PREFACE

International Energy Agency

The International Energy Agency (IEA) was established in 1974 within the framework of the Organization for Economic Co-operation and Development (OECD) to implement an International Energy Programme. A basic aim of the IEA is to foster co-operation among the twenty-one IEA Participating Countries to increase energy security through energy conservation, development of alternative energy sources and energy research development and demonstration (RD&D). This is achieved in part through a programme of collaborative RD&D consisting of forty-two Implementing Agreements, containing a total of over eighty separate energy RD&D projects. This publication forms one element of this programme.

Energy Conservation in Buildings and Community Systems Programme

The IEA sponsors research and development in a number of areas related to energy. In one of these areas, Energy Conservation in Buildings and Community Systems (BCS), the IEA is sponsoring various exercises to predict more accurately the energy use of buildings, including comparison of existing computer programs, building monitoring, comparison of calculation methods, as well as air quality and studies of occupancy. Seventeen countries have elected to participate in this area and have designated contracting parties to the Implementing Agreement covering collaborative research in this area. The designation by governments of a number of private organizations, as well as universities and government laboratories, as contracting parties, has provide a broader range of expertise to tackle the projects in the different technology areas than would have been the case if participation was restricted to governments. The importance of associating industry with government sponsored energy research and development is recognized in the IEA, and every effort is made to encourage this trend.

Overall control of the programme is maintained by an Executive Committee, which not only monitors existing projects but identifies new areas where collaborative effort may be beneficial. The Executive Committee ensures that all projects fit into a pre-determined strategy, without unnecessary overlap or duplication but with effective liaison and communication. The Executive Committee has initiated the following projects to date (completed projects are identified by *).

Annex 1: Load energy determination of buildings *
Annex 2: Ekistics & 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 of schools *
Annex 16: BEMS 1 - User interfaces and system integration
Annex 17: BEMS 2 - Evaluation and emulation techniques
Annex 18: Demand controlled ventilating systems
Annex 19: Low slope roofs systems
Annex 20: Air flow patterns within buildings
Annex 21: Calculation of energy & environmental performance of buildings
Annex 22: Energy efficient communities
Annex 23: Multizone air flow modelling
Annex 24: Heat, air & moisture transport in new and retrofitted insulated envelope parts

Annex 25: Real time simulation of HVAC systems and fault detection
Annex 26: Energy-efficient ventilation of large enclosures
Annex 27: Evaluation and demonstration of domestic ventilation systems
Annex 28: Low-energy cooling systems

Annex 21: Calculation of Energy and Environmental Performance of Buildings

The objectives of Annex 21 are to:

1) develop quality assurance procedures for calculating the energy and environmental performance of buildings by producing guidance on:
    - program and modelling assumptions
    - the appropriate use of calculation methods for a range of design applications
    - the evaluation of calculation methods
2) establish requirements and market needs for calculation procedures in building and environmental services design;
3) propose policy and strategic direction for the development of calculation procedures;
4) propose means to effect technology transfer of calculation procedures into the building and environmental services design profession.

The subtasks of this project are:

A. Documentation of Existing Methods
B. The Appropriate Use of Models
C. Reference Cases and Evaluation Procedures
D. Design Support Environment

The participants in this annex are: Belgium, France, Germany, Italy, the Netherlands, Switzerland and the United Kingdom. Canada, Finland and Sweden also participated in the early part of the project. In addition, Finland, Spain, Sweden and the United States participate in Subtask C as a collaborative research activity between Task 12 Subtask B of the IEA Solar Heating & Cooling Programme.

The UK Building Research Establishment acts as Operating Agent of BCS Annex 21.

Solar Heating and Cooling Programme

Initiated in 1977, the Solar Heating and Cooling (SHC) Programme was one of the first IEA R&D agreements. Its objective is to conduct joint projects between the 20 member countries to advance solar technologies for buildings.

A total of eighteen projects or "Tasks" have been undertaken since the beginning of the Programme. The overall programme is managed by an Executive Committee composed of one representative from each of the member countries, while the leadership and management of the individual Tasks is the responsibility of Operating Agents. These Tasks and their respective Operating Agents are (completed projects are identified by *, tasks in planning stage are identified by #):

Task 1: Investigation of the performance of solar heating and cooling systems - Denmark *
Task 2: Co-ordination of research and development on solar heating and cooling - Japan *
Task 3: Performance testing of solar collectors - United Kingdom *
Task 4: Development of an insulation handbook and instrument package - United States *
Task 5: Use of existing meteorological information for solar energy application - Sweden *
Task 6: Solar heating, cooling, and hot water systems using evacuated collectors - United States *
Task 7: Central solar heating plants with seasonal storage - Sweden *
Task 8: Passive and hybrid solar low energy buildings - United States *
Task 9: Solar radiation and pyranometry studies - Germany *
Task 10: Material research and testing - Japan *
Task 11: Passive and hybrid solar commercial buildings - Switzerland *

Task 12: Building energy analysis and design tools for solar applications - United States
Task 13: Advanced solar low energy buildings - Norway
Task 14: Advanced active solar systems - Canada
Task 15: Advanced central solar heating plants #
Task 16: Photovoltaics in buildings - Germany
Task 17: Measuring and modelling spectral radiation - Germany
Task 18: Advanced glazing materials - United Kingdom
Task 19: Solar air systems - Switzerland
Task 20: Solar retrofit systems - Sweden

Task 12: Building Energy Analysis and Design Tools for Solar Applications

The scope of Task 12 includes:

(1) selection and development of appropriate algorithms for modelling of the interaction of solar energy-related materials, components, and systems with the building in which these solar elements are integrated;

(2) selection of analysis and design tools, and evaluation of the algorithms as to their ability to model the dynamic performance of the solar elements in respect of accuracy and ease of use; and

(3) improvement of the usability of the analysis and design tools, through preparation of common formats and procedures and by standardization of specifications for input/output, default values, and other user-related factors.

The subtasks of this project are:

A) Model Development

B) Model Evaluation and Improvement

C) Model Use

The participants in this task are: Denmark, Finland, Germany, Norway, Spain, Sweden, Switzerland, and the United States. In addition, Belgium, France, Italy, and the United Kingdom participate in Subtask B as a collaborative research activity between Annex 21 Subtask C of the IEA Energy Conservation in Building and Community Systems Program.

Architectural Energy Corporation serves on behalf of the US Department of Energy as Operating Agent of SHC Task 12.

ACKNOWLEDGEMENTS

The authors are grateful to the UK Building Research Establishment for funding this work. We wish to thank the joint IEA BCS Annex 21, Subtask C and SHC Task 12, Subtask B members for their valuable contributions and support:

Viottorio Bocchio / Augusto Mazza, Politecnico di Torino, Italy (BLASTv3.0);
Pascal Dalicieux, EDF, France (CLIM2000v1.1);
Tapio Haapala / Timo Kalema, Tampere University of Technology, Finland (TASEv3.0);
Shirley Hammond, BRE (SERI-RESv12);
Foroutan Parand, BRE (TRNSYSv12.1 & 13.1);
Eduardo Rodriguez, Escuela Superiore Ingnieros Industriales, Seville, Spain (S3PASv2.0);
Peter Verstraete / Rik van de Perre, Vrije Universiteit Brussel, Belgium (TRNSYSv13.1);
Ron Judkoff, NREL, Chair of lEA BCS 21C / SHC 12B experts group
Michael Holtz, Architectural Energy Corporation

We are also indebted to all those who participated, without any dedicated funding, by running their thermal programs. Without these contributions the work would have been far less comprehensive:

Lorenzo Agnoletto, Instituto di Fisica Technica, Udine, Italy (WG6TCv1992);
Don Alexander, UWCC Cardiff (HTB2v1.10);
Angelo Delsante, CSIRO, Australia (CHEETAHv1.2);
Martin Gough / Alan Jones, EDSL (TASv7.54);
Mike Holmes, Arup R&D (ENERGY2v1.0);
Kjeld Johnsen, SBI, Denmark (TSBI3v2.0);
Mike Kennedy, ECOTOPE, USA (SUNCODEv5.7);
Brian Miller / Doug Hittle, Colorado State University, USA (BLASTv3.0);
Peter Moors, DMU (TASv7.54);
Peter Pfrommer, FHT Stuttgart, Germany (HTB2v1.2);
Paul Strachan, ESRU (ESP-rv7.7a);
Glenn Stuart, ASL Sterling (ESP+v2.1);
Andrew Tindale, FACET (3TCv1.0 & APACHEv65.3);
Jeff Thornton, University of Wisconsin, USA (TRNSYSv13.1);
Maria Wall / Petter Wallenten, Lund University, Sweden (DEROB vlth);
Fred Winkelmann, LBL, USA (DOE2v1E).

The permission of the UK Energy Technology Support Unit to use the data from the Energy Monitoring Company test rooms is gratefully acknowledged.

EXECUTIVE SUMMARY

Empirical validation of Detailed Thermal Simulation Programs (DSPs) involves comparing their predictions with actual measurements made in real buildings. It is the most obvious way to evaluate the accuracy of DSPs, but the most difficult to undertake convincingly. The work described in this report was undertaken within a joint group consisting of International Energy Agency (IEA) Energy Conservation in Buildings & Community Systems (BCS) Annex 21 Subtask C and IEA Solar Heating and Cooling (SHC) Task 12 Subtask B. The UK provided the measured data and managed the empirical validation study. This part of the work was conducted by the UK Building Research Establishment (BRE), with De Montfort University and the Energy Monitoring Company (EMC) working under sub-contracts. The work complements the DSP assessment tests based on inter-model comparisons and analytic tests which were also developed by the joint BCS Annex 21/SHC Task 12 group.

The aims of the work were to:

(a) develop well-documented and well-tested empirical validation benchmarks for detailed thermal simulation programs.

(b) provide a 'snapshot' of the ability of DSPs to predict the performance of a few simple buildings under conditions reflecting those which exist when they are used to model real buildings; and

(c) devise and test a strategy for developing empirical validation benchmarks.

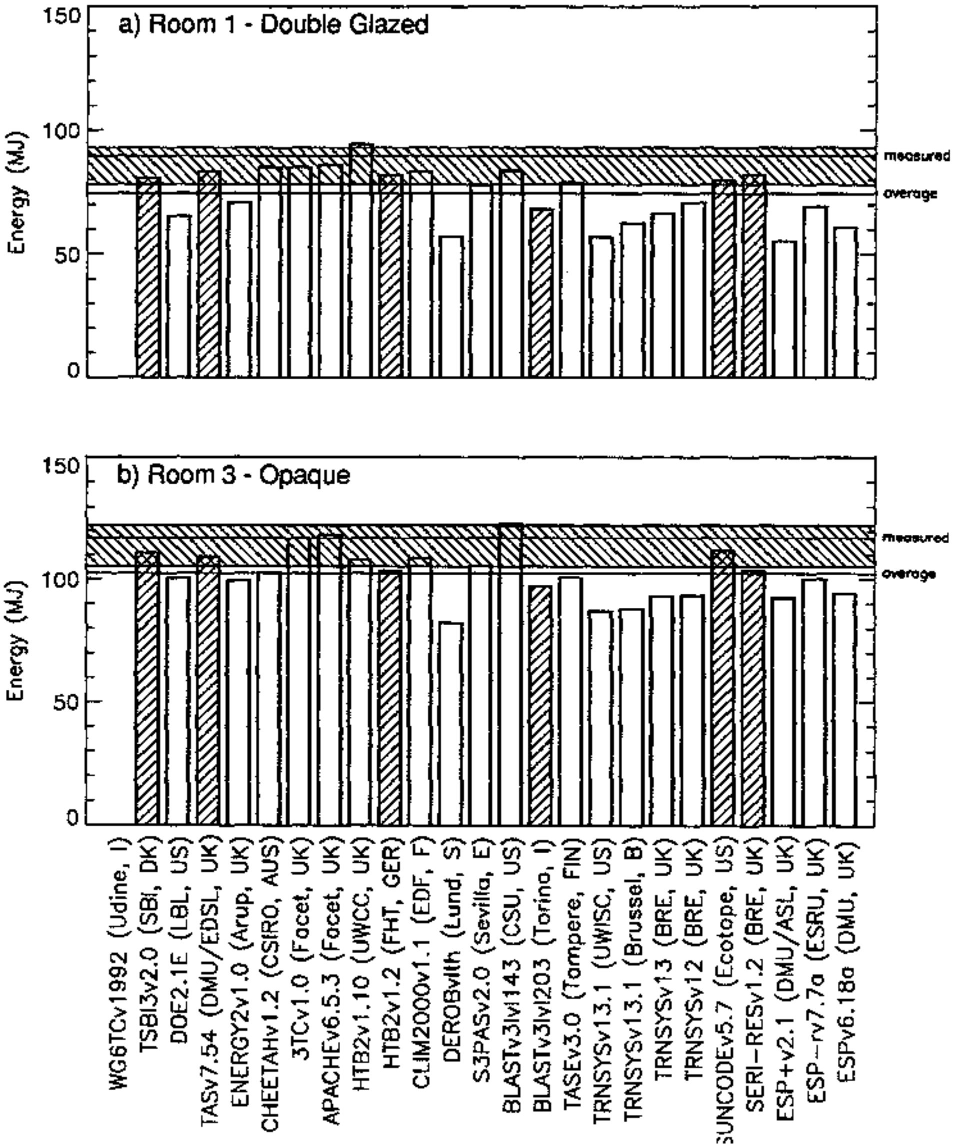
Reviews of previous work were undertaken to assist in directing the management of the exercise and to select suitable data sets upon which to base the empirical validation benchmarks. These reviews are documented in Volume 3 of this final report.

Data sets collected in three of the rooms operated by the EMC were chosen. The rooms were single-zoned and well insulated, they had very low air infiltration rates (less than 0.05 ac/h) and were raised clear of the ground. During the 10-day October period, the rooms were intermittently heated by an oil-filled radiator to 30°C. In the 10-day May period they were unheated (free-floating). Only the South facing facades of the rooms differed, two had different glazing and the third was opaque. The rooms were located on an unobstructed site on the edge of Cranfield airfield near Milton Keynes, UK. Five empirical benchmark tests were produced.

The rooms had been carefully constructed and a detailed Empirical Validation Package suitable for thermal modellers was produced (Volume 2 of this Final Report). To check that the rooms matched this description, after the experiment had been completed one room was dismantled and its construction checked. The overall heat loss coefficient (UA-value) of a room was also measured, to within the experimental error of ±5%, and found to match that calculated using the Site Handbook data (part of Volume 2). Periodic matching trials by the EMC demonstrated that all the rooms had, to within measurement error, the same construction. These tests were in addition to the customary quality assurance procedures adopted by the EMC to ensure the integrity of their data, described in Volume 2 of this report.

Initially, the only modellers involved were 6 of the IEA BCS Annex 21/SHC Task 12 participants. However, invitations to join the exercise were issued in order that a 'snap-shot' of a larger range of simulation program performances might be obtained. This attracted an additional 14 institutions and private companies. Most were either skilled users or the authors, vendors or support offices for the programs. In total, 25 results sets were obtained from 17 genuinely different programs, the remaining results being from alternative versions of some of these. The 17 programs (and country of origin) were: 3TC (UK); APACHE (UK); BLAST (USA); CHEETAH (Australia); CLIMA2000 (France); DEROB (USA); DOE2 (USA); ENERGY2 (UK); ESP (UK); HTB2 (UK); S3PAS (Spain); SERI-RES/SUNCODE (USA); TAS (UK); TASE (Finland); TRNSYS (USA); TSBI3 (Denmark); and WG6TC (Italy).

The program validation exercise was conducted in two phases. In Phase 1 the participants were provided with the Site Handbook and guidance on how to use it to model the rooms. They had no knowledge of the actual measurements, so the predictions were undertaken 'blind'. The primary parameters predicted were:



The hatched horizontal area represents the estimated uncertainty band

Figure El:
Phase 2 - Total heating energy consumption
for the 7-day heated (October) period

(i) The total heating energy consumption for the October heated period.

(ii) The maximum and minimum temperatures recorded in both the October heated period and the May free-floating period.

Also required were:

(iii) The total South facing vertical solar irradiance predicted for both periods;

(iv) The hourly temperatures, and, for the October period, the hourly power demands.

A hot-line service was set up so participants could resolve any modelling difficulties or uncertainties. These hot-line calls, and any other relevant information, were documented in regular Newssheets which were circulated to all the participants. The hot-line operators did not know what the measured performance of the rooms was. A complete analysis of all the results was undertaken after Phase 1.

In Phase 2, the program users were provided with all the measurements and the estimated uncertainties on the program input data, and invited to explore the differences between the program predictions and the measurements. During this phase, 6 participants made legitimate changes to their program input files, to correct mistakes and approximations. They produced a new, second set of results (which are indicated by shading in all figures).

During Phase 2 a number of issues concerning the reliability of the data sets were raised. These were investigated and any additional uncertainties were incorporated with the other, already known, sources of experimental error in order to calculate the total uncertainty in the primary parameters. The total uncertainty was such that if a program lay outside them there was a 99% chance that it was due to inaccurate predictions. This uncertainty was taken into account when assessing the performance of the programs after each phase. (It is shown as a horizontal shaded bar in all figures).

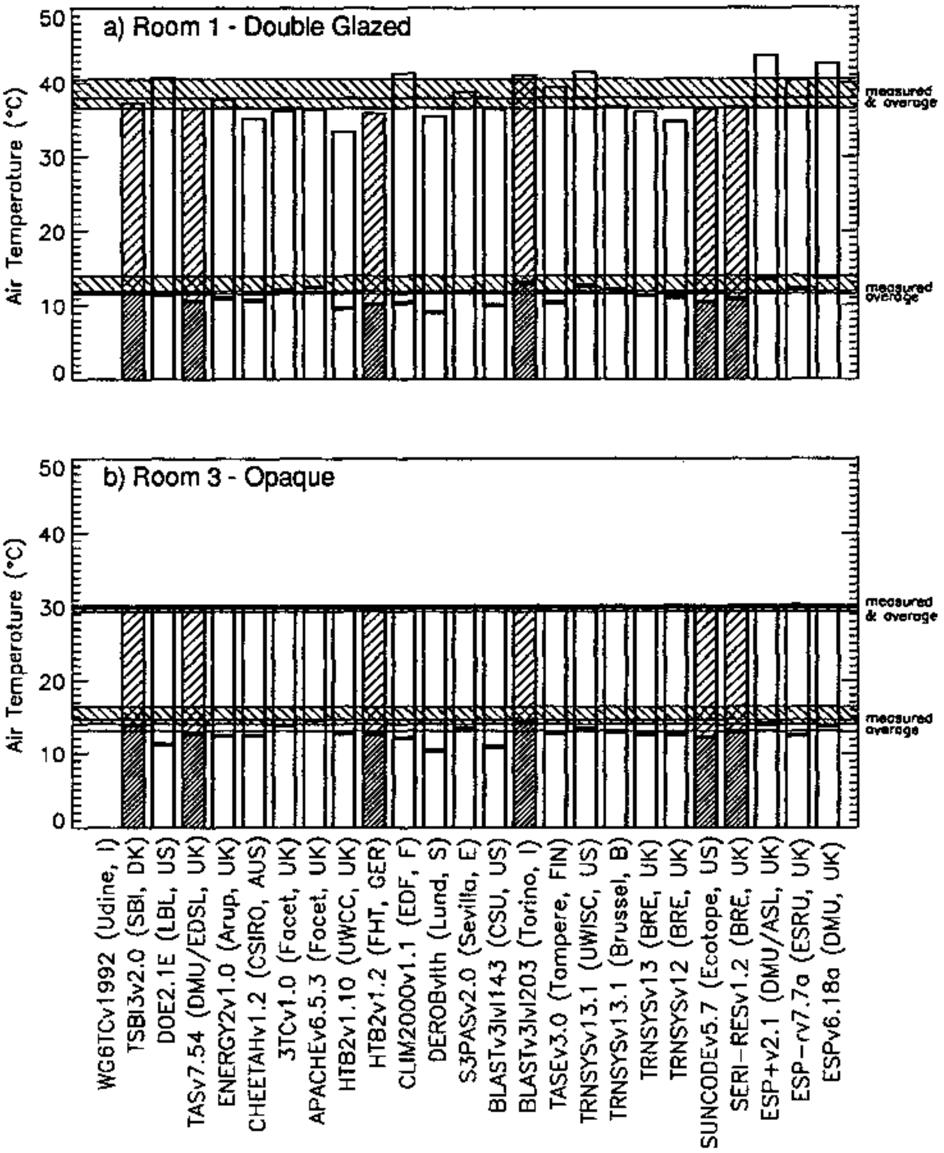
Some results from the heated rooms are shown in Figures El and E2. The following observations can be made.

- Eight of the programs predicted total energy consumptions within the error bars for the opaque room and eleven for the double glazed room.
- There was a tendency for most programs to predict energy consumptions and maximum and minimum temperatures which were lower than the measured values.
- One program predicted heating energy consumption and the maximum and minimum temperatures to within the error bars for both rooms.
- The energy consumption predictions of the programs varied by 40% (of their mean value) in the opaque room and 52% in the double glazed room.
- The predicted 'energy savings' due to replacing the opaque surface with double glazing varied by a factor of 3, from 13% to 40%.
- The predicted peak temperatures varied by 11°C, which was between 3°C and 14°C above the thermostat-set point.

The Phase 2 results from the unheated free-floating rooms (some are shown in Figure E3) reveal the following.

- Six of the programs failed to predict maximum and minimum temperatures in the opaque room which were within the uncertainty bands.
- Five programs predicted maximum and minimum temperatures in the double glazed rooms which were within the uncertainty bands, and only 2 in the single-glazed room.
- There was a general tendency to predict temperatures which were too low.
- The predicted peak temperatures in the double glazed room varied by 8.6°C, minimum temperatures varied by only 2.5°C. The corresponding variations in the opaque room were 2.8°C and 2.4°C.
- All the programs predicted that the peak temperatures in the double glazed room would exceed those in the opaque room. However, the predicted increase varied from 9.8°C to 17.2°C.

The results from this empirical validation exercise are shown to be broadly in line with those from the IEA inter-model comparison exercises based on domestic and commercial buildings. This suggests that the inter-program differences are due to fundamental differences in the programs themselves and not to



The hatched horizontal area represents the estimated uncertainty band

Figure E2:
Phase 2 - Maximum and minimum temperatures during the 7-day heated (October) period

other (user) effects.

Although the results are not shown here, all the programs predicted total south-facing solar irradiances which were within the estimated error bands in the May period, and thirteen programs produced such predictions in the October period. Although the identification of program errors was not the primary intention of the exercise, errors were found in three of the programs and corrected before undertaking Phase 2.

The model user reports produced in Phase 2 identified the following differences between features of the experiment and the assumptions made by their programs: the dynamics and output of the heater; internal heat transfer coefficients; and internal air movement and stratifications. It is argued that these features exist in many real buildings, and that, if they explain why some programs performed worse than others, then this must be viewed as a weakness in those programs.

