Strathprints Institutional Repository

Strachan, Paul and Svehla, Katalin and Heusler, Ingo and Kersken, Matthias (2016) Whole model empirical validation on a full-scale building. Journal of Building Performance Simulation. ISSN 1940-1493 , http://dx.doi.org/10.1080/19401493.2015.1064480

Journal of Building Performance Simulation, 2015
http://dx.doi.org/10.1080/19401493.2015.1064480

Taylor & Francis
Taylor & Francis Group

Whole model empirical validation on a full-scale building

Paul Strachan[a*], Katalin Svehla[a], Ingo Heusler[b] and Matthias Kersken[b]

[a]*ESRU, University of Strathclyde, 75 Montrose Street, Glasgow G1 1XJ, Scotland;* [b]*Fraunhofer Institute for Building Physics IBP, Standort Holzkirchen, Fraunhoferstr. 10, Valley 83626, Germany*

(*Received 1 February 2015; accepted 21 May 2015*)

This paper describes an empirical validation study undertaken on two identical full-size buildings within the scope of the IEA ECB Annex 58 project. Details of the experimental configuration and monitoring are included, together with results from measurements and from predictions made by 21 modelling teams using commercial and research simulation programmes. The two-month, side-by-side experiment was undertaken on buildings with high levels of thermal mass and in a period with high solar gains. The detailed specification and associated measurement data provide a useful empirical validation dataset for programme testing. Results from the modelling demonstrate good agreement between measured data and predictions for a number of programmes, in both absolute predictions of temperatures and heat inputs as well as dynamic response. On the other hand, a significant number of user input errors resulted in poor agreement for other programmes, especially in the blind validation phase of the modelling methodology.

Keywords: empirical validation; IEA Annex 58; dynamic thermal modelling

Introduction

Building energy performance modelling tools are increasingly used in design and regulation compliance. Within building design, there is increasing use of passive technologies in order to reduce energy consumption, active technologies for heating, cooling and electrical energy supply, and thermal and electrical storage. As a result, the complexity of the interactions of heat and mass transfer processes increases. Building response can also become more dynamic, with potential control and overheating problems. It is therefore essential that the thermal simulation programmes used in design are fit for the purpose, and that they are perceived to be so by designers, clients and regulatory authorities.

There have been several large international and national projects that have been successful in establishing, to a certain degree of confidence, the validity of basic heat and mass transfer models and their application for predicting comfort conditions and energy consumption in buildings (e.g. Judkoff and Neymark 2006; Strachan, Kokogiannakis, and Macdonald 2008). For the empirical validation work, the focus has almost exclusively been on relatively simple outdoor test cells, as evidenced in the following section. These validation studies have been useful for uncovering programme errors and limitations of predictive accuracy. However, the question remains as to whether the performance predictions of full-scale buildings can be relied upon. There is much research at present on the so-called performance gap between design predictions and measurement of energy performance. The main causes of differences are likely to be due to factors such as occupant behaviour, workmanship defects, operational settings and control, but it would be useful to determine the extent of uncertainty in the design predictions due to uncertainties in the accuracy and capabilities of the simulation programmes used.

The difficulty in undertaking full-scale empirical validation is due to the fact that all flow paths and boundary conditions must be measured, with the building tested through a range of external boundary conditions and internal operations, in order for the study to be useful for validation. It is believed that there have been no comprehensive full-scale validation datasets produced from full-scale buildings to date. The reason for attempting such an experiment at this time is a combination of factors that should now improve chances of success: namely, widespread availability of sensor and instrumentation equipment, the availability of sophisticated test buildings, knowledge regarding errors in previous experimental programmes and improvements in simulation programmes to model low-energy technologies to assist in the experimental design.

---

*Corresponding author. Email: paul@esru.strath.ac.uk

The aims of the experiment and associated modelling were:

- to obtain and apply high-quality experimental datasets for model validation of the thermal performance of full-scale buildings;
- to develop robust procedures to ensure that the gathered datasets are suitable for validation purposes;
- to apply an iterative validation procedure to compare model predictions with measured data; and
- to promote the study of analysis techniques (particularly sensitivity analyses) to identify causes of discrepancies between measured and predicted energy performance data.

The work was conducted as part of the IEA ECB Annex 58 project "Reliable building energy performance characterization based on full-scale dynamic measurements" (IEA Annex 58 2015).

Previous empirical validation studies

An overall validation methodology for thermal simulation programmes is well established and comprises elements of analytical, inter-programme comparison and empirical tests (Judkoff et al. 1983; Jensen 1993). Inter-programme comparative tests have the advantages that they are relatively easy to apply and that many parameters can be tested. They have been widely used, in particular BESTEST, embedded within ASHRAE Standard 140 (ASHRAE 2011) which resulted from International Energy Agency (IEA) project Annex 21/Task 12 (IEA 1995), particularly in their diagnostic role for detecting programme errors. However, there is the criticism that there is no truth standard in such tests (Judkoff and Neymark 2006). Empirical tests can provide this to a certain degree of accuracy, but gathering high-quality experimental data is expensive and time consuming.

There have been a number of large-scale IEA and European Commission projects over many years that have had empirical validation as the focus (Table 1). At the start of IEA Annex 21 (IEA 1995), a comprehensive worldwide review of existing datasets suitable for empirical validation was reported. The majority of the datasets investigated were found to be of limited use for programme validation – primarily because of missing monitored data of key parameters. Significant attention to detail is required for achieving validation-quality datasets. A key observation from Table 1 and the discussion above is that no high-quality datasets are available at a full-scale building level. More monitored data are becoming available which is beneficial for giving an overall appreciation of the agreement between measured and predicted energy consumption (examples include the CarbonBuzz project (2014), CARB (2010), TSB Retrofit for the Future (2012) and LEED monitoring (Turner and Frankel 2008). However, it must be emphasized that these monitoring studies have a different purpose, and in particular the instrumentation has not been designed to provide the comprehensive coverage required for validation of simulation programmes.

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. 1994; Loutzenhiser

Table 1. IEA and EC projects with a substantial empirical validation component.

Project	Year	Comment
IEA ECBCS Annex 1	1977–1980	A monitored building was used but the accuracy of model inputs was suspect
IEA ECBCS Annex 4	1979–1982	Comparison of predicted with measured data from a commercial office building. Many experimental deficiencies were identified
IEA SHC Task 8	1982–1988	Eleven programmes were compared against test cell data gathered at the Passive Solar Test Facility of the NRC, Canada
EC PASSYS Project	1986–1993	Component level tests were undertaken on outdoor test cells. These included sunspaces, multi-functional facades, ventilated glazing systems and transparent insulation
IEA ECBCS Annex 21	1988–1993	Empirical data from small well-controlled and monitored outdoor test rooms were compared with predictions from 17 different programmes
IEA SHC Task 22	1996–2002	Included empirical studies of fabric elements mounted on test cells, and HVAC component testing
IEA ECBCS Annex 41	2003–2007	Some climate chamber experimental datasets of heat, air and moisture response were used for comparing with model predictions
IEA ECBCS Annex 42	2003–2007	The project developed micro-cogeneration models for incorporation into whole-building simulation programmes
IEA ECBCS Annex 43/ SHC Task 34	2003–2007	Three of the subtasks involved empirical validation. Two were test cell based (shading/daylighting/load interaction and double-façade testing) and one involved laboratory experiments on HVAC components

Notes: ECBCS is Energy Conservation in Buildings and Community Systems (now Energy in Buildings and Communities: EBC); SHC is Solar Heating and Cooling 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 programme the heating and/or cooling as required.
3. *Experimental specification*. Develop the specification which describes 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 programmes 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 are 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 datasets*. The intention is that the resulting specification and datasets will be useful for developers of new programmes and those improving modelling algorithms.

Selection of test building

There are now a number of high-quality outdoor test facilities – these have been documented within IEA Annex 58 (Janssens 2014). Many of these are potentially suitable for validation studies.

At the start of the study, the main requirements were considered to be as follows:

1. Availability of building for structured test sequence with defined operational schedules;
2. Documented building and systems details;
3. High levels of calibrated instrumentation;
4. Ability to isolate parts of the building for initial tests;
5. Options for heating and cooling, for example, electric heaters, conventional boilers, microgeneration, solar thermal and heat pumps;
6. Unoccupied: this was considered necessary to avoid a significant extra set of uncertainties.

A detailed checklist was constructed of the requirements (Table 2) which was circulated to potential experimental teams.

From a short list of four facilities identified within the participant organizations of IEA Annex 58, the Twin Houses at the Fraunhofer Institute for Building Physics (IBP) at Holzkirchen, Germany (Figures 1 and 2), were selected, based on the checklist responses.

These two houses had the added advantage that they were essentially identical, so could be used for side-by-side testing. Pressurization tests were conducted on the two buildings which showed agreement to better than 5%. Measurements were undertaken of the heating power requirements of the two houses to maintain constant internal temperatures. Figure 3 shows the results. 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%. This baseline measurement was undertaken without any natural or mechanical ventilation.

Experimental design

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 experiment described in this paper, there was one significant constraint – the houses were only available in the

Table 2. Information requirements for the full-scale building validation study.

