	Free	1.7	0.8	0.4	2.5	2.7	0.4	1.0	3.5	3.4	1.4	1.0	0.3	0.4	1.5	0.8	1.2	0.6	1.3	2.1	3.8	1.5
N2-O5 LRT	ROLBS	2.0	0.5	0.5	0.4	4.5	0.9	0.7	2.2	3.6	0.7	1.5	0.9	0.5	1.5	0.6	3.2	1.5	2.7	0.5	5.0	1.3
N2-O5 LRT	Free	1.4	0.5	0.4	0.5	3.3	0.6	0.7	1.5	4.1	0.4	1.1	0.5	0.5	1.1	0.6	2.0	1.2	2.3	0.5	2.6	1.5
O5 SBDT	ROLBS	1.4	0.4	0.9	1.4	2.2	1.9	1.5	7.4	0.8	0.9	1.5	1.8	1.5	0.6	1.6	3.0	2.3	4.1	2.1	11.4	0.8
O5 SBDT	Free	0.9	0.3	0.5	1.3	2.0	1.5	0.7	5.4	0.8	0.3	0.8	1.5	1.2	0.5	0.9	1.8	1.3	3.2	1.7	3.8	0.8
N2 SBDT	ROLBS	3.2	0.5	0.6	1.7	3.6	1.0	1.1	5.5	3.4	1.4	0.6	1.0	0.8	0.4	1.0	1.7	0.6	1.7	1.9	6.2	0.8
N2 SBDT	Free	1.9	0.6	0.6	1.3	4.3	1.2	0.5	4.1	3.5	0.8	0.6	1.1	0.9	0.4	0.5	1.6	0.4	1.1	1.5	3.5	0.6
N2-O5 SBDT	ROLBS	1.8	0.7	0.8	0.3	3.1	0.9	0.4	2.0	3.3	0.5	2.0	0.8	0.7	0.9	0.6	1.6	1.7	2.4	0.3	5.3	1.6
N2-O5 SBDT	Free	1.2	0.6	0.5	0.2	2.3	0.5	0.4	1.6	3.7	0.6	1.4	0.4	0.4	0.6	0.4	1.3	1.3	2.1	0.2	2.3	1.4
O5 KITT	ROLBS	1.7	1.4	1.0	0.8	2.5	2.9	0.5	3.8	1.1	2.9	1.9	2.7	2.9	0.8	2.2	4.7	3.0	3.4	1.8	7.6	1.8
O5 KITT	Free	1.1	0.8	0.7	0.9	2.9	2.3	0.5	3.6	1.6	1.3	1.3	1.9	2.3	0.7	1.6	2.2	1.7	2.7	1.0	5.3	1.3
N2 KITT	ROLBS	2.0	0.8	0.7	0.6	4.7	2.5	0.5	2.3	2.3	0.9	1.1	2.4	3.3	0.7	1.7	1.4	2.3	2.6	1.2	6.5	1.1
N2 KITT	Free	1.2	0.5	0.5	0.9	4.4	2.2	0.8	3.3	3.3	1.0	0.7	1.9	2.8	0.7	1.4	0.8	1.1	1.8	0.7	5.5	0.7
N2-O5 KITT	ROLBS	0.3	0.6	0.5	0.2	2.3	0.4	0.4	5.7	1.6	2.5	2.9	0.4	0.4	0.3	0.5	3.4	0.7	1.0	0.6	1.5	0.8
N2-O5 KITT	Free	0.1	0.4	0.3	0.1	1.5	0.2	0.4	6.3	1.7	2.2	1.9	0.3	0.5	0.2	0.3	2.2	0.6	0.9	0.4	0.6	0.6
O5 NBDT	ROLBS	4.0	0.3	0.7	1.3	2.4	1.7	0.9	6.0	0.3	1.4	0.3	1.4	1.3	0.5	1.9	3.4	2.0	1.8	0.2	6.4	0.4
O5 NBDT	Free	2.8	0.3	0.4	1.0	3.0	1.5	0.4	4.7	1.1	0.5	0.3	1.0	1.2	0.5	1.4	1.5	1.2	1.7	0.1	3.8	0.3
N2 NBDT	ROLBS	3.3	0.3	0.5	1.0	4.0	1.7	0.5	4.9	1.8	1.1	0.5	1.6	2.2	0.5	1.6	0.8	1.3	1.2	0.2	4.8	0.4
N2 NBDT	Free	2.3	0.2	0.3	0.6	4.2	1.7	0.4	3.9	2.7	0.3	0.2	1.3	2.0	0.5	1.2	0.8	0.6	1.0	0.2	4.0	0.3
N2-O5 NBDT	ROLBS	0.7	0.1	0.3	0.3	1.6	0.2	0.7	1.3	1.6	0.4	0.8	0.2	0.9	0.1	0.3	2.6	0.7	0.7	0.1	2.2	0.2
N2-O5 NBDT	Free	0.6	0.2	0.2	0.4	1.2	0.3	0.5	0.9	1.6	0.3	0.4	0.4	0.8	0.1	0.3	1.6	0.7	0.7	0.1	0.8	0.2
Fixed temperature periods																						
O5 LRT	30°C	0.4	0.4	0.4	0.5	4.4	0.3	0.5	0.5	0.5	0.4	0.4	0.4	0.6	0.5	0.4	0.7	1.2	0.9	1.0	1.3	0.4
O5 LRT	25°C	0.3	0.2	0.2	0.2	3.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.4	0.3	0.2	0.5	0.5	0.5	0.8	5.0	0.2
N2 LRT	30°C	0.2	0.2	0.2	0.2	0.5	0.2	0.2	0.4	0.2	0.2	0.2	0.2	0.4	0.4	0.2	0.6	0.6	0.2	0.9	0.4	0.2
N2 LRT	25°C	0.2	0.2	0.2	0.3	0.3	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.5	0.4	0.2	0.7	0.6	0.2	0.9	3.6	0.2

Green = <1°C   Yellow = 1<>2°C   Orange = 2<>4°C   Red = 4<>8°C   Purple => 8°C

Table 6: Experiment 1. Blind validation results for the ROLBS sequence and free-floating periods: Temperature shape fit

Shape Fit — **Spearman's Rank Correlation with Experiment**

	Period	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6	Sim 7	Sim 8	Sim 9	Sim 10	Sim 11	Sim 12	Sim 13	Sim 14	Sim 15	Sim 16	Sim 17	Sim 18	Sim 19	Sim 20	Sim 21
Fixed heating periods																						
O5 LRT	ROLBS	0.96	0.99	0.99	0.98	0.89	0.99	0.98	0.76	0.80	0.98	0.98	0.97	0.96	0.97	0.98	0.96	0.93	0.91	0.99	0.25	0.98
O5 LRT	Free	0.95	0.98	0.98	0.98	0.92	0.96	0.95	0.97	0.74	0.98	0.95	0.95	0.93	0.95	0.99	0.94	0.90	0.94	0.96	0.72	0.97
N2 LRT	ROLBS	0.94	0.99	0.99	0.96	0.94	0.98	0.99	0.70	0.85	0.98	0.99	0.98	0.97	0.98	0.98	0.99	0.95	0.97	0.97	0.48	0.97
N2 LRT	Free	0.87	0.98	0.99	0.93	0.84	0.99	0.99	0.78	0.85	0.96	0.97	0.97	0.97	0.95	0.98	0.99	0.91	0.99	0.89	0.82	0.91
O5 SBDT	ROLBS	0.97	0.99	0.97	0.98	0.54	0.99	0.91	0.82	0.91	0.96	0.97	0.98	0.95	0.99	0.98	0.82	0.90	0.84	0.98	0.12	0.96
O5 SBDT	Free	0.97	0.98	0.96	0.94	0.63	0.99	0.87	0.95	0.73	0.98	0.94	0.99	0.97	0.98	0.99	0.93	0.94	0.88	0.98	0.70	0.94
N2 SBDT	ROLBS	0.95	0.99	0.98	0.98	0.74	0.89	0.87	0.68	0.91	0.94	1.00	0.93	0.92	0.98	1.00	0.88	0.96	0.96	0.97	0.43	0.99
N2 SBDT	Free	0.95	0.99	0.98	0.99	0.82	0.99	0.98	0.70	0.92	0.97	0.99	0.99	1.00	0.99	1.00	0.96	0.95	0.98	0.96	0.86	0.98
O5 KITT	ROLBS	0.96	0.99	0.94	0.96	0.66	0.96	0.95	0.83	0.80	0.90	0.98	0.99	0.96	0.97	0.95	0.87	0.93	0.85	0.99	0.24	0.97
O5 KITT	Free	0.96	0.98	0.93	0.91	0.73	0.98	0.94	0.86	0.84	0.89	0.95	1.00	0.98	0.98	0.96	0.90	0.93	0.88	0.99	0.87	0.94
N2 KITT	ROLBS	0.96	1.00	0.92	0.97	0.78	0.92	0.95	0.81	0.81	0.93	0.87	0.96	0.94	0.93	0.91	0.97	0.95	0.95	0.99	0.78	0.99
N2 KITT	Free	0.97	1.00	0.96	0.96	0.76	0.98	0.98	0.92	0.92	0.97	0.95	0.98	0.96	0.97	0.96	0.98	0.96	0.97	1.00	0.84	0.98
O5 NBDT	ROLBS	0.96	0.97	0.96	0.99	0.76	0.94	0.93	0.60	0.95	0.94	1.00	0.99	0.92	0.94	0.95	0.77	0.84	0.72	0.99	-0.09	0.95
O5 NBDT	Free	0.93	0.97	0.98	1.00	0.84	1.00	0.99	0.65	0.99	0.97	0.98	0.99	0.99	0.94	0.96	0.67	0.81	0.74	1.00	0.92	0.93
N2 NBDT	ROLBS	0.98	0.98	0.98	0.99	0.91	0.98	0.98	0.57	0.95	0.95	0.98	0.99	0.98	0.97	0.97	0.94	0.93	0.87	0.99	0.95	0.97
N2 NBDT	Free	0.92	1.00	0.99	1.00	0.88	1.00	1.00	0.63	1.00	1.00	1.00	1.00	0.99	0.98	0.99	0.99	0.95	0.96	1.00	0.93	0.98

Green = >0.90   Yellow = 0.80<>0.90   Orange = 0.70<>0.80   Red = 0.35<>0.70   Purple = <0.35

Table 7: Experiment 1: Blind validation results for the constant temperature periods: Heat input magnitude fit

Magnitude Fit — **Average absolute difference in heat input**

Constant temperature periods

	Period	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6	Sim 7	Sim 8	Sim 9	Sim 10	Sim 11	Sim 12	Sim 13	Sim 14	Sim 15	Sim 16	Sim 17	Sim 18	Sim 19	Sim 20	Sim 21
O5 LRQ	30°C	127	109		149	623	142	231	91	429	97	147	73	291	251		100	157	529	295	640	129
O5 LRQ	25°C	204	77		85	818	112	145	266	367	71	103	78	248	216		52	183	574	245	736	124
N2 LRQ	30°C	142	119		105	140	147	194	187	439	88	135	107	322	343		71	146	353	280	409	86
N2 LRQ	25°C	109	92		109	229	199	129	359	433	92	82	73	394	252		84	150	541	228	817	64
N2-O5 LRQ	30°C	193	145		171	691	156	159	187	333	108	166	92	131	135		139	109	250	162	256	169
N2-O5 LRQ	25°C	219	42		117	629	167	31	98	98	41	35	33	276	42		61	68	67	28	98	97
O5 SBDQ	30°C	83	82		126	94	113	52	493	93	159	139	107	95			122	162	104	35	952	100
O5 SBDQ	25°C	83	83		162	153	146	84	296	173	207	167	137	172			161	184	189	34	332	138
N2 SBDQ	30°C	108	100		156	165	118	76	453	139	238	164	126	124			186	196	100	59	972	124
N2 SBDQ	25°C	90	114		168	250	95	127	248	236	248	174	142	117			167	204	203	70	544	153
N2-O5 SBDQ	30°C	49	39		46	97	35	43	47	98	80	48	32	35			70	47	47	44	302	44
N2-O5 SBDQ	25°C	77	44		32	97	60	53	65	66	52	40	45	57			80	30	37	46	286	39
O5 KITQ	30°C	62	89		153	255	122	97	79	81	120	124	112	113			132	152	111	158	139	110
O5 KITQ	25°C	50	84		142	229	129	114	64	102	74	114	125	134			122	153	166	148	224	118
N2 KITQ	30°C	65	99		164	44	129	116	74	96	217	113	127	130			132	166	129	158	161	120
N2 KITQ	25°C	48	98		165	102	120	139	82	141	265	121	137	147			124	172	181	158	247	133
N2-O5 KITQ	30°C	7	13		12	284	16	23	20	22	98	23	15	18			12	14	19	7	23	13
N2-O5 KITQ	25°C	6	21		24	321	24	45	41	41	198	19	22	14			51	19	21	14	53	22
O5 NBDQ	30°C	190	19		47	25	32	24	116	30	79	52	24	30			59	110	33	82	44	27
O5 NBDQ	25°C	191	18		35	63	42	39	75	47	71	50	51	49			37	95	95	68	184	46
N2 NBDQ	30°C	174	42		76	42	55	21	99	54	102	65	55	62			65	130	51	90	55	54
N2 NBDQ	25°C	166	41		75	104	34	49	70	101	96	71	77	83			99	120	122	85	237	74
N2-O5 NBDQ	30°C	16	26		29	19	23	32	32	33	23	14	31	32			18	23	30	8	33	27
N2-O5 NBDQ	25°C	26	32		42	47	24	56	60	60	31	28	34	35			82	27	33	21	102	33
Fixed heating periods																						
O5 LRQ	ROLBS	241	66		243	12	0	239	520	338	75	238	13	26			65	240	63	239	953	69
O5 LRQ	Free	0	0		0	0	0	0	0	0	0	0	1	2			0	2	0	0	516	0
N2 LRQ	ROLBS	240	67		242	11	0	237	542	310	75	237	14	28			67	239	63	238	953	69
N2 LRQ	Free	0	0		0	0	0	0	0	0	0	0	1	2			0	2	0	0	516	0

Where difference is 240W in ROLBS period, ROLBS heat input was modelled as casual gain

Green = <100 W; Yellow = 100<>200W; Orange = 200<>300 W; Red = 300<>500 W; Purple = >500W

Table 8 shows the Spearman's rank correlation coefficient between the measurements and predicted heat inputs for the constant temperature periods. Agreement is poorer for this metric, particularly for the south bedroom. The reasons for this are not yet clear.