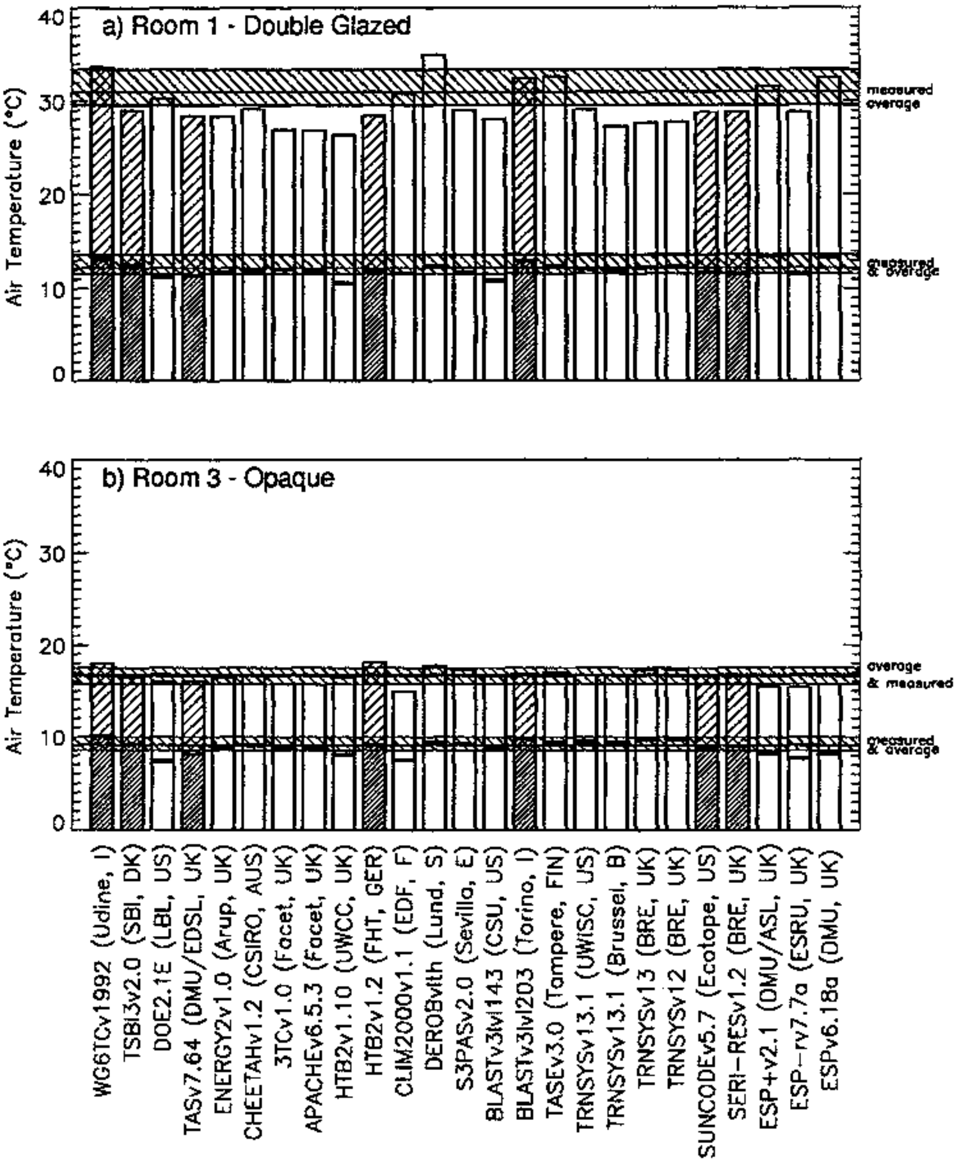
The heat output from the test room heaters had a large radiant component. The assumption, inherent in a number of programs, that heater output is convective, is consistent with the observed under-prediction of heating energy demand. It may be useful to complement this study with one in which convective heating and well-mixed zone air is employed. The issues of internal convection and air circulation, particularly over poorly insulated (glazed) surfaces, may also be fruitful areas for future research.

It has been argued that, where the predictions of programs lie outside the estimated error bands then this should be taken as a sign that the program is likely to contain errors. These may be

- the omission of important sub-models or algorithms;
- the use of inappropriate algorithms; or
- coding errors.

As a result of the work the following main conclusions can be drawn.

- This was the largest empirical validation exercise ever undertaken.
- A well documented benchmark has been developed. It is an exemplar of how to document an empirical validation benchmark.
- By incorporating a blind phase, added value is given to an empirical validation exercise. In such a phase it is important that a mechanism is established for providing modellers with all the information they need.
- Although none of the programs predicted primary parameters which were within the estimated uncertainty bands on all occasions, some of the programs clearly performed better than others.
- Some programs performed particularly well overall whilst others did not perform as well. Other programs performed well on some tasks and not on others.
- Many programs are likely to contain internal errors, i.e., they may omit the modelling of important processes; they may contain inappropriate algorithms; or they may contain coding errors.
- This work, and the other activities of the IE A BCS Annex 21/SHC Task 12 group, has clearly demonstrated that there is considerable scope for improving the predictive accuracy of DSPs.



The hatched horizontal area represents the estimated uncertainty band

Figure E3:
Phase 2 - Maximum and minimum temperatures during the 7-day free-floating (May) period

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Organisation

Detailed simulation programs (DSPs) are now extremely powerful, relatively inexpensive to purchase or license, and have the potential to dramatically improve the energy and environmental performance of buildings. However, their use is still limited albeit growing. To encourage wider use, the two key areas to be tackled concern their accuracy and their usability. These aspects are closely related, and both are currently being studied in International Energy Agency (IEA) Energy Conservation in Building and Community Systems (BCS) Annex 21. Usability is being addressed in Subtask A (program documentation), Subtask B (formalising and documenting building performance assessment methods) and Subtask D (design support environments).

This report describes the empirical validation work undertaken under the auspices of the group formed by combining lEA BCS Annex 21 Subtask C and lEA SHC Task 12 Subtask B. The work was directed by the UK Building Research Establishment (BRE), and managed by the Environmental Computer Aided Design and Performance (ECADAP) group in the School of the Built Environment at De Montfort University Leicester, and by the Energy Monitoring Company (EMC), Newport Pagnell, UK. The latter two participated via sub-contracts from the BRE.

This report is part of a 3 volume set, produced by the UK participants:

Volume 1: Final Report
Volume 2: Empirical Validation Package
Volume 3: Working Reports

This empirical validation work complements the work using other evaluation techniques undertaken within the lEA BCS Annex 21/SHC Task 12 group. These activities resulted in the production of a set of Building Energy Simulation Tests (BESTESTs), based on inter-model comparisons. These tests, based on domestic scale buildings, are structured such that reasons for poor predictions from a program can be diagnosed. Other tests based on inter-model comparisons relate to commercial buildings. Some work was also undertaken to develop analytic tests.

1.2. Empirical Validation

In empirical validation, program predictions are compared with the measured performance of a full scale building or test room. The appeal of such comparisons as a way of extending the credibility of building thermal simulation tools is immediately apparent. Other validation techniques exist for determining whether a simulation code correctly carries out the calculations that its author intended (debugging and verification). However, all models, of necessity, contain simplifications and assumptions: empirical validation is the only way in which the overall effect of these on the accuracy of a simulation tool can be assessed. It is difficult to carry out empirical validation convincingly (Bloomfield et al(1)). This is because uncertainty, or errors, exist in the validation process. In previous work these have been divided into two categories and termed 'internal' and 'external errors' (2,3).

Internal errors are those embedded in the DSPs, some examples are:

- not modelling thermal phenomena, perhaps because they were incorrectly thought to be unimportant;
- approximating complex, possibly ill-defined but important, phenomena, for example by empirical correlations with measurements;
- simplifying the known physical processes which occur in the real world, for example to ease (or speed up) calculations;
- coding errors.

DSPs are likely to include internal errors of all four types. For example, they often simplify fabric conduction by assuming it to be 1-dimensional and they usually approximate convective heat transfer at surfaces as a simple empirically derived coefficient. These are legitimate and necessary procedures to enable useful DSPs to be developed. Nevertheless, it is still important to test whether they are acceptable or whether they lead to unacceptably inaccurate predictions.

Phase	Date	Event
Introduction	Nov. '89	Inception of IEA BCS Annex 21 and SHC Task 12.
	'90	Seville meeting - co-ordinated effort by BCS Annex 21 Subtask C and SHC Task 12 Subtask B agreed.
	Sep./Oct. '90	BRE meeting - Empirical Validation proposed.
	March '91	Monitored data survey.
	April '91	Boulder meeting - diverse range of available data sets discussed.
	May '91	Critical appraisal of previous IEA Empirical Validation (Working Report, IEA21RN180/91).
	Oct. '91	Summary and appraisal of High Quality Data Sets (Working Report,IEA21RN162/92).
	Nov. '91	Torino meeting - EMC Test Rooms selected for further investigation.
	March '92	Copenhagen meeting - Empirical Validation exercise agreed in 2 Phases: blind and non-blind.
Phase 1	March / April '92	Invitations sent to possible participants. Distribution of Validation Guidebook, Site Handbook, Weather Data and Emprical Validation Report Form.
Blind	May '92	Newssheet 1 distributed - 11 participants, deadline for first set of results set to June 1.
	June '92	Newssheet 2 distributed - Results received from all original participants. Number of participants had grown to 20. Individual results deadlines set for participants who had joined later.
	July '92	Newssheet 3 distributed including personal feedback on first results set. All results received from confirmed participants.
	August '92	Revised results received where appropriate.
	Sept. '92	Newssheet 4 distributed - 22 different results sets had been received from 19 participants. Intention to publish work stated.
	Sept./Oct.'92	Portland meeting - decision to invite several more institutions to participate who had so far been unable to do so or had not been invited. Some results shown at meeting. Revealing of results outside IEA group delayed until end of December to keep the exercise blind. Publication of results discussed.
	Nov. '92	Newssheet 5 distributed giving summary of Portland meeting. Intention to publish work re-stated.
	Jan. '93	Newssheet 6 distributed - 25 different results sets received. Comments invited on key sections of proposed publication.
	March '93	Newssheet 7 distributed - revised version of key sections of proposed publication circulated following comments.
Phase 2	March '93	Madrid meeting - decision to release measured data to give participants opportunity for follow-up work. Three-page modellers report to be sought.
Non-blind	April '93	Newssheet 8 distributed with: EMC Quality Assurance report; further Phase 1 results; measured data; format for 3-page modellers report; and model information proforma.
	July '93	Newssheet 9 distributed containig uncertainties in all model input data.
	July '93	Newssheet 10 distributed - timescales for finalising the Empirical Validation exercise.
	Aug. '93	Newssheet 11 distributed - included draft of part 1 of final IEA Empirical Validation report.
	Sept. '93	Newssheet 12 distributed - personal feedback to participants who submitted Model User Reports, reminder to others.
	Sept. '93	Fontainbleau meeting - draft of final IEA Empirical Validation report discussed.

Table 1: Key Events in the Life of the IEA BCS 21C/SHC 12B Empirical Validation Exercise

For empirical validation to be of any use, it must be capable of revealing the existence of such internal errors. The achievement of this is however difficult due to the existence of external errors. Some examples are:

- mistakes made when interpreting the real world building into program input data;
- the uncertainty in all the input parameters used by the DSPs to describe the building operation and the weather conditions;
- the inevitable uncertainty in the monitored performance of the building.

External errors can be much greater than, and hence swamp, any internal errors. For empirical validation to be successful, external errors must be eliminated or controlled, and remaining errors must be quantified. This remaining error can then be fully accounted for in any comparisons between predictions and measurements. Failure to tackle the difficulties associated with external errors has meant that most previous empirical validation exercises would have been incapable of revealing the existences of even substantial internal errors (4).

A methodology for avoiding such problems was devised as part of the joint UK Building Research Establishment (BRE) and Science and Engineering Research Council (SERC) validation exercise (Lomas (4), Bloomfield (5)). This methodology built on the earlier work of the Solar Energy Research Institute (now known as NREL) (3). The approach has been applied successfully by others (EMC (6), PASSYS (7)) and it was adopted as the basis of the IEA work.

This report describes the background to the lEA project, the management of the work and the main results. Two associated volumes contain the Empirical Validation Package (8) and the Working Reports (9). The Empirical Validation Package is intended to enable program users and developers, who did not participate in this lEA exercise, to evaluate programs of interest to them.

The Package consists of:

(i) a Site Handbook describing the test rooms and the experimental uncertainties;
(ii) a Validation Guide explaining how to use the handbook and undertake the validation;
(iii) a Quality Assurance report which outlines the procedures undertaken to ensure that the data were high quality; and
(iv) a Data Disk containing hourly measurements of weather data, energy consumption, internal air and surface temperatures, and external South facing solar irradiance.

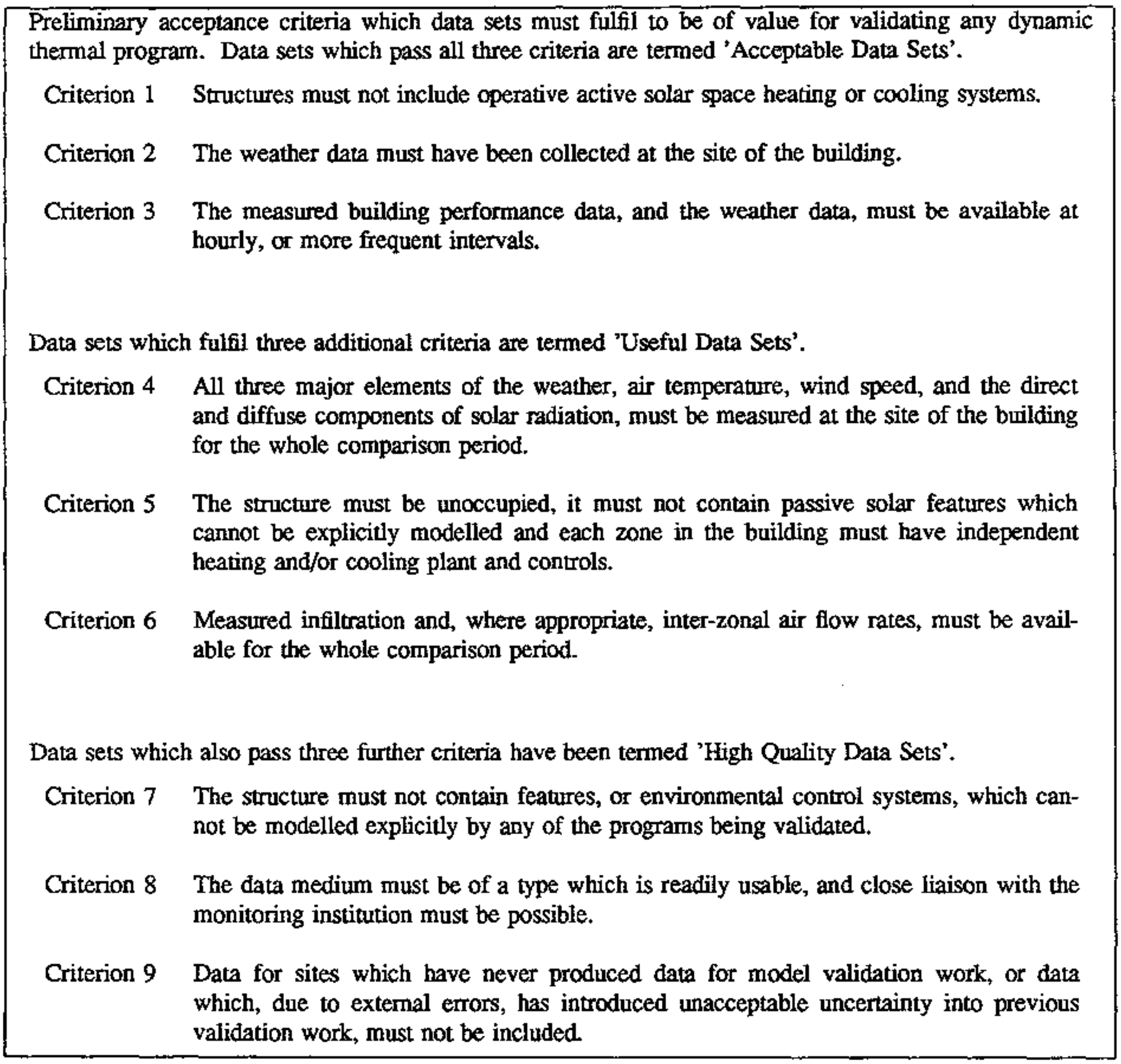
The Working Reports (9) document the background to the work and other progressional details. The main events which occurred during the lifetime of this research are outlined in Table 1.

13. Aims and Objectives

At present few properly documented whole program validation benchmarks exist (10), and even fewer (perhaps none) have been tested on a wide range of programs. Given this situation, it was decided that the aims of the lEA work should be to:

(a) develop well documented and well tested empirical validation benchmarks for detailed thermal simulation
(b) assess the ability of a number of DSPs to predict the performance of a few simple buildings under conditions reflecting those which exist when used to model real buildings; and
(c) devise and test a strategy for developing empirical validation benchmarks.

It was not an aim of this MA work to try and discover why programs performed well or why they performed badly in the validation test. Rather, the work was intended to produce a 'snapshot' of the capabilities of state-of-the-art DSPs when used under conditions which mimic a real design situation. When programs perfom well, confidence in the predictive abilities will quite rightly have been increased, and the credibility of the programs will have been enhanced. When programs perform poorly, it would be appropriate to explore the reasons to ascertain whether aspects of the program need modifying, for example to remove coding errors or to replace algorithms with more appropriate formulations.



Preliminary acceptance criteria which data sets must fulfil to be of value for validating any dynamic thermal program. Data sets which pass all three criteria are termed 'Acceptable Data Sets'.

Criterion 1	Structures must not include operative active solar space heating or cooling systems.
Criterion 2	The weather data must have been collected at the site of the building.
Criterion 3	The measured building performance data, and the weather data, must be available at hourly, or more frequent intervals.

Data sets which fulfil three additional criteria are termed 'Useful Data Sets'.

Criterion 4	All three major elements of the weather, air temperature, wind speed, and the direct and diffuse components of solar radiation, must be measured at the site of the building for the whole comparison period.
Criterion 5	The structure must be unoccupied, it must not contain passive solar features which cannot be explicitly modelled and each zone in the building must have independent heating and/or cooling plant and controls.
Criterion 6	Measured infiltration and, where appropriate, inter-zonal air flow rates, must be available for the whole comparison period.

Data sets which also pass three further criteria have been termed 'High Quality Data Sets'.

Criterion 7	The structure must not contain features, or environmental control systems, which cannot be modelled explicitly by any of the programs being validated.
Criterion 8	The data medium must be of a type which is readily usable, and close liaison with the monitoring institution must be possible.
Criterion 9	Data for sites which have never produced data for model validation work, or data which, due to external errors, has introduced unacceptable uncertainty into previous validation work, must not be included.

Table 2: Criteria for Classifying Data Sets (after Lomas (9))

Neither was it the aim to ascertain whether a particular program is adequate for a specific design task. Whether a program is, or is not, adequate, depends on the task in question, and the program user has to judge this. Clearly, however, a program which performs badly is unlikely to be as useful as one which performs well. The results of this exercise will assist program users to make such judgements.

Within lEA BCS Annex 21/SHC Task 12, the Building Energy Simulation Tests (BESTEST) have been developed (13). These are based on inter-model comparisons and seek to expose gross errors in programs. The tests include diagnostics to identify the faulty algorithms. Inter-model comparisons based on a commercial building have also been developed in IEA BCS Annex 21/SHC Task 12 (14).

2. REVIEW OF PREVIOUS WORK

2.1. Project Management Strategy

To assist in project planning, the work conducted in IEA SHC Task VIII (Mørck (11)) was reviewed. This led to important messages about how to conduct empirical validation within an international context (9). These may be summarised as follows:

Management Requirement 1: A strong centralised project management should be responsible for

(a) ensuring that the agreed methodology and time-scales are followed;

(b) interfacing between the data collection team and the modellers to ensure that the same information is available to all modellers and that this information is consistent; and

(c) analysing the results.

Management Requirement 2: A thorough review and assessment of acceptable data sets should establish those which would be suitable as the basis for validation benchmarks.

Management Requirement 3: The initial predictions should be made 'blind', that is, program users should be unaware of the actual measurements. They should all be given the same detailed information about the building, the operating conditions and the weather data. The model/data comparisons should be made by an independent third party, which is not responsible for program predictions.

This requirement led to a 2-Phase empirical validation exercise. In Phase 1, all the predictions were made blind, compared with the measurements, and the results documented (Section 5). In Phase 2, the program users were provided with the measured data and encouraged to investigate the reasons for any discrepancies observed in the blind phase (Section 6).

2.2. Selection of Data Sets

Management requirements 1 and 3 relate to procedure, however, further preliminary work had to be conducted to comply with requirement 2. This led to the following recommendations concerning the choice of the data sets.

Data Requirement 1: The data set(s) must be high quality, i.e. fulfill all nine of the criteria given in Table 2. This was seen as far more important than trying to cover a range of buildings and weather conditions and, in the process, accepting inferior data.

Data Requirement 2: The data must be available for use both within the IEA project and for subsequent use by others.