Building description and location	References/reports available with building and instrumentation description
Availability	Availability of the building for testing for an extended period
Building construction	Dimensional details and orientation
	Construction materials and layer thicknesses
	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
	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 ventilation
	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 set point 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
	Manufacturer's data available for the heating/cooling system
	Measured performance data available for the heating/cooling system
	Performance data for renewable technologies available
Instrumentation	Air temperatures in rooms: number of sensors and location and whether sensors are shielded
	Surface temperatures: number and location
	Electrical power consumption
	Delivered heating
	Delivered cooling
	Instrumentation for heating/cooling plant (flow rates, return/supply temperatures, etc.)
	Ventilation
	Other instrumentation
Climate and other boundary conditions	Air temperature
	Solar radiation – global horizontal, diffuse horizontal, total vertical
	Wind speed
	Wind direction
	Relative humidity
	Longwave radiation
	Ground reflectivity
	Ground temperature

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. The experimental design was undertaken by modelling the houses using a representative climate for Munich (Munich IWEC 2014) with the following aims:

1. To ensure that the mechanical ventilation rate was sufficient to prevent significant overheating above the heating set point.
2. To determine the heater capacities necessary for maintaining a suitable set point.
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. In the experiment, all blinds are up all the time, except for the south windows. The southern blinds are closed on one house permanently (house N2) and are closed only for the initialization and the constant temperature scenarios on the other house (house O5).

View of West

View of East

View of South

View of North-east

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

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 2D 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, with linear thermal transmittances included in the specification.

The experimental configuration is shown in Figure 4.

To reduce complexity, the temperatures in the cellar and attic spaces were measured and treated as boundary conditions.

The experiment was undertaken over a period of two months (1 August to 26 September 2013). The schedule is shown schematically in Figure 5. It was divided into five different periods. The control in these periods was chosen to reflect common conditions in buildings as well as ensuring that 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. These measured temperatures are provided as inputs for the modelling, with heating power to achieve these measured temperatures being predicted.

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 spilt of 30%:70%). The use of a pseudo-random sequence of heat injections ensures that the solar and heat inputs are uncorrelated, which helps 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

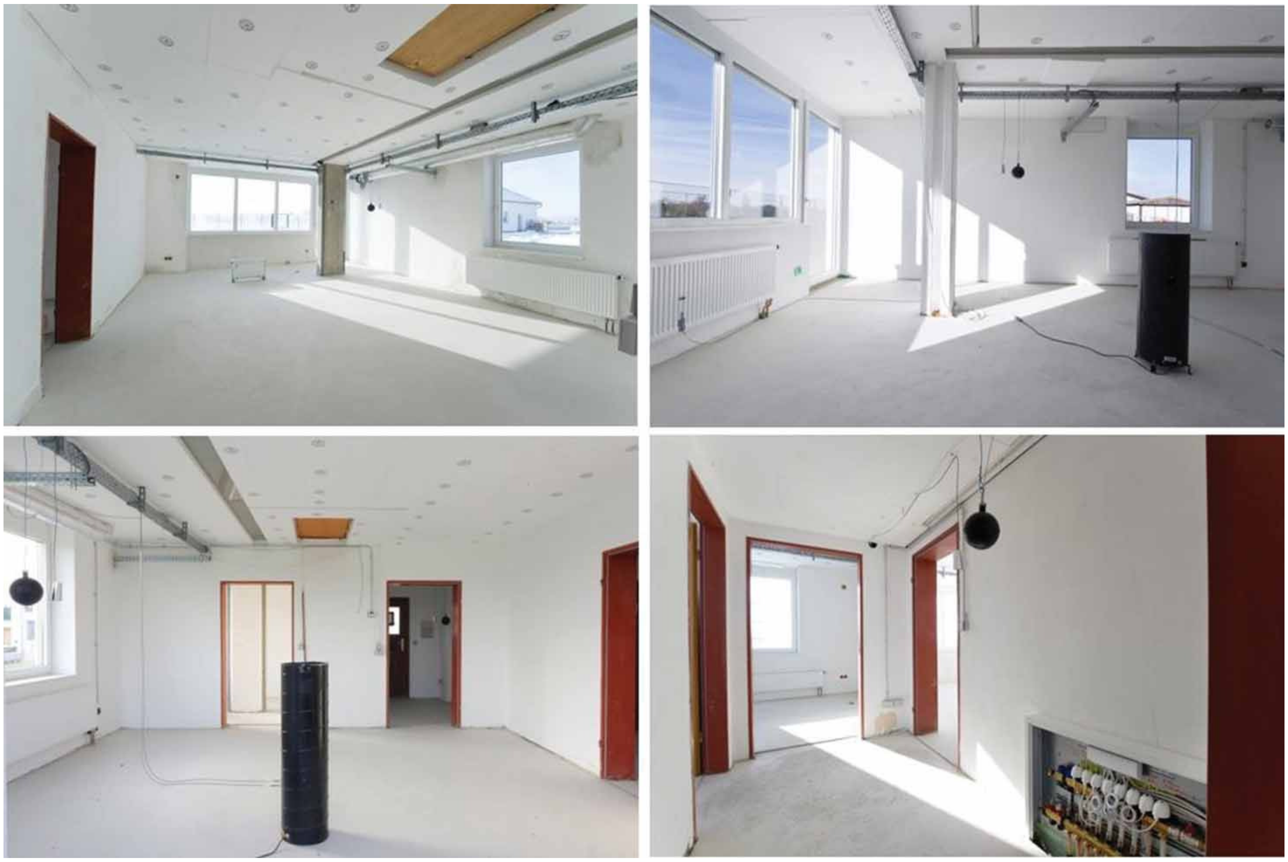


Figure 2. Internal views of the Twin Houses.

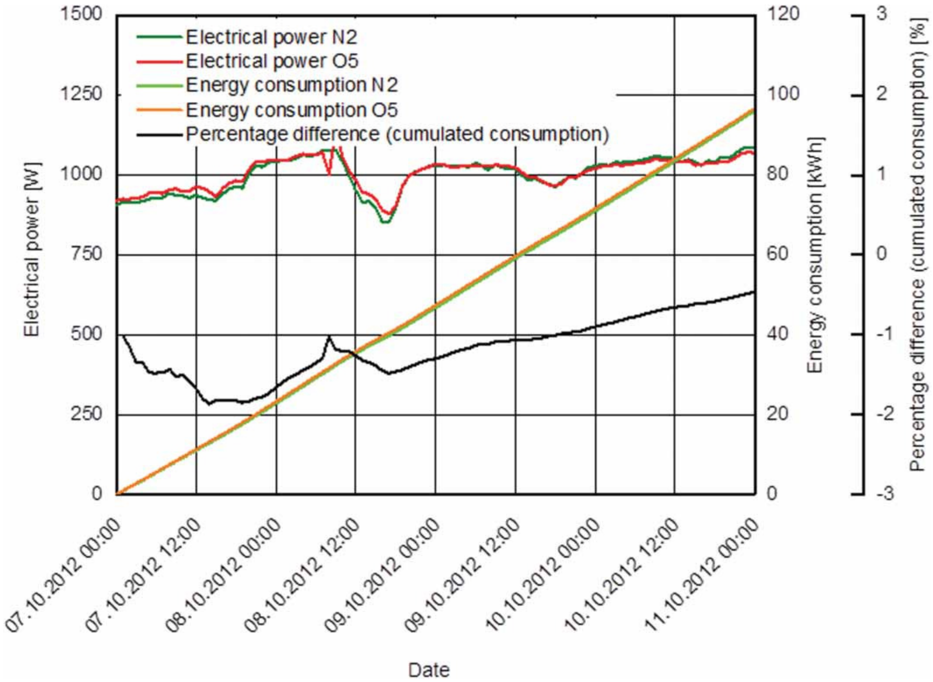


Figure 3. Base line measurement of the Twin Houses.

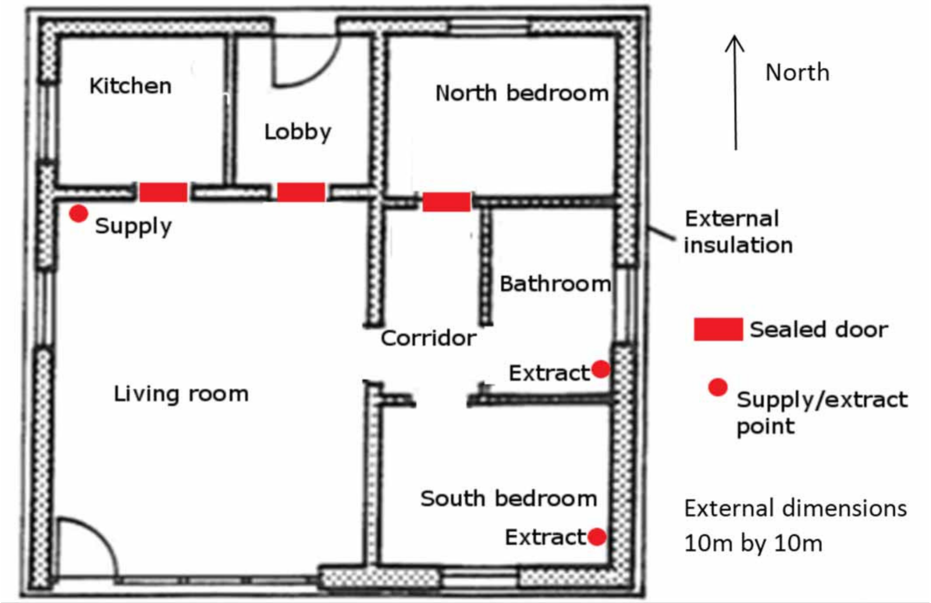


Figure 4. Experimental layout.