Table 8: Experiment 1. Blind validation results for constant temperature periods: Heat input shape fit

Shape Fit — **Spearman's Rank Correlation with Experiment**

Constant temperature periods

	Period	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6	Sim 7	Sim 8	Sim 9	Sim 10	Sim 11	Sim 12	Sim 13	Sim 14	Sim 15	Sim 16	Sim 17	Sim 18	Sim 19	Sim 20	Sim 21
O5 LRQ	30 C	0.959	0.985		0.971	-0.010	0.955	0.866	0.938	0.415	0.970	0.916	0.986	0.973	0.981		0.952	0.964	0.865	0.970	0.760	0.982
O5 LRQ	25 C	0.695	0.962		0.857	0.093	0.894	0.876	0.636	0.286	0.906	0.924	0.962	0.817	0.975		0.931	0.875	0.681	0.978	0.277	0.636
N2 LRQ	30 C	0.966	0.985		0.987	0.868	0.987	0.800	0.898	0.401	0.942	0.977	0.989	0.988	0.978		0.967	0.963	0.883	0.994	0.767	0.971
N2 LRQ	25 C	0.910	0.965		0.952	0.846	0.919	0.951	0.608	0.187	0.889	0.955	0.975	0.935	0.981		0.955	0.822	0.784	0.959	0.090	0.960
O5 SBDQ	30 C	0.899	0.868		0.947	0.792	0.933	0.747	0.859	0.376	0.868	0.857	0.947	0.921			0.900	0.912	0.757	0.944	0.348	0.943
O5 SBDQ	25 C	0.256	0.541		0.363	0.127	0.152	0.616	0.023	0.307	0.059	0.520	0.501	0.054			0.214	0.300	-0.105	0.426	-0.022	0.417
N2 SBDQ	30 C	0.187	0.223		0.460	0.054	0.314	-0.192	0.229	-0.002	0.057	0.267	0.369	0.359			0.045	0.599	0.069	0.362	-0.062	0.322
N2 SBDQ	25 C	0.572	0.761		0.767	0.182	0.734	0.773	0.087	0.445	0.295	0.731	0.726	0.736			0.433	0.491	0.349	0.551	-0.379	0.643
O5 KITQ	30 C	0.888	0.910		0.780	0.763	0.965	0.854	0.912	0.421	0.924	0.781	0.966	0.963			0.893	0.947	0.862	0.885	0.610	0.887
O5 KITQ	25 C	0.909	0.906		0.755	0.731	0.871	0.875	0.721	0.658	0.911	0.782	0.885	0.877			0.900	0.852	0.737	0.863	-0.120	0.850
N2 KITQ	30 C	0.882	0.822		0.800	0.872	0.923	0.884	0.918	0.455	0.000	0.814	0.966	0.967			0.895	0.944	0.873	0.919	0.625	0.945
N2 KITQ	25 C	0.947	0.846		0.798	0.759	0.828	0.914	0.681	0.723	0.000	0.685	0.883	0.904			0.884	0.882	0.773	0.883	0.002	0.916
O5 NBDQ	30 C	0.893	0.833		0.932	0.764	0.973	0.727	0.857	0.406	0.967	0.837	0.924	0.898			0.824	0.934	0.865	0.956	0.534	0.959
O5 NBDQ	25 C	0.871	0.890		0.870	0.316	0.941	0.901	0.580	0.572	0.890	0.774	0.783	0.774			0.889	0.795	0.664	0.922	0.070	0.949
N2 NBDQ	30 C	0.897	0.834		0.942	0.661	0.960	0.773	0.854	0.465	0.957	0.789	0.926	0.901			0.873	0.932	0.865	0.958	0.690	0.961
N2 NBDQ	25 C	0.818	0.905		0.931	0.214	0.952	0.964	0.523	0.664	0.931	0.767	0.825	0.839			0.909	0.833	0.681	0.962	-0.291	0.945

Green = >0.90; Yellow = 0.80<>0.90; Orange = 0.70<>0.80; Red = 0.35<>0.70; Purple = <0.35

5.3 Experiment 1 Re-modelling

A total of 18 submissions were made in this phase of the exercise. Note that this included some additional contributions from HFT Stuttgart (INSEL program), RWTH Aachen (Modelica; version: 3.2; Dymola 2014) and from University of Liege (TRNSYS) that had not participated in the blind validation.

Figures 23 and 24 are given as examples. These are for the free-floating period (where the data was not released to modellers). There are some anomalous programs with poor agreement, but qualitatively, the agreement in magnitude and shape is good.

As for the blind validation results, an overall comparison between the different models was made using the same two metrics for the magnitude and shape fits between the time series data. Tables 9 to 12 correspond to Tables 4 to 7, but for the re-modelled submissions. As might be expected, there are fewer anomalous results and the overall level of agreement is improved.

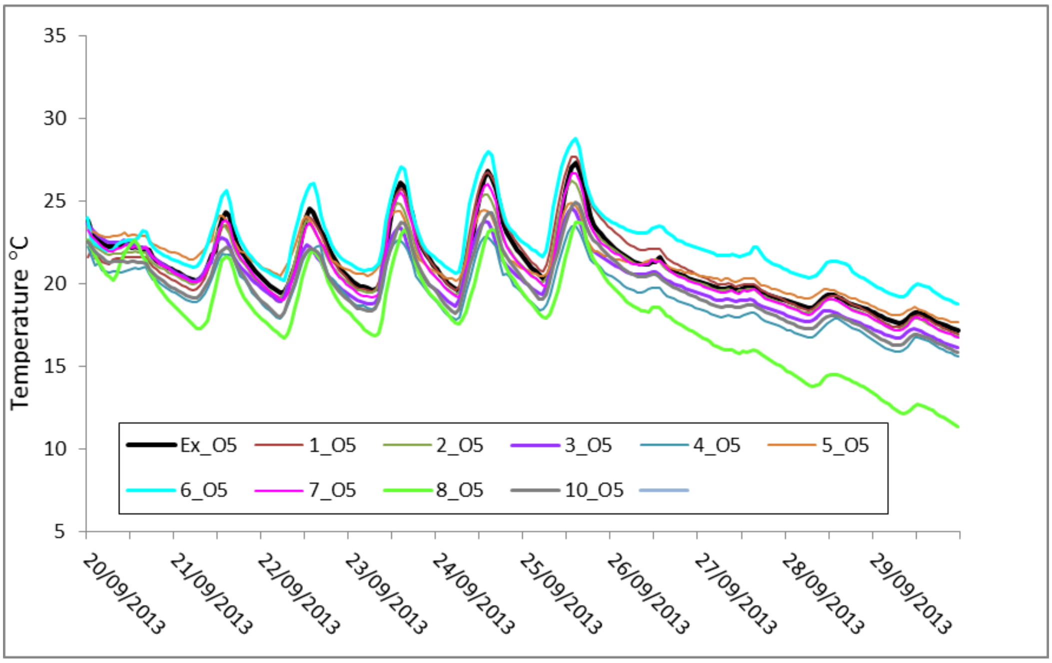

Figure 23: Experiment 1 Re-modelling. Living room temperature: Free-float period: 9 models + experimental data: House O5

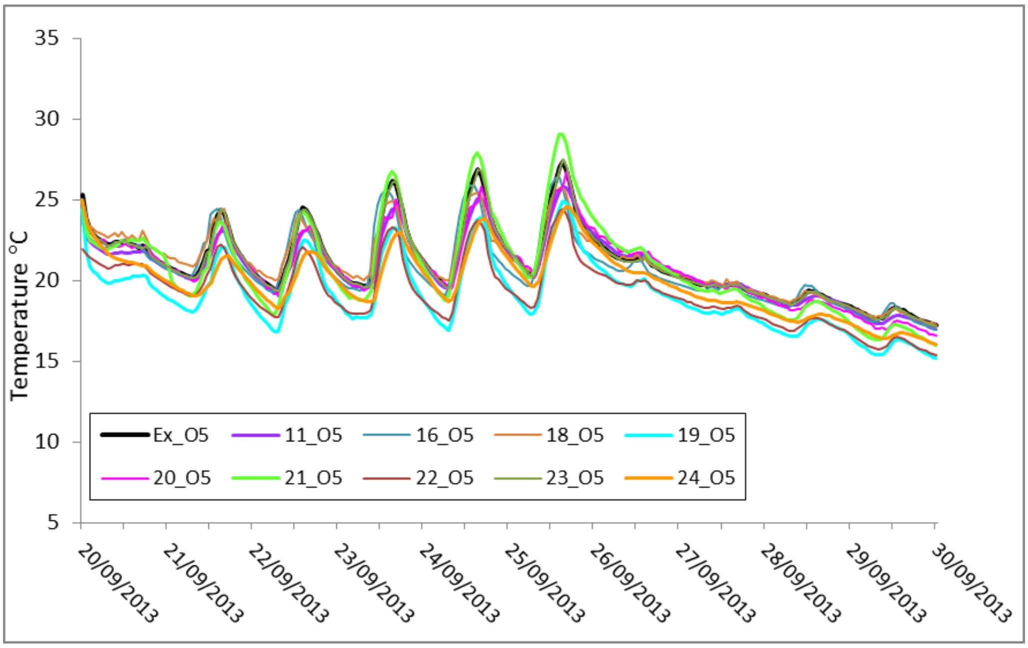

Figure 24: Experiment 1 Re-modelling. Living room temperature: Free-float period: another 9 models + experimental data: House O5

Table 9: Experiment 1. Re-modelling results for the ROLBS sequence and free-floating periods: Temperature magnitude fit

Magnitude Fit	**Average absolute difference in temperature**																		
Fixed heating periods																			
	Period	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6	Sim 7	Sim 8	Sim 10	Sim 11	Sim 16	Sim 18	Sim 19	Sim 20	Sim 21	Sim 22	Sim 23	Sim 24
O5 LRT	ROLBS	1.0	0.4	1.1	1.3	0.8	2.2	0.4	3.4	0.8	0.3	0.8	0.6	1.8	0.4	1.1	1.3	0.1	0.9
O5 LRT	Free	0.4	0.4	1.0	1.8	0.7	1.4	0.3	3.1	1.3	0.3	0.6	0.4	2.0	0.4	0.7	1.8	0.1	1.2
N2 LRT	ROLBS	0.7	0.6	1.1	0.9	0.6	0.6	0.6	2.7	0.8	0.4	0.4	0.5	1.8	0.6	1.0	0.4	0.1	0.5
N2 LRT	Free	1.1	0.6	0.8	1.0	0.5	0.2	0.5	2.1	1.2	0.5	0.4	0.4	1.8	0.9	1.1	0.4	0.0	0.7
N2-O5 LRT	ROLBS	1.6	0.4	0.2	0.6	0.5	2.6	0.3	1.0	0.2	0.5	0.6	0.3	0.4	0.7	1.2	1.3	0.1	0.5
N2-O5 LRT	Free	1.1	0.4	0.3	0.8	0.7	1.2	0.3	1.3	0.2	0.5	0.6	0.3	0.4	0.7	0.8	1.5	0.1	0.5
O5 SBDT	ROLBS	2.2	0.4	0.9	1.2	0.8	2.6	0.5	8.7	0.4	0.3	0.3	1.4	1.8	1.5	0.5	1.0	0.0	0.9
O5 SBDT	Free	1.4	0.3	1.0	1.2	0.8	2.0	0.4	5.8	0.3	0.2	0.3	1.0	1.5	1.5	0.5	1.2	0.0	1.2
N2 SBDT	ROLBS	0.9	0.2	0.9	0.9	0.7	0.6	0.5	5.8	0.3	0.2	0.6	0.6	1.6	0.6	0.5	0.6	0.0	0.8
N2 SBDT	Free	1.1	0.3	1.1	0.7	1.0	0.9	0.5	3.5	0.6	0.3	0.6	0.5	1.1	1.2	0.6	0.7	0.0	1.0
N2-O5 SBDT	ROLBS	1.9	0.5	0.4	0.3	0.4	2.5	0.4	2.9	0.5	0.4	0.6	1.3	0.3	1.4	1.0	1.6	0.0	0.3
N2-O5 SBDT	Free	0.7	0.4	0.2	0.6	0.5	1.1	0.3	2.3	0.4	0.4	0.6	1.1	0.4	0.8	1.0	1.8	0.0	0.3
O5 KITT	ROLBS	0.8	0.3	1.0	2.0	0.7	0.8	0.4	2.4	0.6	0.5	0.7	0.4	2.9	0.8	1.3	1.0	0.2	0.5
O5 KITT	Free	0.5	0.2	1.0	1.5	0.5	1.1	0.3	1.6	0.8	0.2	0.5	0.4	3.6	0.7	0.8	1.5	0.1	0.6
N2 KITT	ROLBS	0.5	0.5	1.1	2.1	0.6	0.6	0.3	2.5	0.6	0.4	0.7	0.4	1.4	0.5	1.7	0.5	0.0	0.5
N2 KITT	Free	0.3	0.4	1.0	1.4	0.5	1.0	0.3	1.6	0.9	0.2	0.5	0.4	1.3	0.7	0.7	0.3	0.0	0.5
N2-O5 KITT	ROLBS	1.2	0.3	0.1	0.1	0.3	0.7	0.4	0.4	0.4	0.1	0.3	0.1	1.5	0.7	0.6	1.4	0.2	0.3
N2-O5 KITT	Free	0.7	0.2	0.1	0.2	0.3	0.2	0.3	0.5	0.2	0.2	0.2	0.1	2.2	0.4	0.2	1.6	0.2	0.3
O5 NBDT	ROLBS	1.2	0.1	1.2	1.3	0.5	1.1	0.4	5.2	0.5	0.3	0.4	2.1	0.2	0.4	0.6	0.9	0.0	0.7
O5 NBDT	Free	1.0	0.1	1.0	1.2	0.5	1.2	0.4	3.5	0.2	0.1	0.3	1.0	0.5	0.3	0.4	1.2	0.0	0.4
N2 NBDT	ROLBS	0.7	0.2	1.1	0.9	0.8	0.7	0.7	4.5	0.6	0.2	1.5	0.9	0.5	0.3	0.3	1.1	0.0	0.7
N2 NBDT	Free	0.7	0.2	1.0	0.7	1.0	1.2	0.7	2.9	0.2	0.3	0.6	0.5	0.5	0.5	0.3	0.8	0.0	0.4
N2-O5 NBDT	ROLBS	0.6	0.1	0.1	0.4	0.5	0.5	0.3	0.8	0.1	0.4	1.1	1.4	0.4	0.5	0.3	2.0	0.0	0.1
N2-O5 NBDT	Free	0.3	0.1	0.1	0.5	0.4	0.2	0.3	0.7	0.1	0.3	0.4	1.3	0.2	0.4	0.2	2.0	0.0	0.2
Fixed temperature periods																			
O5 LRT	30°C	0.3	0.4	0.4	0.4	0.4	0.4	0.3	0.3	0.4	0.4	0.5	0.4	0.9	0.3	0.4	0.4	0.3	0.4
O5 LRT	25°C	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.3	0.2	0.6	0.2	0.2	0.2	0.2	0.2
N2 LRT	30°C	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.7	0.2	0.3	0.2	0.2	0.2	0.2	0.2
N2 LRT	25°C	0.2	0.2	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.7	0.2	0.3	0.2	0.2	0.2	0.2	0.2

Green = <1°C   Yellow = 1><2°C   Orange = 2<>4°C   Red = 4<>8°C   Purple => 8°C

Table 10: Experiment 1. Re-modelling results for the ROLBS sequence and free-floating periods: Temperature shape fit