Data Requirement 3: Ideally, the site from which the data was collected should be active. This would allow participants to have first-hand experience of the building and the monitoring (which itself would lead to more accurate use of the programs). Any necessary peripheral investigations could also be undertaken and extra experiments could be commissioned. Furthermore, the monitoring team would be

		STRUCTURE	Primary Features				Secondary Features			HEATING Type				MONITORING			OPERATION					AIR FLOW Intra-zone		Ventilation			BUILDING MONITORING	Temperature							Ventilation		Others					WEATHER MONITORING	Temperature			Radiation						Wind		VALIDATION Data		Comparisons			
Monitoring Institution [Location of Structure] Contact (and Number) [Major References]	'Name' and Code Number	Volume (m^3)	Opaque, Heavyweight	Opaque, Lightweight	Direct Gain, Lightweight	Direct Gain, Heavyweight	Raised off Ground	Night Insulation	Clerestory Window	None	Fan Convector	Natural Convector	Panel Radiator	Start Date (Month, Year)	Duration (Days)	Recording Frequency (Mins)	Free Floating	Intermittent, T'stat Control[1]	Continuous, T'stat Control	Impulses	Pseudo-Random	Stratified	De-stratified/Mixed	Negligible	Natural	Mechanically Induced	Number of Sensors	Air	Globe	Opaque Surface	Window Surface	Intra-fabric	Below Floor	Above Ceiling	Continuous Flow Meter	Continuous Gas Decay	Auxiliary Heating Power	Surface Heat Flux	Interior Insolation	Interior Longwave	Some thermo-physical Props.	Sensors	Air, Dry Bulb	Air, Wet Bulb	Ground	Global Horizontal	Direct Normal	Diffuse Horizontal	Vertical Glazing Plane	Ground Reflected	Net Longwave	Speed	Direction	High Quality Datasets	Site Handbook Available	ESP	HTB2	SERIES	DOE-2
Polytechnic Central London [Peterborough UK] Kevin Lomas (0533-551551) [A1-A5]	Test Cell 1	10.4			*		*			*				2:84	9	60	*					*		*			10	*				*									*	6	*			*		*	*			*	*	4	*	*	*	*	
		10.4			*		*	*		*				4:84	9	60																																											
	Test Cell 2	10.4			*		*			*				2:84	9	60																																											
		10.4			*		*			*				4:84	9	60																																											
Energy Monitoring Co. [Cranfield UK] R. Hitchin (071-736-1212)	British Gas Test Cell	9.7	*				*			*[3]			*	:85	16	3/20[2]	*[3]		*[3]			*				*	26	*		*					*		*	*				6	*			*		*			*	*	*	10					
													*	:85	35[4]	3/20[2]			*																																								
												*		:85	21[5]	3/20[2]			*																																								
										*[3]		*[3]		:85	16	3/20[2]	*[3]		*[3]																																								
		9.7	*				*				*			12:88	21	5	*[6]	*[6]	*[6]			*				*	54	*		*		*			*		*	*		*		6	*			*		*				*	*	5					
														1:89	9	5	*[6]	*[6]	*[6]																																								
														2:89	9	5	*[6]	*[6]	*[6]																																								
														2:89	1	5		*																																									
														3:89	20	5					*																																						
		9.7	*				*					*		2:90	13.5	5	*[7]	*[7]	*[7]			*				*	59	*		*		*			*		*	*		*		6	*			*		*				*	*	1					
														8:87a	10	60		*																																									
	Room 0	7.9				*	*					*		8:87b	10	60			*																																			12	*		*	*	
	Room 1	7.9				*	*						*	10:87a	10	60		*				*		*			20	*	*	*			*	*			*	*																					
	Room 2	7.9	*				*					*		10:87b	10	60			*																							10	*			*		*				*	*						
	Room 3	7.9	*				*						*	1:88	10	60		*				*				*	20	*	*	*			*	*	*		*	*																24	*		*	*	
	Room 4	7.9				*	*					*		2:88	10	60		*				*		*			20	*	*	*			*	*			*	*																					
	Room 5	7.9				*	*						*	3:90	48	60		*				*		*			20	*	*	*			*	*			*	*																					
														5:90	48	60	*					*		*			20	*	*	*			*	*				*				13	*	*		*		*	*		*	*	*	12	*				
National Bureau Standards [Washington DC USA] Kevin Lomas (0533-55155)	Control Room 3	87.1				*					*			1:84	21	60			*				*		*		26	*		*			*			*	*					26	*		*	*	*		*	*	*	*	*	4	*	*	*	*	*
														2:84	25	60			*				*		*																																		
	Direct Gain Room 4	87.1				*					*			1:84	21	60			*				*		*		226	*	*	*	*	*	*			*	*	*	*		*																		
														2:84	25	60			*				*		*																																		
Univ. of Strathclyde [Glasgow, Scotland] P Strachan (41-552-4400)	Cell 1	31.3		*			*			*				2:89	32	60	*			*		*			*		2	*									*				*	6	*	*		*		*				*	*	2		*			
	Cell 2			*			*					*	*[8]	2:89	32	60	*			*		*					2	*																															

[1] Cycle of heating to a set point then free floating in between

[2] 3 minutes to first heater 'turn off' then 20 minutes

[3] Free floating followed by continuous heating

[4] 5 at 7 days each

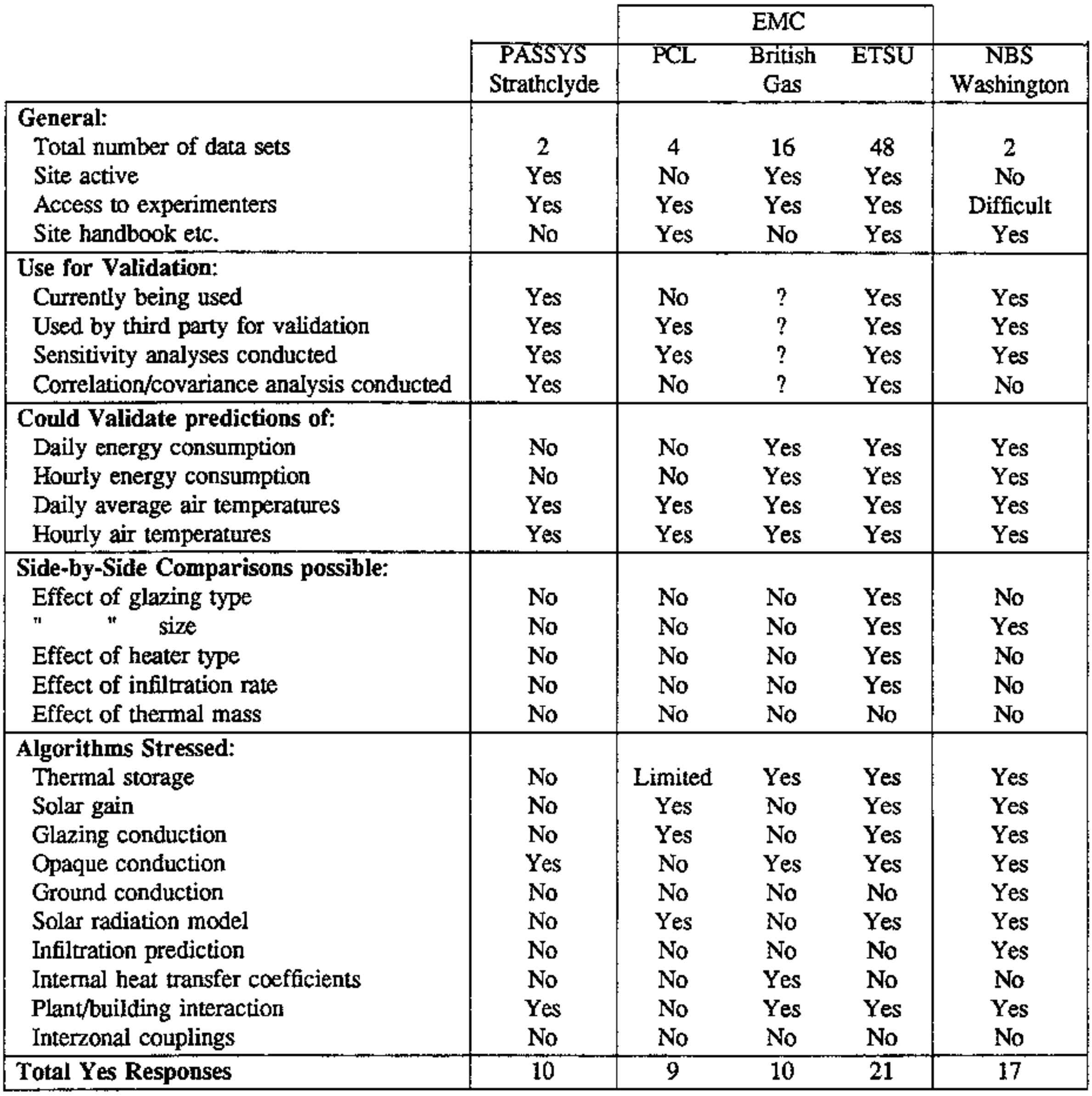
[5] 3 at 7 days each

[6] Period of continuous heating followed by periodic heating followed by free floating period

[7] Complex operation including 28 hour cycle of 6 hours continuous heating, 2 hours free float, 6 hours heating, 14 hours free float

[8] Exact type of radiant source unknown

Table 3: Compiled Information about High Quality Data Sets available in the UK (after (10))



		EMC			
	PASSYS Strathclyde	PCL	British Gas	ETSU	NBS Washington
General:					
Total number of data sets	2	4	16	48	2
Site active	Yes	No	Yes	Yes	No
Access to experimenters	Yes	Yes	Yes	Yes	Difficult
Site handbook etc.	No	Yes	No	Yes	Yes
Use for Validation:					
Currently being used	Yes	No	?	Yes	Yes
Used by third party for validation	Yes	Yes	?	Yes	Yes
Sensitivity analyses conducted	Yes	Yes	?	Yes	Yes
Correlation/covariance analysis conducted	Yes	No	?	Yes	No
Could Validate predictions of:					
Daily energy consumption	No	No	Yes	Yes	Yes
Hourly energy consumption	No	No	Yes	Yes	Yes
Daily average air temperatures	Yes	Yes	Yes	Yes	Yes
Hourly air temperatures	Yes	Yes	Yes	Yes	Yes
Side-by-Side Comparisons possible:					
Effect of glazing type	No	No	No	Yes	No
" " size	No	No	No	Yes	Yes
Effect of heater type	No	No	No	Yes	No
Effect of infiltration rate	No	No	No	Yes	No
Effect of thermal mass	No	No	No	No	No
Algorithms Stressed:					
Thermal storage	No	Limited	Yes	Yes	Yes
Solar gain	No	Yes	No	Yes	Yes
Glazing conduction	No	Yes	No	Yes	Yes
Opaque conduction	Yes	No	Yes	Yes	Yes
Ground conduction	No	No	No	No	Yes
Solar radiation model	No	Yes	No	Yes	Yes
Infiltration prediction	No	No	No	No	Yes
Internal heat transfer coefficients	No	No	Yes	No	No
Plant/building interaction	Yes	No	Yes	Yes	Yes
Interzonal couplings	No	No	No	No	No
Total Yes Responses	10	9	10	21	17

Table 4: Details of High Quality Data Sets for Empirical Validation

Program	Version	Country of Origin	Operating Institution
APACHE[5]	6.5.2	UK	Facet Ltd., UK[1]
BLAST[5]	3lvl143	USA	Colorado State University, USA[1]
BLAST[5]	3.0lvl203	USA	Politecnico di Torino, Italy[2]
CHEETAH[5]	1.2	Australia	CSIRO, Australia[1]
CLIM2000[4]	1.1	France	Electricité de France[1]
DEROB[4]	LTH	USA	Lund Institute of Technology, Sweden[2]
DOE[5]	2.1E	USA	LBL, USA[1]
ENERGY2[6]	1.0	UK	Arup R&D, UK[1]
ESP[4]	6.18a	UK	DMU Leicester, UK[2]
ESP-r[4]	7.7a	UK	ESRU, Univ. of Strathclyde, UK[1]
ESP+[5]	2.1	UK	DMU Leicester[3] / ASL Sterling[1], UK
HTB2[4]	1.2	UK	FHT Stuttgart, Germany[2]
HTB2[4]	1.10	UK	University of Wales College of Cardiff, UK[1]
SERI-RES[4]	1.2	USA	BRE, UK[2]
SUNCODE[5]	5.7	USA	Ecotope, USA[1]
S3PAS[4]	2.0	Spain	Escuela Superiore Ingenieros Industriales, Sevilla, Spain[1]
TASE[4]	3.0	Finland	Tampere University of Technology, Finland[1]
TAS[5]	7.54	UK	DMU Leicester[3] / EDSL[1], UK
TRNSYS[5]	13.1	USA	University of Wisconsin, Madison, USA[1]
TRNSYS[5]	12	USA	BRE, UK[2]
TRNSYS[5]	13	USA	BRE, UK[2]
TRNSYS[5]	13.1	USA	Vrije Universiteit Brussel, Belgium[2]
TSBI3[4]	2.0	Denmark	Danish Building Research Institute[1]
WG6TC[6]	1992	Italy	Institute di Fisica Technica, Udine, Italy[1]
3TC[5]	1.0	UK	Facet Ltd., UK[1]

[1] Program authors / vendors / support office; [2] Experienced users of program; [3] DMU ran program on behalf of vendors, input files checked by vendors in Phase 1, vendors ran programs in Phase 2; [4] Research program; [6] Commercial program; [6] Program under development

Table 5: List of Programs and Participants

available to assist in resolving any uncertainties and ambiguities.

Data Requirement 4: The actual monitored performance of the buildings must not be widely known; otherwise 'blind comparisons' could not be assured.

These requirements precipitated a wide ranging review and assessment of available data sets. The existing information, which followed from a previous world wide review (10), was the basis for this process.

To ensure that the previous survey had not overlooked any recently collected data sets, a new survey was conducted by distributing a questionnaire to principal researchers in the validation field (see (9)). Only one additional facility capable of producing high quality data was revealed (Fauconnier et al (15)) but the data was not immediately available. (This illustrates the lack of research currently being conducted in the empirical validation field.) Therefore, the 72 data sets identified by the previous review (10) as being both available and of high quality were examined further (Table 3).

There were potential problems with all the data sets except those from the Energy Monitoring Company (EMC) rooms (9). These rooms had produced 48 high quality data sets, so a diverse range of validation tests could be developed (Table 4). The IEA participants therefore selected these rooms as the basis for their work.

This review highlights two important points.

(i) Whilst data sets may be classified as high quality (see Table 2), they may still not be ideal for a particular empirical validation exercise. The 9 criteria in Table 2 may thus be seen as the minimum requirements for data which are to be used for DSP validation.

(ii) There is a need to collect and archive further very high quality data suitable for validating a wide range of thermal programs.

2.3. Documentation

Most previous validation work has had little benefit beyond the small group of experts directly involved - primarily due to limited documentation of the validation activities, procedures and results. To overcome this difficulty a number of documentation requirements were prescribed.

Documentation Requirement 1: The methodology by which the benchmark is conceived must be clearly stated before beginning the work. Modifications (and reasons for them) should be described and recommendations about the approach to be used in future should ensue.

Documentation Requirement 2: The benchmarks must contain:

(a) a description of the building and its operating conditions;

(b) the weather data; and

(c) the procedure to be followed when using these for validation.

To facilitate the non-blind phase, the following were needed:

(d) the measured building performance; and

(e) the uncertainties in all the building description and measured data.

The documentation had to be clear and unambiguous and serve the needs of all the modellers working within the IEA project, as well as modellers who may subsequently use the data for validation.

Program	Phase 1	Phase2		
		MUR[#]	New results files	Input files
WG6TCv1992 (Udine, I)	○		○	
TSBI3v2.0 (DBRI, DK)	○	○	○	○
DOE2.1E (LBL, US)	●	●		
TASv7.54 (DMU/EDSL, UK)	○	○	○	○
ENERGY2v1.0 (Arup, UK)	●			
CHEETAHv1.2 (CSIRO, AUS)	●			
3TCv1.0 (Facet, UK)	●			
APACHEv6.5.2 (Facet, UK)	●			
HTB2v1.9 (UWCC, UK)	●			
HTB2v1.2 (FHT, GER)	○	○	○	○
CLIM2000v1.1 (EDF, F)	●			
DEROBvlth (Lund, S)	●	●		
S3PASv2.0 (Sevilla, E)	●			
BLASTv3lvl143 (CSU, US)	●			
BLASTv3.0lvl203 (Torino, I)	○	○	○	○
TASEv3.0 (Tampere, FIN)	●	●		
TRNSYSv13.1 (UWISC, US)	●			
TRNSYSv13.1 (Brussel, B)	●	●		
TRNSYSv13 (BRE, UK)	●			
TRNSYSv12 (BRE, UK)	●			
SUNCODEv5.7 (Ecotope, US)	○		○	○
SERI-RESv1.2 (BRE, UK)	○	○	○	○
ESP+v2.1 (DMU/ASL, UK)	●	●		
ESP-Rv7.7a (ESRU, UK)	●	●		
ESPv6.18a (DMU, UK)	●			

The names of programs for which Phase 2 results were produced are shown in *italics* throughout the report. Different symbols are used for tabular entries and in bar charts to distinguish entries for these programs from those for which only Phase 1 results were produced (see section 5.2).

[#] Model User Report

Table 7: Contributions from each program/user combination

3. PROGRAMME OF WORK

3.1. Participants

Initially, the only modellers involved were 6 of the IEA BCS Annex 21/SHC Task 12 participants who between them planned to run 9 programs. In view of the difficulties of keeping measured results unknown after completion of this project and the need for such studies to be performed blind, it was decided to invite other institutions to participate together with the IEA group. All participants agreed in advance that their results from the blind phase could be published.

Invitations to join the exercise attracted an additional 14 institutions and private companies (Table 5). Most were either skilled users or the authors, vendors or support offices for the programs. In total, 25 results sets were eventually obtained from 17 genuinely different programs (the remaining results were from alternative versions of some of these; one program, WG6TC, was only applicable to temperature predictions in unheated buildings). The programs originated from 6 different European countries, with 4 of USA origin plus 1 from Australia, and the program description questionnaire indicated that they employed a diverse range of algorithms for the key heat transfer processes (Table 6). This was the most extensive empirical exercise ever undertaken, and it is encouraging to see so many program users and authors willing to participate. 10 of the results sets were from research versions of programs, 13 from commercial programs and 2 from prototypes (Tables 5&6). The feedback from such an expert group, using a diverse range of programs, helped to produce a complete, unambiguous and widely applicable validation package. It also ensured that no possible sources of experimental error were overlooked, in particular, by prompting extensive quality assurance checks (see Section 7.1, Appx.E and Volume 2, part 3: Quality Assurance Report) and rigorous attention to the treatment of experimental uncertainty (external errors, see Section 4.4&7.1).

3.2. Management

The UK Building Research Establishment (BRE) directed the empirical validation work, with the De Montfort University (DMU) team being responsible for communicating with the participants, including the operation of a support hotline. Technical matters concerning the datasets were dealt with by the Energy Monitoring Company (EMC), ensuring the close contact between the participants using the data and the team responsible for collecting the data. The overall structure of the work programme was developed jointly by all three of these groups and implemented, with suitable modifications, following each meeting of the IEA participants (Table 1).

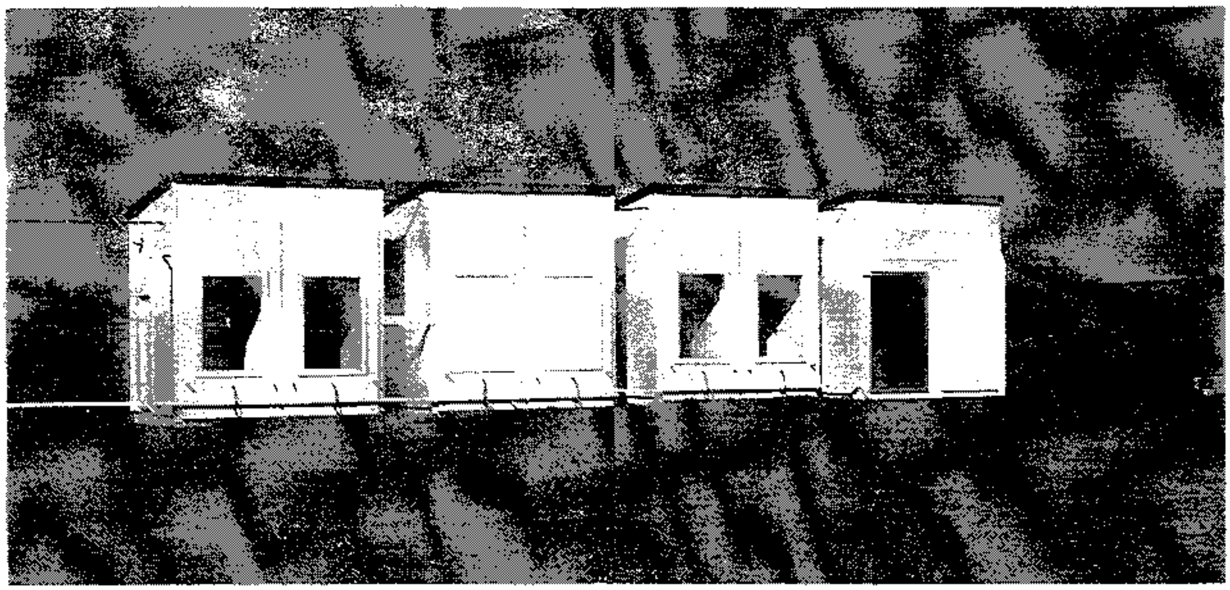
The work programme had three distinct Phases; the later Phases began only when the previous Phase was completed.

In Phase 1, all the predictions had to be made blind, i.e. without any knowledge of what the actual measured performance of the EMC test rooms was. The modellers were provided with a detailed description of the test rooms and the prevailing weather data. The modellers could obtain (from DMU) any further information they needed consistent with the 'blind' philosophy. Even the team at DMU did not know what the measured performance was. This prevented the DMU group biasing the results through their feedback. The issue of quality assurance was highlighted, and procedures were investigated to reduce the likelihood of the modellers making mistakes (external errors).

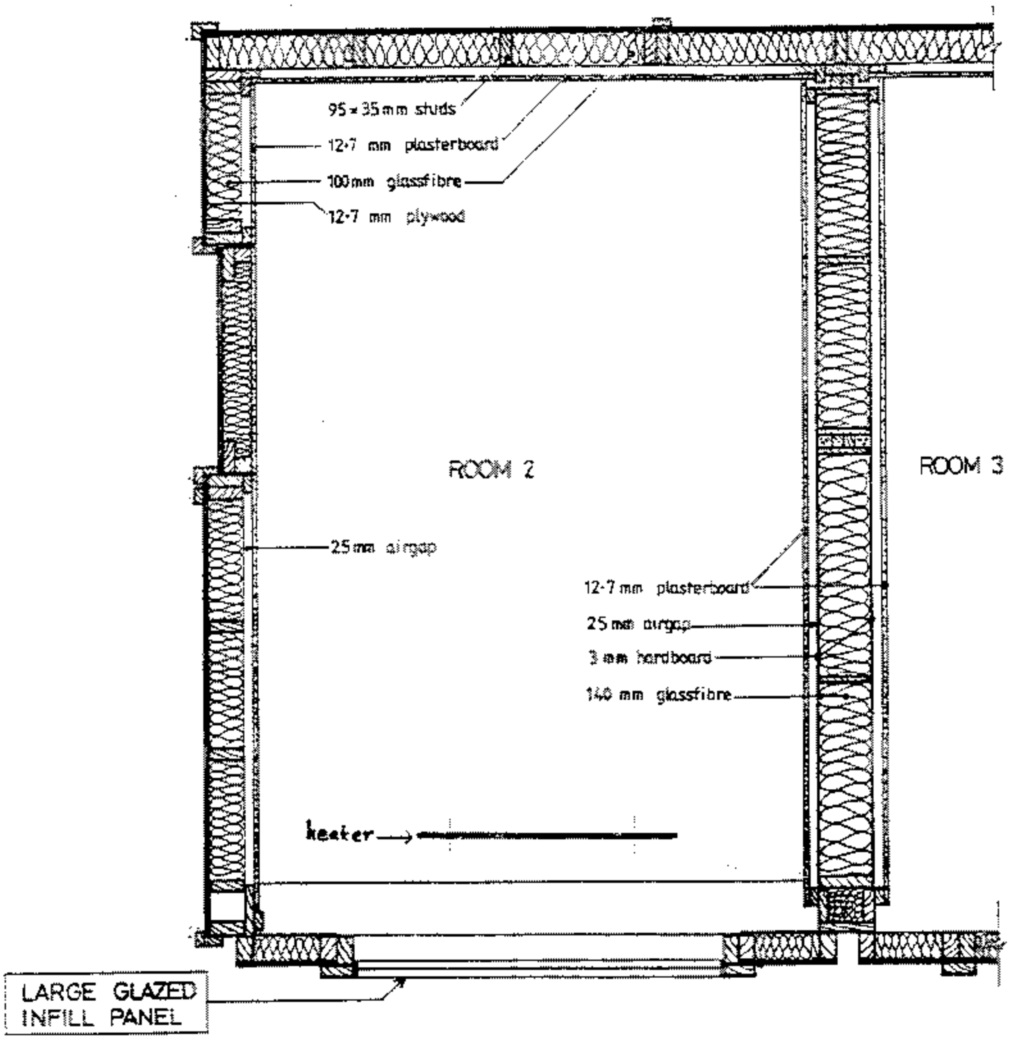
The Phase 1 conditions mimicked those which prevail when programs are used in a real building design project, because there was no possibility of modellers manouvering a fit between the measurements and the predictions. The goal of Phase 1 was to produce a 'clean test' of the relative capabilities of the programs for the specific building and operating conditions.