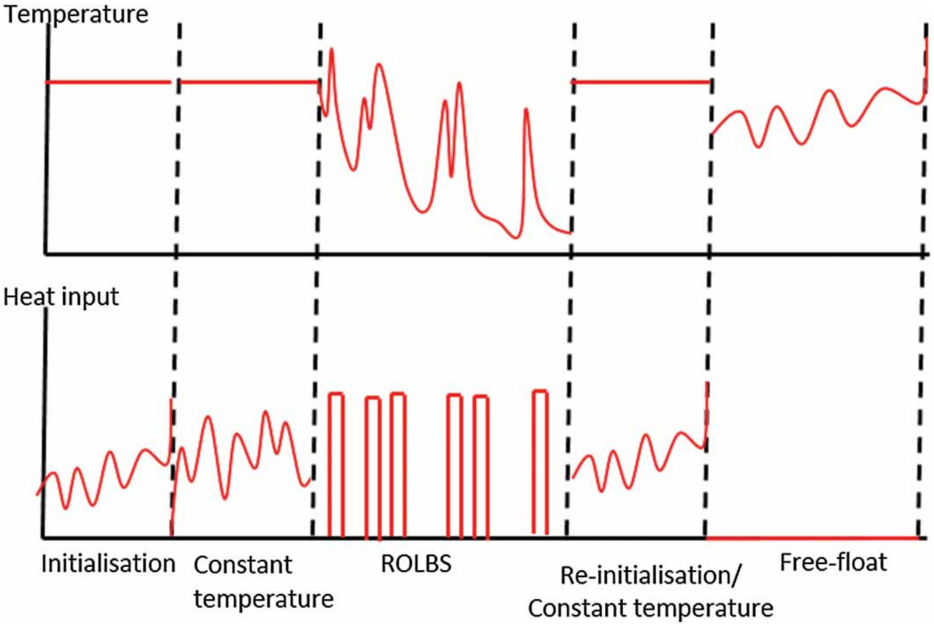


Figure 5. Schematic of the test schedule.

of thermal mass. All other rooms were without heating power in this period to increase the interaction between the rooms. In this case, the ROLBS sequence of heat inputs is provided for the modelling, with resulting temperatures being predicted.

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). Again the measured indoor air temperatures are provided for modelling, with the resulting heating power being calculated.

Period 5: In this 7-day period, there were no artificial heat injections. Modelling teams were required to predict the resulting temperatures given only the external climate for this free-float period.

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.

The southern rooms of the ground floor were ventilated as can be seen in Figure 4. A balanced ventilation system was implemented, with supply air entering the living room with a volume flow rate of 120 m^3/h and extracted through the bathroom and the south bedroom with a flow of 60 m^3/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 m^3/h, which was achieved with a standard deviation of only ± 0.2 m^3/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 set-up the disc valves in both rooms were adjusted using a second, temporary flow meter. All duct joints were sealed carefully using tape to minimize pressure losses throughout the ducts' length. The supply air temperature was measured after the fan, so the fan's waste heat was included in this temperature. 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 programmes.

Infiltration was measured by pressurization tests before and after the experiment. To give an idea of the magnitude of the leakage, before the experiment the two houses were found to have 1.62 and 1.54 ac/h, respectively, at 50 Pa pressure difference.

Instrumentation

For validation-quality datasets, it is necessary to have a comprehensive calibrated suite of sensors that measure all important flow paths. Both Twin Houses are equipped with a building management system. All sensors are sampled once per second. These measurement data are averaged and stored with a 1-minute frequency. These data were averaged and provided to modelling teams in both 10-minutely and hourly averaged formats (with the 1-minutely data available on request).

Inside both Twin Houses, the sensors listed in Table 3 were used. These sensors were calibrated 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 on-site weather station were provided as boundary conditions. These sensors are calibrated regularly as recommended by the manufacturer.

Validation experiment 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

Table 3. Sensors and accuracy.

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	$\pm 2.0\%$
Air temperatures in the cellar and attic spaces	± 0.14 K	Ground temperatures, depth of 0, 50, 100 and 200 cm	
Relative humidity in the living room	$\pm 2.3\%$	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)	$\pm 1.0°$
Heating power of the six heated rooms	$\pm 1.5\%$	Solar radiation: global, diffuse and vertical (north, east, south, west)	$\pm 2.0\%$
Supply and exhaust fan power	$\pm 1.5\%$	Longwave radiation (horizontal, west vertical)	< 34 W/m^2
Ventilation flow rates	± 3.5 m^3/h		
Heat flux at the west facade	± 0.65 W/m^2		
West wall temperatures: internal, external and between layers	± 0.14 K		

- Ventilation system details
- Electrical heater details

Details of all sensors and their calibration were provided, together with all boundary conditions (attic and cellar temperatures and weather data). The datasets collected were continuous apart from a few hours of missing sensor data due to a logging failure. As part of the quality assurance of data, these missing data periods were filled by interpolation and were considered of minor importance.

Modelling teams were then requested to make predictions of the temperatures and heating power for the various experimental periods, and provide these in a standard format together with a modelling report outlining the simulation programme used and any assumptions made. In the course of the work, an email hotline was set up to answer questions arising during the modelling – clarifications were then posted to all teams. For example, more accurate measurements of glazing and frame areas were provided.

Modelling teams

Modelling teams participating in this blind validation phase are listed in Table 4. There is a good range of programmes, both research and commercial.

Modeller reports

Each submission was accompanied by a filled-in questionnaire covering the main algorithms used within the simulation programme, and many included a detailed modelling report on assumptions made and sensitivity studies undertaken. Table 5 gives the details of the questionnaire.

Table 5. Model questionnaire.

Organization
Modeller
Simulation programme & version
Simulation time step
Number of zones
Solar diffuse sky model
Shading due to: Window reveal/surrounding buildings
Window modelling: Use transmission and absorption properties/use solar heat gain factor/use angle dependency
Blind modelling: Optical and thermal properties adjusted when blind operated/only optical properties adjusted
Internal convection coefficients: Fixed or dependent on delta-T and/or air change rate
Thermal bridges included
Boundary conditions: Fixed temperature/time-varying measured data
Temperature used for temperature control in the living room: Average of three measured values/middle height sensor
Ventilation: Was infiltration superimposed on ventilation rate? If so, what infiltration rate was used?
Internal solar radiation distribution: Calculated or assumed
Internal longwave exchange: View factor calculations or area/emissivity weighted
External longwave model

The following observations were made on the reports received:

- Most modellers modelled each space as a separate room. A few modellers combined the south rooms (as 1 or 2 zones) and the north rooms (as 1 zone).

Table 4. Experiment 1 blind validation: participating modellers.

Organization	Country	Programme
CIEMAT	Spain	TRNSYS
Czech Technical University 1	Czech Republic	Matlab
Czech Technical University 2	Czech Republic	Matlab_Simulink
Danish Technical University	Denmark	ESP-r Release 12
University of Gent	Belgium	TRNSYS Version 16
Hong Kong City University 1	Hong Kong	eQuest Version 3.65
Hong Kong City University 2	Hong Kong	EnergyPlus Version 8.0.0
IES	UK	IESVE Version 2013.2.0.3
Equa Solutions	Sweden	IDA-ICE 4.6 Beta 19
Fraunhofer Institute for Building Physics 1	Germany	TRNSYS Version 17
Fraunhofer Institute for Building Physics 2	Germany	WUFI Plus 2.5.3.9
University of Liege_HEPL	Belgium	EES
University of Liege_JCG	Belgium	EES
University of Liege_Ulg	Belgium	Modelica: no library
Politecnico di Milano	Italy	EnergyPlus Version 8.1.0
University of Strathclyde	UK	ESP-r Release 12
University Innsbruck	Austria	Dynbil Version 0.8.1
University of Leuven 1	Belgium	Modelica_model_1: IDEAS library Build 01.12.2013
University of Leuven 2	Belgium	Modelica_model_2: IDEAS library Build 23.12.2013
University of Leuven 3	Belgium	Modelica_model_3: IDEAS library Build 23.12.2013
University of Leuven 4	Belgium	TRNSYS Version 17

Table 6. Modelling of internal convection coefficients.