Shape Fit — **Spearman's Rank Correlation with Experiment**

Fixed heating periods

	Period	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6	Sim 7	Sim 8	Sim 10	Sim 11	Sim 16	Sim 18	Sim 19	Sim 20	Sim 21	Sim 22	Sim 23	Sim 24
O5 LRT	ROLBS	0.98	0.99	0.98	0.98	0.97	0.95	0.99	0.94	0.98	0.99	0.97	0.98	0.99	0.99	0.98	0.99	1.00	0.98
O5 LRT	Free	0.97	0.99	0.97	0.99	0.94	0.93	0.99	0.98	0.99	0.99	0.95	0.98	0.98	0.98	0.98	0.98	1.00	0.97
N2 LRT	ROLBS	0.99	0.99	0.98	0.98	0.95	0.97	0.98	0.90	0.99	1.00	0.99	0.98	0.98	0.99	0.98	0.98	1.00	0.99
N2 LRT	Free	0.99	0.99	0.99	0.97	0.97	0.99	0.99	0.85	0.99	1.00	0.99	0.97	0.93	0.99	0.94	0.99	1.00	0.96
O5 SBDT	ROLBS	0.96	0.99	0.98	0.97	0.95	0.94	0.99	0.89	0.99	1.00	0.98	0.84	0.99	0.89	0.98	0.94	1.00	0.98
O5 SBDT	Free	0.98	0.98	0.97	0.92	0.91	0.95	0.98	0.96	0.99	1.00	0.96	0.79	0.99	0.79	0.97	0.94	1.00	0.98
N2 SBDT	ROLBS	0.83	0.99	0.99	0.97	0.89	0.86	0.91	0.80	0.98	0.99	0.90	0.88	0.97	0.98	0.99	0.93	1.00	0.97
N2 SBDT	Free	0.97	0.99	0.98	0.99	0.98	0.98	0.98	0.74	0.99	1.00	0.98	0.98	0.98	0.98	0.97	0.99	1.00	0.91
O5 KITT	ROLBS	0.99	0.99	0.94	0.95	0.88	0.94	0.96	0.92	0.98	0.98	0.94	0.98	0.99	0.95	0.98	0.98	0.99	0.98
O5 KITT	Free	0.98	0.98	0.92	0.94	0.92	0.98	0.96	0.95	0.98	0.99	0.94	0.98	0.98	0.97	0.82	0.99	0.99	0.94
N2 KITT	ROLBS	0.99	0.98	0.94	0.96	0.91	0.90	0.99	0.86	0.98	0.98	0.95	0.98	0.97	0.99	0.94	0.98	1.00	0.97
N2 KITT	Free	0.99	0.99	0.96	0.98	0.97	0.99	0.98	0.88	0.99	1.00	0.96	0.98	0.97	0.99	0.92	0.98	1.00	0.93
O5 NBDT	ROLBS	0.96	0.99	0.96	0.97	0.91	0.93	0.97	0.79	0.98	0.95	0.97	0.83	1.00	0.91	0.98	0.96	1.00	0.92
O5 NBDT	Free	1.00	1.00	0.96	1.00	0.99	1.00	1.00	0.82	0.99	1.00	0.99	0.78	0.99	0.99	0.96	1.00	1.00	0.88
N2 NBDT	ROLBS	0.98	0.99	0.95	1.00	0.98	0.97	0.99	0.65	0.98	0.99	0.85	0.90	0.99	0.98	1.00	1.00	1.00	0.93
N2 NBDT	Free	1.00	1.00	0.98	1.00	1.00	1.00	1.00	0.74	1.00	1.00	0.93	0.99	0.99	1.00	0.99	1.00	1.00	0.96

Green = >0.90 **Yellow = 0.80<>0.90** **Orange = 0.70<>0.80** **Red = 0.35<>0.70** **Purple = <0.35**

Table 11: Experiment 1. Re-modelling results for the constant temperature periods: Heat input magnitude fit

Magnitude Fit — **Average absolute difference in heat input** — *Sim3 was a 2-zone model; Sim 24 modelled all S rooms as one zone*

Constant temperature periods

	Period	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6	Sim 7	Sim 8	Sim 10	Sim 11	Sim 16	Sim 18	Sim 19	Sim 20	Sim 21	Sim 22	Sim 23	Sim 24
O5 LRQ	30°C	253	108		137	122	181	46	240	130	130	141	267	189	265	148	122	177	
O5 LRQ	25°C	341	54		55	94	106	69	391	78	27	104	257	137	129	64	234	267	
N2 LRQ	30°C	121	74		76	110	170	76	297	64	49	91	286	222	342	112	39	96	
N2 LRQ	25°C	159	72		106	192	227	61	452	96	35	102	267	207	463	134	90	55	
N2-O5 LRQ	30°C	174	130		122	134	241	99	169	122	149	100	168	191	410	169	139	158	
N2-O5 LRQ	25°C	185	38		100	107	252	27	74	43	34	37	43	91	466	132	147	246	
O5 SBDQ	30°C	128	70		82	78	100	93	411	113	46	89	91	42	117	117	45	90	
O5 SBDQ	25°C	214	81		70	95	133	118	249	120	54	134	125	40	106	150	50	193	
N2 SBDQ	30°C	143	87		110	106	75	149	387	150	63	150	154	0	905	145	121	104	
N2 SBDQ	25°C	176	103		120	75	54	154	209	152	79	159	162	9	1026	171	85	135	
N2-O5 SBDQ	30°C	35	29		42	37	50	56	42	41	39	73	82	42	1005	44	85	35	
N2-O5 SBDQ	25°C	42	37		56	52	95	46	46	44	37	40	47	46	1132	38	117	59	
O5 KITQ	30°C	25	26		27	54	48	59	58	23	22	40	23	31	23	70	23	28	21
O5 KITQ	25°C	45	25		23	66	30	43	84	26	27	33	26	47	33	62	60	25	29
N2 KITQ	30°C	26	23		30	52	49	43	48	27	20	42	30	12	140	82	25	25	23
N2 KITQ	25°C	37	24		35	67	53	44	98	35	30	34	38	10	133	82	21	21	45
N2-O5 KITQ	30°C	5	4		10	6	7	17	16	6	7	5	9	30	125	14	33	7	10
N2-O5 KITQ	25°C	14	9		31	13	34	7	16	13	16	8	15	39	148	24	64	7	18
O5 NBDQ	30°C	54	15		10	22	22	31	30	64	12	61	35	24	64	35	24	104	44
O5 NBDQ	25°C	91	12		31	42	22	20	27	51	14	36	35	13	30	46	68	142	45
N2 NBDQ	30°C	82	24		32	34	8	59	31	91	31	62	14	0	69	60	83	128	78
N2 NBDQ	25°C	102	22		25	35	40	48	51	81	27	42	29	3	69	74	47	144	90
N2-O5 NBDQ	30°C	29	26		23	21	22	28	6	27	26	20	32	24	17	26	66	24	33
N2-O5 NBDQ	25°C	24	25		46	17	39	32	36	32	29	15	31	15	66	32	109	20	46

Fixed heating periods — *Where difference is 240W in ROLBS period, ROLBS heat input was modelled as casual gain*

	Period	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6	Sim 7	Sim 8	Sim 10	Sim 11	Sim 16	Sim 18	Sim 19	Sim 20	Sim 21	Sim 22	Sim 23	Sim 24
O5 LRQ	ROLBS	242	58		242	58	0	242	102	65	242	70	70	0	146	70	71	13	242
O5 LRQ	Free	0	0		0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
N2 LRQ	ROLBS	240	67		241	70	0	237	111	75	237	80	74	1	1	238	75	13	238
N2 LRQ	Free	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Green = <100 W **Yellow = 100><200W** **Orange = 200<>300 W** **Red = 300<>500 W** **Purple = >500W**

Table 12: Experiment 1. Re-modelling results for the constant temperature periods: Heat input shape fit

Shape Fit		**Spearman's Rank Correlation with Experiment**						*Sim3 was a 2-zone model; Sim 24 modelled all S rooms as one zone*											
Constant temperature periods																			
	Period	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6	Sim 7	Sim 8	Sim 10	Sim 11	Sim 16	Sim 18	Sim 19	Sim 20	Sim 21	Sim 22	Sim 23	Sim 24
O5 SBDQ	30 C	0.872	0.841		0.936	0.802	0.825	0.965	0.855	0.944	0.938	0.792	0.870	0.945	0.713	0.945	0.913	0.786	
O5 SBDQ	25 C	0.208	0.579		0.517	0.189	-0.276	0.525	0.290	0.389	0.497	0.161	0.264	0.433	0.367	0.386	0.486	0.343	
N2 SBDQ	30 C	-0.404	0.222		0.423	0.272	0.324	0.588	0.412	0.390	0.404	0.032	0.166	1.000	0.320	0.364	0.532	-0.141	
N2 SBDQ	25 C	0.839	0.816		0.687	0.751	0.732	0.720	0.379	0.649	0.576	0.336	0.355	1.000	0.440	0.633	0.707	0.737	
O5 SBDQ	30 C	0.872	0.841		0.936	0.802	0.825	0.965	0.855	0.944	0.938	0.792	0.870	0.945	0.713	0.945	0.913	0.786	
O5 SBDQ	25 C	0.208	0.579		0.517	0.189	-0.276	0.525	0.290	0.389	0.497	0.161	0.264	0.433	0.367	0.386	0.486	0.343	
N2 SBDQ	30 C	-0.404	0.222		0.423	0.272	0.324	0.588	0.412	0.390	0.404	0.032	0.166	1.000	0.320	0.364	0.532	-0.141	
N2 SBDQ	25 C	0.839	0.816		0.687	0.751	0.732	0.720	0.379	0.649	0.576	0.336	0.355	1.000	0.440	0.633	0.707	0.737	
O5 KITQ	30 C	0.939	0.933		0.911	0.875	0.803	0.981	0.853	0.957	0.941	0.892	0.981	0.961	0.974	0.894	0.965	0.915	0.940
O5 KITQ	25 C	0.944	0.952		0.851	0.913	0.932	0.910	0.738	0.919	0.911	0.842	0.941	0.840	0.847	0.805	0.918	0.891	0.891
N2 KITQ	30 C	0.922	0.942		0.903	0.878	0.868	0.974	0.850	0.955	0.942	0.887	0.985	0.985	0.900	0.885	0.978	0.916	0.945
N2 KITQ	25 C	0.957	0.958		0.877	0.927	0.905	0.935	0.731	0.938	0.930	0.855	0.947	0.985	0.896	0.827	0.939	0.936	0.903
O5 NBDQ	30 C	0.628	0.775		0.948	0.693	0.922	0.989	0.704	0.973	0.934	0.889	0.940	0.957	0.897	0.970	0.832	0.779	0.866
O5 NBDQ	25 C	0.958	0.889		0.913	0.635	0.945	0.977	0.647	0.939	0.882	0.843	0.912	0.940	0.754	0.968	0.869	0.841	0.904
N2 NBDQ	30 C	0.629	0.812		0.935	0.686	0.935	0.984	0.685	0.965	0.936	0.866	0.918	1.000	0.920	0.969	0.920	0.746	0.899
N2 NBDQ	25 C	0.954	0.932		0.962	0.756	0.877	0.984	0.556	0.955	0.874	0.891	0.875	1.000	0.906	0.970	0.952	0.943	0.850
			Green = >0.90			**Yellow = 0.80<>0.90**			**Orange = 0.70<>0.80**			**Red = 0.35<>0.70**			**Purple = <0.35**				

Table 13 plots the cumulative heat input for the experiment. This table shows the total heating energy for the constant temperature heating periods: Period 2 (30 °C) and Period 4 (25 °C) for the combinations of the main rooms: south zone (living room and south bedroom) and north zone (kitchen and north bedroom). The percentage difference between the measured and predicted total heat input is also shown. For the case of the differences between the two houses (O5 and N2), this percentage can be large, but absolute differences can be small.

Figure 25 is a plot of the data as a 6-hour moving average across the whole experiment, in this case the living room absolute temperature difference between measurement and predictions for House O5. This was found useful for getting an overview of the results by removing short period fluctuations and transients during the change between periods. As can be seen, simulation 2 shows good agreement throughout, whereas simulation 3 shows diurnal variations corresponding with solar radiation levels. Simulation 23 shows almost perfect agreement – it is suspected that the measured data was used as input in this case, so results are treated with caution (there was no blind validation for this program).

Table 13: Total heating inputs for Experiment 1 (re-modelling phase)

Total heating input to South (living room, south bedroom) and North (kitchen and north bedroom)

Constant temperature periods

	Period	Experiment	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6	Sim 7	Sim 8	Sim 10	Sim 11	Sim 16	Sim 18	Sim 19	Sim 20	Sim 21	Sim 22	Sim 23	Sim 24
O5 S zone (kWh)	2 (30 C)	138.9	101.0	134.5	150.6	120.0	139.8	119.5	127.2	176.5	120.0	138.9	111.4	97.0	165.4	137.3	118.8	149.0	119.6	151.8
O5 S Zone Difference with Measured (%)			-27%	-3%	8%	-14%	1%	-14%	-8%	27%	-14%	0%	-20%	-30%	19%	-1%	-14%	7%	-14%	9%
N2 S zone (kWh)		193.4	156.1	190.2	200.8	165.3	191.8	209.5	158.4	207.7	161.2	190.4	155.9	120.1	230.8	401.8	158.1	170.9	178.3	194.8
N2 S Zone Difference with Measured (%)			-19%	-2%	4%	-15%	-1%	8%	-18%	7%	-17%	-2%	-19%	-38%	19%	108%	-18%	-12%	-8%	1%
N2-O5 (kWh)		54.5	55.1	55.7	50.2	45.3	52.1	89.9	31.2	31.2	41.2	51.5	44.5	23.1	65.4	264.5	39.4	21.9	58.7	42.9
N2-O5 Difference with Measured (%)			1%	2%	-8%	-17%	-5%	65%	-43%	-43%	-24%	-6%	-18%	-58%	20%	385%	-28%	-60%	8%	-21%
O5 N zone (kWh)		60.8	50.1	65.6	0.0	56.2	69.1	71.6	45.8	52.6	46.9	62.5	50.4	53.8	59.0	47.5	43.9	59.6	46.0	52.3
O5 N Zone Difference with Measured (%)			-18%	8%	-100%	-8%	14%	18%	-25%	-13%	-23%	3%	-17%	-11%	-3%	-22%	-28%	-2%	-24%	-14%
N2 N zone (kWh)		69.7	53.7	69.8	92.1	60.7	73.9	77.2	52.9	63.3	50.5	66.1	58.7	64.8	71.8	34.6	46.4	52.1	50.0	54.1
N2 N Zone Difference with Measured (%)			-23%	0%	32%	-13%	6%	11%	-24%	-9%	-28%	-5%	-16%	-7%	3%	-50%	-33%	-25%	-28%	-22%
N2-O5 (kWh)		9.0	3.6	4.2	92.1	4.6	4.8	5.6	7.1	10.7	3.6	3.6	8.3	11.0	12.8	-12.8	2.4	-7.5	4.0	1.8
N2-O5 Difference with Measured (%)			-60%	-54%	928%	-49%	-46%	-38%	-21%	20%	-60%	-60%	-7%	23%	43%	-243%	-73%	-183%	-55%	-80%
O5 S zone (kWh)	4 (25 C)	159.1	79.2	151.2	171.8	151.8	155.3	135.7	140.2	137.7	132.7	151.8	126.3	104.1	173.7	143.5	133.4	195.1	92.9	194.0
O5 S Zone Difference with Measured (%)			-50%	-5%	8%	-5%	-2%	-15%	-12%	-13%	-17%	-5%	-21%	-35%	9%	-10%	-16%	23%	-42%	22%
N2 S zone (kWh)		182.6	134.8	179.4	192.0	160.5	199.5	208.6	164.6	145.9	150.1	180.4	146.4	122.1	213.6	397.0	146.4	181.1	160.0	215.6
N2 S Zone Difference with Measured (%)			-26%	-2%	5%	-12%	9%	14%	-10%	-20%	-18%	-1%	-20%	-33%	17%	117%	-20%	-1%	-12%	18%
N2-O5 (kWh)		23.5	55.6	28.2	20.3	8.7	44.2	72.9	24.4	8.2	17.4	28.6	20.0	18.0	39.9	253.5	13.0	-14.0	67.0	21.6
N2-O5 Difference with Measured (%)			136%	20%	-14%	-63%	88%	210%	4%	-65%	-26%	22%	-15%	-23%	70%	979%	-45%	-160%	185%	-8%
O5 N zone (kWh)		58.8	39.4	62.4	0.0	63.6	71.2	61.5	55.0	45.5	48.4	62.5	52.9	54.9	64.6	55.8	43.4	75.6	36.4	49.2
O5 N Zone Difference with Measured (%)			-33%	6%	-100%	8%	21%	5%	-6%	-23%	-18%	6%	-10%	-7%	10%	-5%	-26%	28%	-38%	-16%
N2 N zone (kWh)		73.4	53.9	74.1	100.8	67.9	84.4	85.0	66.0	53.1	56.9	72.8	65.7	70.3	75.3	45.6	51.1	65.2	50.9	54.9
N2 N Zone Difference with Measured (%)			-27%	1%	37%	-7%	15%	16%	-10%	-28%	-22%	-1%	-10%	-4%	3%	-38%	-30%	-11%	-31%	-25%
N2-O5 (kWh)		14.6	14.5	11.7	100.8	4.4	13.1	23.5	10.9	7.5	8.5	10.3	12.8	15.5	10.7	-10.1	7.7	-10.4	14.5	5.7
N2-O5 Difference with Measured (%)			0%	-20%	592%	-70%	-10%	61%	-25%	-48%	-41%	-29%	-12%	6%	-27%	-170%	-47%	-171%	0%	-61%
O5 Total (kWh)	2 (30 C)	199.7	151.1	200.1	150.6	176.2	208.9	191.2	173.0	229.1	166.9	201.5	161.8	150.8	224.4	184.8	162.7	208.6	165.6	204.1
O5 Total Difference with Measured (%)			-24%	0%	-25%	-12%	5%	-4%	-13%	15%	-16%	1%	-19%	-24%	12%	-7%	-19%	4%	-17%	2%
N2 Total (kWh)		263.2	209.8	260.0	292.9	226.1	265.8	286.7	211.3	271.0	211.7	256.5	214.6	184.9	302.5	436.4	204.5	223.0	228.4	248.8
N2 Total Difference with Measured (%)			-20%	-1%	11%	-14%	1%	9%	-20%	3%	-20%	-3%	-18%	-30%	15%	66%	-22%	-15%	-13%	-5%
N2-O5 (kWh)		63.5	58.7	59.8	142.3	49.9	56.9	95.5	38.3	41.9	44.8	55.0	52.8	34.1	78.2	251.6	41.8	14.4	62.7	44.7
N2-O5 Difference with Measured (%)			-8%	-6%	124%	-21%	-10%	50%	-40%	-34%	-29%	-13%	-17%	-46%	23%	296%	-34%	-77%	-1%	-30%
O5 Total (kWh)	4 (25 C)	217.9	118.6	213.6	171.8	215.3	226.6	197.2	195.3	183.2	181.1	214.3	179.2	159.0	238.3	199.3	176.8	270.7	129.3	243.2
O5 Total Difference with Measured (%)			-46%	-2%	-21%	-1%	4%	-10%	-10%	-16%	-17%	-2%	-18%	-27%	9%	-9%	-19%	24%	-41%	12%
N2 Total (kWh)		256.0	188.7	253.5	292.8	228.5	283.9	293.5	230.6	199.0	207.0	253.2	212.1	192.5	288.9	442.7	197.5	246.3	210.8	270.5
N2 Total Difference with Measured (%)			-26%	-1%	14%	-11%	11%	15%	-10%	-22%	-19%	-1%	-17%	-25%	13%	73%	-23%	-4%	-18%	6%
N2-O5 (kWh)		38.1	70.1	39.9	121.0	13.1	57.3	96.3	35.3	15.7	25.9	38.9	32.8	33.5	50.6	243.4	20.7	-24.4	81.6	27.3
N2-O5 Difference with Measured (%)			84%	5%	218%	-66%	51%	153%	-7%	-59%	-32%	2%	-14%	-12%	33%	539%	-46%	-164%	114%	-28%