All participants produced results for Phase 1, and these were documented prior to commencing Phase 2 (Table 7).

In Phase 2, modellers were provided with the measured data and the estimated uncertainty (external errors) in the building description and climate data. This enabled them to refine their predictions, either because they had made mistakes or because new information about the test rooms (which was hitherto thought to be unimportant and had not been requested from the hotline during Phase 1) was



External view of the test rooms



Plan view of test room showing construction

Figure 1: The EMC Test Rooms

made available. Additionally, modellers could explore the reasons for any divergence between the measurements and the predictions. This could entail detailed analysis of the empirical validation results, using analytic tests (e.g. Bland (12)) or diagnostic inter-model comparisons. Modellers were asked to submit a three page Model User Report (MUR) outlining these studies and the results. They were also asked to provide new results if appropriate together with any new input files.

During this Phase, further quality assurance checks were undertaken (over and above those customarily adopted by the EMC) to make absolutely sure that the rooms were built as described in the site handbook. This included the dismantling of one room in the presence of an independent observer.

Following this Phase, 7 participants produced new results and submitted these to DMU for plotting and analysis. In all but one case, these differed only slightly from the Phase 2 results, and they differed because legitimate changes to the input files had been made (Table 7).

In Phase 3 the work was reported, via this final IEA report (Volume 1) and the associated volumes - the empirical validation package (Volume 2) and the working reports (Volume 3). The final empirical validation package includes the minor additions, deletions and modifications to the room descriptions which were circulated to the participants in Phase 1 as a result of requests to the hotline. The package also contains a diskette of the weather data and the measured thermal performance, and recommendations about how to use the complete package. These recommendations reflect the experiences gained from using the package in this IEA exercise.

At all stages of the work, all the participants were kept informed of progress, any future plans and the emerging results (after Phase 1), via a Newssheet. During the course of the work, 14 such Newssheets were produced (Table 1) (see Volume 3).

33. The EMC Test Rooms

The EMC test rooms were built in pairs, separated by a heavily insulated party wall. The outer shells are of stud-frame construction covered by plasterboard with a concrete slab floor (Fig. 1). The monitored spaces are well insulated and extremely well sealed to reduce infiltration to less than 0.05 air-changes per hour. An 'attic' space was less well insulated and poorly sealed so infiltration occurred. A well insulated ceiling limited the heat flow between this space and the monitored room below. The rooms occupy an unobstructed site at Cranfield airfield, and have been used in numerous monitoring projects, primarily for the UK Energy Technology Support Unit (e.g. (6)). The rooms are representative of typical lightweight UK domestic rooms in terms of the level of insulation, the amount of thermal mass and the window-to-floor area ratio. However, they stress the solar gain and fabric heat loss processes because they have very low infiltration rates, a large surface-area-to-volume ratio and no incidental internal heat gains.

In the IEA work, data from three rooms (rooms 1, 3 and 5) were used for each of two 10-day periods. In the first period, beginning in October 1987, the rooms were heated by a panel radiator from 06:00 to 18:00 to a set-point temperature of 30°C (Table 8). In the second period, beginning in May 1990, the rooms were unheated (free-floating). In both periods Room 3 housed an opaque panel in the South facing front wall whereas the other two rooms had (different) glazed facades (Table 8).

The climatic data typically required by DSPs were collected and the thermo-physical properties of the construction materials were defined. The key building performance parameters measured were the hourly heating energy consumption and the room air temperature at three levels. Also recorded were the total hourly South facing vertical solar irradiances on the glazed facade and the temperatures of the inner surfaces of the floor, back wall and ceiling. The rooms are described in greater detail elsewhere (8) and in Volume 2 of the IEA report on Empirical Validation.

Prediction Weather Period	**Room**	**Glazing Type**	**Glazing Area**	**Heating**
May 24 to 30 1990	1	Double	1.5m^2	None
	3	Opaque	-	None
	5	Single	1.5m^2	None
October 20 to 26 1987	1	Double	1.5m^2	06:00 - 18:00
	3	Opaque	-	06:00 - 18:00
	5	Double[1]	0.75m^2	06:00 - 18:00

[1] Predictions made for 1.5m^2 of single glazing

Table 8: Synoptic Description of the EMC Test Rooms

4. RESULTS ANALYSIS

4.1. Primary Parameters

Each participant was required to predict the hourly values of the air temperature, the temperaturers of the floor, back wall and ceiling, the South-facing global solar irradiance and, for the heated rooms (October period) the hourly energy consumption. These data were required for all three rooms in both weather periods for the last 7 days of the 10-day collection period. Data from the first 3 days was provided so modellers could pre-condition the programs. Given the large number of participants, this produced a considerable amount of data. Therefore, analysis concentrated on 4 primary parameters:

- the maximum air temperature during each 7-day period;
- the minimum air temperature during each 7-day period;
- the total heating energy consumption (October period); and
- the total south-facing global solar irradiances in each 7-day period.

Because 25 sets of results were available, they could be analysed, either by comparing the programs in an inter-model comparison exercise, or, by comparing each program with the measurements - an empirical validation exercise. The large number of results did, however, preclude detailed scrutiny of all the hourly results by DMU. The individual modellers were expected to do this themselves for their own program in Phase 2.

4.2. Differences in Predictions

When programs are used for designing buildings it is often more important that they predict the changes in building performance which will occur as a result of a design alteration, or to predict the difference between the thermal performance of a proposed design and a 'reference' design, than that they produce accurate predictions of absolute performance. The correct prediction of these differences or trends is important to ensure that any design changes drive the building towards the desired objective - usually that it is more energy efficient or less likely to overheat.

For the heated (October) period therefore, the predicted differences in energy consumption between the opaque and double glazed buildings have been compared with the measured values. Similarly, the predicted differences in the peak air temperature has been compared with the measured differences for the free-floating (May) period.

Within the time scales of the IEA project it was not possible to resolve the theoretical issues surrounding the calculation of the total uncertainty (experimental plus modelling) for difference analyses.

4.3. Hourly Results

The hourly predictions of all the required parameters were produced in the same consistent format. These could therefore be readily analyzed by modem computer based statistical and visualization packages, in this case PV-Wave (16) was chosen. Simple statistical measures were used to quantify the differences between the measurements and predictions:

Difference $$D_t = X_t - R_t \tag{1}$$

Maximum Difference $$\hat{D} = Max\ D_t \tag{2}$$

Minimum Difference $$\check{D} = Min\ D_t \tag{3}$$

Mean Difference $$\bar{D} = \sum_{t=1}^{N} D_t / N \tag{4}$$

Mean Absolute Difference $$|\bar{D}| = \sum^{N} |D_t / N| \tag{5}$$

Parameter	Nominal Value	Uncertainty
Site Details		
Latidute	52.07°N	±.05°
Longitude	0.63°W	±.05°
Altitude	100 m	±5 m
Ground Reflectivity	0.20	±0.05
Glazing Orientation	9° W of S	±0.5°
Test Room Surface Finishes		
External surface absoprtivities	0.16	-0.06 +0.14
Internal floor absorptivity	0.50	±0.10
Internal and external emissivity	0.90	±0.05
Material Properties		
Styrofoam conductivity	0.027 W/mK	-0.002 +0.006 W/mK
Concrete heat capacity	1840 kJ/K	±184 kJ/K
Rockwool conductivity	0.043 W/mK	±0.003 W/mK
Rockwool thickness	Various	±10 mm
Plasterboard heat capacity	937 kJ/K	±94 kJ/K
Wood conductivity	0.125 W/mK	±0.025 W/mK
Glazing Properties		
Glazed area	1.50 m^2	±0.02 m^2
Glass extinction coefficient	0.030 mm^{-1}	±0.005 mm^{-1}
Glazing cleanliness	1.00	-0.02 +0.00
Test Room Heater Characteristics		
Power output	680 W	±40 W
Rad./conv. split	60/40	±10/10
Time constant	22 min.	±2 min.
Setpoint	30°C	±0.2°C
Experimental Design		
Test room ventilation rate	0.00 ac/h	-0.00 +0.05 ac/h
Edge effects (Wood A, B & C conductivity)	Various	-0 +50%
Stratification of air	Variable	max. ±1.2°C for dbl.glazed heated room
Climate Data		
External air temperature	Variable	±0.2°C
Solar radiation	Variable	±5%
Wind speed	Variable	±5%
Wind direction	Variable	±5°
Program Specific Parameters		
Equivalent angle for diffuse radiation	60°	±5°
External surface coefficient	16.6 W/m^{2}K	-6.6 +13.4 W/mm^{2}K
Internal surface coefficient	8.33 W/m^{2}K	-2.33 +1.67 W/m^{2}K
Double glazing U-value	3.4 W/m^{2}K	-0.2 +0.4 W/m^{2}K
Single glazing U-value	5.6 W/m^{2}K	-0.65 +1.1 W/m^{2}K

Table 9: Input parameter uncertainties used for generating the error bands

Root Mean Square Difference $$\sqrt{D^2} = \sqrt{\sum_{t=1}^{N} D_t^2/N} \tag{6}$$

where: X_t = Predicted values at hour t
R_t = Reference (measured value) at hour t
N = Total hours in comparison period

The last three statistics provide alternative measures of the overall agreement between the measured and predicted values, whereas the first three are spot values.

The hourly energy consumption during the seven 12-hour heating periods was evaluated in this way for the October period, along with the hourly internal air temperatures for the whole seven day May period.

4.4. Dealing with Uncertainty

The above comparisons do not take account of the inherent uncertainty in the measured performance of the rooms or the data describing them which are entered into the programs (external errors). It is crucial to account for these in order to assess whether the divergences between a program and the measurements are likely to be due to this inherent uncertainty or to internal error(s) in the program combined, perhaps, with mistakes made by the program user.

Much effort was therefore directed towards sensitivity analysis. Ideally, the sensitivity analysis should be undertaken for each program, however, within the time-scales of the IEA exercise this was impractical and so the DMU group estimated the total uncertainties with just one program - SERI-RES. For all but SERI-RES, the total uncertainty bands are therefore only estimates. However, because SERI-RES requires program users to specify some highly uncertain parameters (in particular window U-value and internal and external surface coefficients), the total uncertainty was likely to be greater, rather than smaller, than that which would be appropriate for most other programs. The uncertainty due to measurement errors was added to the total for the SERI-RES input parameters.

Initially, the uncertainties were based on those used in earlier UK work (6). During Phase 2 of the exercise, participants pointed out other sources of uncertainty *(see* Section 6.2), which had not already been accounted for. Although these had a very small influence, they were subsequently incorporated in the total uncertainty calculation. They have also been incorporated in the final version of the Validation Package (8). The total uncertainties shown in this report include the contributions from all these sources (Table 9).

The total uncertainties in each of the prime parameters of interest were estimated using the Monte Carlo Analysis (MCA) technique (Lomas and Eppel (17)). In MCA all the uncertain inputs to a program must be assigned a definite probability distribution. For each simulation one value is selected at random for each input based on its probability of occurrence. For inputs with uncertainties which are normally distributed, values near the modal value are more likely to be selected than extreme values. The predictions produced by this unique set of input values are saved and the process is repeated many times, using a different and unique set of inputs on each occasion. Upon completion many values of each predicted parameter will have been obtained and these will have a particular distribution. Subject to certain assumptions about the input uncertainties and the program, the values predicted for a particular output parameter *(p)* will have a distribution close to normal. Thus, the total uncertainty in the predictions may be expressed by the standard deviation (s):

$$s = \sqrt{\frac{1}{N-1}\left(\sum_{n=1}^{N} p_n^2 - N\bar{p}^2\right)} \tag{7}$$

where: n = simulation number
N = total number of simulations
$\bar{p}$ = mean value of output parameter p

Period	Room Number	Description	Parameter	Measured Value	Uncertainty band		Uncertainty band width	Width as % measured
					From	To		
October	1	Double glaz.	Energy [MJ]	89.3	92.7	78.1	14.6	16
			Max Temp [°C]	37.8	40.5	36.5	4.0	11
			Min Temp [°C]	11.9	13.9	11.5	2.4	20
	3	Opaque	Energy [MJ]	117.1	122.3	105.3	17.0	15
			Max Temp [°C]	29.8	30.2	29.4	0.8	3
			Min Temp [°C]	14.6	16.4	14.0	2.4	16
	5	Single glaz.	n/a	n/a	n/a	n/a	n/a	n/a
	n/a	Svfr	Irrad. [MJ/m^2]	81.1	85.5	76.7	8.8	11
May	1	Double glaz.	Max Temp [°C]	31.0	33.4	29.6	3.8	12
			Min Temp [°C]	12.2	13.6	11.6	2.0	16
	3	Opaque	Max Temp [°C]	16.8	17.5	15.7	1.8	11
			Min Temp [°C]	9.2	10.0	8.6	1.4	15
	5	Single glaz.	Max Temp [°C]	32.6	35.0	31.2	3.8	12
			Min Temp [°C]	12.1	13.6	11.6	2.0	17
	n/a	Svfr	Irrad. [MJ/m^2]	82.8	88.8	76.8	12.0	15

Table 10: Measured values of primary parameters and their total uncertainty

An estimate of s can be obtained after any number of simulations and the accuracy of this estimate can be determined using the χ^2-distribution to calculate a confidence interval around s (e.g. (18)). The accuracy of s depends only on the number of simulations undertaken and not on the number of uncertain input parameters. Irrespective of the number of inputs and outputs to the program, only marginal improvements in accuracy are obtained after 60 to 80 simulations. Nevertheless, the analysis was undertaken with 100 simulations.

Since all the inputs are perturbed simultaneously the method can fully account for any interactions between the inputs and, in particular, any synergistic effects. It is not necessary to assume that the effects of the inputs are superposable: any non-linearities in the input/output relationships are fully accounted for. The obvious disadvantage is that, because the inputs are varied simultaneously, the sensitivities of the predictions to the individual input parameter changes are not divulged.

Additional Differential Sensitivity Analysis studies were undertaken for a few parameters which became important in Phase 2; these were included in the total uncertainty bands (see Section 7.1).

The total uncertainties are such that if the predictions of a DSP lie outside the upper and lower error bounds there is a very high probability that there are 'internal' errors in the DSP.

Most of the external uncertainty (total error) bars are not centred around the measured values. This is because a number of the parameters have error bars which are not centred around the nominal (basic) values specified in the Site Handbook (see Volume 2). For example, the glass was specified as absolutely clean, whereas in fact it could have been dirty (reducing the solar transmission) - it could not have been cleaner. The error bars include any contribution from measurement errors and the effects of temperature stratification during the unheated periods (see section 5).

The range for the upper to the lower error bound represents about 2.33 s. For the 7-day energy consumption in the double glazed heated room (Room 3), this range was -6% to +10%. This is close to, but larger than the value which had been calculated previously using differential sensitivity analysis (DSA) by the EMC. Similar results, but slightly larger for MCA rather than DSA, have been observed before and, in this study the uncertainty bounds have been further enlarged by including uncertainties due to parameters not previously included (e.g. relative humidity, stratification, diffuse solar radiation measurements and additional edge and corner effects) (Table 10).

5. PHASE 1: BLIND COMPARISONS

5.1. Management

A detailed 'Validation Package' was distributed which consisted of: a description of the test room; a diskette of the measured weather data; a description of the simulations to be undertaken, the format in which the results were to be presented, and a program description questionnaire (8). The predictions required from the participants were the hourly values of the air and surface temperatures, the heater power input and the south-facing vertical solar irradiance. These data were required for all three rooms in both weather periods for the last 7 days of the collection period. The measured building performance data was retained by the EMC, not even the DMU team had access to the data; thus the exercise was truly 'blind'.

The program users were asked "to model each of the three rooms in as much detail as the simulation program will allow" (Validation Guidebook). To help achieve this, a direct "hotline" to DMU was established in order that all the modellers could immediately resolve any uncertainties which they encountered. All enquiries and the responses were logged.

To ensure that all the modellers had access to all the relevant information a newssheet was circulated. This listed all the current participants, the significant hotline enquiries made, and the responses given. The newssheet exposed a number of subtle modelling aspects which some of the participants had clearly not appreciated - it therefore proved to be a useful learning tool. In all, 7 Newssheets were produced during the 12 month duration of the blind phase of the work.

A small number of errors was revealed in the Site Handbook and the Validation Guide. The most significant, for Phase 1, was that during the October period, Room 5 was specified as containing 1.5m ^{2}

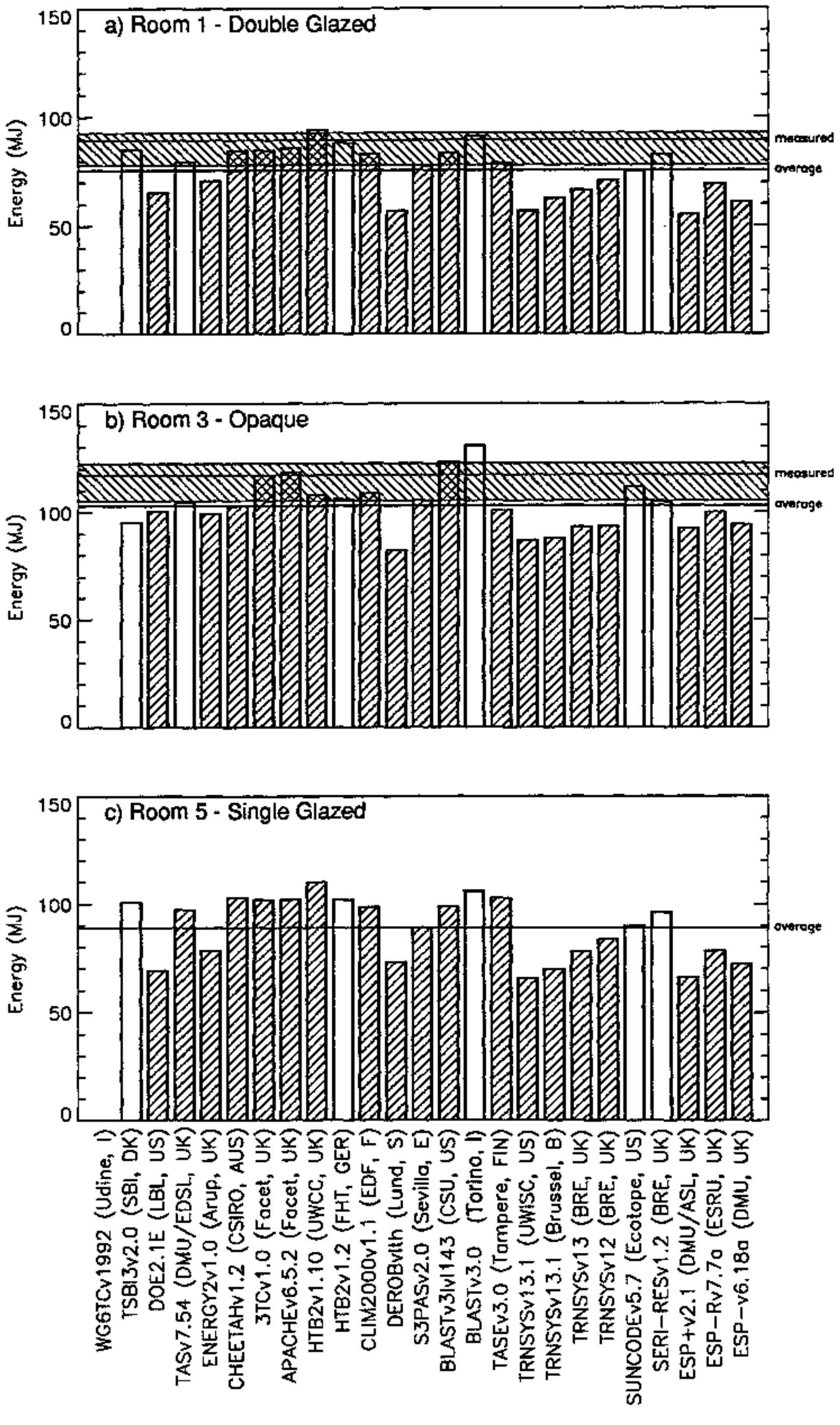


Figure 2:
Phase 1 - Total heating energy consumption
for the 7-day heated (October) period

of single glazing, whereas in fact it had 0.75m^2 of double glazing. This has been rectified in the final version of the Site Handbook (8). The predictions in this case were therefore made for a single glazed room, so they could not be compared with measured values. However, they were still used as an inter-model comparison. Weaknesses in the data included missing relative humidity data for the October period. It also proved necessary to construct hour-centred weather data (from the ½ -hour-centred data originally provided) for some programs. The uncertainties introduced by these factors (and others identified during Phase 2) were incorporated within the total uncertainty band describing all the external errors (Section 4.4).

A time-table was defined, so that the modellers were given a similar opportunity to produce the best possible results from their programs. Each modeller submitted the first set of results to the DMU team along with the input files which they created and a completed proforma describing the key features of their program.

The onus was on the modellers to conduct appropriate quality assurance checks (9); however, the DMU team also inspected the input files to uncover any obvious errors. They fed the results of this inspection back to the modellers. Many of the errors uncovered were minor, however, in two or three cases serious modelling errors had been made. Careful data checking by program users is clearly essential to ensure the quality of DSP predictions.

Following the feedback, most participants sent a second and final set of results to the DMU team, and it is these results which form the basis of the Phase 1 analysis and which are discussed in Section 5.2.

It was not possible for the authors/vendors of TAS (EDSL) and ESP+ (ASL Sterling) to participate in quite the way that is described above. In the case of EDSL, because they had previously worked with EMC using data from the test rooms, and so they may not have been considered to be working 'blind'. In the case of ASL Sterling, they did not have sufficient resources to participate at the time. To ensure that these important UK programs were included, DMU provided an initial set of input files and the first set of results for each program. These were then sent to the vendors for them to check the files and correct any errors. The two vendors then sent a second and final set of results to the DMU team (in the case of ESP+, no hourly results files were obtained). Thereafter these two programs were treated in the same way as the other 23.