Method	Internal convection coefficients
1	Khalifa and Marshall (1990) correlations when heater on; Alamdari and Hammond (1983) correlations when off
2	Horizontal surfaces: based on temperature difference and heat flow direction. Vertical surfaces: a constant value of 2.5 W/m^2 K
3	Fixed value of internal surface resistance (0.13 m^{2}K/W for walls, 0.10 m^{2}K/W for ceiling, 0.17 m^{2}K/W for floor). These values include longwave radiation contribution as well. The model merges convection and longwave radiation at interior surfaces
4	Dependent on air change rate
5	Fixed coefficients used (EN ISO 6946: 2007)
6	Assumed fixed surface coefficients for combined radiation/convection
7	Variable, dependent on temperature difference
8	Dependent on temperature difference (natural convection assumed)
9	Fixed coefficients (all internal surfaces = 3 W/m^{2}K)
10	Assumed buoyancy-driven convection – Alamdari and Hammond (1983) correlations

- There was a large divergence of techniques for modelling thermal bridges. Some programmes did not provide for thermal bridge input. In some cases, these were omitted; in others, additional heat loss surfaces were introduced with thermophysical properties adjusted to match the provided linear heat loss coefficients. None of the modeller reports indicated that it was possible to directly input linear heat loss coefficients for internal bridges (between the internal walls and the cellar and attic).
- There was a large variation in modelling the distribution of solar transmission and distribution. Some programmes used supplied total solar energy transmittance (*g*-values); others used the detailed angle-dependent transmission/absorption/reflection data.
- Modelling of internal convection coefficients was also variable. Table 6 shows the distinct modelling methods that were reported.

Results of blind validation

Figure 6 shows the prevailing external air temperature and global horizontal irradiation for the experimental period after the initialization.

Some representative graphs are presented of modelling predictions and measured data to indicate the variability (Figures 7–10). These examples are blind validation results (i.e. modelling teams had not seen the measured data) for the living room of house O5 which had open blinds.

Figures 7 and 8 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 programmes showed qualitatively good agreement with measurements.

Figures 9 and 10 show the predicted and measured living room temperatures in the same house during the ROLBS input sequence. Again, a few models are clearly erroneous, whereas others follow the trends well.

Even where models showed good qualitative fit, there could be big differences in the degree of agreement between different periods and between temperature and heat input predictions. To give an overall comparison between the different models, two metrics were used to summarize the level of agreement.

1. The magnitude fit was defined as the absolute average difference between measurement and prediction for each experimental period in each

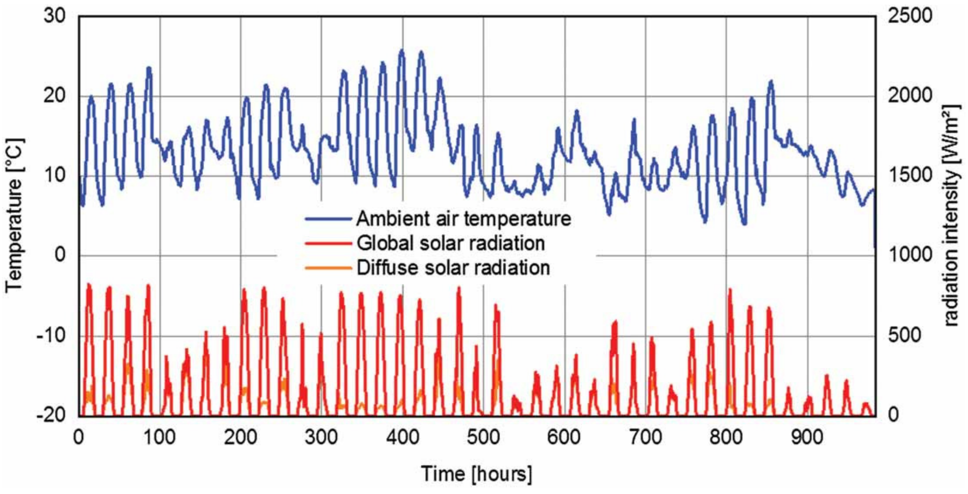


Figure 6. Ambient temperature and global horizontal irradiation.

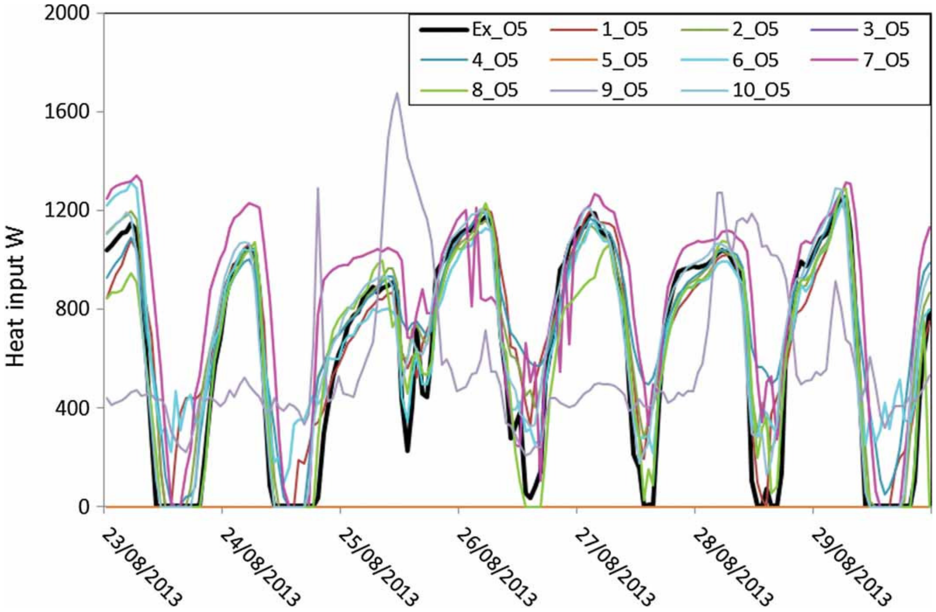


Figure 7. Living room heat input: constant temperature phase (30°C): models 1–10 + experimental data: House O5.

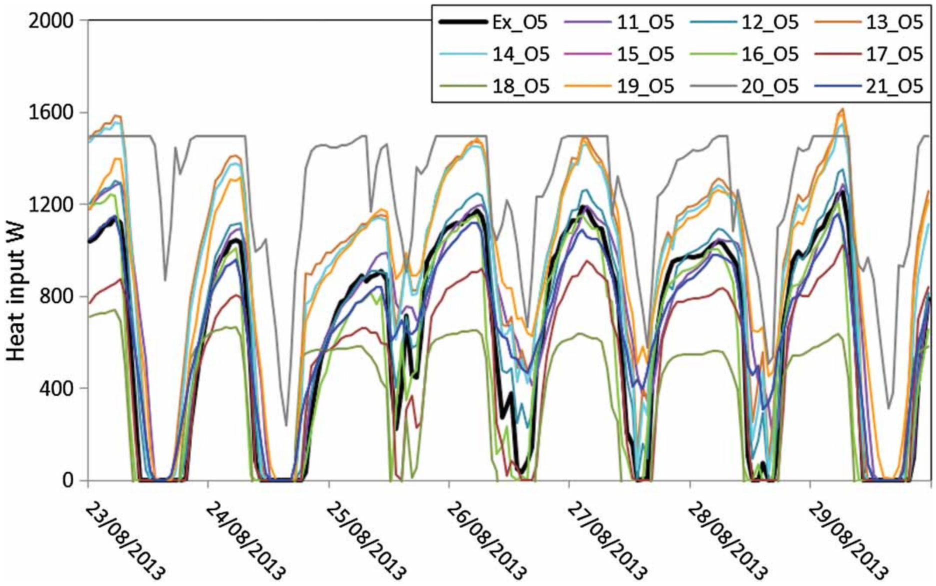


Figure 8. Living room heat input: constant temperature phase (30°C): models 11–21 + experimental data: House O5.

room. For the few programmes where rooms were combined, the same predicted temperatures were used in all the rooms.

2. The level of correspondence in the shape of the profile was given by Spearman's rank correlation coefficient (Kendall and Gibbons 1990) between predictions and measurements.

Table 7 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 of 1–2°C; orange in the range of 2–4°C; red in the range of 4–8°C and purple showing outliers > 8°C.

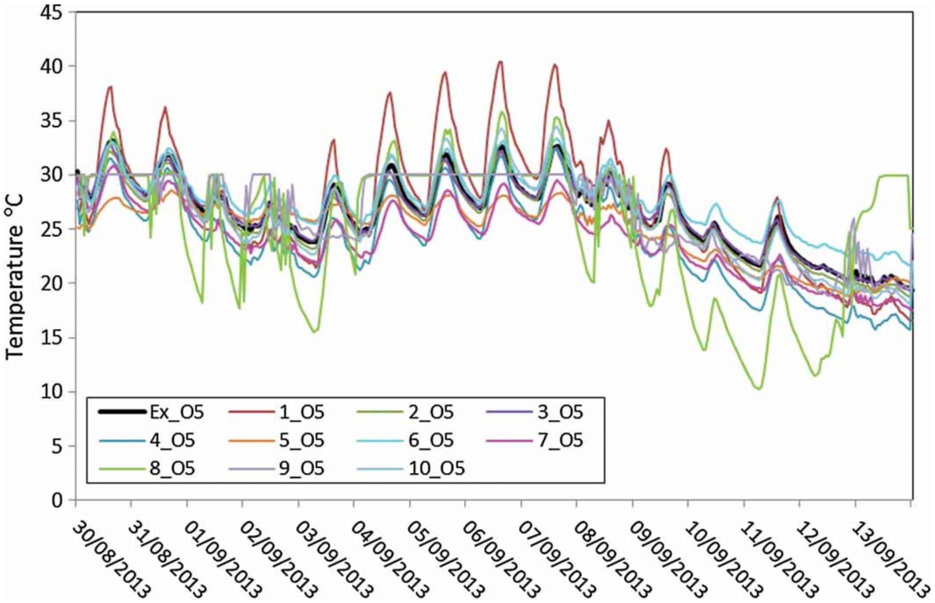


Figure 9. Living room temperature: ROLBS sequence: models 1–10 + experimental data: House O5.