Green < +/-5%   Yellow < +/-10%   Orange < +/-20%   Red < +/-40%   Purple > +/-40%

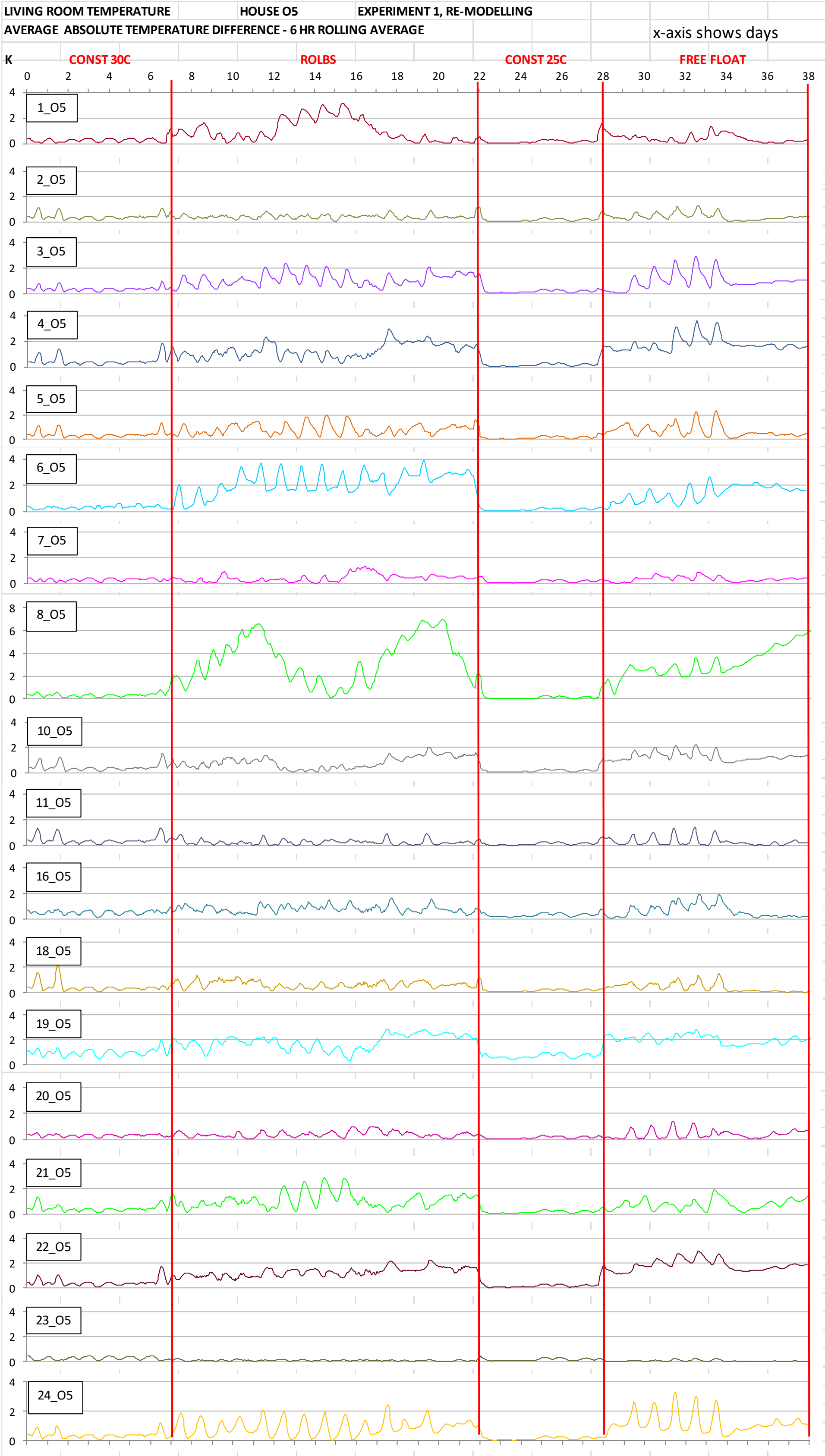

Figure 25: 6-hour moving average of living room absolute temperature difference (Experiment 1 Re-modelling, house O5)

5.4 EXPERIMENT 2

Some example graphs are presented, together with summary tables as for Experiment 1. In this case, 13 datasets were submitted by modelling teams.

Figures 26 and 27 show the measured and predicted living room temperatures for the constant temperature heating period, Figures 28 and 29 show the living room temperatures for the ROLBS sequence, and Figures 30 and 31 show the living room temperatures for the free-float period.

Tables 14 to 17 show the summary results for Experiment 2, using the same metrics and banding as for Experiment 1, except this time there is only one house tested, so the differences between houses are not included. Table 18 shows the total heat input to all ground floor rooms, separated into a south zone (living room, south bedroom, bathroom and corridor) and north zone (kitchen, north bedroom, lobby). It should be noted that a few teams used the provided measured data for the northern rooms for Experiment 2 as a boundary condition, so there is no error for these rooms (e.g. result sets 11 and16). Labelling is as for the Experiment 1 results, except that the corridor temperature and heat inputs (CORT and CORQ respectively), lobby temperature and heat inputs (LOBT and LOBQ respectively), and the bathroom temperature and heat inputs (BATT and BATQ respectively) are included.

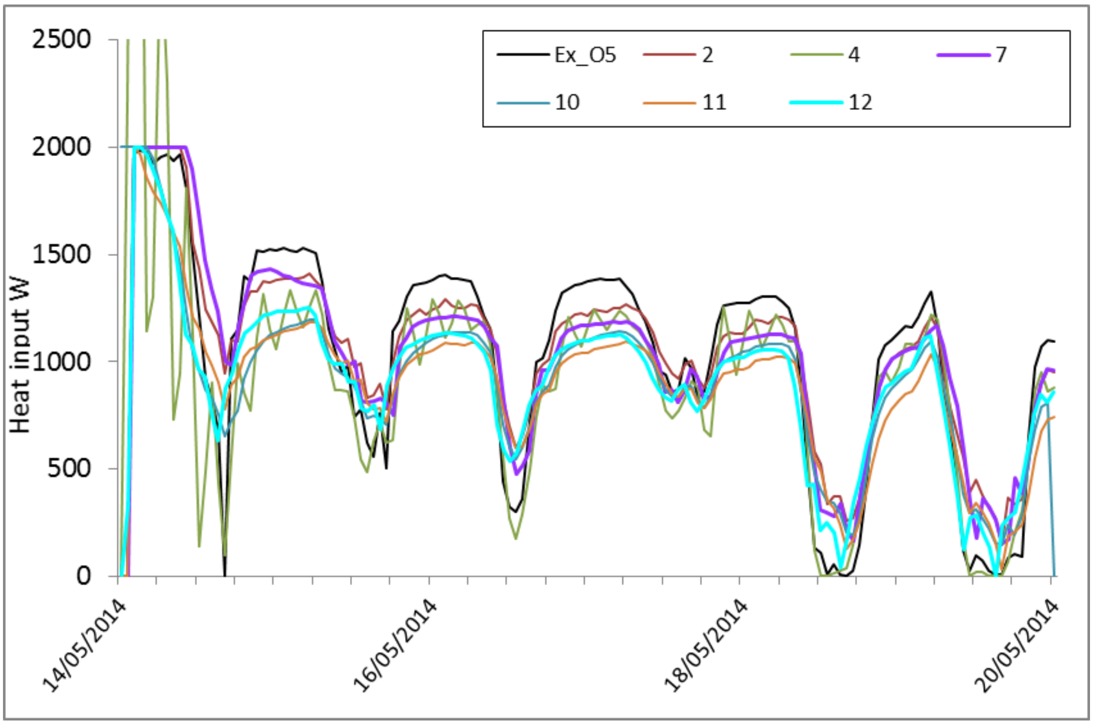

Figure 26: Experiment 2: Living room heat input: constant temperature phase (30°C): 6 models + experimental data: House O5

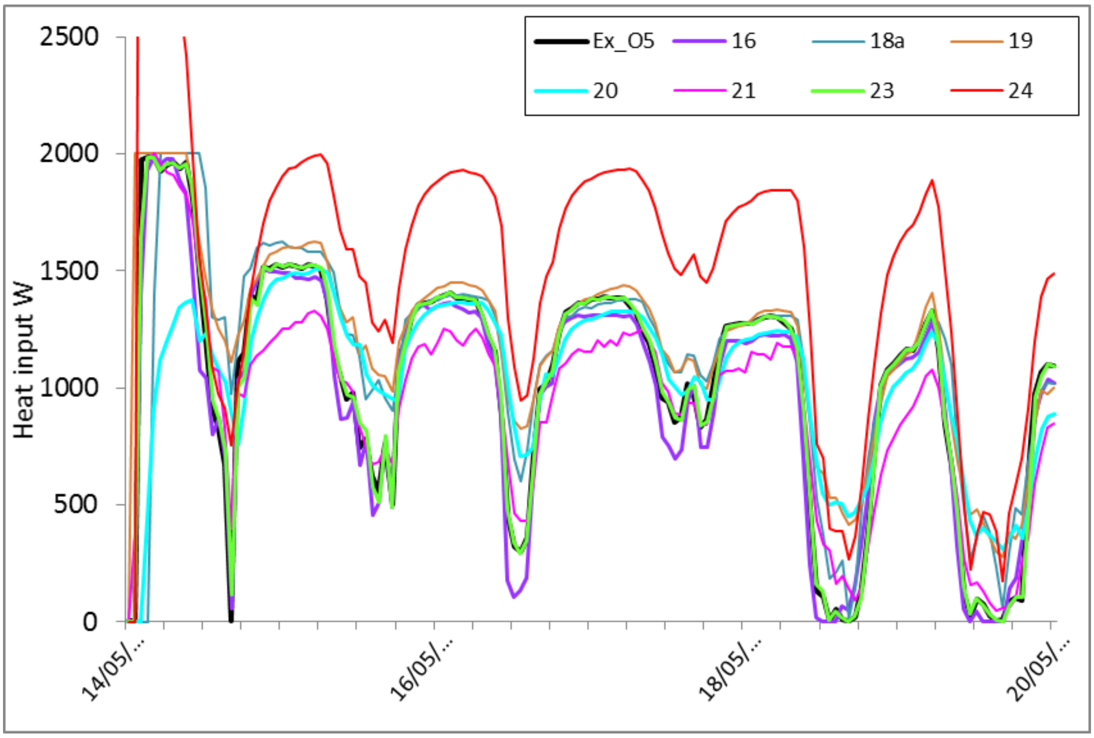

Figure 27: Experiment 2: Living room heat input: constant temperature phase (30°C): the other 7 models + experimental data: House O5

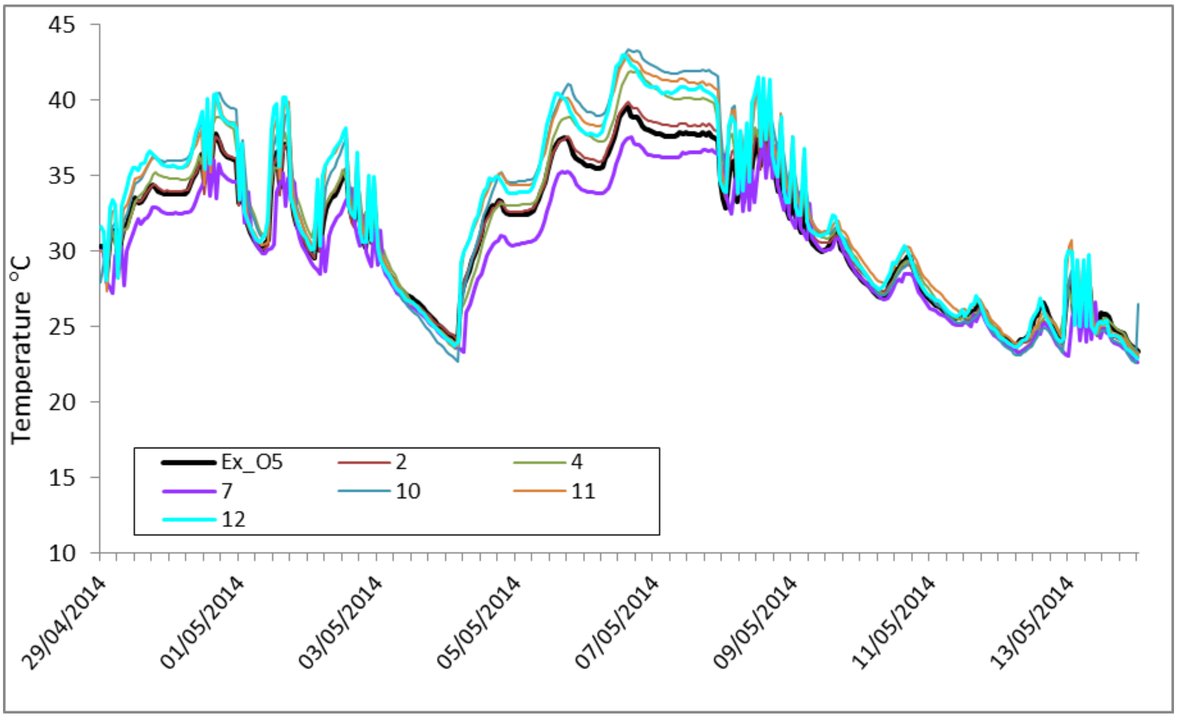

Figure 28: Experiment 2: Living room temperature: ROLBS sequence: 6 models + experimental data: House O5