This two stage process sought to reduce to a minimum the possibility that input data errors remained. Thus, any differences between the predictions and the measurements are very likely to be due to the programs themselves. If, however, despite these checking procedures, the skilled users could not produce error free input descriptions for a relatively simple building, then it is suggested that the input data structure of the program and/or the in-house quality assurance procedures should be reviewed.

5.2. Phase 1 Results

All the participants produced results for Phase 1 of the exercise, however, following their Phase 2 investigations, six participants felt that a small number of legitimate changes to their input files should be made (see Section 6). In this report the Phase 2 results from these 7 programs are clearly identified (see below), to distinguish them from the corresponding Phase 1 results from the other programs. It has been argued that the Phase 1 results should be ignored for those programs where Phase 2 results were produced (i.e. BLASTv3.01v1143, HTB2v1.2, SERI-RESv1.2, SUNCODEv5.7, TASv7.54, TSBI3v2.0, and WG6TCv1992). However, counter arguments include:

- in Phase 1 all programs were tested under the same conditions and with the same information available to all participants;
- only in Phase 1 were the programs tested in a way which precluded (fortuitous) maneuvering of the input data by program users;
- only in Phase I did the conditions mimic those which prevail in a design situation - it is the performance under these conditions which is of most interest to potential users; and

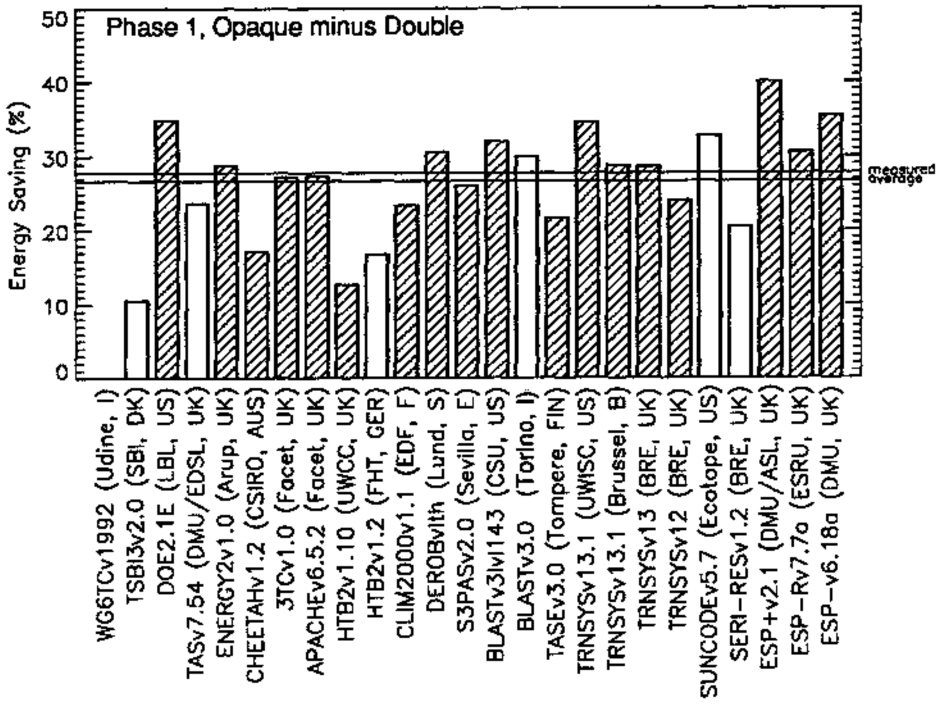


Figure 3:
Phase 1 - Difference in total heating energy between
the opaque room (Room 3) and the double glazed room (Room 1)

- One of the objectives of the study was to obtain a snapshot of program performance under realistic 'blind' conditions. All participants agreed in advance to the publication of Phase 1 results.

In this section therefore, Phase 1 results are shown for all programs including those for which Phase 2 results were subsequently submitted. To avoid any misunderstandings, however, the latter are identified in all the plots throughout the report, either by using an open rather than a hatched bar in bar charts, or by placing an * in the legend. Where the programs are mentioned in the text or in tables, *italics* are used.

The primary parameters predicted by each program, plus the measurements and the error bands around the measurements can be found in Appendix B. These are plotted in Figures 2, 4, 6 and 9. The results of the simple statistical analysis used to quantify the differences between the measurements and the predictions is also included in Appendix B. Typical hourly profiles of air temperature and energy consumption are also produced (Figures 5 and 8).

5.2.1. Heated Rooms

Results for the total seven-day energy consumption for the October 1987 period, when the rooms were heated to 30°C during the day, are given in Figure 2. A number of features are worth highlighting.

(a) Seven of the programs produced predictions which lay within the error bands for the opaque room and eleven for the glazed room. Four programs (3TCv1.0, APACHEv6.5.3, CLIM2000v1.1 and *HTB2v12)* are within the bounds for both cases. This illustrates that accurate program performance under one set of conditions cannot always be taken to mean that the program will perform well under different circumstances.

(b) Even for the completely opaque room, the predictions of the programs vary significantly (from 83MJ to 131MJ, i.e. by about 47% of the mean value, 103M3). This is only marginally less than the percentage range for the room with the double glazed window (51%) and for the room modelled with single glazing (50%). Given such a wide variation, it is impossible for the predictions of more than about 8 programs to fall within the 99-percentile uncertainty bound around the measured values, irrespective of what the actual measured values were.

(c) It is worth noting that the average result for all the programs is between 13% (Room 3 - opaque) and 15% (Room 1 - double glazed) lower than the measured value and outside the estimated error bounds. Possible reasons for this underprediction are explored later (e.g. section 7.2).

Considering the predictions for the savings in energy consumption which are made due to having south facing double glazing (Room 1) rather than an opaque room (Room 3), a number of observations can be made (Figure 3).

(d) The average saving predicted by all the programs is very close to the actual measured savings.

(e) The predictions of 3TCv1.0, APACHEv6.5.3, ENERGY2v1.0, TRNSYSv13.1 (Brussel) and TRNSYSv13.1 (BRE) were within about 1% of the measured energy savings. For the last three named programs this occurs because the absolute energy predictions were underpredicted in a consistent fashion in both Room 1 and Room 3.

(f) The predicted energy savings vary considerably, from a low of 11% for *TSBI3v2.0* to 40% for ESP+v2.1. Ignoring the TSBI3v2.0 results, for which Phase 2 results were produced, the variability is from 13% for HTB2v1.10 to 40%. The measured saving was 24%.

The results for the maximum and minimum temperatures predicted in the three rooms for the same period are given in Figure 4.

(g) The programs predict minimum temperatures in the opaque room which vary by about 4°C. In the glazed rooms the predicted peak temperatures vary by 11°C, or from 3°C to 14°C above the setpoint. Such large ranges are worrying as the programs are often called on to predict peak temperatures in real design studies.

(h) In general, the programs under-predicted the minimum room air temperature and no program produced maximum and minimum temperatures which were always within the estimated error bands.

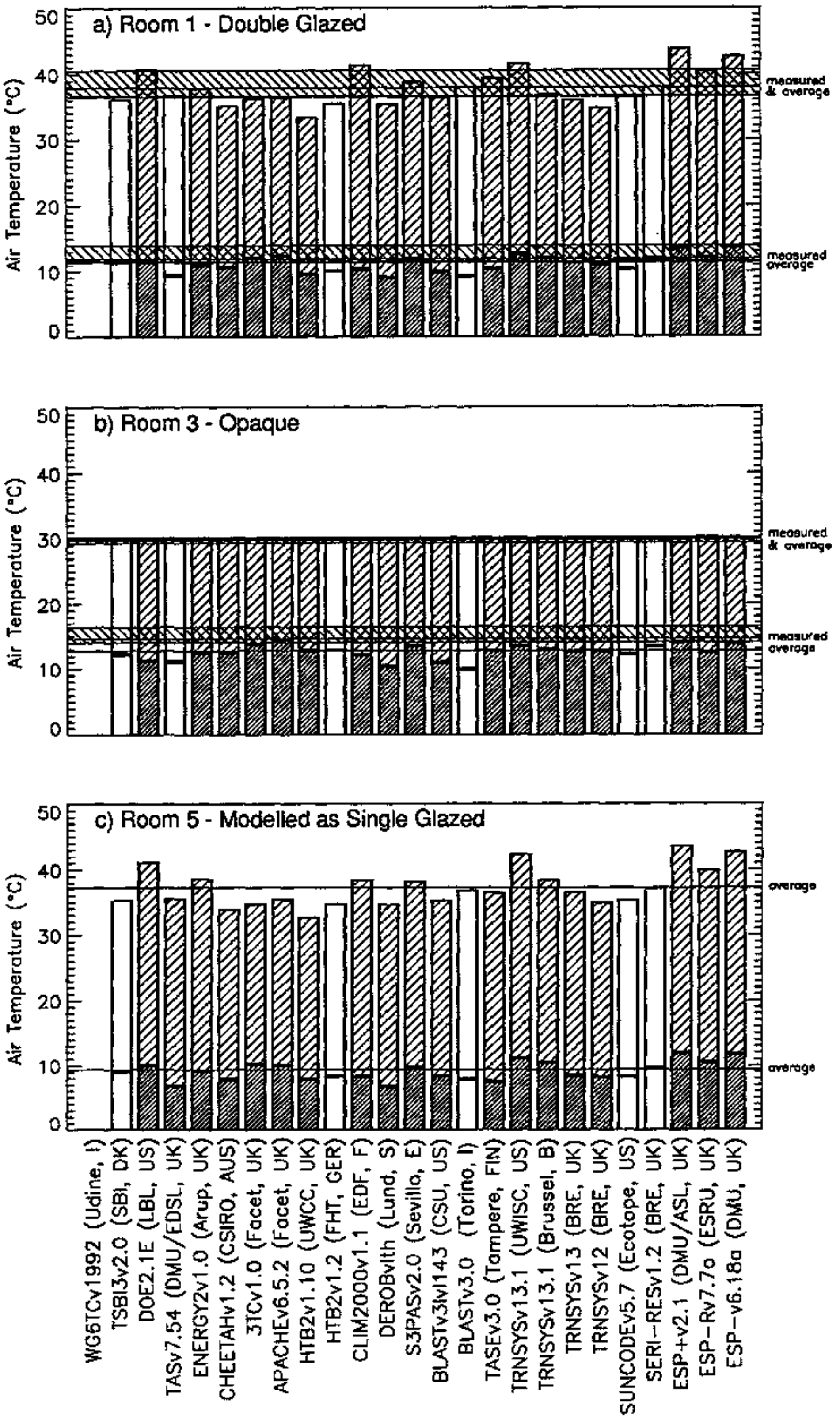


Figure 4:
Phase 1 - Maximum and minimum temperatures during the 7-day heated (October) period

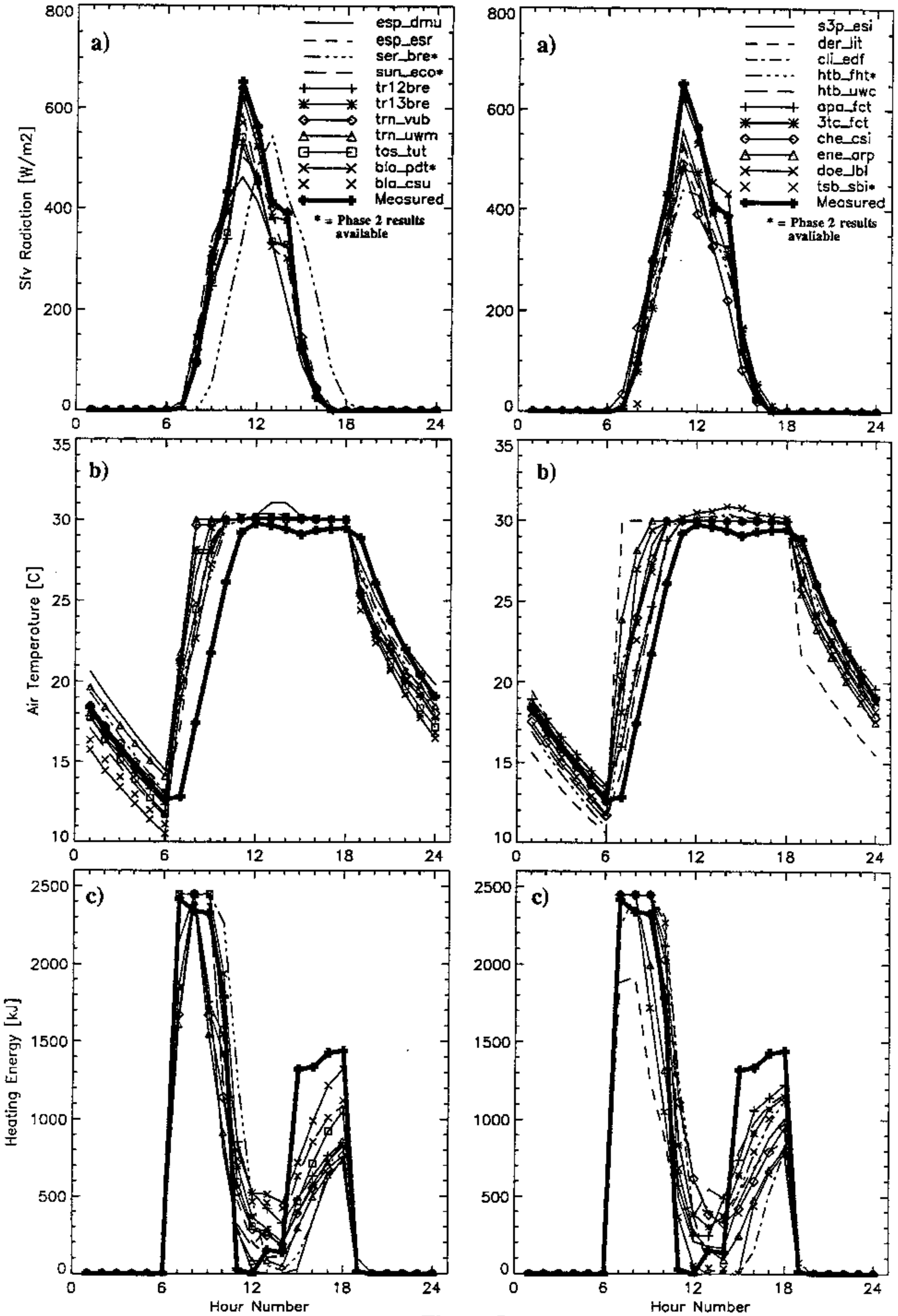


Figure 5:
Phase 1 - Typical hourly predictions for one day in the heated (October) period (Day no. 296, double glazing)

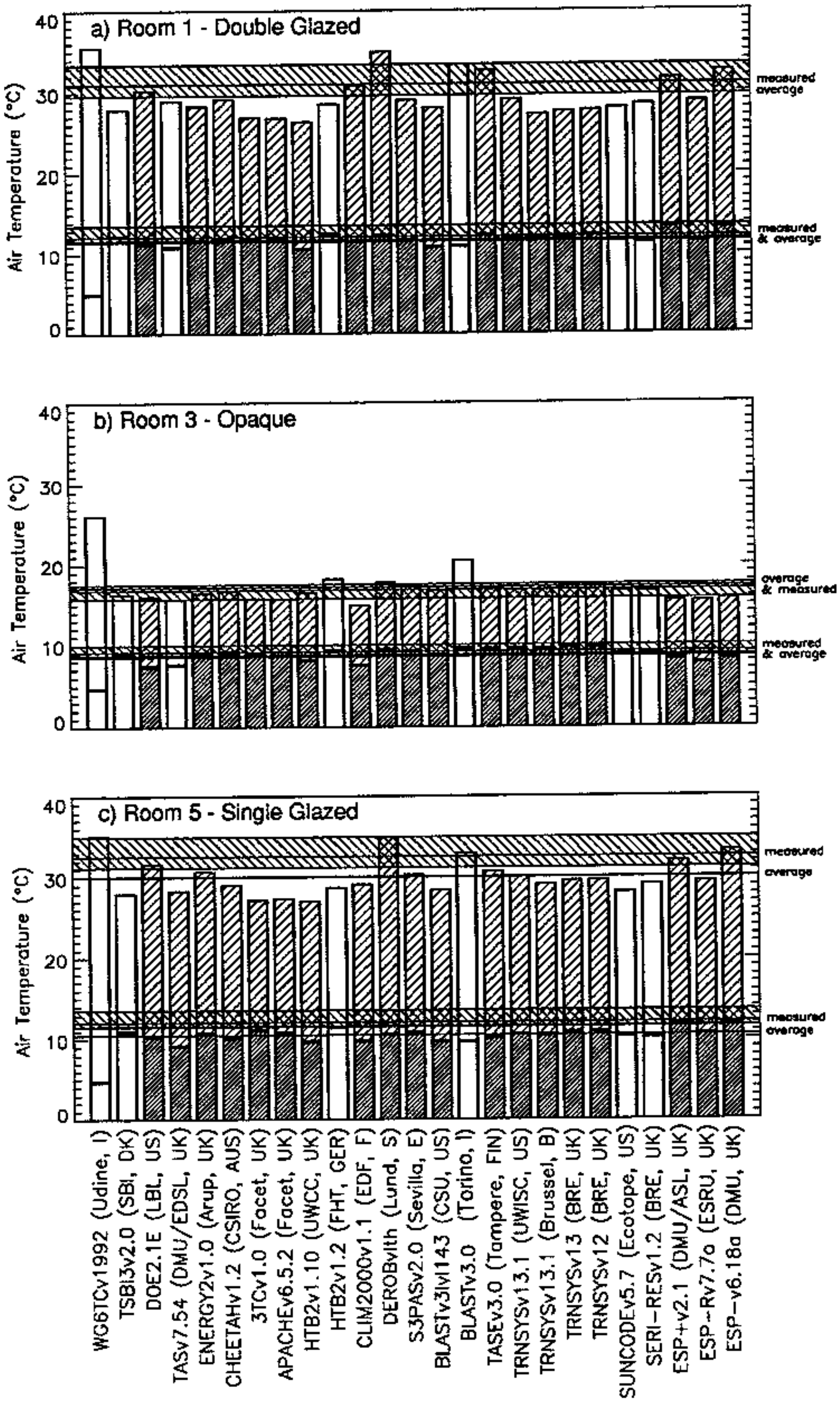


Figure 6:
Phase 1 - Maximum and minimum temperatures during the 7-day free-floating (May) period

Irrespective of the type of South facing facade, the hourly predictions during each day of the October (heated) period showed similar behaviour, this is illustrated in Figure 5.

(i) All the DSPs predicted a more rapid rise in the air temperature at the start of the heating period than that which was actually measured (e.g. Fig.5b). The DSPs also predicted a faster decrease in the air temperature at the end of the heating period; DEROBv1th exhibited the most extreme divergence of this type; the predictions of APACHEv6.5.3 and HTB2v1.10 (and *HTBv1.2)* were the closest to the measurements.

(j) Because the programs predicted a rapid rise in the air temperature, the set-point was reached earlier (in one case 3 hours earlier) than was in fact the case (Fig.5b). As a result, the predicted power output from the heater tended to decline much more rapidly than the measured power output (Fig.5c). This could be caused, in part, by the assumption, in the programs, that the room heater and controls are ideal, whereas in fact they have associated time lags and delays (8). These control issues alone may not, however, explain the large differences between the measured and predicted total energy consumptions (Fig.2).

(k) During the middle of the heating period, during periods of high solar gains (Fig.5a), the actual room seems to make greater use of the gains (the power demand is very small) than the programs assume (for which the heating demand is greater). During the last four hours of the heating period, the programs predict much lower power demands than was actually measured (Fig.5c).

(1) This general tendency to underpredict the heating demands (after the initial start-up phase) also occurred in the opaque room and is the main reason for the lower total energy consumption predictions of the programs.

(m) The statistical analyses (Appx.B, Table B2) indicate that APACHEv6.5.3 and 3TCv1.0 gave the closest overall agreement between the measured and predicted hourly heating energy values for both the double glazed and opaque cases. The *BLASTv3.0* results showed good agreement for the double glazed room, and the *SUNCODEv5.7* results for the opaque room. For both the double glazed and the opaque rooms, DEROBvIth and TRNSYSv13.1 (UWISC and Brussels) showed marked differences from the measurements.

5.2.2. Free-floating Rooms

The results for the maximum and minimum air temperatures during the 7-day free-floating May period are given in Figure 6. The following aspects are worth noting.

(a) For all 3 rooms the program *WG6TCv1992,* which was still under development, produced results which were significantly different from those of the other programs and from the measurements (Figs. 6a, 6c, 6b and Appx. B). The Phase 1 results for *WG6TCv1992* are ignored in the observations which follow (Phase 2 results are shown later).

Considering initially the temperatures in the glazed rooms (Figs.5a,c, 6b and Appx.A, Table A2), the following observations may be made.

(b) Even ignoring *WG6TC,* the predicted peak temperatures in the double glazed room varied by 8.6°C, in the single glazed room the range was 7.9°C. Most of the values were lower than those measured and outside the estimated uncertainty bands. The predicted minimum temperatures were less variable (less than 3°C). Two of the programs (ESP+v2.1 and ESPv6.18a) predicted both maximum and minimum temperatures inside the estimated error bands for both glazed rooms (Figs. 6a and 6c).

(c) For the double glazed room, DEROBvIth generated peak temperatures and hourly temperature trends which were markedly different from both the measurements and the other predictions; the *SERI-RESvl 2* results suggest a phase shift (Fig.8b).

Considering the hourly temperature predictions (Figure 8), the following may be noted.

(d) The hourly temperatures predicted by *WG6TCv1992* and DEROBv1th are considerably higher than the values predicted by the other programs.