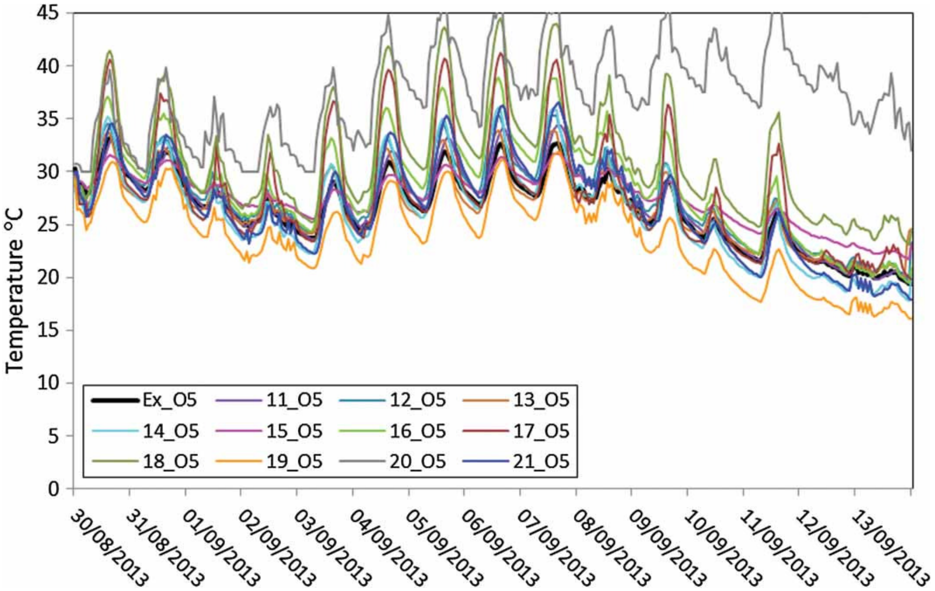


Figure 10. Living room temperature: ROLBS sequence: models 11–21 + experimental data: House O5.

As seen in the timeline comparisons, some submissions are clearly erroneous, but others show good levels of agreement overall. No programme predicted temperature in every room and every period within 1°C although two simulations came close. The bottom of the table shows the same data 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 programme inputs. The differences occur mainly because in the experiment there were a few times during the constant temperature periods when the set point was exceeded – especially in the living room with its large south-facing windows, and in most cases modellers assumed the fixed set point rather than using the measured temperatures.

Table 8 shows 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.

Table 9 shows the difference between the model predictions of heating to maintain the set point and the measurements in the constant temperature periods: period 2 at 30°C

Table 7. Blind validation 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 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 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 8. Blind validation 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 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 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**

and period 4 at 25°C. In this case, green represents agreement of better than 100 W, yellow is 100–200 W, orange is 200–300 W, red is 300–500 W and purple is > 500 W. The data at the bottom of the table show 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 programmes included the ROLBS heat inputs as casual gains rather than heater inputs, which accounts for those where the difference is around 240 W. Simulation results 5, 8, 9 and 20, however, show large errors which

Table 9. 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		*Where difference is 240W in ROLBS period, ROLBS heat input was modelled as casual gain*																				
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

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

Table 10. Blind validation results for the 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**

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 10 shows Spearman's rank correlation coefficient between the measurements and predicted temperatures for the constant temperature periods.

Results of re-modelling

After the blind validation phase, all measurements were supplied to the modelling teams. They were encouraged to compare their predictions with measurements, adjust their models if user errors or model deficiencies were identified and then resubmit, with a clear report of what changes had been made in order to ensure no tuning of models occurred.

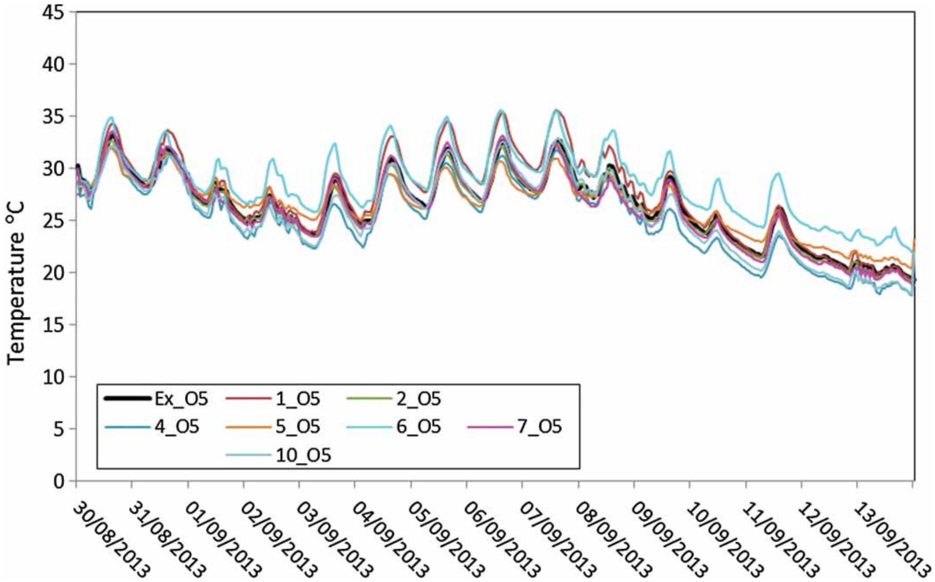


Figure 11. Living room temperature: ROLBS sequence: models 1–10 + experimental data: House O5.

As a result of comparing the blind validation results with measurements, a few specification and experimental errors were identified, so the teams were also supplied 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.
- 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).

A total of 14 submissions were made in this phase of the exercise. (Additional contributions were subsequently received from HFT Stuttgart using the INSEL programme and from the University of Liege with TRNSYS, not included in this analysis.) A representative example of the improved agreement is shown in Figure 11 – there are some anomalous programmes 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 11–14 correspond to Tables 7–10, but for the re-modelled submissions.

An additional metric was generated for this re-modelled data. Table 15 shows the total heating energy for the constant temperature heating periods: period 2 (30°C) and period 4 (25°C) for the combined rooms: living room, south bedroom, kitchen and north bedroom.

Discussion of results

For the comparisons shown in this paper, the experimental data uncertainties are small. As shown in Table 3, the individual calibrated shielded temperature sensors have an accuracy of better than 0.15°C and the heating power accuracy is ± 1.5%. However, some stratification was observed in the living room where the topmost temperature sensor recorded between 1°C and 2°C higher than the middle and lower sensors. Some modellers used the average of the three sensors; others used the middle sensor in order to represent the well-mixed room assumption of all the models used in this exercise. So, a reasonable estimate of the room-averaged measured temperature accuracy is in the order of 0.5–1°C.

Regarding the overall validation exercise:

- The results submitted cover a large range of capabilities in terms of the programmes used (simplified to detailed) and user capability (individual Ph.D. researchers to commercial companies undertaking internal QA before submitting results).
- Not all submissions can be classified as programme validation: the full capability of a programme is not always used. For example, a few modellers