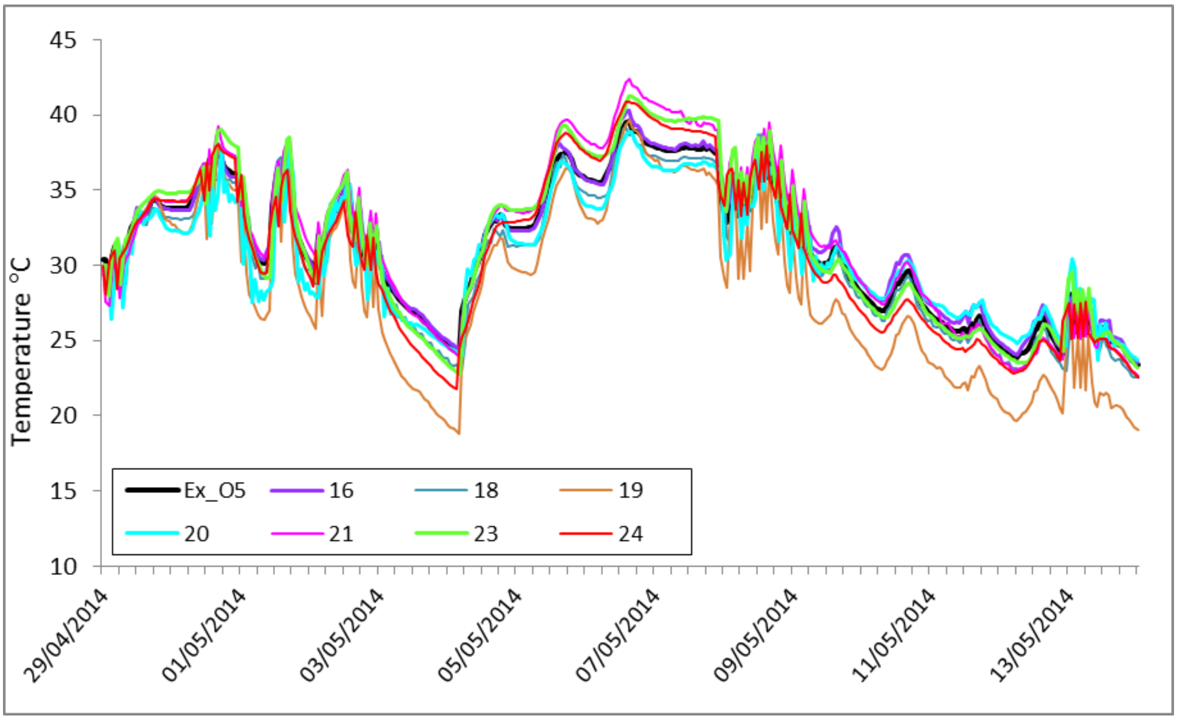

Figure 29: Experiment 2: Living room temperature: ROLBS sequence: the other 7 models + experimental data: House O5

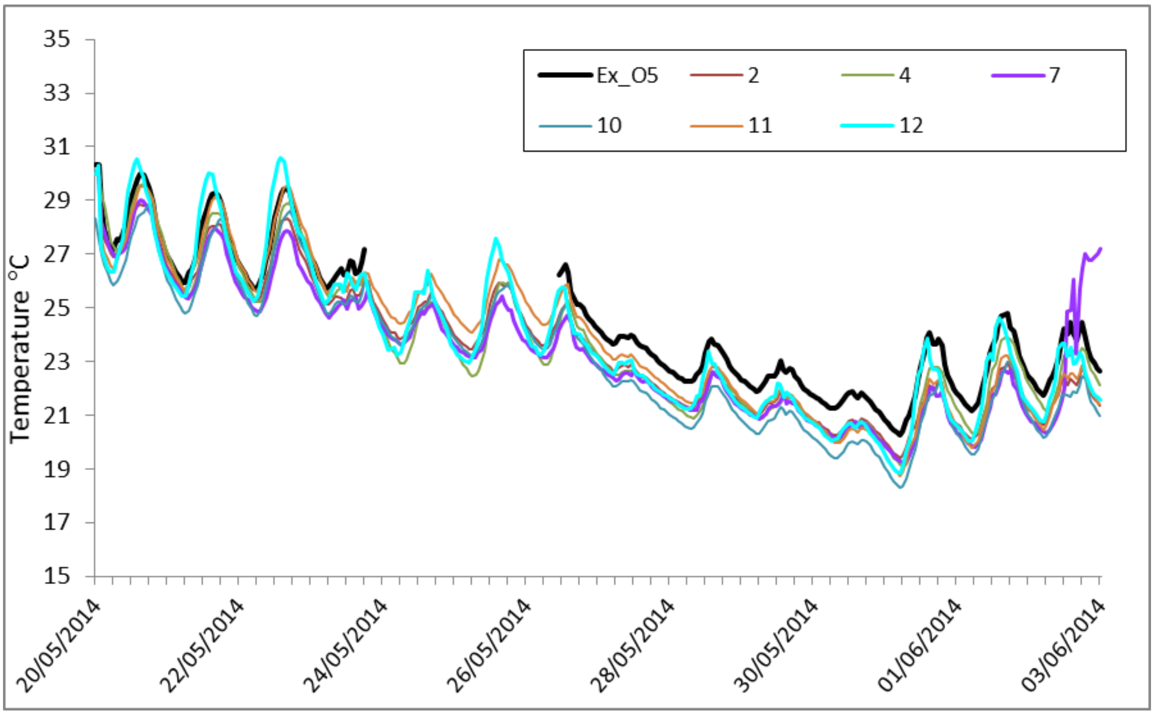

Figure 30: Experiment 2: Living room temperature: Free-float period: 6 models + experimental data: House O5

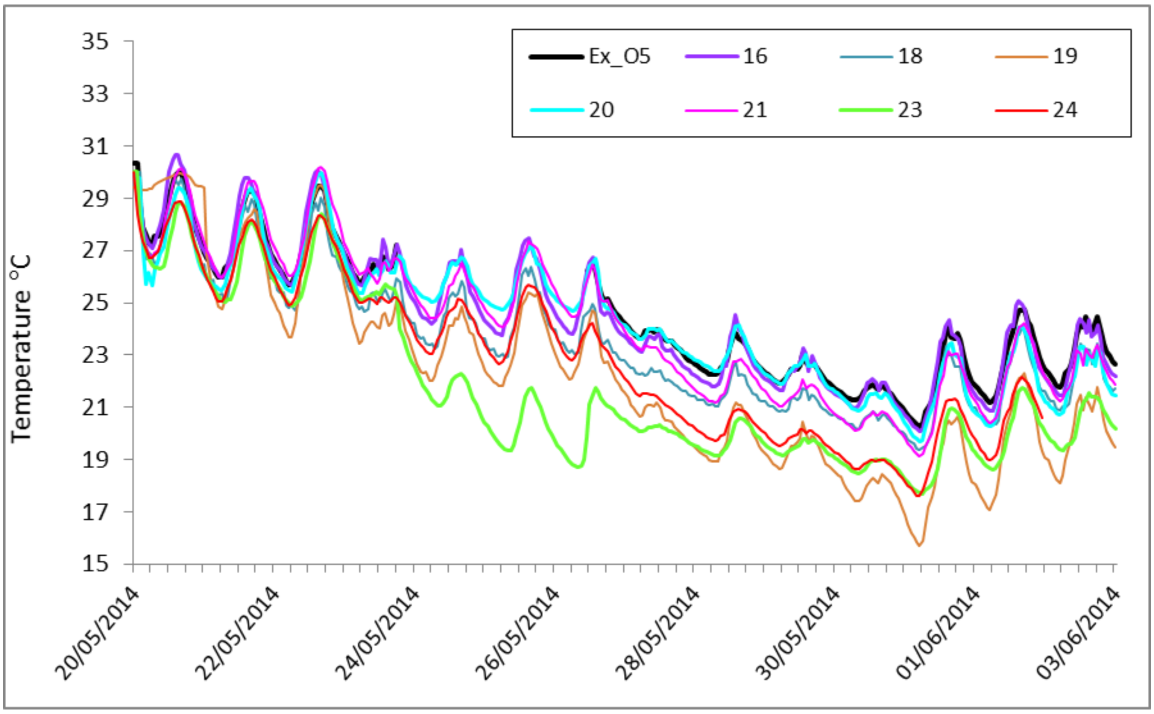

Figure 31: Experiment 2: Living room temperature: Free-float period: the other 7 models + experimental data: House O5

Detailed plots are useful for checking whether there are offsets in solar or internal heat injections. As can be seen in Figure 32 on the left, most programs follow the heat injections correctly but simulation 24 is out of phase by 1 hour. In Figure 32 on the right, temperature peaks are aligned with the measured data for most simulations, but program 12 in particular has a time-shift in the maximum temperature compared to the measured data.

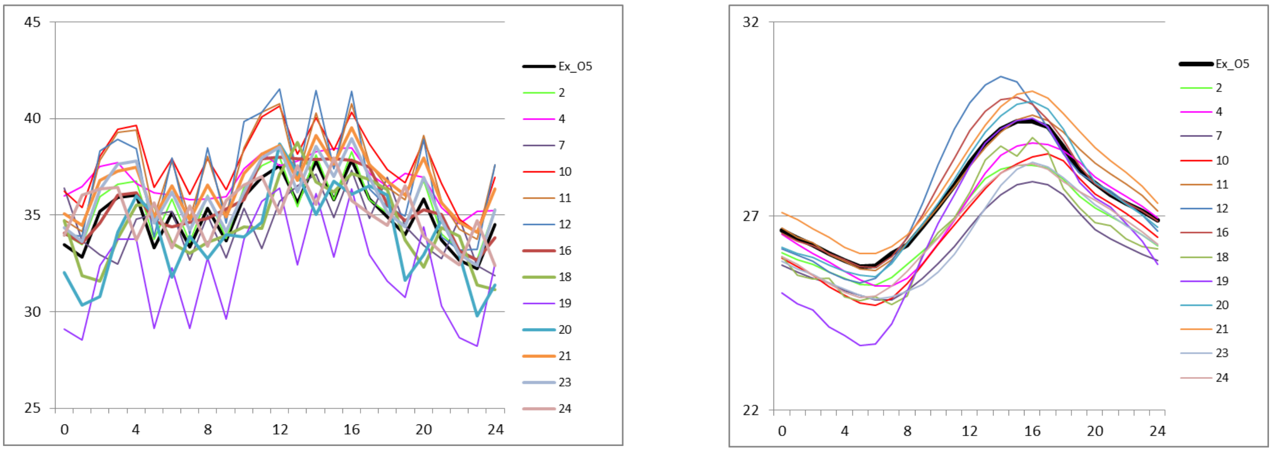

Figure 32: Experiment 2: plots of individual days: ROLBS on left and free-float on right

Table 14: Experiment 2: ROLBS sequence and free-floating periods: temperature magnitude fit

Magnitude Fit	**Average absolute difference in temperature**													**EXPERIMENT 2**
Fixed heating periods														
	Period	Sim 2	Sim 4	Sim 7	Sim 10	Sim 11	Sim 12	Sim 16	Sim 18	Sim 19	Sim 20	Sim 21	Sim 23	Sim 24
O5 LRT	ROLBS	0.3	1.0	1.3	1.7	1.5	1.4	0.5	1.0	2.7	1.2	1.0	0.8	1.3
O5 LRT	Free	1.0	0.8	1.3	1.5	0.8	0.8	0.3	1.0	2.5	0.4	0.7	2.4	1.9
O5 SBDT	ROLBS	0.9	1.4	1.4	3.1	1.9	3.2	0.7	0.5	1.4	1.4	1.6	1.9	1.3
O5 SBDT	Free	0.4	0.8	0.7	0.7	0.6	0.5	0.3	0.9	1.9	0.7	0.4	1.5	1.8
O5 BATT	ROLBS	0.9	1.4	1.2	3.4	2.3	3.6	0.6	0.7	1.4	1.3	1.4	2.6	1.2
O5 BATT	Free	0.4	1.0	0.7	0.7	0.6	0.7	0.2	1.2	1.6	0.5	0.5	0.9	2.0
O5 CORT	ROLBS	0.4	1.1	2.5	1.4	1.7	1.4	0.5	1.3	2.6	0.8	1.0	1.4	1.3
O5 CORT	Free	0.6	1.0	1.0	0.9	0.7	0.9	0.3	1.1	1.7	0.4	0.6	2.0	1.9
O5 KITT	ROLBS	0.2	0.3	0.8	1.1	0.0	0.7	0.0	0.2	0.5	0.1	0.0	0.3	0.3
O5 KITT	Free	0.1	0.2	0.3	0.1	0.0	0.2	0.0	0.1	0.4	0.2	0.0	0.2	0.2
O5 NBDT	ROLBS	0.2	0.6	0.9	1.2	0.0	0.7	0.0	0.3	0.5	0.3	0.0	0.5	0.5
O5 NBDT	Free	0.1	0.5	0.3	0.1	0.0	0.4	0.0	0.2	0.4	0.2	0.0	0.4	0.4
O5 LOBT	ROLBS	0.2	0.3	0.9	1.6	0.0	0.4	0.0	0.6	0.6	0.3	0.0	0.4	0.4
O5 LOBT	Free	0.1	0.3	0.3	0.2	0.0	0.2	0.0	0.2	0.3	0.2	0.0	0.3	0.2
Fixed temperature periods														
O5 LRT	30°C (1)	0.4	0.3	0.4	0.4	0.4	0.9	0.4	0.4	0.9	0.4	0.3	0.4	0.4
O5 LRT	30°C (2)	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.5	1.0	0.4	0.3	0.3	0.4
	Green = <1°C			**Yellow = 1><2°C**			**Orange = 2<>4°C**			**Red = 4<>8°C**			**Purple => 8°C**	

Table 15: Experiment 2: ROLBS sequence and free-floating periods: temperature shape fit