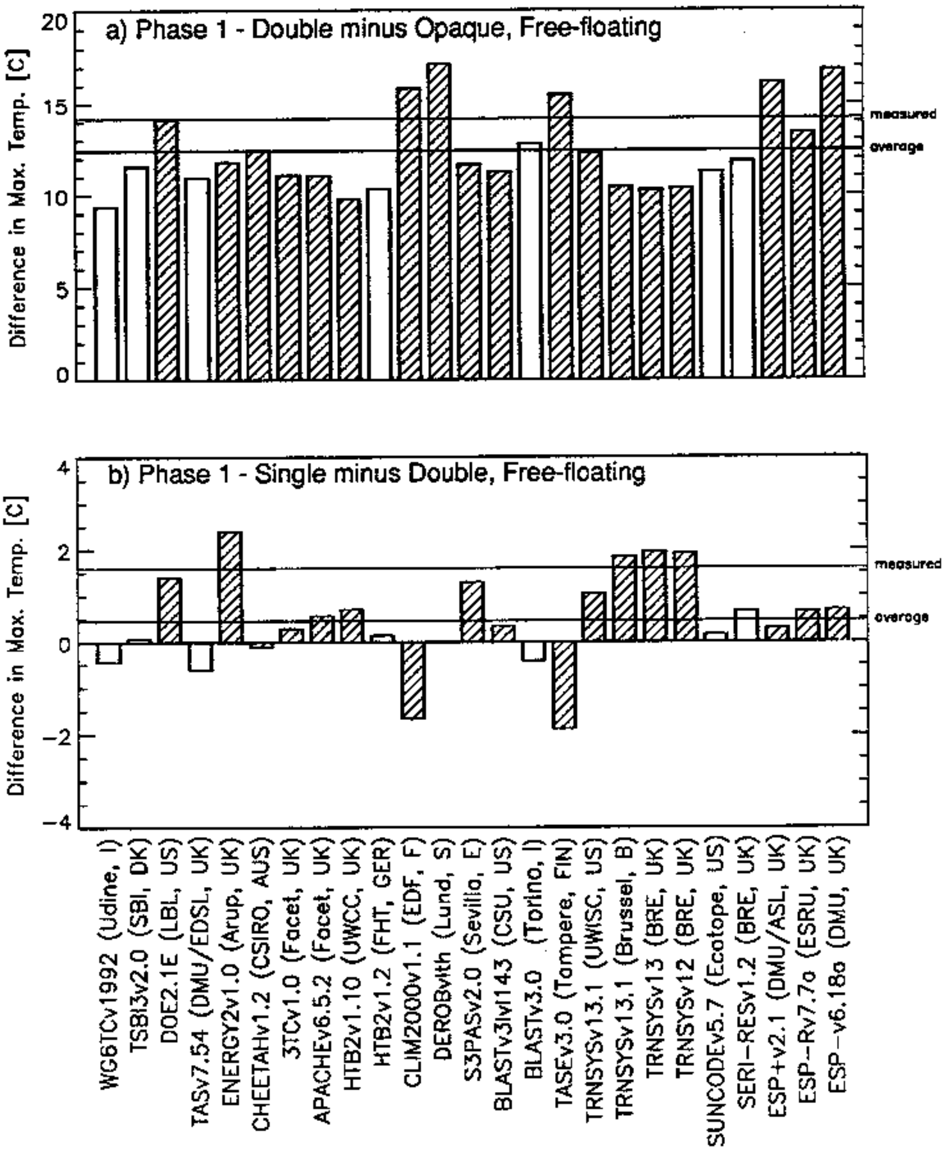


Figure 7:
Phase 1 - Differences in maximum temperatures between different rooms during the free-floating (May) period

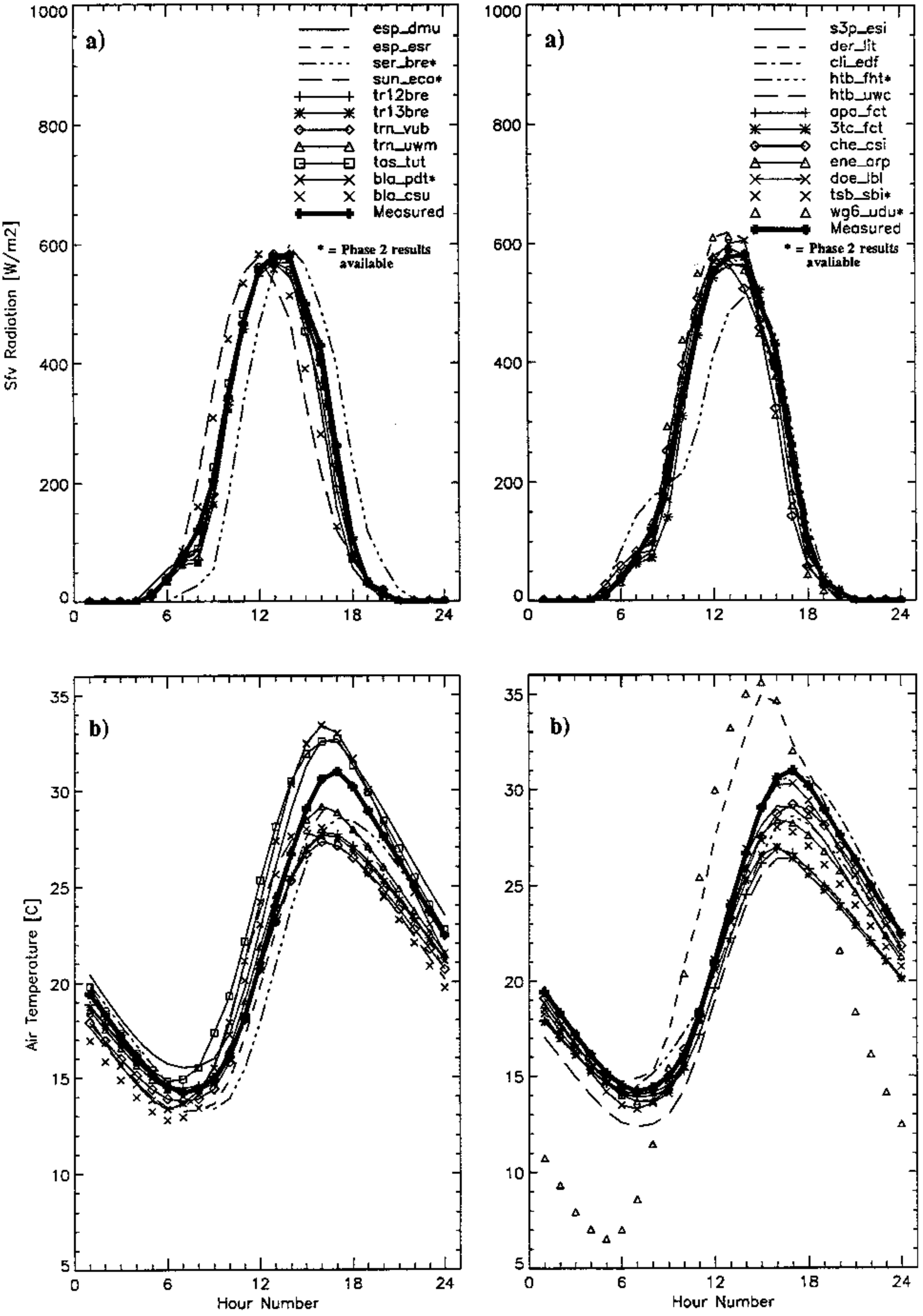


Figure 8:
Phase 1 - Typical hourly predictions for one day in the free-floating (May) period (Day no.147, double glazing)

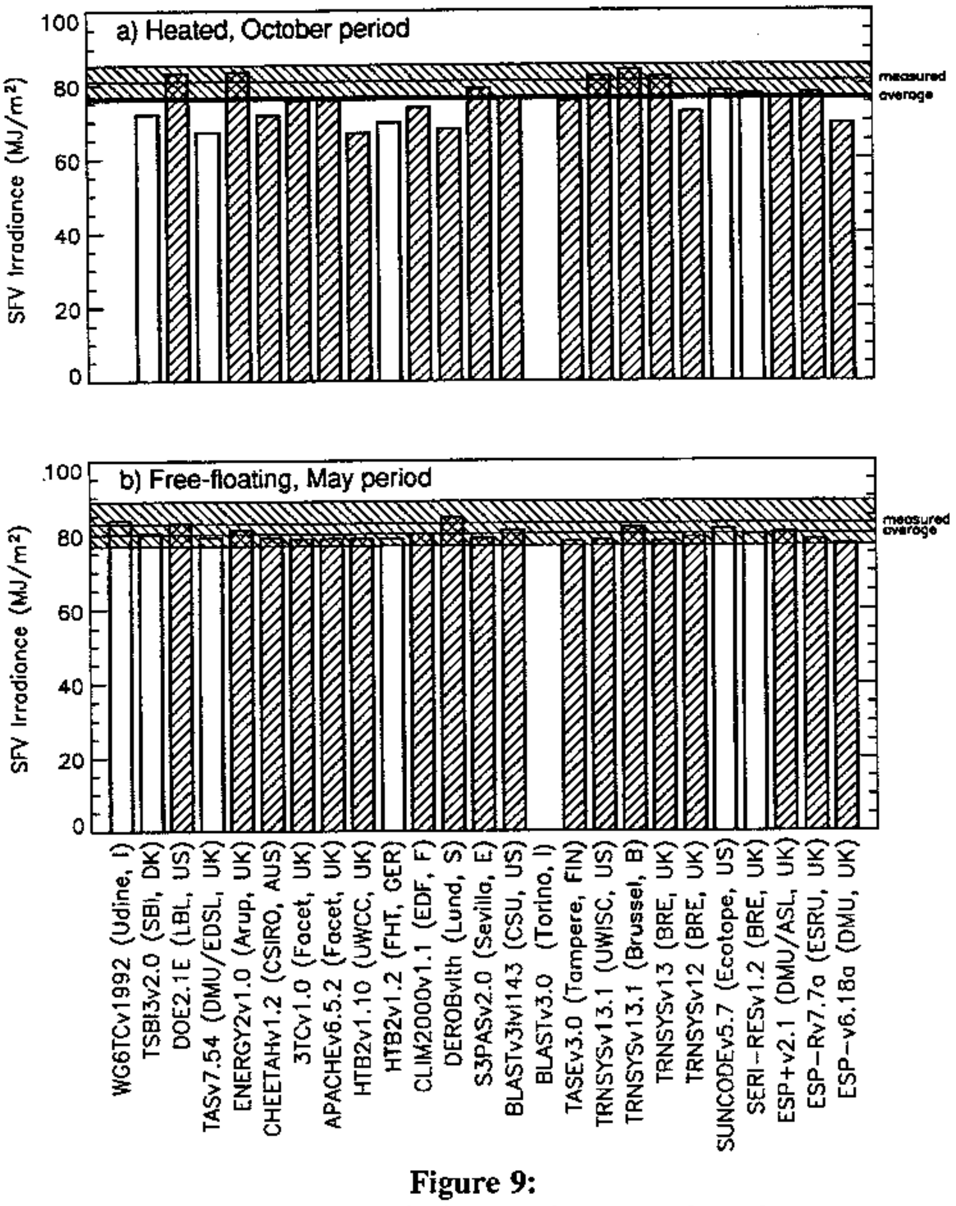


Figure 9:
Phase 1 - Total south facing vertical solar irradiances for the two 7-day periods

(e) The values predicted by 3TCv1.0, APACHEv6.5.3 and HTB2v1.10 are lower than those of most of the other programs. This is reflected in the lower maximum temperatures which they predict (Figure 6).

Reflecting now on the results for the opaque room (Fig.6b and Appx.B, Table B3), the following can be noted.

(f) Eleven programs predicted both maximum and minimum temperatures for the opaque room which were within the error bars. This represents much better performance than for the glazed rooms (Fig.6b). Overall, the deviations from the measured values were smaller than for the glazed rooms (Appx.B, Table B3).

(g) In the opaque room, *BLASTv3.01v1143* produced a higher peak temperature (Fig.6b) than the other programs (except *WG6TCv1992),* and the predictions of CLIM2000v1.1 and the ESP programs were noticeably lower. Overall, these programs differed from the measurements noticeably more than did the other programs (Appx.B, Table B3).

The differences in temperature between the peak air temperature in the double glazed room and in the opaque room are shown in Figure 7a. A similar plot for the differences between the temperature in the single glazed room and the double glazed room is shown in Figure 7b. The following observations can be made.

(h) All the programs correctly predicted that the peak temperatures would be higher in the double glazed room than the opaque room (Fig.7a).

(i) The predicted amount by which the maximum temperature in the double glazed room would exceed that in the opaque room varied from 9.8°C (HTB2v1.10) to 17°C (DEROBv1th); a range of 60% (compared to the average predicted temperature increase).

(j) Two programs (ESP-rv7.7a and DOE2v1E) predicted a maximum temperature increase in the double glazed room, compared to the opaque room, which was within 1°C of the actual measurement (i.e.14.2°C). A further six programs *(BLASTv3.01v1143,* CHEETAHv1.2, CLIM2000v1.1, ESP+v2.1, TASEv3.0 and TRNSYSv13.1 (UWISC) were within 2°C of the measurements.

(k) The maximum air temperature in the single glazed room was measured to be only 1.6°C higher than in the double glazed room. All except six of the programs correctly predicted the trend towards higher temperatures in the single glazed room.

(1) Eleven program/user combinations predicted the higher temperature in the single glazed room within 1°C of the actual measured increase. They were DOE2v1E, ENERGY2v1.0, ESPv6.18a, ESP-rv7.7a, HTB2v1.10, S3PASv2.0, *SERI-RESv1.2* and the 4 TRNSYS results.

5.23. South-facing Solar Irradiances

(a) The South facing solar irradiance predictions for the October period, when the sun was lower in the sky, and the sky conditions were changing rapidly, were more variable than in the May period (Fig.9).

(b) In the October period, most programs tended to predict total irradiances which were less than the measured values. (This cannot therefore explain the general under-prediction of heating energy consumption). All together, 11 programs produced predictions for the October period which were within the estimated error bands.

(c) The hourly plots (Fig.5&8) and the statistics (Appx.B, Table B4) illustrate the marked divergence of *HTB2v1.2, SERI-RESv1.2* and *SUNCODEv5.7* and, to a lesser extent, BLASTv3.01v1143, CHEETAHv1.2, ESPv6.18a and HTB2v1.10, from the measurements.

(d) For the May period, all the programs predicted values within the estimated error bands.

53. Discussion of Phase 1 Results

53.1. Inter-model Variability

Before considering the actual measurements it is worth discussing just the variability in the predictions between the programs. The causes of this variability could be:

(a) remaining errors made by the program users.

(b) ambiguities in the descriptions of the rooms provided, leading to substantial differences in the way the same aspect was modelled in different programs; or

(c) fundamental differences in the algorithms and sub-models employed by the DSPs.

The first two causes are external errors, the last one covers internal errors. They are discussed in turn.

The likelihood that errors by program users are having a significant effect on the results was minimised by making sure that highly skilled users were ultimately responsible for producing the results and/or conducting the quality assurance checks on the inputs (section 5.1), in a short all the possible steps to avoid user errors. It is worth noting however, that if such errors do remain it is likely that they would also exist in a real design situation when less rigorous data checking may take place, the buildings will be more complex and tight (commercial) time pressures will exist.

The likelihood of ambiguities in the room descriptions and data was minimised by the provision of a hotline. This ensured all the modellers had immediate access to any information which they believed necessary in order to resolve uncertainties. The newssheets also ensured that all participants were made aware of any further information.

Some programs will be sensitive to relative humidity, which was not measured during the October period. Also, some programs required hour-centred weather data, rather than the ½-hour-centred data originally provided. For the October period, the original 5-minute data was not available, so the hour-centred data had to be constructed by interpolation; this led to some smoothing. The effect of both of those factors is small (section 6.2), and affects only a handful of programs. These factors explain only a small part of the observed inter-model variability in the results.

The foregoing arguments suggest that some of the inter-model variability may be due to remaining external errors, but that most of the variation is much more likely to be due to fundamental internal calculation differences between the DSPs. Some features of the observed variability add further credence to this proposal. The variation between the results obtained for some programs, but with different operators and different version numbers, is much less than the variation between all the programs. Furthermore, all the users of one program predicted broadly similar performance. For example, for the heating energy in the double glazed room, the range for all programs was 48%, but all 3 ESP results lay below the average from all 24 programs (the range between the 3 results was 17%), as did the four TRNSYS results (range 20%). The two BLAST results were above the mean (range 12%), as were the two HTB2 results (range 12%). The two SERI-RES results straddled the mean (range 12%). These programs also produced broadly similar performance in the other two heated cases. The remaining differences (i.e. 12% to 20%) could be due to differences in the program version and/or user effects. This issue should become clearer following Phase 2.

53.2. Comparisons with Measurements

The general tendency of the programs to under-predict the heating energy relative to the measurements inevitably led to the suspicion that there was a basic flaw in the description of the test rooms, as a result of which the program users might have been led towards modelling a room which had an overall heat loss rate which was less than that of the actual rooms. If this was the case, one would expect the programs to predict internal temperatures which were generally higher than the measured temperatures. In fact, in both the heated period (at times when solar gain drove the temperature above the setpoint) and in the free-floating period, there was a tendency for the programs to under-predict the internal temperatures. Despite this contradiction, the quality assurance procedures undertaken by the EMC team were reviewed and further tests undertaken; these entailed:

Program	Modifications
WG6TCv1992 (Udine, I)	*- coding error was corrected*
TSBI3v2.0 (DBRI, DK)	**- incorrect heater schedule had been used (heater off on last day for opaque case). This was corrected** - horizon altitude was incorrectly modelled - direct normal and diffuse horizontal irradiance data preferred
TASv7.54 (DMU/EDSL, UK)	- construction details refined - shading effect by neighbouring test room modelled - more detailed modelling of interior solar distribution - infiltration rates of roofspace and floorspace were adjusted until their air temperatures matched the measured temperatures - internal clock adjusted by ½hour for consistency with measured data
HTB2v1.2 (FHT, GER)	*- solar calculation routine was originally in error. This was corrected* - timing convention in error for May period. This was corrected.
BLASTv3.0lvl203 (Torino, I)	**- ceiling insulation had been omitted. This was added** *- the partition wall had been modelled as an external wall. This was corrected.* - adjacent cell shading was not modelled **- roof absorptivity was in error. This was corrected**
SUNCODEv5.7 (Ecotope, US)	**- building orientation was incorrectly specified (as 9° east of south rather than 9° west). This was corrected**
SERI-RESv1.2 (BRE, UK)	- inconsistency in climate data was corrected

italics indicate program improvement
bold indicates input error correction
others indicate modelling improvement

Table 11: Modifications to modelling procedures to produce the Phase 2 results

- decommissioning and partly dismantling one room; and
- measuring the heat-loss coefficient (UA-value) of one room.

The dismantling of the room was witnessed by DMU so that independence from the test cell operator EMC could be assured. The purpose of dismantling one room was to check that the specification given in the Site Handbook matched that of the actual rooms and, most importantly, to ensure that all the insulation was dry and in position (see Validation Package, Quality Assurance report). This process did not reveal any faults in the construction of the moms. In other words, they were constructed as described in the Site Handbook.

Overall, the available evidence at the end of Phase 1 is that differences between many of the programs and the measurements (that exceed the error bands) cannot be explained by experimental errors, which have gone undetected. Nevertheless, the overall UA-value of the test cells was calculated whilst Phase 2 of the exercise was underway, and further sensitivity analyses were also undertaken to account for the small additional experimental uncertainties which were identified by modellers during Phase 2. (The error bands used here for the analysis of the Phase 1 results include these additional small uncertainties.)

6. PHASE 2: NON-BLIND INVESTIGATIONS

6.1. Management

The primary purpose of Phase 2 was to give all participants the opportunity to explore the reasons for any divergence between the predictions of their programs and the measurements. To facilitate this, all program users were sent a diskette containing all the data actually measured in the rooms during both periods, namely: the air temperature at three levels, the temperature of the floor, back wall and ceiling, the South facing vertical solar irradiances, the heater power input (October period only) and the roofspace and floorspace temperatures. The participants were asked to provide a three page report explaining their investigations. They were invited to:

- explore the reasons for any divergence between the predictions and the measurements;
- undertake sensitivity analyses with their own programs;
- provide a new set of predictions where modifications to the room descriptions or the program had been made;
- comment on the IEA empirical validation exercise; and
- provide further descriptive information about their program.

It was hoped that these reports would permit improvements to be made to the Validation Package, and that they would also help the formulation of recommendations about the conduct of future validation exercises. They would also help in directing program development work by highlighting perceived areas of weaknesses in the current generation of DSPs. The participants were given about four months to undertake this work.

The modeller's reports were submitted to the DMU team who checked each report to ensure that the information was clear, complete and logical. The feedback to each participant sought only to clarify such ambiguities. No attempt was made to direct or constrain the arguments they contained except where the contents were demonstrably factually incorrect. Model User Reports (MURs) were obtained from 11 participants (Table 7), and the final versions of all reports are reproduced in Appendix F. Seven new hourly results sets were obtained. Four of the modellers used exactly the same program versions as in Phase 1. An updated version of *BLAST* (v3.01v1143 -> v3.01v1203) was used to produce the Phase 2 results. The *WG6TCv1992* user corrected a coding error for Phase 2, and a slightly modified version of *TASv7.54* was used (Table 11).

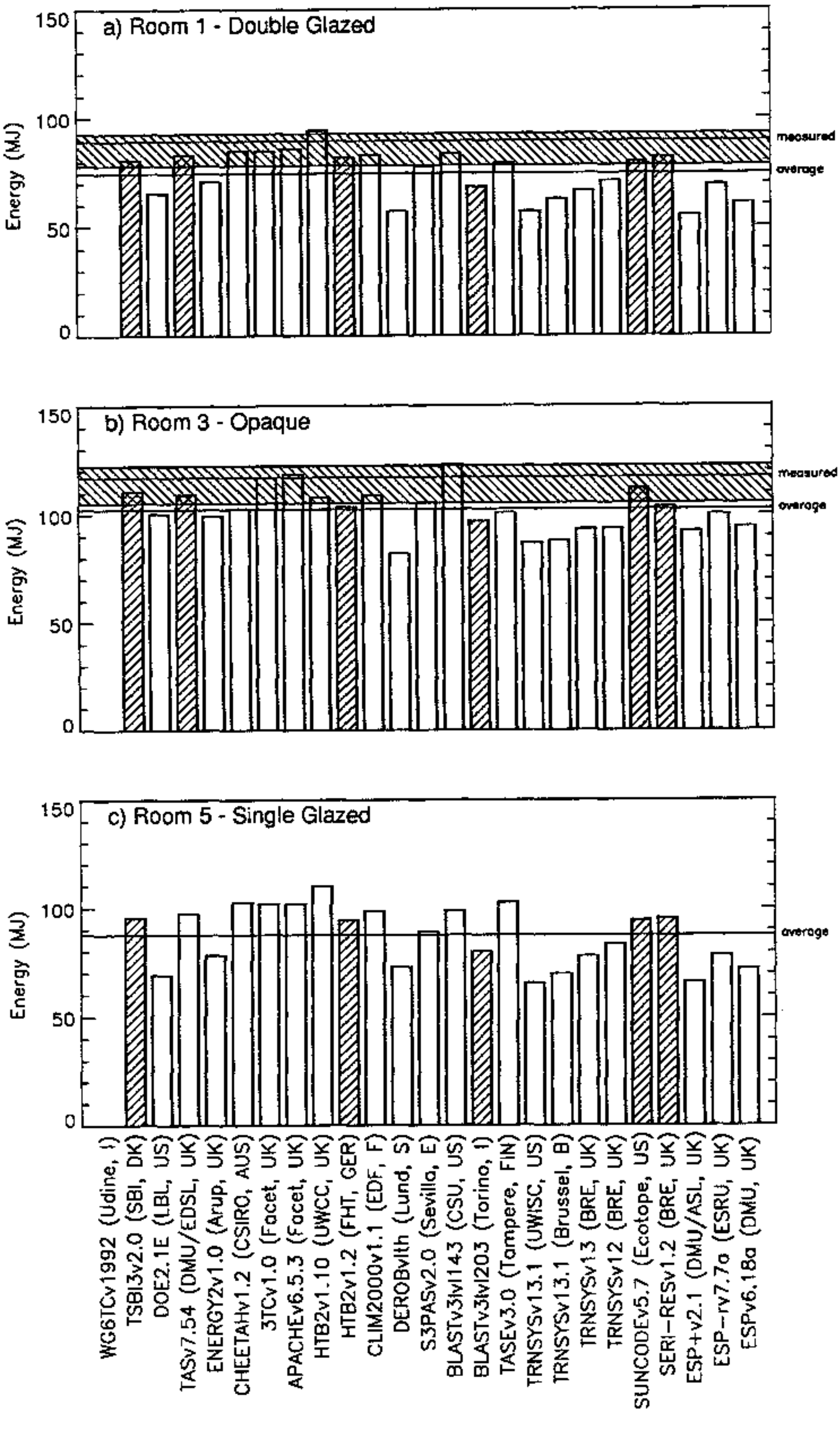


Figure 10:
Phase 2 - Total heating energy consumption
for the 7-day heated (October) period

6.2. Refining Uncertainty Bands

The error bars originally produced in Phase 1 had been estimated using SERI-RES and a list of uncertain inputs taken from previous work (6). Some participants felt that these bars may have been inappropriate for their particular program. Therefore, a table of the estimated uncertainties in all the important program input parameters (24 in total) was circulated. This was intended to enable the program users to calculate the error bars with their own programs. Only one participant (using TASEv3.0) attempted such an analysis (see Appx.F7).