Table 11. 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 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 LRT	ROLBS	1.0	0.4		1.3	0.8	2.2	0.4	3.4		0.8	0.3					1.0		0.6	2.5	0.4	1.1
O5 LRT	Free	0.4	0.4		1.8	0.7	1.4	0.3	3.1		1.3	0.3					0.7		0.4	2.4	0.4	0.7
N2 LRT	ROLBS	0.7	0.6		0.9	0.6	0.6	0.6	2.7		0.8	0.4					0.4		0.5	1.8	0.6	1.3
N2 LRT	Free	1.1	0.6		1.0	0.5	0.2	0.5	2.1		1.2	0.5					0.4		0.4	1.8	0.9	1.5
N2-O5 LRT	ROLBS	1.6	0.4		0.6	0.5	2.6	0.3	1.0		0.2	0.5					1.0		0.3	0.7	0.7	1.3
N2-O5 LRT	Free	1.1	0.4		0.8	0.7	1.2	0.3	1.3		0.2	0.5					0.7		0.3	0.6	0.7	1.3
O5 SBDT	ROLBS	2.2	0.4		1.2	0.8	2.6	0.5	7.5		0.4	0.3					1.0		1.4	2.0	1.0	0.5
O5 SBDT	Free	1.4	0.3		1.2	0.8	2.0	0.4	5.0		0.3	0.2					0.7		1.0	1.6	0.7	0.5
N2 SBDT	ROLBS	0.9	0.2		0.9	0.7	0.6	0.5	5.8		0.3	0.2					0.6		0.6	1.6	0.6	0.9
N2 SBDT	Free	1.1	0.3		0.7	1.0	0.9	0.5	3.5		0.6	0.3					0.6		0.5	1.1	1.2	0.6
N2-O5 SBDT	ROLBS	1.9	0.5		0.3	0.4	2.5	0.4	1.7		0.5	0.4					0.9		1.3	0.4	1.0	1.3
N2-O5 SBDT	Free	0.7	0.4		0.6	0.5	1.1	0.3	1.5		0.4	0.4					1.0		1.1	0.5	1.0	1.1
O5 KITT	ROLBS	0.8	0.3		2.0	0.7	0.8	0.4	2.4		0.6	0.5					0.7		0.4	3.4	0.4	1.3
O5 KITT	Free	0.5	0.2		1.5	0.5	1.1	0.3	1.6		0.8	0.2					0.5		0.4	3.8	0.4	0.8
N2 KITT	ROLBS	0.5	0.5		2.1	0.6	0.6	0.3	2.5		0.6	0.4					0.7		0.4	1.4	0.5	2.0
N2 KITT	Free	0.3	0.4		1.4	0.5	1.0	0.3	1.6		0.9	0.2					0.5		0.4	1.3	0.7	0.5
N2-O5 KITT	ROLBS	1.2	0.3		0.1	0.3	0.7	0.4	0.4		0.4	0.1					0.4		0.1	1.9	0.2	0.6
N2-O5 KITT	Free	0.7	0.2		0.2	0.3	0.2	0.3	0.5		0.2	0.2					0.3		0.1	2.4	0.3	0.3
O5 NBDT	ROLBS	1.2	0.1		1.3	0.5	1.1	0.4	5.2		0.5	0.3					2.9		2.1	0.9	0.5	0.6
O5 NBDT	Free	1.0	0.1		1.2	0.5	1.2	0.4	3.5		0.2	0.1					1.6		1.0	0.9	0.3	0.4
N2 NBDT	ROLBS	0.7	0.2		0.9	0.8	0.7	0.7	4.5		0.6	0.2					1.5		0.9	0.5	0.3	0.6
N2 NBDT	Free	0.7	0.2		0.7	1.0	1.2	0.7	2.9		0.2	0.3					0.6		0.5	0.5	0.5	0.5
N2-O5 NBDT	ROLBS	0.6	0.1		0.4	0.5	0.5	0.3	0.8		0.1	0.4					1.5		1.4	0.4	0.3	0.2
N2-O5 NBDT	Free	0.3	0.1		0.5	0.4	0.2	0.3	0.7		0.1	0.3					1.2		1.3	0.4	0.5	0.1
Fixed temperature periods																						
O5 LRT	30°C	0.3	0.4		0.4	0.4	0.4	0.3	0.3		0.4	0.4					0.5		0.4	1.0	0.3	0.4
O5 LRT	25°C	0.2	0.2		0.2	0.2	0.2	0.2	0.1		0.2	0.2					0.3		0.2	0.7	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					0.7		0.2	0.3	0.2	0.2
N2 LRT	25°C	0.2	0.2		0.3	0.2	0.2	0.2	0.2		0.2	0.2					0.7		0.2	0.3	0.2	0.2

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

Table 12. 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

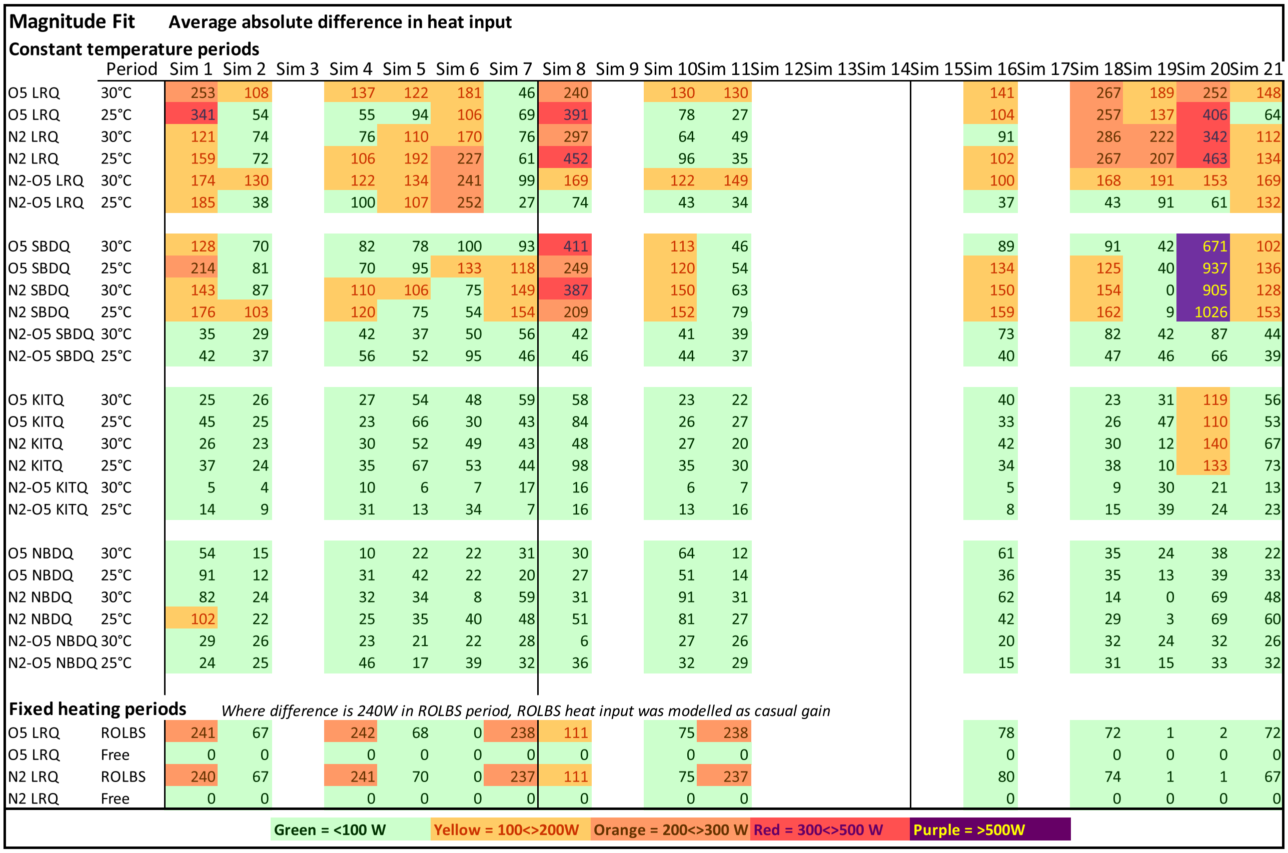
	Period	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6	Sim 7	Sim 8	Sim 9	Sim 10	Sim 11	Sim 12	Sim 13	Sim 14	Sim 15	Sim 16	Sim 17	Sim 18	Sim 19	Sim 20	Sim 21
O5 LRT	ROLBS	0.98	0.99		0.98	0.97	0.95	0.99	0.94		0.98	0.99					0.96		0.98	0.99	0.99	0.98
O5 LRT	Free	0.97	0.99		0.99	0.94	0.93	0.99	0.98		0.99	0.99					0.95		0.98	0.98	0.98	0.98
N2 LRT	ROLBS	0.99	0.99		0.98	0.95	0.97	0.98	0.90		0.99	1.00					0.99		0.98	0.98	0.99	0.97
N2 LRT	Free	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	0.89
O5 SBDT	ROLBS	0.96	0.99		0.97	0.95	0.94	0.99	0.92		0.99	1.00					0.88		0.84	0.98	0.98	0.98
O5 SBDT	Free	0.98	0.98		0.92	0.91	0.95	0.98	0.99		0.99	1.00					0.80		0.79	0.98	0.98	0.97
N2 SBDT	ROLBS	0.83	0.99		0.97	0.89	0.86	0.91	0.80		0.98	0.99					0.90		0.88	0.97	0.98	0.98
N2 SBDT	Free	0.97	0.99		0.99	0.98	0.98	0.98	0.74		0.99	1.00					0.98		0.98	0.98	0.98	0.98
O5 KITT	ROLBS	0.99	0.99		0.95	0.88	0.94	0.96	0.92		0.98	0.98					0.94		0.98	0.99	0.98	0.98
O5 KITT	Free	0.98	0.98		0.94	0.92	0.98	0.96	0.95		0.98	0.99					0.94		0.98	0.98	0.98	0.82
N2 KITT	ROLBS	0.99	0.98		0.96	0.91	0.90	0.99	0.86		0.98	0.98					0.95		0.98	0.97	0.99	0.95
N2 KITT	Free	0.99	0.99		0.98	0.97	0.99	0.98	0.88		0.99	1.00					0.96		0.98	0.97	0.99	0.92
O5 NBDT	ROLBS	0.96	0.99		0.97	0.91	0.93	0.97	0.79		0.98	0.95					0.72		0.83	0.98	0.96	0.98
O5 NBDT	Free	1.00	1.00		1.00	0.99	1.00	1.00	0.82		0.99	1.00					0.67		0.78	0.98	0.97	0.96
N2 NBDT	ROLBS	0.98	0.99		1.00	0.98	0.97	0.99	0.65		0.98	0.99					0.85		0.90	0.99	0.98	0.99
N2 NBDT	Free	1.00	1.00		1.00	1.00	1.00	1.00	0.74		1.00	1.00					0.93		0.99	0.99	1.00	0.99

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

combined rooms even though the programme used was capable of modelling all spaces in the building. In other cases, combined surface convective and radiative coefficients were used, although the programme was capable of separate coefficients being specified.