Shape Fit		**Spearman's Rank Correlation with Experiment**												**EXPERIMENT 2**
Fixed heating periods														
	Period	Sim 2	Sim 4	Sim 7	Sim 10	Sim 11	Sim 12	Sim 16	Sim 18	Sim 19	Sim 20	Sim 21	Sim 23	Sim 24
O5 LRT	ROLBS	1.00	0.99	0.96	0.99	0.99	0.99	0.99	0.98	0.99	0.95	0.99	0.99	0.97
O5 LRT	Free	0.98	0.97	0.95	0.98	0.97	0.97	0.98	0.99	0.98	0.94	0.98	0.95	0.98
O5 SBDT	ROLBS	0.99	0.97	0.94	0.98	0.99	0.99	0.99	0.98	0.99	0.96	0.99	0.99	0.99
O5 SBDT	Free	0.96	0.96	0.89	0.95	0.98	0.97	0.98	0.99	0.97	0.94	0.98	0.97	0.91
O5 BATT	ROLBS	0.99	0.98	0.96	0.98	0.99	0.99	0.99	0.98	0.99	0.96	0.99	0.99	0.98
O5 BATT	Free	0.98	0.93	0.92	0.95	0.98	0.93	0.98	0.95	0.99	0.91	0.95	0.96	0.89
O5 CORT	ROLBS	1.00	0.95	0.85	0.94	0.99	0.99	0.99	0.98	0.99	0.97	0.99	0.95	0.99
O5 CORT	Free	0.98	0.97	0.87	0.95	0.98	0.94	0.97	0.98	0.99	0.92	0.98	0.95	0.95
O5 KITT	ROLBS	0.99	0.96	0.91	0.98	1.00	0.88	1.00	0.98	0.89	0.96	0.99	0.96	0.96
O5 KITT	Free	0.78	0.55	0.53	0.62	0.88	0.57	0.90	0.56	0.60	0.72	0.73	0.50	0.58
O5 NBDT	ROLBS	0.98	0.96	0.96	0.96	1.00	0.97	1.00	0.99	0.97	0.91	1.00	0.93	0.97
O5 NBDT	Free	0.70	0.43	0.63	0.76	0.91	0.64	0.91	0.59	0.78	0.83	0.90	0.43	0.63
O5 LOBT	ROLBS	0.99	0.98	0.97	0.98	1.00	0.95	1.00	0.98	0.90	0.94	1.00	0.95	0.97
O5 LOBT	Free	0.65	0.54	0.56	0.56	0.87	0.59	0.86	0.58	0.60	0.74	0.72	0.38	0.49
	Green = >0.90			**Yellow = 0.80><0.90**			**Orange = 0.70<>0.80**			**Red = 0.35<>0.70**			**Purple = <0.35**	

Table 16: Experiment 2: constant temperature periods: Heat input magnitude fit

Magnitude Fit		**Average absolute difference in heat input**												**EXPERIMENT 2**
Constant temperature periods														
	Period	Sim 2	Sim 4	Sim 7	Sim 10	Sim 11	Sim 12	Sim 16	Sim 18	Sim 19	Sim 20	Sim 21	Sim 23	Sim 24
O5 LRQ	30°C (1)	117	113	160	206	205	18	88	166	190	224	184	16	646
O5 LRQ	30°C (2)	150	252	173	232	238	204	75	183	183	224	151	22	562
O5 SBDQ	30°C (1)	50	41	39	120	54	5	43	44	56	83	97	4	248
O5 SBDQ	30°C (2)	56	86	45	117	53	114	57	64	64	132	96	5	310
O5 BATQ	30°C (1)	48	47	58	83	80	5	44	63	45	89	89	4	203
O5 BATQ	30°C (2)	56	86	45	117	53	114	57	64	64	132	96	5	310
O5 KITQ	30°C (1)	1	8	21	11	29	33	11	10	10	21	38	0	11
O5 KITQ	30°C (2)	6	10	30	9	35	47	26	8	11	16	39	0	5
O5 NBDQ	30°C (1)	0	10	11	11	0	11	11	11	10	13	20	192	11
O5 NBDQ	30°C (2)	0	11	11	11	0	14	11	12	12	7	24	244	11
O5 LOQ	30°C (1)	1	6	7	1	0	4	1	1	1	8	8	0	3
O5 LOQ	30°C (2)	4	11	16	2	0	14	3	2	1	6	10	0	8
Fixed heating periods				Where difference is 860W in ROLBS period, ROLBS heat input was modelled as casual gain										
O5 LRQ	ROLBS	0	859	857	29	858	44	65	239	7	860	30	44	859
O5 LRQ	Free	0	7	5	8	5	1	1	10	57	8	4	1	6
	Green = <100 W			**Yellow = 100><200W**			**Orange = 200<>300 W**			**Red = 300<>500 W**			**Purple = >500W**	

Table 17: Experiment 2: constant temperature periods: Heat input shape fit

Shape Fit		**Spearman's Rank Correlation with Experiment**												**EXPERIMENT 2**
Constant temperature periods														
	Period	Sim 2	Sim 4	Sim 7	Sim 10	Sim 11	Sim 12	Sim 16	Sim 18	Sim 19	Sim 20	Sim 21	Sim 23	Sim 24
O5 LRQ	30°C (1)	0.847	0.876	0.820	0.791	0.800	0.985	0.904	0.821	0.814	0.830	0.854	0.988	0.853
O5 LRQ	30°C (2)	0.951	0.838	0.907	0.907	0.923	0.975	0.982	0.856	0.907	0.794	0.941	0.997	0.923
O5 SBDQ	30°C (1)	0.582	0.655	0.695	0.338	0.644	0.909	0.757	0.683	0.618	0.725	0.627	0.984	0.000
O5 SBDQ	30°C (2)	0.814	0.582	0.821	0.644	0.850	0.761	0.801	0.758	0.810	0.670	0.830	0.986	0.000
O5 BATQ	30°C (1)	0.756	0.705	0.746	0.629	0.750	0.950	0.579	0.456	0.704	0.361	0.332	0.990	0.000
O5 BATQ	30°C (2)	0.852	0.687	0.815	0.749	0.870	0.897	0.841	0.739	0.799	0.589	0.698	0.993	0.000
O5 KITQ	30°C (1)	0.920	0.381	0.558	#DIV/0!	-0.441	0.552	#DIV/0!	0.430	0.384	0.646	0.557	0.994	0.530
O5 KITQ	30°C (2)	0.903	0.876	0.934	0.922	0.124	0.872	0.144	0.962	0.946	0.830	0.925	0.998	0.958
O5 NBDQ	30°C (1)	0.997	0.130	#DIV/0!	#DIV/0!	0.992	0.190	#DIV/0!	#DIV/0!	0.231	0.404	0.450	0.386	#DIV/0!
O5 NBDQ	30°C (2)	1.000	-0.004	0.095	#DIV/0!	0.994	-0.040	-0.005	0.021	-0.040	0.163	-0.176	0.125	#DIV/0!
O5 LOQ	30°C (1)	0.839	-0.206	-0.164	#DIV/0!	0.980	-0.046	#DIV/0!	#DIV/0!	#DIV/0!	-0.453	-0.203	0.972	0.184
O5 LOQ	30°C (2)	0.753	0.310	0.422	0.397	0.992	0.255	0.144	0.420	0.503	-0.569	0.288	0.985	0.550
	Green = >0.90			**Yellow = 0.80><0.90**			**Orange = 0.70<>0.80**			**Red = 0.35<>0.70**			**Purple = <0.35**	

Table 18: Experiment 2: Total heat input

Total heating input to South (living room, south bedroom, bathroom) and North (kitchen and north bedroom)															
Constant temperature periods															
	Period	Experiment	Sim 2	Sim 4	Sim 7	Sim 10	Sim 11	Sim 12	Sim 16	Sim 18a	Sim 19	Sim 20	Sim 21	Sim 23	Sim 24
O5 S zone (kWh)	2 (30 C)	152.8	150.6	152.9	171.6	116.1	129.4	152.5	158.6	180.0	180.3	153.5	124.5	154.0	176.4
O5 S Zone Difference with Measured (%)			-1.5%	0.1%	12.3%	-24.0%	-15.3%	-0.2%	3.8%	17.8%	18.0%	0.4%	-18.5%	0.8%	15.4%
O5 N zone (kWh)		2.8	3.0	1.4	4.1	0.0	-0.7	5.4	0.0	0.4	0.9	7.2	10.4	26.0	1.5
O5 N Zone Difference with Measured (%)			6.1%	-49.4%	45.7%	-100.0%	-124.3%	92.9%	-100.0%	-84.9%	-66.4%	160.3%	275.3%	833.6%	-45.6%
O5 S zone (kWh)	4 (30 C)	228.8	224.9	206.9	231.4	187.4	189.8	183.9	217.4	248.2	257.7	199.0	194.5	230.2	228.1
O5 S Zone Difference with Measured (%)			-1.7%	-9.5%	1.2%	-18.1%	-17.0%	-19.6%	-5.0%	8.5%	12.7%	-13.0%	-15.0%	0.6%	-0.3%
O5 N zone (kWh)		5.7	6.6	4.7	10.3	2.5	0.9	15.1	0.0	3.7	4.8	6.3	15.8	40.8	4.8
O5 N Zone Difference with Measured (%)			15.2%	-18.4%	80.0%	-55.4%	-84.5%	163.5%	-99.7%	-35.9%	-16.1%	11.1%	176.4%	614.4%	-15.3%
O5 Total (kWh)	2 (30 C)	155.6	153.5	154.3	175.7	116.1	128.7	157.9	158.6	180.4	181.2	160.7	135.0	179.9	177.9
O5 Total Difference with Measured (%)			-1.3%	-0.8%	12.9%	-25.4%	-17.3%	1.5%	1.9%	15.9%	16.5%	3.3%	-13.3%	15.6%	14.3%
O5 Total (kWh)	4 (30 C)	234.5	231.5	211.6	241.7	189.9	190.7	199.0	217.4	251.8	262.5	205.3	210.3	271.0	233.0
O5 Total Difference with Measured (%)			-1.3%	-9.8%	3.1%	-19.0%	-18.7%	-15.1%	-7.3%	7.4%	12.0%	-12.4%	-10.3%	15.6%	-0.6%
		Green < +/-5%		**Yellow < +/-10%**			**Orange < +/-20%**			**Red < +/-40%**			**Purple > +/-40%**		

The data in Table 18 was also plotted in a histogram that summarises the experimental and prediction data (Figure 33). Here, the totals for the south zone and north zone for the constant period sequences (periods 2 and 4) are presented.

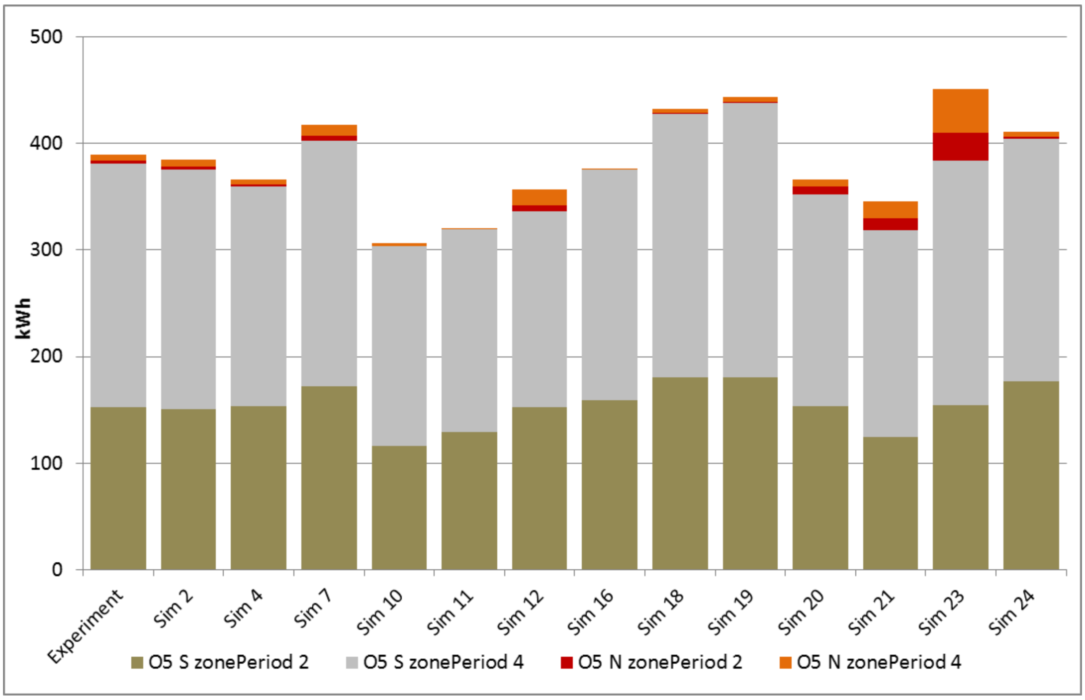

Figure 33: Experiment 2: cumulative heat inputs

An analysis was made of the predicted and measured solar irradiation on the four facades. Figure 34 is a summary histogram showing the cumulative solar irradiation over the whole of the analysis period (24th April to 2nd June). For most programs, agreement is good, although simulations 10 and 18 show significant error. No values were submitted values for the north and east walls in the case of Simulation 12.

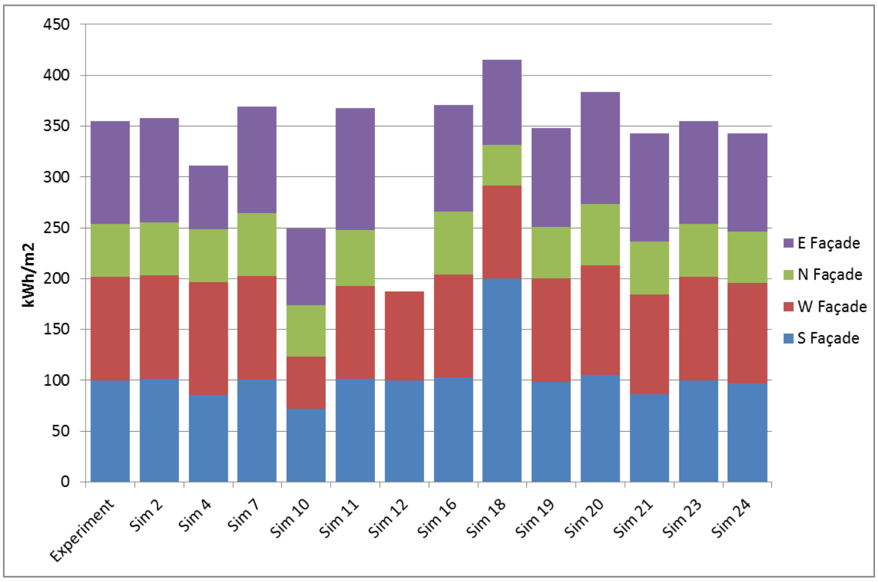

Figure 34: Experiment 2: cumulative solar irradiation

An analysis was also made of the correlations between the measured and predicted temperatures. Figure 35 show two examples for living room temperatures. Simulation 2 shows less scatter and no significant divergence across the temperature range; simulation 4 shows more scatter and a temperature dependent difference. This type of analysis may be useful to the modellers to identify possible causes for differences.