Other participants identified areas in which they believed the information supplied in Phase 1 was either incorrect, deficient or more uncertain than the management team had originally thought. The important points, which impact on the analysis of program performance are:

- at any time, the air temperature varies throughout the room;
- the edge and corner effects had been underestimated;
- the relative humidity data was either missing (October period) or possibly inaccurate (May period); and
- the solar radiation data were inconsistent.

The uncertainty in each of these aspects was estimated (Appx.D), and the impact of the individual uncertainties on the primary parameters was estimated (for each room and weather period) using Differential Sensitivity Analysis (DSA). Because SERI-RES does not utilize relative humidity to calculate external longwave radiation (i.e. is completely insensitive to it) ESP was used in the DSA for this parameter. All these individual sensitivities were added to those already calculated (section 4.4) to produce the total uncertainty bands. (The largest increase in the total uncertainty occurred for the maximum temperature in the double glazed heated room, where the error band width changed from 3.2°C to 4°C. For all other parameters the change was either zero or very small). It is these updated bands which have been plotted throughout this report (including section 5 which dealt only with the Phase 1 results).

The refined list of uncertainties for all parameters (Table 9) has been incorporated in the Validation Package (see Volume 2 of this final IEA report). The issues raised by the model user reports are summarised in section 7 of this report, along with the results of the salient sensitivity analyses. Each of the points is discussed in greater detail in Appendix D.

6.3. Phase 2 Results

New, Phase 2, results were produced for all the 6 cases for *BLASTv3.01v1203, HTB2v1.2, SERI-RESv1.2, SUNCODEv5.7* and TSBI3v2.0,. *TASv7.54* results were produced for the double glazed and opaque buildings. For the unheated, free-floating buildings results were obtained from WG6TCv1992.

The new results were produced by the participants for a number of reasons (Table 11). These were mostly because there were legitimate reasons for modifying the original Phase 1 input files, either because mistakes had been made or because the room modelling could be improved upon.

In the heated rooms, for all the programs except *BLASTv3.01v1203,* the new Phase 2 results differed only slightly from the original Phase 1 results. For example, in the double glazed room the energy consumptions changed by less than 5MJ, and in the opaque room by less than 16MJ. For *BLASTv3.01v1203, the* changes were much larger, resulting in energy consumption predictions which were up to 36MJ lower than for Phase 1.

For the free-floating cases, the new results for *BLASTv3.01v1203* and *WG6TCv1992* differed most noticeably, e.g. a reduction of 8°C in the maximum temperature predicted by *WG6TC* during the May period.

The results of Phase 2 have been plotted in the same way as for Phase 1 (Figures 10 to 17). In all these figures the new, Phase 2, results replace those produced in Phase 1. The programs for which Phase 2 results exist are identified throughout (see section 5.2).

Overall observations are made following the style adopted for Phase 1 (section 5.2). These change very little even though some individual results have changed.

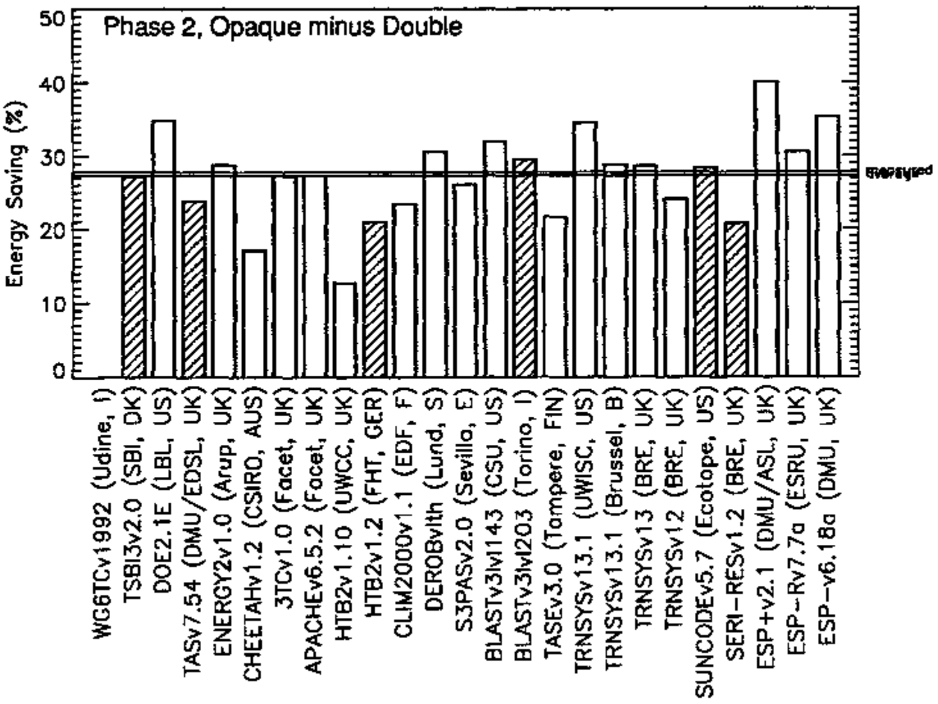


Figure 11:
Phase 2 - Difference in total heating energy between the opaque room (Room 3) and the double glazed room (Room 1)

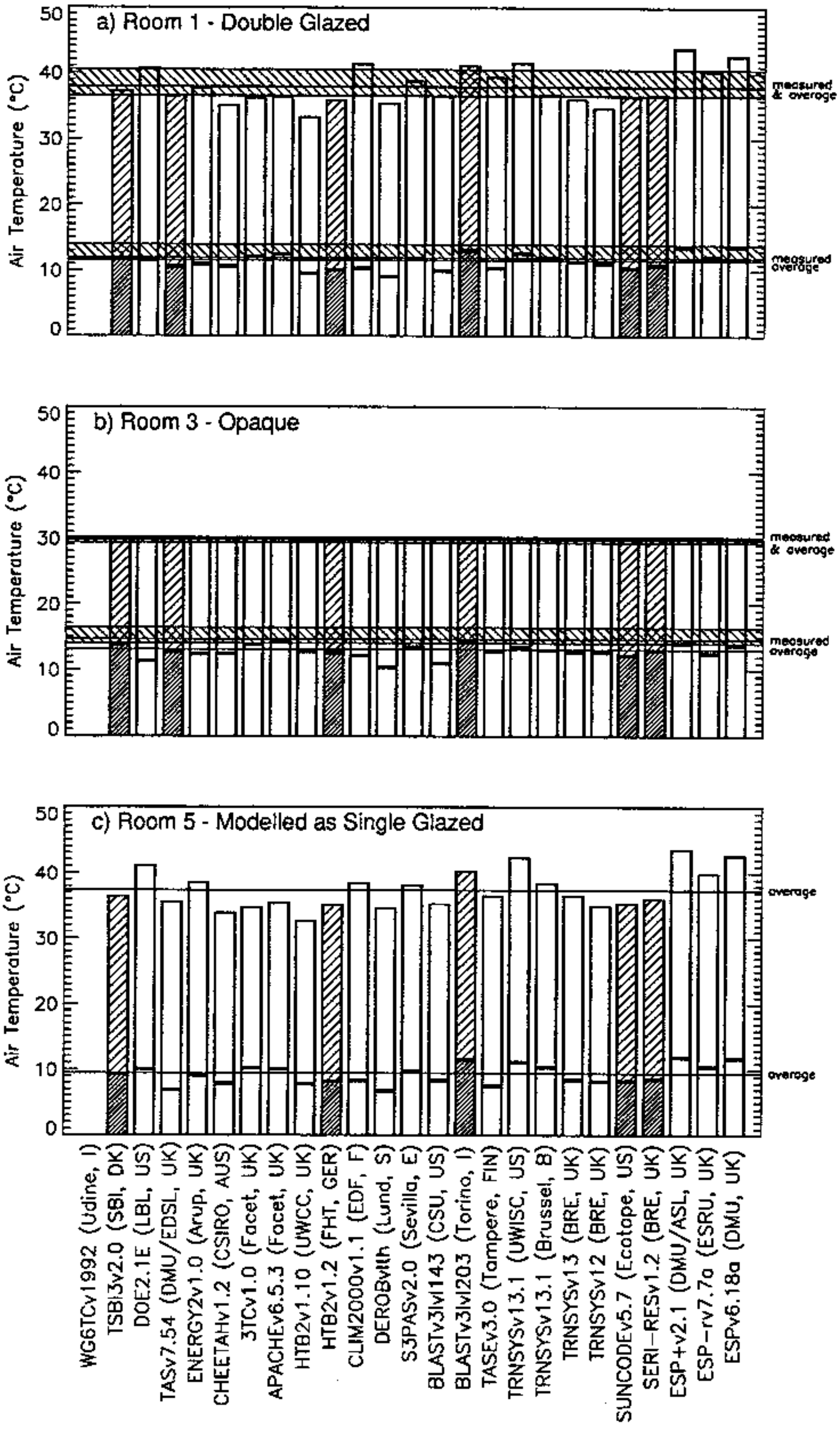


Figure 12:
Phase 2 - Maximum and minimum temperatures during the 7-day heated (October) period

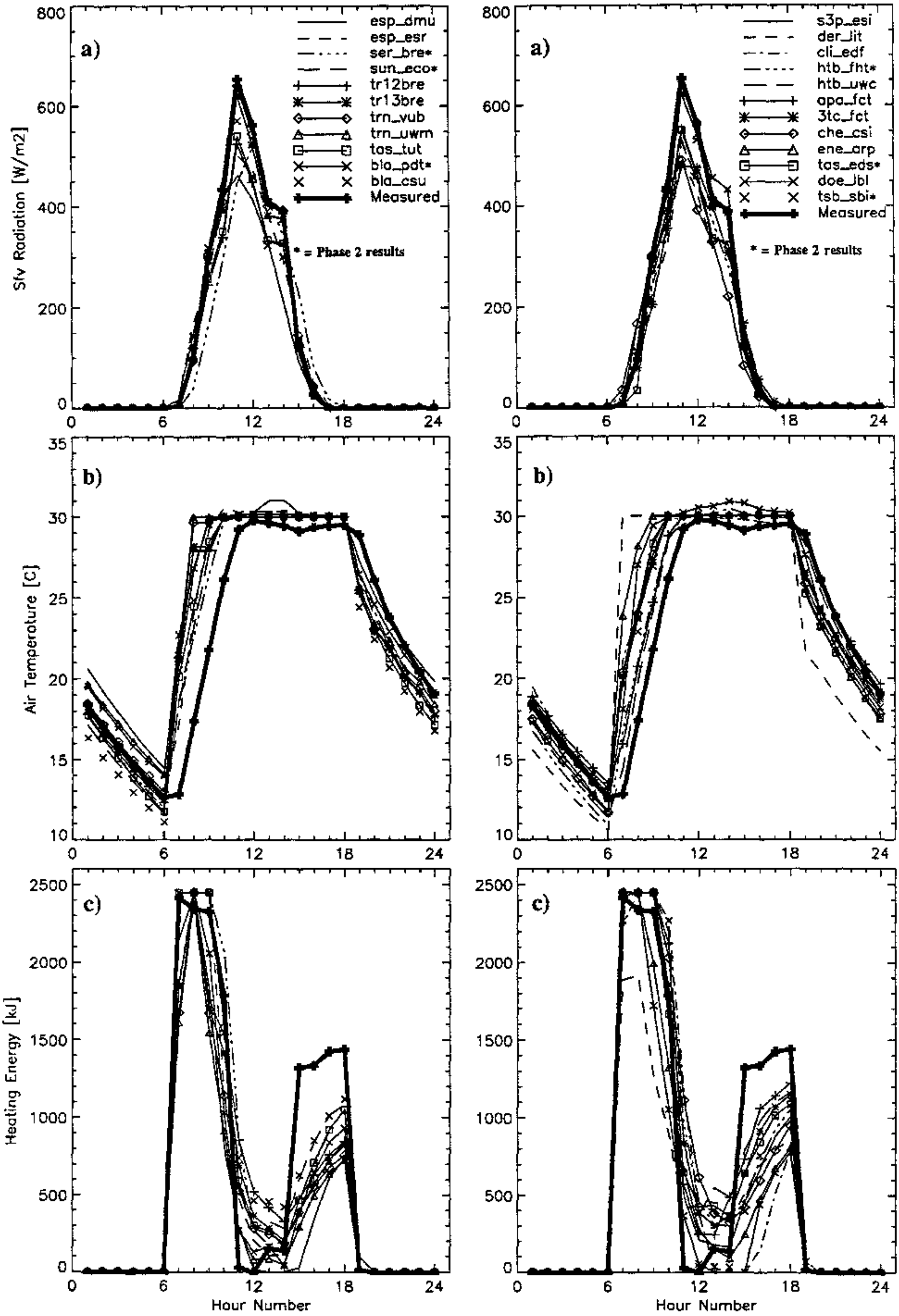


Figure 13:
Phase 2 - Typical hourly predictions for one day in the heated (October) period (Day no. 296, double glazing)

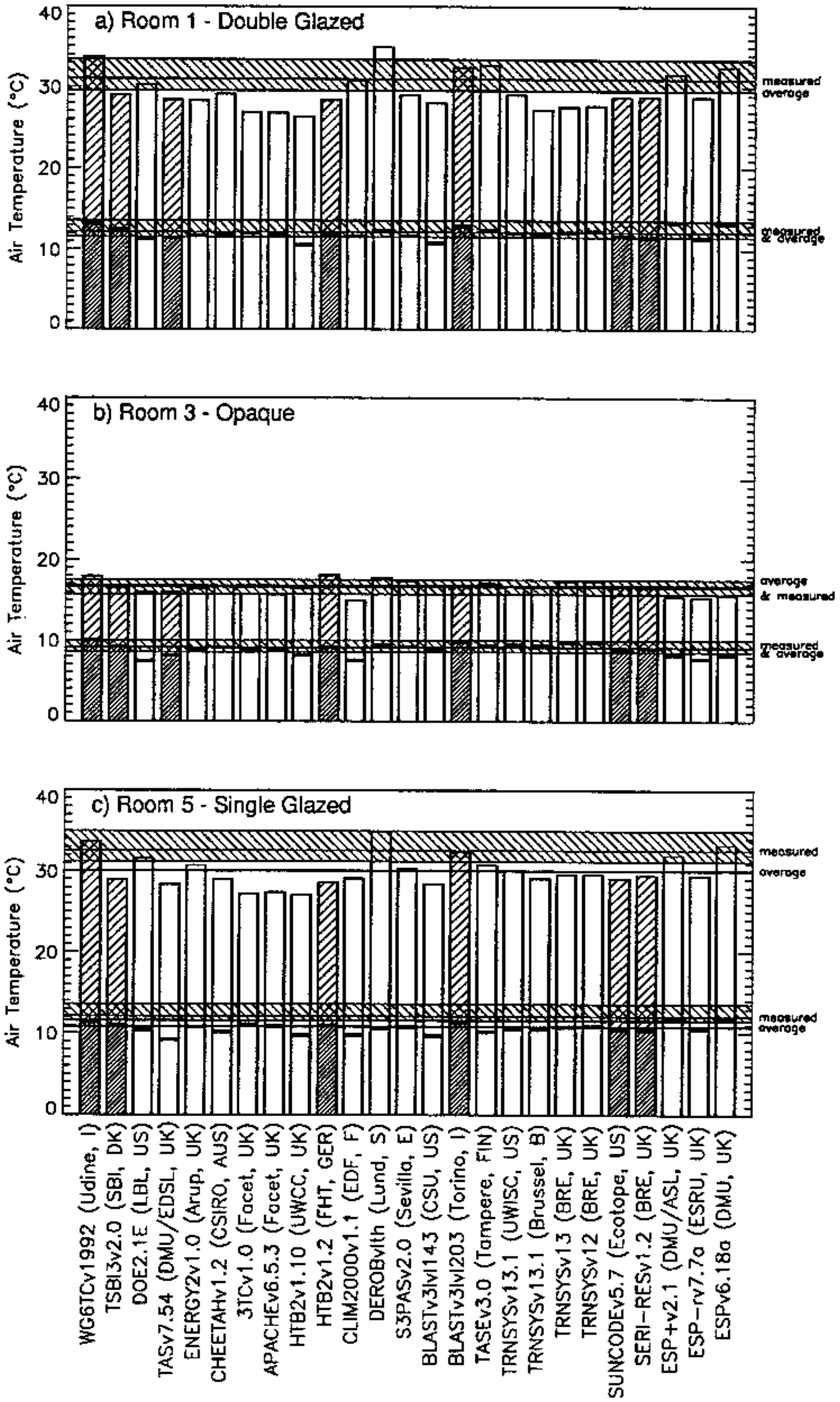


Figure 14:
Phase 2 - Maximum and minimum temperatures during the 7-day free-floating (May) period

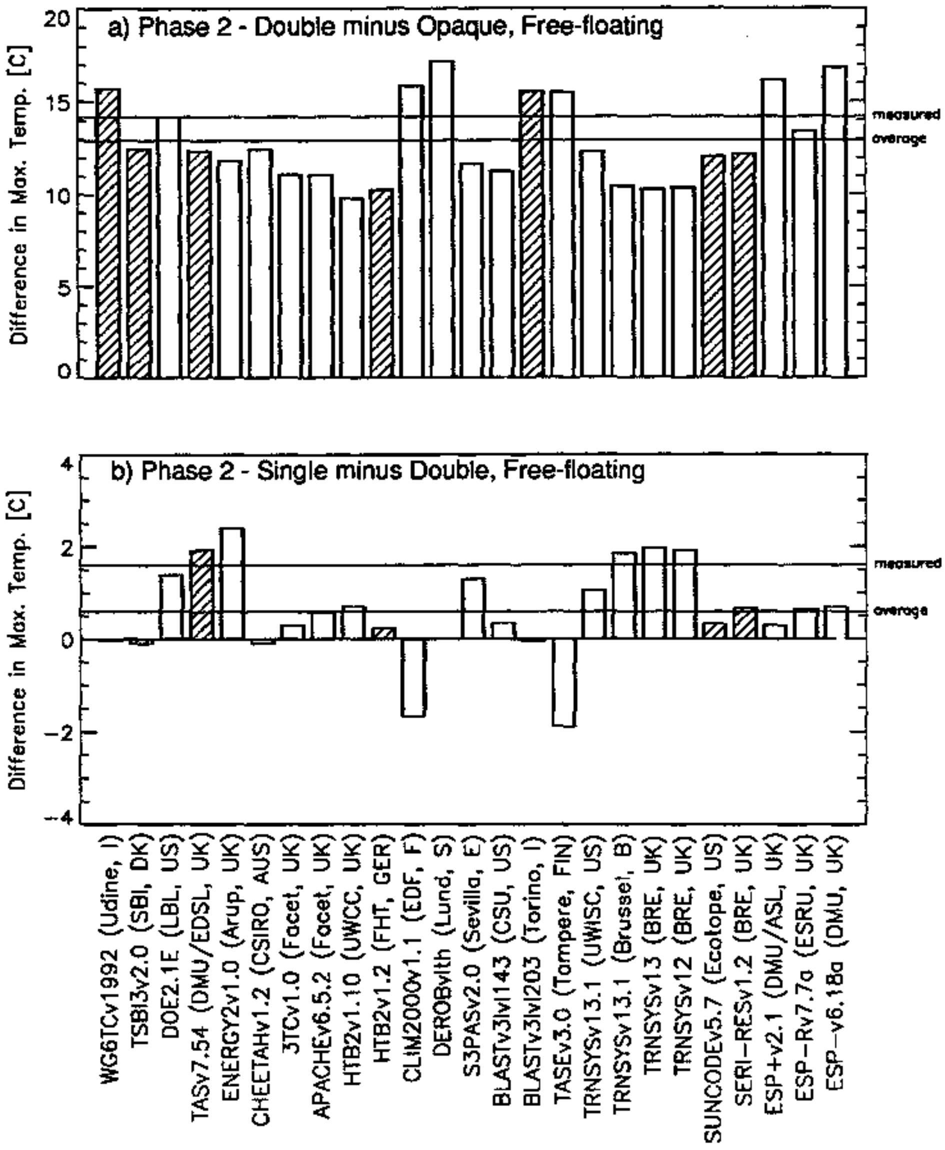


Figure 15:
Phase 2 - Differences in maximum temperatures between different rooms during the free-floating (May) period