- There are no clear-cut programmes which are markedly better than others. However, programmes that were closest to the measured data tended to be those undertaking detailed solar modelling.
- As a result of the exercise, model flaws in the internal treatment of the sky longwave thermal radiation

Table 13. Re-modelling results for the constant temperature periods: Heat input magnitude fit.



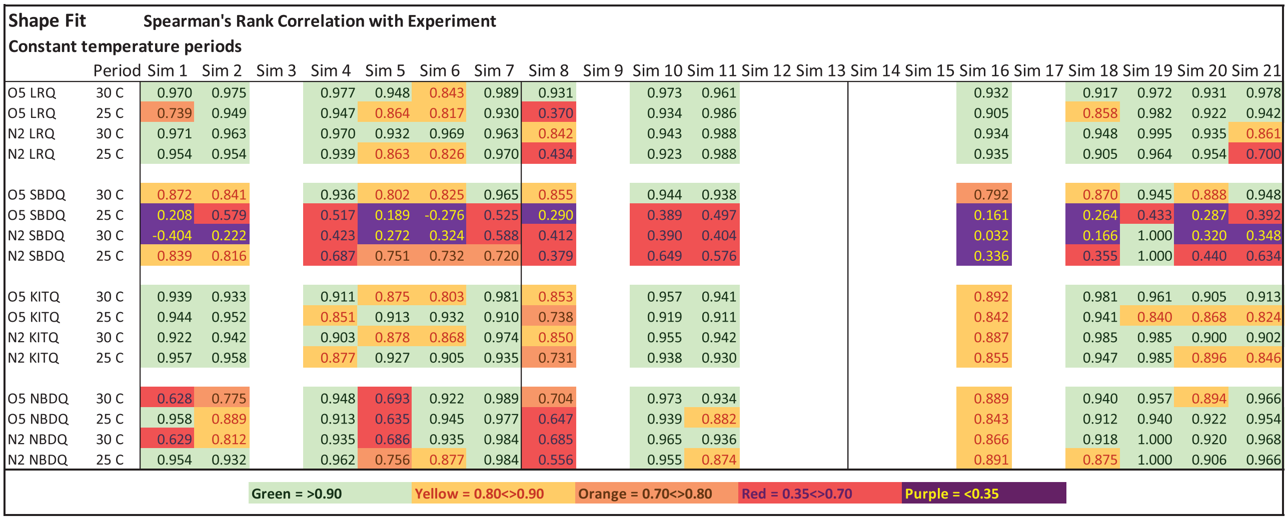
Magnitude Fit **Average absolute difference in heat input**

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	253	108		137	122	181	46	240		130	130					141		267	189	252	148
O5 LRQ	25°C	341	54		55	94	106	69	391		78	27					104		257	137	406	64
N2 LRQ	30°C	121	74		76	110	170	76	297		64	49					91		286	222	342	112
N2 LRQ	25°C	159	72		106	192	227	61	452		96	35					102		267	207	463	134
N2-O5 LRQ	30°C	174	130		122	134	241	99	169		122	149					100		168	191	153	169
N2-O5 LRQ	25°C	185	38		100	107	252	27	74		43	34					37		43	91	61	132
O5 SBDQ	30°C	128	70		82	78	100	93	411		113	46					89		91	42	671	102
O5 SBDQ	25°C	214	81		70	95	133	118	249		120	54					134		125	40	937	136
N2 SBDQ	30°C	143	87		110	106	75	149	387		150	63					150		154	0	905	128
N2 SBDQ	25°C	176	103		120	75	54	154	209		152	79					159		162	9	1026	153
N2-O5 SBDQ	30°C	35	29		42	37	50	56	42		41	39					73		82	42	87	44
N2-O5 SBDQ	25°C	42	37		56	52	95	46	46		44	37					40		47	46	66	39
O5 KITQ	30°C	25	26		27	54	48	59	58		23	22					40		23	31	119	56
O5 KITQ	25°C	45	25		23	66	30	43	84		26	27					33		26	47	110	53
N2 KITQ	30°C	26	23		30	52	49	43	48		27	20					42		30	12	140	67
N2 KITQ	25°C	37	24		35	67	53	44	98		35	30					34		38	10	133	73
N2-O5 KITQ	30°C	5	4		10	6	7	17	16		6	7					5		9	30	21	13
N2-O5 KITQ	25°C	14	9		31	13	34	7	16		13	16					8		15	39	24	23
O5 NBDQ	30°C	54	15		10	22	22	31	30		64	12					61		35	24	38	22
O5 NBDQ	25°C	91	12		31	42	22	20	27		51	14					36		35	13	39	33
N2 NBDQ	30°C	82	24		32	34	8	59	31		91	31					62		14	0	69	48
N2 NBDQ	25°C	102	22		25	35	40	48	51		81	27					42		29	3	69	60
N2-O5 NBDQ	30°C	29	26		23	21	22	28	6		27	26					20		32	24	32	26
N2-O5 NBDQ	25°C	24	25		46	17	39	32	36		32	29					15		31	15	33	32
Fixed heating periods		*Where difference is 240W in ROLBS period, ROLBS heat input was modelled as casual gain*																				
O5 LRQ	ROLBS	241	67		242	68	0	238	111		75	238					78		72	1	2	72
O5 LRQ	Free	0	0		0	0	0	0	0		0	0					0		0	0	0	0
N2 LRQ	ROLBS	240	67		241	70	0	237	111		75	237					80		74	1	1	67
N2 LRQ	Free	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 14. Re-modelling results for the 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	0.970	0.975		0.977	0.948	0.843	0.989	0.931		0.973	0.961					0.932		0.917	0.972	0.931	0.978
O5 LRQ	25 C	0.739	0.949		0.947	0.864	0.817	0.930	0.370		0.934	0.986					0.905		0.858	0.982	0.922	0.942
N2 LRQ	30 C	0.971	0.963		0.970	0.932	0.969	0.963	0.842		0.943	0.988					0.934		0.948	0.995	0.935	0.861
N2 LRQ	25 C	0.954	0.954		0.939	0.863	0.826	0.970	0.434		0.923	0.988					0.935		0.905	0.964	0.954	0.700
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	0.888	0.948
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	0.287	0.392
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	0.348
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	0.634
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	0.905	0.913
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	0.868	0.824
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	0.902
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	0.846
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	0.894	0.966
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	0.922	0.954
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	0.968
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	0.966

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

were identified by programme authors in two different programmes, and deficiencies in thermal bridge modelling was noted by others.

Regarding the blind validation:

- 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 high level of agreement between measurements and predictions. 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 was an interesting result which led to the identification of the heat losses in the kitchen to the

Table 15. Re-modelling results for the constant temperature periods: Total heat input.

Total heating input to living room, south bedroom, 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 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 (kWh)	30 C	199.7	151.1	200.1		176.2	208.9	191.2	173.0	229.1		166.9	201.5					161.8		150.8	224.4	177.9	170.4
O5 Difference with Measured (%)			-24.4	0.2		-11.8	4.6	-4.3	-13.4	14.7		-16.4	0.9					-19.0		-24.5	12.4	-10.9	-14.6
N2 (kWh)	30 C	263.2	209.8	260.0		226.1	265.8	286.7	211.3	271.0		211.7	256.5					214.6		184.9	302.5	235.5	212.2
N2 Difference with Measured (%)			-20.3	-1.2		-14.1	1.0	8.9	-19.7	3.0		-19.6	-2.5					-18.5		-29.7	15.0	-10.5	-19.4
N2-O5 (kWh)	30 C	63.5	58.7	59.8		49.9	56.9	95.5	38.3	41.9		44.8	55.0					52.8		34.1	78.2	57.7	41.8
N2-O5 Difference with Measured (%)			-7.5	-5.8		-21.4	-10.4	50.5	-39.7	-34.0		-29.4	-13.3					-16.8		-46.3	23.2	-9.2	-34.2
O5 (kWh)	25 C	217.9	126.9	228.9		231.2	243.2	211.1	208.6	203.5		193.2	230.4					191.7		172.0	255.1	221.1	193.6
O5 Difference with Measured (%)			-41.8	5.0		6.1	11.6	-3.1	-4.3	-6.6		-11.4	5.7					-12.0		-21.1	17.1	1.5	-11.2
N2 (kWh)	25 C	256.0	201.9	272.4		245.7	305.2	315.1	247.0	221.0		221.3	272.8					227.7		208.8	310.1	251.7	217.0
N2 Difference with Measured (%)			-21.1	6.4		-4.0	19.2	23.1	-3.5	-13.7		-13.6	6.6					-11.1		-18.4	21.1	-1.7	-15.2
N2-O5 (kWh)	25 C	38.1	81.8	46.6		15.3	66.8	112.4	41.2	18.4		30.2	45.4					38.3		39.1	59.0	34.3	24.4
N2-O5 Difference with Measured (%)			114.8	22.3		-59.8	75.6	195.3	8.3	-51.7		-20.6	19.4					0.6		2.6	55.1	-9.8	-35.9

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

uninsulated ductwork as a deficiency in the model specification.