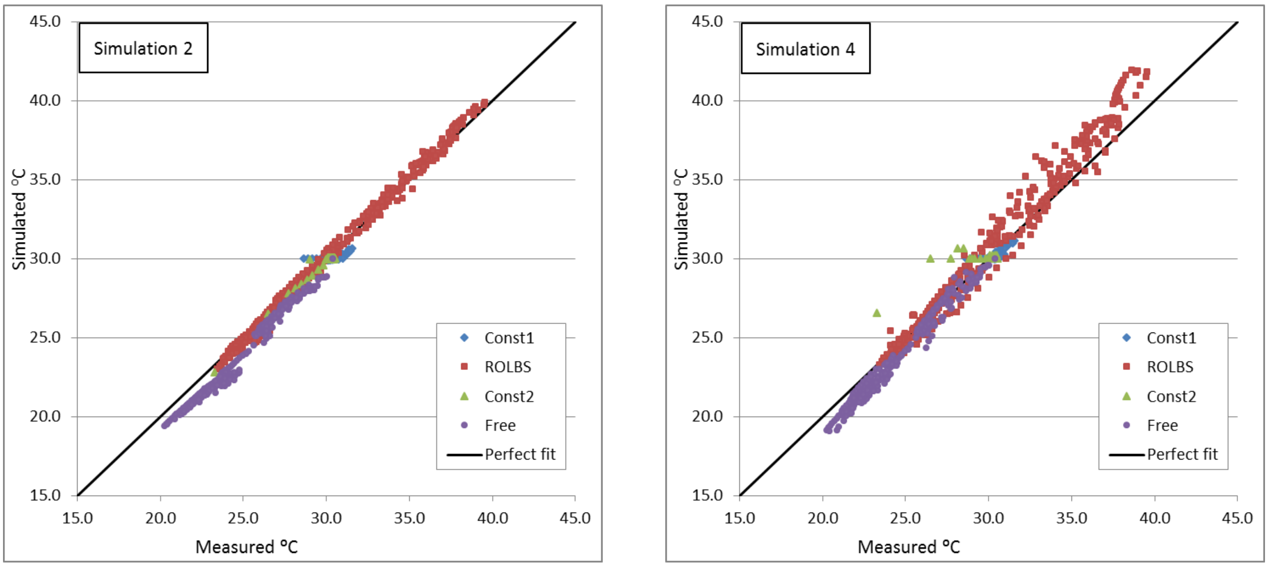

Figure 35: Experiment 2: correlation between measured and predicted temperatures

5.5 UNCERTAINTY ANALYSIS

The data presented in the graphs and tables do not include uncertainty bands, although this is important for judging whether the simulations can represent the performance of the real building. As mentioned in the section on instrumentation, experimental uncertainty in the room-averaged air temperature is estimated to be in the order of 1°C.

Although care was taken to ensure models inputs were as accurate as possible, there will still be uncertainty due to sensor uncertainty, uncertainty in thermophysical properties, glazing optical properties, appropriate selection of internal convection coefficients etc. The effect of these uncertainties on the modelling predictions will be model-dependent, and for this reason modelling teams were encouraged to use sensitivity techniques to identify important parameters and to assess overall modelling uncertainty bands. As an example, one of the modelling teams undertook a detailed sensitivity analysis for one of the programs, using the Morris method (Morris 1991) to identify important parameters and then undertaking Monte Carlo simulations (Robert and Casella, 2004) to produce modelling uncertainty bands. This was briefly described in Strachan et al (2015). Figure 36 shows the M scores for an analysis of the living room constant temperature period. A high M score, displayed by both the height of the line and the size of the circle in Figure 36, indicates the most important parameters affecting the model predictions. Figure 37 shows the mean values and 95% confidence bands for the living room measured and predicted internal temperature during the ROLBS sequence. A more comprehensive analysis is being prepared for publication.

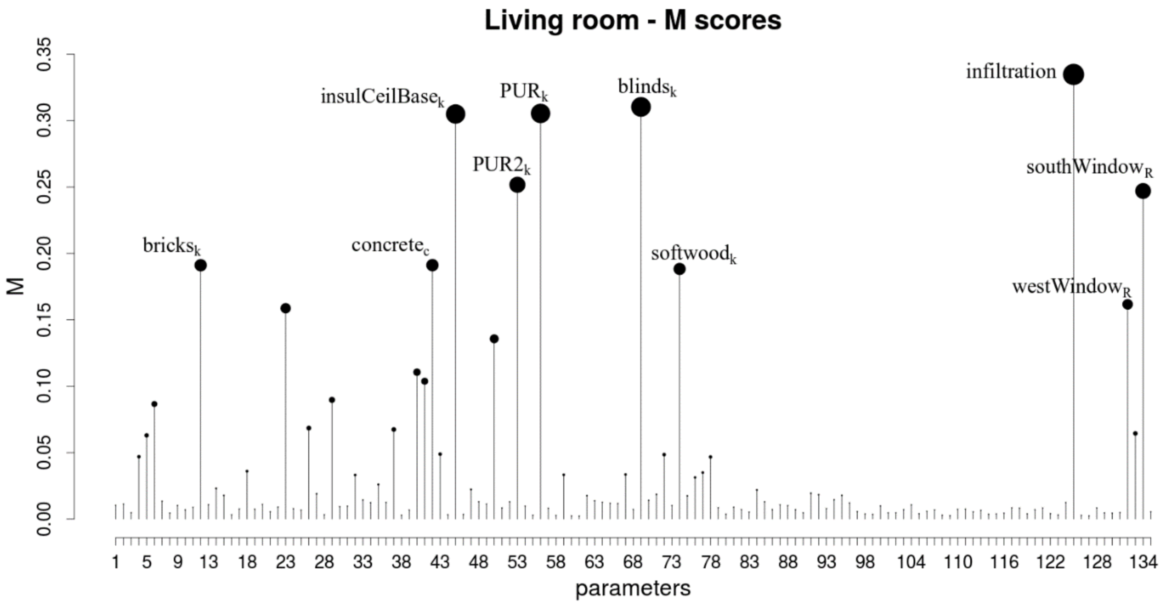

Figure 36: Mi scores for the living room constant temperature period. R, k and c indicate thermal resistance, conductivity and specific heat

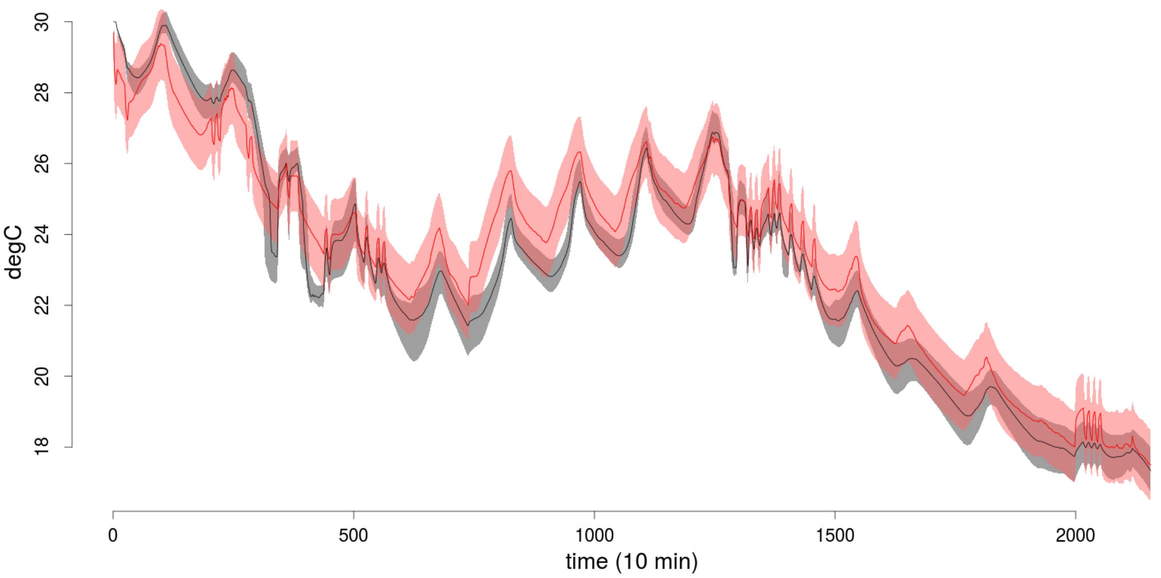

Figure 37: Mean values and 95% confidence bands for the living room measured (red) and predicted (black) internal temperature: ROLBS

6 Conclusions

6.1 Outcomes from the project

The datasets and experimental specification developed in this subtask is considered to be of high quality and arguably the best currently available for empirical validation based on real buildings. The experiment was undertaken by an experienced experimental team using a well-instrumented test facility. In addition, there has been a high level of engagement from modellers (over 20 sets of modelling predictions; 16 organisations; 12 different programs, both research and commercial), with the developed specification being implemented and thoroughly tested. Significant additional measurements and improvements to the specification have been made in the course of the study.

A comprehensive archive has been created of the experimental data and specification documents that is available for others to test their existing programs and for developers of new programs such as those being developed within IEA Annex 60 using Modelica.

6.2 Analysis of results obtained

The results presented have been anonymised so that it is not possible to identify any particular result set with the modelling team or program. The reasons for this are: firstly, that some of the programs are commercial and anonymity was promised to encourage participation; secondly, that some of the discrepancies even at the re-modelling stage, are likely to be as a result of modeller error and thus do not necessarily offer a definitive assessment of a program; and thirdly, that if results were published, it is more likely that teams would undertake tuning of the re-modelled submissions.

The summary scoring tables presented in this report present the overall results. In addition, rankings of the simulation programs were undertaken for some key metrics (Tables 19, 20, 21). Some programs did not model all rooms separately, in which case they are not included for some metrics.

Table 19: Experiment 1 Blind validation: Rank order of simulation predictions against measured data for various metrics

Rank	Temperature Magnitude	Temperature Shape	Heat Magnitude	Heat Shape
1	Sim 2	Sim 2	Sim 2	Sim 12
2	Sim 3	Sim 19	Sim 12	Sim 19
3	Sim 14	Sim 12	Sim 21	Sim 8
4	Sim 7	Sim 15	Sim 7	Sim 21
5	Sim 21	Sim 6	Sim 11	Sim 2
6	Sim 10	Sim 4	Sim 16	Sim 4
7	Sim 15	Sim 11	Sim 6	Sim 17
8	Sim 19	Sim 3	Sim 4	Sim 13
9	Sim 11	Sim 14	Sim 1	Sim 1
10	Sim 4	Sim 21	Sim 19	Sim 16
11	Sim 12	Sim 13	Sim 17	Sim 11
12	Sim 6	Sim 10	Sim 10	Sim 7
13	Sim 13	Sim 7	Sim 13	Sim 18
14	Sim 17	Sim 1	Sim 8	Sim 10
15	Sim 1	Sim 17	Sim 9	Sim 8
16	Sim 16	Sim 16	Sim 18	Sim 5
17	Sim 18	Sim 18	Sim 5	Sim 9

18	Sim 9	Sim 9	Sim 20	Sim 20
19	Sim 5	Sim 5		
20	Sim 8	Sim 8		
21	Sim 20	Sim 20		

Table 20: Experiment 1 Re-modelling: Rank order of simulation predictions against measured data for various metrics

Rank	Temperature Magnitude	Temperature Shape	Heat Magnitude	Heat Shape	Total heat
1	Sim 23	Sim 23	Sim 11	Sim 19	Sim 2
2	Sim 11	Sim 11	Sim 2	Sim 7	Sim 11
3	Sim 2	Sim 2	Sim 4	Sim 22	Sim 22
4	Sim 7	Sim 10	Sim 7	Sim 4	Sim 3
5	Sim 10	Sim 19	Sim 19	Sim 10	Sim 24
6	Sim 16	Sim 7	Sim 16	Sim 11	Sim 6
7	Sim 24	Sim 22	Sim 10	Sim 2	Sim 5
8	Sim 5	Sim 4	Sim 5	Sim 21	Sim 8
9	Sim 20	Sim 1	Sim 22	Sim 18	Sim 4
10	Sim 21	Sim 3	Sim 21	Sim 20	Sim 19
11	Sim 18	Sim 20	Sim 6	Sim 23	Sim 7
12	Sim 3	Sim 21	Sim 18	Sim 6	Sim 16
13	Sim 1	Sim 6	Sim 23	Sim 5	Sim 10
14	Sim 4	Sim 24	Sim 1	Sim 1	Sim 21
15	Sim 6	Sim 16	Sim 8	Sim 8	Sim 23
16	Sim 22	Sim 5	Sim 20	Sim 16	Sim 18
17	Sim 19	Sim 18			Sim 1
18	Sim 8	Sim 8			Sim 20

Table 21: Experiment 2: Rank order of simulation predictions against measured data for various metrics

Rank	Temperature Magnitude	Temperature Shape	Heat Magnitude	Total heat	Incident solar irradiation (AvAbsDiff S and W facades)
1	Sim 16	Sim 16	Sim 16	Sim 2	Sim 2
2	Sim 2	Sim 11	Sim 2	Sim 16	Sim 24
3	Sim 21	Sim 21	Sim 23	Sim 24	Sim 23
4	Sim 20	Sim 2	Sim 12	Sim 20	Sim 19
5	Sim 18	Sim 20	Sim 7	Sim 4	Sim 20
6	Sim 11	Sim 19	Sim 18	Sim 7	Sim 21
7	Sim 4	Sim 10	Sim 19	Sim 12	Sim 11
8	Sim 7	Sim 18	Sim 4	Sim 18	Sim 4
9	Sim 24	Sim 12	Sim 11	Sim 21	Sim 12
10	Sim 12	Sim 24	Sim 21	Sim 19	Sim 7
11	Sim 23	Sim 4	Sim 10	Sim 23	Sim 10
12	Sim 10	Sim 23	Sim 20	Sim 11	Sim 16
13	Sim 19	Sim 7	Sim 24	Sim 10	Sim 18

Key findings were as follows.

- Some programs performed very well, even in the blind validation phase of Experiment 1 and free-float period of Experiment 2. Simulation program 2 did particularly well throughout, with the absolute average temperature prediction in most rooms within 1°C of the measured data, the absolute average heat input usually within 100W of measured data, and the total heat input within 5% of the measured heat input. It is interesting to note the importance of choice of metrics – Simulation 16 performed well in Experiment 2 except for the average absolute difference in solar radiation. The total radiation prediction was reasonable (Figure 34), but a time offset in the peaks resulted in the relatively poor result for the average absolute difference.
- Most programs showed more variability in their ranking position, depending on the metric. However, as shown in Table 5, and particularly Tables 9 and 14, a good proportion (in the order of 40%) of the predictions of average absolute temperatures in the various spaces and experimental phases was within 1°C of measurements. The fact that several programs were able to predict satisfactorily over a large range of climatic conditions indicates that modelling programs are capable of modelling reality, given appropriate care with inputs, at least for the buildings modelled in this exercise.
- Modelling resulted in detecting the need for additional experimental information: specifically kitchen duct losses, additional sensors for surface heat flux and temperature, and measurements of ground reflectivity and surface absorptivity.
- In the blind validation phase, without any knowledge of the correct heat injections (for the constant temperature periods) or internal temperatures (for the ROLBS and free-float sequences), there are several examples of a reasonable level of agreement between measurements and predictions (e.g. Simulations 2, 3, 10). In some cases, the agreement in terms of average absolute difference in temperatures was better than 1°C in all spaces except the kitchen. This interesting result led to the identification of the heat losses from the kitchen to the uninsulated ductwork as a deficiency in the model specification.
- In the re-modelling of Experiment 1, there are some anomalous programs with poor agreement, but qualitatively, the agreement in both the absolute predictions of temperatures and heat inputs, and the dynamic response, is good for the majority of programs. This holds for both Twin Houses and the differences between them. Because solar gains are a dominant heat transfer process in these experiments, this indicates that the prediction of solar radiation on the different facades and the solar transmission through the glazing is well represented. The good agreement in dynamic response indicates acceptable modelling of the large thermal mass in these buildings.
- Most of the modelling reports submitted with the re-modelling mentioned user errors in the input which had been corrected (in addition to implementing the new information provided regarding thermal bridges, internal absorptivity, supply air temperature in the living room and kitchen ductwork heat losses). These errors varied from minor input error to more significant errors such as not limiting the heat inputs.
- For the re-modelling results, no one simulation result set came out in the top four for every metric used in the comparisons (based on summing the outcomes for all periods and all rooms). The South Bedroom heat input was worst in terms of the agreement between measured and predicted but no obvious reason could be found.
- For the re-modelled results, the total heating inputs to the four rooms analysed (living room, south bedroom, kitchen and north bedroom) showed large variations in the level of agreement between predicted and measured. Simulation number 6 is interesting – the level of agreement for the two houses was generally good, but the level of agreement for the difference between the two houses was relatively poor. The reason is that the predictions for house O5 (blinds up) were lower than measured, and the predictions for house N2 (blinds down) were higher than predicted. This would suggest

a problem with modelling the solar transmission as this is the essential difference between the two houses.

- Conclusions from Experiment 2 are similar to those of the re-modelled datasets of Experiment 1. There are some anomalous programs with poor agreement, but qualitatively, the agreement in both the absolute predictions of temperatures and heat inputs, and the dynamic response, is good for the majority of programs.

Only a subset of graphs are displayed in this project, but the aim was to demonstrate some display options that modelling teams can use to investigate differences between their programs and measurements. Some of the various ways of displaying the data shown in the results section are as follows.

1. Time series data for the experimental periods. These graphs (e.g. Figure 17) are useful for giving a visual impression of the overall level of agreement, whether there are significant time shifts or offsets.
2. Time series data of individual days. These graphs (e.g. Figure 32) are useful for identifying the magnitude of any phase shifts – either due to timing convention misunderstandings of the measured data or within the simulation program, or due to the modelled response to heat inputs from solar radiation and internal heat gains.
3. Time series data with moving average. These graphs (e.g. Figure 25) can be useful for eliminating the short period fluctuations occurring at transitions between the experimental periods and other short duration events, to give an overall view of results over the whole experiment.
4. Plots of correlations between measured and predicted values. The graphs (e.g. Figure 35) can be useful to identify consistent differences over the experimental period.
5. Aggregated histograms. These graphs (e.g. Figure 33) can (in addition to the summary tables such as Table 4) provide an insight into whether the discrepancies between predictions and measurements are in particular phases of the overall experiment. They are also an indicator of the quality of the static energy balance of the model.
6. Analysis of individual heat transfer paths. The comprehensive datasets collected in these experiments can be analysed by the various modelling teams. One example is given in Figure 34, where the measured and predicted solar radiation on the building facades are compared. It is expected that detailed investigations can take place using measured surface temperatures and heat fluxes, air temperatures at various heights in the rooms, globe temperatures, and even some data not directly used in these experiments such as the ground temperatures at various depths.

Other useful analysis techniques not covered in detail in this report include statistical analysis of the data to provide overall building performance characteristics such as total heat loss coefficients, solar aperture, which can be compared to the buildings specification. This is covered in the IEA EBC Annex 58 Final Report Subtask 3 (2015).

6.3 USEFULNESS

The exercise has already proved useful for modellers. In particular:

- In one case, an incorrect sky temperature calculation was identified, leading to errors in the external long-wave radiation transfer. The program developers, in this case a commercial program, recognised that their algorithm needed to be improved.
- Another modelling team used results to identify deficiencies in the modelling of external longwave heat transfer.
- Several modellers mentioned the need for simulation program capabilities to be enhanced to facilitate easy modelling of thermal bridges. For many programs, additional constructions had to be created with properties set to provide the additional heat losses. As buildings' fabric improves and wall thicknesses increase with increased

insulation levels, thermal bridge losses can assume a greater relative significance, and the ease of modelling should be improved.

As mentioned above, an important outcome is the archive of measured data and the specifications. These datasets have already proved useful:

- Model developers have used them to test program development. Two comments illustrate this benefit:
    - HFT Stuttgart: “We will certainly continue to improve the building simulation parts in INSEL - one of the reasons why the Twin House validation study has been so highly welcome by us.”
    - University of Bordeaux: “... I have started to use the Fraunhofer validation set to validate the Modelica Buildings library model”
- Participants in IEA EBC Annex 58 from the University of Leuven used them as a training exercise for students to indicate the importance of modelling assumptions and the necessary rigour in setting up models.

The ASHRAE Standard 140 committee has expressed a strong interest in using the dataset and specification documentation. A presentation was made to the Standard 140 committee in June 2015, with a request for an update at the January 2016 ASHRAE conference.

It is expected that additional analysis of the datasets will be undertaken to explore in more detail appropriate convective heat transfer coefficients, the radiative/convective split of the radiators, stratification, air circulation in the rooms, wall surface temperatures and fluxes, etc.

It is worth recording the views at the end of the exercise from leaders of two well-known commercial simulation programs.

Per Sahlin, CEO, EQUA Simulation AB:

“Validation must go on! As models grow increasingly complex, there is a risk that one starts trusting them just because some sub-models have been around for a while and many people seem to be using them. In this situation, nothing is more refreshing than to take part in a well-formulated and well-measured empirical validation study. It inevitably brings you back to where you should always be – thinking about the fundamentals. The Annex 58 empirical study gave us two new interesting insights. First, the impact of thermal bridges in internal floors and, secondly, the importance of correct optical parameters of the window recess. It has also provided an excellent data set for development of new methods for automatic model tuning. We understand that funding bodies may think that yet another experimental exercise may seem mundane, especially in view of other more spectacular developments in the energy field. However, keep in mind that almost all new and promising energy systems depend on having good simulation models to design them with, and without constant validation projects, these models will be built on very questionable foundations.”

Craig Wheatley, Chief Technology Officer, IES:

“Increasingly building owners and operators are realising the potential of energy efficiency to their triple bottom line, and awareness of the impact of energy usage within the world's buildings upon climate change has risen substantially. In order that energy efficiency measures can be successfully adopted by the mainstream it is important that the tools that are available that assess energy performance are validated against real world scenarios. The continued efforts of the IEA Annex 58 are important as it provides real world benchmarks for validation that gives confidence to mainstream users that will encourage increased usage in these tools which will lead to better performing lower energy usage buildings.”

6.4 Publications Arising from Subtask

The datasets and experimental specification are available under an Open Access licence. For Experiment 1, these can be obtained from:

http://dx.doi.org/10.15129/8a86bbbb-7be8-4a87-be76-0372985ea228

For Experiment 2, the link is:

http://dx.doi.org/10.15129/94559779-e781-4318-8842-80a2b1201668

The following are publications to date arising from the study.

1. Strachan P, Svehla K, Heusler I and Kersken M, Whole Model Empirical Validation on a Full-Scale Building, Journal of Building Performance Simulation, 2015, http://dx.doi.org/10.1080/19401493.2015.1064480.
2. Kersken M, Heusler I, Strachan P and Sinnesbichler H, Entwicklung eines neuen, messdatenbasierten Validierungsszenarios für die dynamische Gebäudesimulation, Bauphysik, Volume 37, Issue 3, pages 153–158, June 2015, DOI: 10.1002/bapi.201510021
3. Strachan P, Monari F, Kersken M and Heusler I, IEA Annex 58: Full-scale Empirical Validation of Detailed Thermal Simulation Programs, 6th International Building Physics Conference, IBPC 2015, June 2015, Torino, Italy.
4. Masy G, Rehab I, André P, Georges E, Randaxhe F, Lemort V and Lebrun J, Lessons Learned from Heat Balance Analysis for Holzkirchen Twin Houses Experiment, 6th International Building Physics Conference, IBPC 2015, June 2015, Torino, Italy.
5. Kersken M, Heusler I and Strachan P, Full Scale Empirical Validation for Building Energy Simulation Programs, 9th International Conference on System Simulation in Buildings, Liege, December 10-12, 2014.
6. Masy G, Delarbre F, Lebrun J, Georges E, Randaxhe F, Lemort V, Rehab I and André P, Back from Holzkirchen full scale dynamic testing experiment, 9th International Conference on System Simulation in Buildings, Liege, December 10-12, 2014.
7. Kersken M, Heusler I and Strachan P, Erstellung Eines Neuen, Messdatengestützten Validierungs-Szenarios Für Gebäude-Simulationsprogramme, 5th German-Austrian IBPSA Conference, Bausim 2014, Aachen, Germany, Sept 2014.
8. Strachan P, Hand J, Svehla K, Heusler I and Kersken M, A Full-Scale Empirical Validation Study Applied to Thermal Simulation Programs, Building Simulation '15, Hyderabad, India, December 2015.
9. Monari F and Strachan P, Characterization of an airflow network model by sensitivity analysis: parameter screening, fixing, prioritizing and mapping, Journal of Building Performance Simulation, 2016, http://dx.doi.org/10.1080/19401493.2015.1110621.

6.5 External Interactions

There has been good engagement with researchers external to participants in IEA EBC Annex 58. At the September 2014 Annex meeting in LBNL, Berkeley, one day was held jointly with IEA EBC Annex 60 that included a presentation of the Twin House Experiments. Several of the participants in Annex 60 submitted results using models built with Modelica.

Some commercial program vendors not involved as participants in IEA EBC Annex 58 were approached at the start of the study and invited to participate. As can be seen from the list in Table 3, the response was good, usefully extending the range of programs involved.

In addition, some external groups requested access to the datasets and specification following conference presentations, and have used them as the basis for testing. Some of these teams then submitted datasets (although not in time for the blind validation phase).

6.6 LESSONS FOR FUTURE EMPIRICAL VALIDATION STUDIES

The empirical validation study described in this report is considered to be of high quality. It was undertaken by a large number of experienced experimentalists and modellers, and the focus was on full-scale houses with high solar penetration and high thermal mass. Nevertheless, there are limitations, notably:

1. the Twin Houses are relatively simple buildings and only the ground floor rooms were considered;
2. only the building fabric and ventilation were tested, with electric heating, so heating and cooling systems were not included;
3. there were no occupants.

There is a clear need for further studies, extending the range of building types and including HVAC systems. Including occupants in the study would clearly be the ideal target, although this would need significant increases in monitoring levels.

In terms of the validation methodology, the weakest point is thought to be in ensuring that the re-modelled submitted predictions (after the measured data has been distributed) do not include any "parameter tuning" or calibration, and that changes are documented. If the models are calibrated, they no longer serve to test the simulation program, which is the objective of the validation exercise (although calibration techniques can be useful for identifying influential parameters which can help to isolate the possible causes for observed differences between measurements and predictions). In this study, modelling teams were asked only to correct input errors before resubmitting the data, although they were also encouraged to use sensitivity analyses to explore possible reasons for differences between measurements and predictions. The level of documentation submitted was variable. In future studies, it may be advisable to ask teams to submit their models for both the blind and remodelling predictions so that changes could be checked; this may also be useful for comparing models created by different modelling teams using the same simulation program. It is also recommended that a more extensive template for the modelling reports is issued to improve the consistency of the reporting.

Regarding the experimentation, the main recommendations would be to make additional air temperature measurements to check on stratification and the degree of room mixing, to include multi-zone tracer gas measurements, using different tracer gases in each room, to monitor the inter-zonal airflow, and to use an uninterruptible power supply to guard against power failure for the data logging system.

Despite the detailed experimental dataset and specification, some other uncertainties of the specification still remain, particularly regarding the thermal bridges within the construction of the Twin Houses. It would be interesting to undertake a similar experiment on new buildings where particular attention has been paid to construction quality and the avoidance of thermal bridges (e.g. a certified Passivhaus design).

There are currently no standards for what constitutes "good" agreement for the predictions of temperatures, heat inputs etc. To some extent, this will be dependent on the experiment and accuracy of measurements. However, there could be some discussion by modelling teams in the experimental design stage to set evaluation bands before the results are known. Also, if more empirical validation studies are undertaken, it may be possible to develop appropriate limits on acceptability.

The time and effort to conduct these empirical validation experiments was substantial, by the experimental team, the modellers and the analysis team. Although additional such studies are needed to improve confidence in the ability of simulation programs to predict accurately the energy and environmental performance of buildings, a large resource is required to succeed in terms of both time and money.

Acknowledgements

The authors of this report would like to recognise the efforts of a large number of modellers from many countries who have undertaken the modelling work, written modellers reports and submitted predictions. In addition, Gabriel Rojas-Kopeinig (University of Innsbruck) is thanked for undertaking the analysis of the heat exchange from the ductwork. Glenn Reynders (University of Leuven), Filippo Monari (University of Strathclyde), Roberto Garay (Tecnalia) and Kyung Hun Woo (Danish Technical University) are thanked for their inputs to the thermal bridge analyses.

References

IEA EBC Annex 21 (1995), Environmental Performance, http://www.iea-ebc.org/projects/completed-projects/ebc-annex-21/

IEA EBC Annex 41 (2007), Whole Building Heat, Air and Moisture Response, http://www.ecbcs.org/annexes/annex41.htm

IEA EBC Annex 42 (2007) The Simulation of Building-Integrated Fuel Cell and Other Cogeneration Systems (COGEN-SIM), http://www.iea-ebc.org/projects/completed-projects/ebc-annex-42/

IEA EBC Annex 43 (2007), Testing and Validation of Building Energy Simulation Tools, http://www.iea-ebc.org/projects/completed-projects/ebc-annex-43/

IEA EBC Annex 58 Final Report Subtask 3: Dynamic Data Analysis and Performance Characterisation, 2015.

Judkoff R and Neymark J (2006), Model Validation and Testing: The Methodological Foundation of ASHRAE Standard 140, ASHRAE Conference, Quebec City, Canada.

Judkoff R, Wortman D, O'Doherty B and Burch J, A Methodology for Validating Building Energy Analysis Simulations, NREL/TP-550-42059 1983/2008 National Renewable Energy Laboratory Technical Report.

Kalyanova O, Heiselberg P, Felsmann C, Poirazis H, Strachan P, and Wijsman A (2009), An Empirical Validation of Building Simulation Software for Modelling of Double Skin Façades, Proceedings of the 11th IBPSA Conference (International Building Performance Simulation Association), Glasgow, UK, July, pp 1107-1114.

Kendall M and Gibbons J D (1990), Rank Correlation Methods, 5th edition, Edward Arnold, London, pp 69-77.

Lomas K J, Eppel H, Martin C J, Bloomfield D P (1997), Empirical Validation of Building Energy Simulation Programs, Energy and Buildings, 26(3), pp 253-275.

Loutzenhiser P G, Manz H, Felsmann C, Strachan P and Maxwell G M (2007), An Empirical Validation of Modeling Solar Gain through a Glazing Unit with External and Internal Shading Screens, Applied Thermal Engineering, Vol 27, Issues 2-3, pp528-538.

Manz H, Loutzenhiser P, Frank T, Strachan P, Bundi R and Maxwell G (2006), Series of Experiments for Empirical Validation of Solar Gain Modelling in Building Energy Simulation Codes - Experimental Setup, Test Cell Characterization, Specifications and Uncertainty Analysis, Building and Environment 41(12), pp1784-1797.

Morris, M D (1991), Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics, 33 (2), pp161-174. http://dx.doi.org/10.2307/1269043.

Munich IWEC (2014), available from:
http://apps1.eere.energy.gov/buildings/energyplus/weatherdata_about.cfm

PHLuft (2014), http://www.passiv.de/old/04_pub/Literatur/PHLuft/PHL_F.htm

Robert C P and Casella G (2004), Monte Carlo Statistical Methods, New York, Springer.

Strachan P, Monari F, Kersken M and Heusler I, IEA Annex 58: Full-scale Empirical Validation of Detailed Thermal Simulation Programs, 6th International Building Physics Conference, IBPC 2015, Torino, Italy.

THERM (2014), http://windows.lbl.gov/software/therm/therm.html van Dijk, H A L and Tellez, F M (1995), Measurement and Data Analysis Procedures, Final Report of the JOULE II COMPASS Project (JOU2-CT92-0216).

Williamson T J (1995), A Confirmation Technique for Thermal Performance Simulation Models, Building Simulation '95, Madison, Wisconsin, U.S.A, August, pp 268-275.