- There are clearly several user input errors – in a few cases there is little correspondence with measured data, with the most probable explanation that the heat input scheduling was incorrect, and in one case a timing error either in model input or in output. The use of the summary tables (Tables 7–10) makes it easy to identify which prediction sets differ significantly from the measurements (and are likely to be due to user input error) – in this case, simulations 5, 8, 9 and 20 are obvious outliers.

Regarding the re-modelling:

- 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.
- The majority of the re-submitted results show good agreement in both the absolute predictions of temperatures and heat inputs, and the dynamic response. This holds for both Twin Houses and the differences between them. Given that 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.
- 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). Out of the 14 re-modelled simulation result sets, numbers 11 and 2 were consistently ranked first and second for overall agreement with the experiment in three tables: temperature magnitude and shape for fixed heat input periods; and heat input magnitude for constant temperature periods. But neither was in the top four for heat input shape. Numbers 7 and 10 also ranked among the best four except for temperature shape.
- Only one (no. 8) came out in the worst four in all four tables. No other simulation was in the worst four more than twice.
- One (no. 20) came among the best four in one category and the worst four in another.
- The heat input shape fit comes out better in the re-modelling than in the blind validation. Interestingly, the South Bedroom heat input was the worst modelled room but no obvious reason could be found.
- 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 predictions and measurements. Again, numbers 11, 2 and 20 were the best performers. 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

This paper has reported on an empirical validation study on full-size buildings under the auspices of IEA Annex 58. The specification for the validation experiment has been scrutinized and implemented by a large number of modellers (21 individual modellers or modelling teams) using a large variety of simulation programmes, and it has been refined following inputs from modellers for additional requested information. This final specification, together with the measured data, constitutes a high-quality empirical validation dataset on a full-scale, multi-zone building. The detailed experimental specification and experimental dataset, summarized in this paper, is provided as supplementary material available via the journal website. It is intended to be suitable for programme developers to test their programmes, as well as provide a template for organizing future empirical validation experiments.

The dataset collected comprises almost two months of experimental minute data. Detailed meteorological data are uninterrupted for this period, and the building data have only a few short gaps. Both 10-minutely and hourly averaged data are available to the modellers, with interpolation for any missing data, to provide a complete dataset. The experiments and data are for two identical buildings which were operated in the same manner, except for external solar shading differences, to provide a useful side-by-side experiment with high and variable levels of solar radiation in buildings with high amounts of thermal mass.

Although the specification and datasets are believed to be the most comprehensive yet available for a full-scale, multi-zone building, there are of course some limitations. The experiments had, by necessity, to be undertaken in summer months, so fabric losses were not tested through a large range of temperature differences. Similarly, the magnitude of internal heat injections had to be limited. Only one mechanical ventilation case was tested, and system and occupancy factors were deliberately excluded to reduce complexity. A further experiment was conducted at a cooler time of year on one of the Twin Houses with a lower ventilation rate, larger heat injections in another ROLBS sequence and additional sensors – this will be reported at a later date.

The modelling results showed a large range in levels of agreement with the experimental data. Some programmes showed excellent agreement, even at the blind validation stage. Overall, the better simulations seem to be better across all the rooms, test periods and different performance metrics, but not invariably. Given the extensive dataset, the fact that comparisons are made for several rooms in both houses and in terms of differences between the houses in the side-by-side experiment, some confidence can be expressed that these programmes can accurately model this building configuration.

Other submitted predictions showed poor agreement in the blind validation stage, largely and perhaps not unexpectedly, caused by user error, although in such cases it is not possible to say definitively whether the differences are caused by user error or programme deficiencies. Only one result set seemed to be consistently the worst. However, the number of input errors, given such a comparatively simple building, shows that much more work is needed by developers of simulation programmes to reduce errors. This is certainly not a new finding, but it does seem that the greater use of simulation programmes has not resulted in sufficient user training, or feedback and checking within programme interfaces. It is recommended that future studies are undertaken that focus on the types and impacts of user errors on larger scale building designs, with a view to informing programme developers.

Feedback from modellers demonstrated the importance of such experiments, and has led to improvements being made to programmes. In several cases, the treatment of thermal bridges was mentioned as requiring more attention. In many programmes, it is difficult to model thermal bridges – modellers need to calculate modified thermophysical properties or add additional constructions to represent the edge losses. Even programmes in which linear thermal transmittances could be defined were found to be unable to include thermal bridges associated with internal partitions. In one case, an incorrect sky temperature calculation was identified, leading to errors in the external longwave radiation transfer. Similarly, another modelling team using external temperature for the longwave heat transfer found, by analysing energy balances, the significant error that this assumption introduced.

Judging by the modelling reports, there are significant differences in modelling approaches between programmes, particularly for glazing transmission and internal convection. Several modelling teams are currently investigating this in more detail, using the measured surface and air temperatures, together with detailed sensitivity analyses and identification techniques.

The time and effort to conduct this empirical validation experiment was substantial, by the experimental team, the modellers and the analysis team. It is recognized that the experiment was conducted on a simple unoccupied building. Similar datasets are needed from other, larger building types, but it would require a high level of resourcing to undertake such an experiment with a similar level of detail as the experiment described in this paper.

Acknowledgements

The authors would like to recognize 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.

Funding

The UK authors acknowledge support from the Engineering and Physical Sciences Research Council (Ref: EP/K01532X/1). The German authors acknowledge support from the Federal Ministry for Economic Affairs and Energy (Ref: 03ET1144A).

Supplemental data

Supplemental data for this article can be accessed here. http://dx.doi.org/10.15129/8a86bbbb-7be8-4a87-be76-0372985ea228

References

Alamdari, F., and G. P. Hammond. 1983. "Improved Data Correlations for Buoyancy-Driven Convection in Rooms." *Building Services Engineering Research and Technology* 4 (3): 106–112.

ASHRAE Standard 140–2011. 2011. "Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs." ASHRAE, Atlanta, GA.

CARB. 2010. "Special Issue on Carbon Reduction in Existing Buildings."*Building Research & Information* 38 (1): 1–129.

CarbonBuzz. 2014. www.carbonbuzz.org.

van Dijk, H. A. L., and F. M. Tellez. 1995. *Measurement and Data Analysis Procedures*. Final Report of the JOULE II COMPASS Project (JOU2-CT92-0216).

EN ISO 6946. 2007. *Building Components or Building Elements – Calculation of Thermal Transmittance*.

IEA Annex 21. 1995. http://www.ecbcs.org/annexes/annex21.htm.

IEA Annex 58. 2015. *Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements*. http://www.kuleuven.be/bwf/projects/annex58.

Janssens, A. 2014. *State of the Art of Full Scale Test Facilities for Evaluation of Building Energy Performances*. IEA Annex 58 Subtask 1 Report. Available from http://www.kuleuven.be/bwf/projects/annex58/index.htm.

Jensen, S. O., ed. 1993. *Validation of Building Energy Simulation Programs, Part I and II*. Research Report PASSYS Subgroup Model Validation and Development, CEC, Brussels, EUR 15115 EN.

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

Judkoff, R., D. Wortman, R. O'Doherty, and J. Burch. 1983. *A Methodology for Validating Building Energy Analysis Simulations*. SERI/TR-254-1508, Golden, CO: SERI (now NREL).

Kalyanova, O., P. Heiselberg, C. Felsmann, H. Poirazis, P. Strachan, and A. Wijsman. 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, July, 27–30.

Kendall, M., and J. D. Gibbons. 1990. *Rank Correlation Methods*. 5th ed., 69–77. London: Edward Arnold.

Khalifa, A. J. N., and R. H. Marshall. 1990. "Validation of Heat Transfer Coefficients on Interior Building Surfaces Using a Real-Sized Indoor Test Cell." *International Journal of Heat Mass Transfer* 33 (10): 2219–2236.

Lomas, K. J., H. Eppel, C. Martin, and D. Bloomfield. 1994. *Empirical Validation of Thermal Building Simulation Programs using Test Room Data*, IEA Annex 21/Task 12 Project, Final Report, Vols. 1–3.

Loutzenhiser, P. G., H. Manz, C. Felsmann, P. Strachan, and G. M. Maxwell. 2007. "An Empirical Validation of Modeling Solar Gain through a Glazing Unit with External and Internal Shading Screens." *Applied Thermal Engineering* 27 (2–3): 528–538.

Munich IWEC. 2014. http://apps1.eere.energy.gov/buildings/energyplus/weatherdata_about.cfm.

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

Strachan, P., G. Kokogiannakis, and I. Macdonald. 2008. "History and Development of Validation with the ESP-r Simulation Program." *Building and Environment* 43 (4): 601–609.

THERM. 2014. http://windows.lbl.gov/software/therm/therm.html.

TSB Retrofit. 2012. https://retrofit.innovateuk.org/.

Turner, C., and M. Frankel. 2008. *Energy Performance of LEED for New Construction Buildings*. New Buildings Institute.