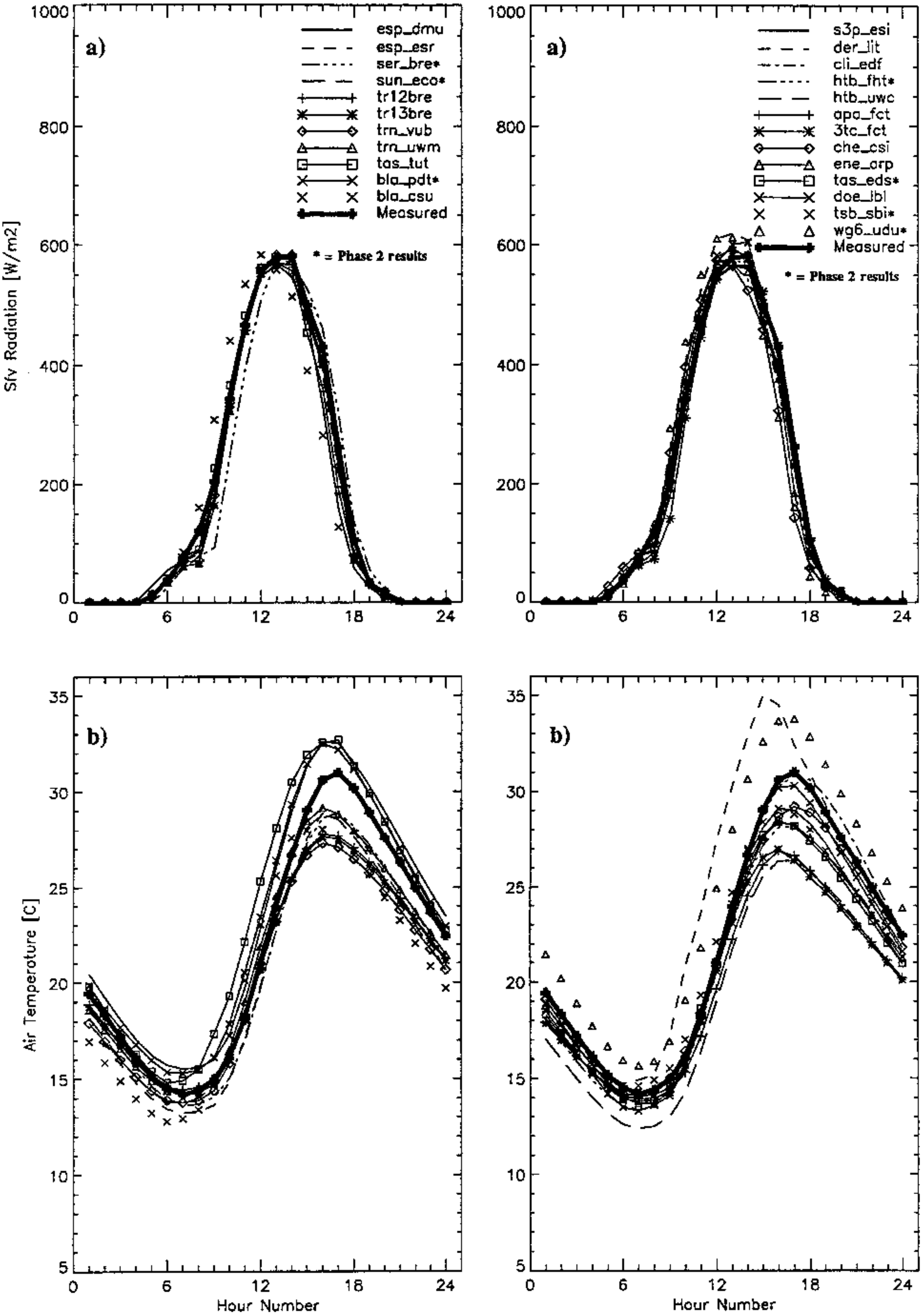


Figure 16:
Phase 2 - Typical hourly predictions for one day in the free-floating (May) period (Day no.147, double glazing)

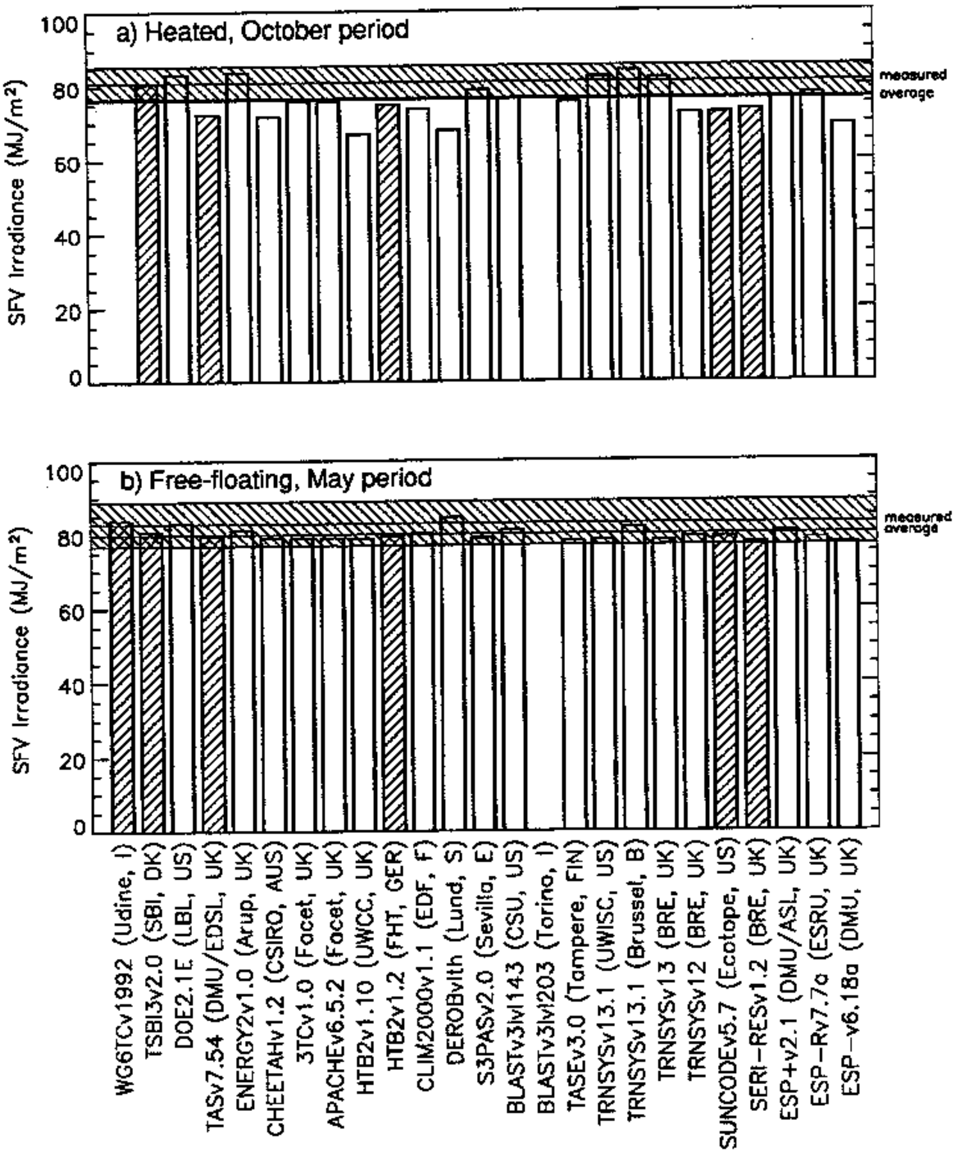


Figure 17:
Phase 2 - Total south facing vertical solar irradiances
for the two 7-day periods

6.3.1. Heated Rooms

(a) The overall variability in the heating energy consumption for the opaque room improved from 47% to 40% compared with Phase 1. For the glazed rooms it remained unchanged at about 50%.

(b) One more program, making 8 in all, predicts energy consumptions within the error bands for the opaque rooms, the number of programs within the error bands for the double glazed room remains unchanged at 11. Now six programs (3TCv1.0, APACHEv6.5.3, CLIM2000v1.1, *SUN-CODEvS.7, TAS7.54* and *TSBI3v2.0)* produce results inside the error bands for both rooms.

(c) DEROBv1th and all versions of ESP and TRNSYS tend to predict energy consumptions which are lower than the measurements and, particularly in the opaque room, noticeably lower than the predictions of the other programs. A possible explanation for the performance of DEROB and TRNSYS is discussed later (section 7.2).

(d) The variability in the predicted energy savings (due to substituting double glazing for an opaque wall) has changed very little. In addition to 3TCv1.0, APACHEv6.5.3, ENERGY2v1.0, TRNSYSv13.1 (Brussel) and TRSNSYSv13.1 (BRE), two more programs, *SUNCODEv5.7* and *TSBI3v2.0* now also produce predicted savings within about 1% of the measured value.

(e) The inter-program variation in the maximum and minimum temperatures is the same as for Phase 1. One program, TSBI3v2.0 now produces maximum and minimum temperatures which are within the estimated error bands for both rooms.

(f) The only noticeable difference between the hourly results produced in Phase 1 (Fig.5) and the Phase 2 plots (Fig.13) is that the phase shift in *SERI RESv12* is now less marked.

(g) The statistical analyses (Appx.C, Table C2) indicate that now, for both the opaque and double glazed rooms, the closest overall agreement between the hourly measured and predicted heating energy values are produced by: 3TCv1.0, APACHEv6.5.3, *BLASTv3.01v1203, SUNCODEv5.7* and *TASv7.54.* In the opaque room, *TSBI3v2.0* also performed well in this regard.

6.3.2. Free-floating rooms

(a) The results for the program *WG6TCv1992* for the opaque room are considerably improved compared to those which were obtained in Phase 1, but still slightly higher. They are now closer to those obtained by the other programs. The improvement arises because a coding error was found in the program and corrected (and not because the input data was changed). The search for the error was prompted by the poor Phase 1 performance in this exercise (the error had not been revealed by previous tests).

The hourly results for *WG6TCv1992* (and DEROBv1th) are still noticeably different from the measurements and the results of the other programs (Fig.16). These differences are reflected in the higher maximum air temperatures predicted by these two programs (Figs.14a&c).

The inter-program variation of the predicted maximum temperatures in the glazed rooms is the same as that for Phase 1: 8.6°C in the double glazed room and 7.9°C in the single glazed room. Ignoring the results for *WG6TCv1992* and DEROBv1th the ranges become: 6.3°C for the double glazed room and 6.2°C for the single glazed room.

The minimum temperatures in the glazed rooms (with or without the *WG6TCv1992* and DEROBv1th results) varied by the same amount as in Phase 1, i.e. 2.5°C in the double glazed room and 2.1°C in the single glazed room.

Two programs, *BLASTv3.01v1203* and ESPv6.18a predicted all but one of the maximum and minimum temperatures to within the error bands. ESP+v2.1 predicted maximum and minimum temperatures within the error bands for the two glazed rooms.

In the opaque room (with or without the *WG6TCv1992* and DEROBvIth results) the maximum temperatures varied by 2.8°C and the minimum temperatures (ignoring *WG6TCv1992)* by 2.4°C.

In Phase 1, 11 programs predicted both maximum and minimum temperatures in the opaque room which were within the estimated error bands. Following Phase 2, this increased to 15 programs.

Only 6 programs, *WG6TCv1992*, *HTB2v1.2*, CLIM2000v1.1, DEROBv1th, ESP+v2.1 and ESP-rv7.7a failed to produce predictions of the maximum temperature within the estimated error bands.

The differences in the maximum temperature between the double glazed and opaque room are shown in Figure 15a with the corresponding plot for the difference between the single glazed room and the double glazed room in Figure 15b.

(h) As for Phase 1, all the programs correctly predicted that the temperatures in the double glazed room would exceed those in the opaque room. The same two programs (ESP-rv7.7a and DOE2v1E) predicted temperature differences within 1°C of the measured difference (i.e. 14.2°C). In Phase 1, 7 programs produced predictions within 2°C of the measured differences, following Phase 2 a further four programs *(WG6TCv1992, TSBI3v2.0, TASv7.54, SERI-RESv1.2* were within 2°C of the measurements.

(i) All the programs except 5 correctly predicted that the temperatures in the single glazed room would exceed those in the double glazed room. The same eleven program / user combinations as in Phase 1, plus the new *TASv7 54* results, were within 1°C of the measured temperature difference (+1.6°C). The predictions of CLIM2000v1.1 and TASEv3.0 differed markedly from the measured value and the predictions of the other programs.

6.3.3. South-facing Solar Irradiances

(a) As for Phase 1, all the programs predicted values which were within the estimated error bands for the May period.

(b) In addition to the 11 results from Phase 1, two more programs produced results within the estimated error bands for the October period *(SER1-RESv1.2* and *SUNCODEv5.7).*

(c) The hourly predictions produced in Phase 2 by *HTB2v1.2, SERI-RESv12 and SUNCODEv5.7* are in much closer agreement with the hourly measured values than was obtained in Phase 1 (Figs.13a, 16c).

7. ISSUES RAISED BY MODEL USER REPORTS

Prior to drawing conclusions about the performance of each program it *is* necessary to examine thoroughly some of the points raised by the model user reports (MURs). These are all contained in Appendix F. The points raised fall into two main categories:

- experimental uncertainty, and
- modelling issues.

These two areas are discussed in detail elsewhere (Appendix D) so it is only necessary to present an overview here.

7.1. Experimental Uncertainty

Some of the model user reports questioned the reliability of the experimental data which were provided. The concerns raised cover.

- overall simplifications in the building specification (F11);
- reliability of room fabric heat loss characteristics (F10);
- uncertainty in relative humidity data;
- reliability of solar radiation data; and
- uncertainty in air-temperatures.

Each of these is addressed in turn in this section.

As part of the effort to reduce the impact of inter-user variability, and to reduce the likelihood of user errors, the rooms were described (Vol. 2) by a series of tables and diagrams rather than by the original constructional/working drawings. This approach meant that some features, deemed to be thermally unimportant, were either omitted or simplified. Wherever specific instances of this were highlighted in the MURs, the issue was considered and where necessary sensitivity analyses were conducted to assess the thermal consequences of the simplifications. It was concluded that no thermally significant data had been omitted (see Appendix D).

As noted earlier (section 5.3.2) decommissioning studies indicated that the rooms were built as described in the Site Handbook (minor simplifications excepted). Nevertheless, the accuracy of the material properties and the estimates of the edge effects, was questioned.

As a further check, the overall heat loss rate (UA-value) of the opaque room was calculated from the results of steady-state heating (co-heating) trials conducted at the same time as the October heating period used for the validation. (A detailed description of this investigation is given in Appendix E). The trials enabled 10 measurements of the overall UA-value to be made. These gave a consistent result of 10.94 W/K±5%, i.e. 10.39 to 11.49 W/K. The value calculated using the data in the Site Handbook and standard CIBSE values for surface coefficients and air-gap resistances was 10.13 W/K.

In the co-heating trials the internal air temperature was constant at around 30°C. The conductivity of the insulation would therefore be greater than the standard 'book' value quoted in the Handbook. After accounting for this, the calculated UA-value was 10.55 W/K with an uncertainty of -7 to +8%, i.e. the value lies between 9.81 and 11.39 W/K. The measured and calculated UA-values are therefore, to within the accuracy of the measurement and experiments, in very good agreement. This adds further support to the proposition that the thermal properties of the rooms were as described in the Site Handbook and certainly could not explain the differences between measured and average predicted values.

Neither the measurements of the UA-value nor its calculation is exact, so small differences between actual thermal characteristics of the rooms and the characteristics deduced from the Site Handbook could remain. However, these differences must be small. In this regard it was noted (by modellers, and in the Site Handbook), that the preferential heat loss through corners and edges could be slightly greater than that which was already accounted for in the Site Handbook description. (This could account for the 4% difference in the calculated and measured UA values). To account for this, a large positive uncertainty was introduced in the calculation of error bands. (The error bands shown in this report take this effect into account). The Site Handbook (Vol. 2) contains the updated edge/comer effect corrections.

The lack of relative humidity data for the heated period will add uncertainty into the predictions of some of the programs, e.g. ESP and Tas, (which use it for calculation of long-wave radiation to the sky-vault), for other programs (most of those used in the IEA exercise) it will have no impact on predictions (the programs do not use it). To account for this omission, a very large uncertainty was introduced for this parameter (for both the heated and unheated periods). The consequential uncertainty in predictions was estimated using ESP and included in all the error bands shown in this report.

The accuracy of the solar radiation data has been questioned. The technical issues surrounding this matter are complex. However, there is no reason to suppose that any of the measurements are in error by more than the uncertainties originally quoted and used in the generation of the error bands.

Finally, it was suggested that the uncertainty attributed to the measured room air temperature and the heater set-point temperature was too small. This suggestion arises because virtually all the programs assume that the room air is at the same uniform temperature, whereas, in practice, it varies with location. The attributed measurement uncertainty was therefore reviewed and in the case of the maximum and minimum temperatures (primary prediction parameters) increased. The error bars plotted in this report include these increased uncertainties.

Overall, none of the investigations into the experimentation, either by the EMC, DMU or the numerous participants in the IEA exercise, have identified any serious omissions. A small number of additional areas of experimental uncertainty, which had not been identified prior to undertaking the exercise, have been exposed, but these have been fully accounted for in the error bands used in this report (and tabulated in the Site Handbook, Vol. 2).

Program	Heated, October Period						Free-floating, May Period						Sfvr		Number of Prim. Params. within Bands[2,3]
	Double glazed			Opaque			Double gl.		Single gl.		Opaque		Oct.	May	
	E	$\hat{T}$	$\check{T}$	E	$\hat{T}$	$\check{T}$	$\hat{T}$	$\check{T}$	$\hat{T}$	$\check{T}$	$\hat{T}$	$\check{T}$			
WG6TCv1992 (Udine, I)	n/a	n/a	n/a	n/a	n/a	n/a	–	○	○	–	–	–	n/a	●	2 (out of 6)
TSBI3v2.0 (DBRI, DK)	○	○	○	○	○	○	-	○	-	-	○	○	○	○	9
DOE2.1E (LBL, US)	–	–	●	–	●	–	●	–	●	–	●	–	●	●	5
TASv7.54 (DMU/EDSL, UK)	○	–	–	○	○	–	–	○	–[1]	–[1]	○	–	–	○	5
ENERGY2v1.0 (Arup, UK)	–	●	–	–	●	–	–	●	–	–	●	●	●	●	5
CHEETAHv1.2 (CSIRO, AUS)	●	–	–	–	●	–	–	●	–	–	●	●	–	●	5
3TCv1.0 (Facet, UK)	●	–	●	●	●	●	–	●	–	–	●	●	–	●	8
APACHEv6.5.3 (Facet, UK)	●	–	●	●	●	●	–	●	–	–	●	●	–	●	8
HTB2v1.10 (UWCC, UK)	–	–	–	●	●	–	–	–	–	–	●	–	–	●	3
HTB2v1.2 (FHT, GER)	○	–	–	–	○	–	–	○	–	–	–	○	–	○	4
CLIM2000v1.1 (EDF, F)	●	–	–	●	●	–	●	●	–	–	–	–	–	●	5
DEROBvlth (Lund, S)	–	–	–	–	●	–	–	●	–	–	–	●	–	●	3
S3PASv2.0 (Sevilla, E)	–	●	●	●	●	–	–	●	–	–	●	●	●	●	7
BLASTv3.0lvl143 (CSU, US)	●	–	–	–	●	–	–	–	–	–	●	●	–	●	4
BLASTv3.0lvl203 (Torino, I)	–	–	○	–	○	○	○	○	○	–	○	○	n/a	n/a	8
TASEv3.0 (Tampere, FIN)	●	●	–	–	●	–	●	●	–	–	●	●	–	●	7
TRNSYSv13.1 (UWISC, US)	–	–	●	–	●	–	–	●	–	–	●	●	●	●	5
TRNSYSv13.1 (Brussel, B)	–	●	●	–	●	–	–	●	–	–	●	●	●	●	6
TRNSYSv13 (BRE, UK)	–	–	●	–	●	–	–	●	–	–	●	●	●	●	5
TRNSYSv12 (BRE, UK)	–	–	–	–	●	–	–	●	–	–	●	●	–	●	4
SUNCODEv5.7 (Ecotope, US)	○	–	–	○	○	–	–	○	–	–	○	○	–	○	6
SERI-RESv1.2 (BRE, UK)	○	○	–	–	○	–	–	○	–	–	○	○	–	○	6
ESP+v2.1 (DMU/ASL, UK)	–	–	●	–	●	●	●	●	●	●	–	–	●	●	7
ESP-Rv7.7a (ESRU, UK)	–	●	●	–	–	–	–	●	–	–	–	–	●	●	3
ESPv6.18a (DMU, UK)	–	–	–	–	●	–	●	●	●	●	●	–	–	●	6

italics and ○ indicate Phase 2 results

[1] Phase 1 result; [2] excludes South facing vertical irradiance, Svfr; [3] Maximum possible score = 12

Table 12: Program Predictions within Error Bands after Phase 2

7.2. **Modelling Issues**

Modelling issues are defined here to cover aspects of the experimentation which could not be modelled accurately by the programs. It may be argued that, because these features cannot be modelled, the experiment represents an unfair test of the program's performance. However, an alternative, and probably more relevant view is that, provided the features are real thermal phenomena which actually occur in the buildings which the programs address in practice, it is legitimate to evaluate performance when these phenomena operate. Should the programs fail to perform well because they ignore the phenomena then this could correctly be interpreted as a weakness of the program which ought to be rectified. The relevance of predictions for a test room to real building analyses is discussed in section 10.

The main issues raised cover:

- dynamics and output of the heater;
- internal surface convective heat transfer, and
- internal air movement and stratification.

The heater characteristics were typical of those used in domestic UK buildings and were fully described in the Site Handbook. The heaters used a mix of radiant and convective output and had associated time delays due to their inherent thermal inertia. Few of the programs could model the dynamics and some could not model a heater with radiant output (Appendix A). Both of these factors must be viewed as limitations of the programs. The inability to model radiant output is the most serious and is likely to lead to under-prediction of heating energy demands. Programs which appear to have 'convective only' heater models are: DOE2v1E, CHEETAHv1.2, DEROBv1th, ENERGY2v1.0 and TRNSYS (Appendix A). The programs tended to underpredict the energy consumption of the rooms. The modelling of heater dynamics is probably less important, at least for the prediction of total heating energy demand (e.g. Appendix F10).

The users of ESP-r v7.7a and ESP+ v2.1 have shown that by using high values for the internal convective heat transfer coefficients (7Wm $^{-2}$K^{-1}) the models produce heating energy results which are much closer to the measurements. However, at least one other program, TASv7.54, uses the same default Alamdari and Hammond equations (19) for internal convection and yet produced results much closer to the measurements than ESP and, in Phase 2, within the error bars. Most other programs use convection values around 3Wm [illegible]
than ESP. All considerations based on the experimental regime, the theoretical basis of the Alamdari and Hammond equations, and internal convective heat transfer measurements at the EMC site, suggest that abnormally high coefficient values are unlikely to be appropriate.

The TASv7.54 user (Appendix F3) originally suggested that the air in the unheated room may consist of a circulating outer loop around a relatively stagnant inner core. More detailed 3-dimensional CH) analyses by DMU did not support this highly idealised view for the heated period (Appendix D), however, the analysis did show more rapid air movement around the walls than in the middle of the room. The temperature measurements themselves corroborated the temperature gradient of 2°C between ceiling and floor predicted by the CFD analysis (Appendix D). No CFD analyses have been conducted by DMU for the unheated periods; however, the characteristics of the room response and the 3 vertical temperature measurements are broadly in line with the circulating outer loop and slower moving core hypothesis proposed in Appendix F3.

The importance of this issue is still a matter of debate. On the one hand stratification is a well known phenomena in real buildings as is the occurrence of rising plumes over emitters and down drafts close to windows. The real question, however, is the extent to which these phenomena may be having a disproportionately large effect in the small test rooms compared to real buildings. At present, this question cannot be answered.

One thing that is apparent is that, although none of the programs consider internal air flow in detail, a number of them produce predictions of individual primary prediction parameters which are close to the measurements and a small number perform well for the whole range of comparisons (section 8). This may be fortuitous, on the other hand, a more obvious explanation is that air circulation details do not need to be considered in order to produce reliable predictions for the test rooms.

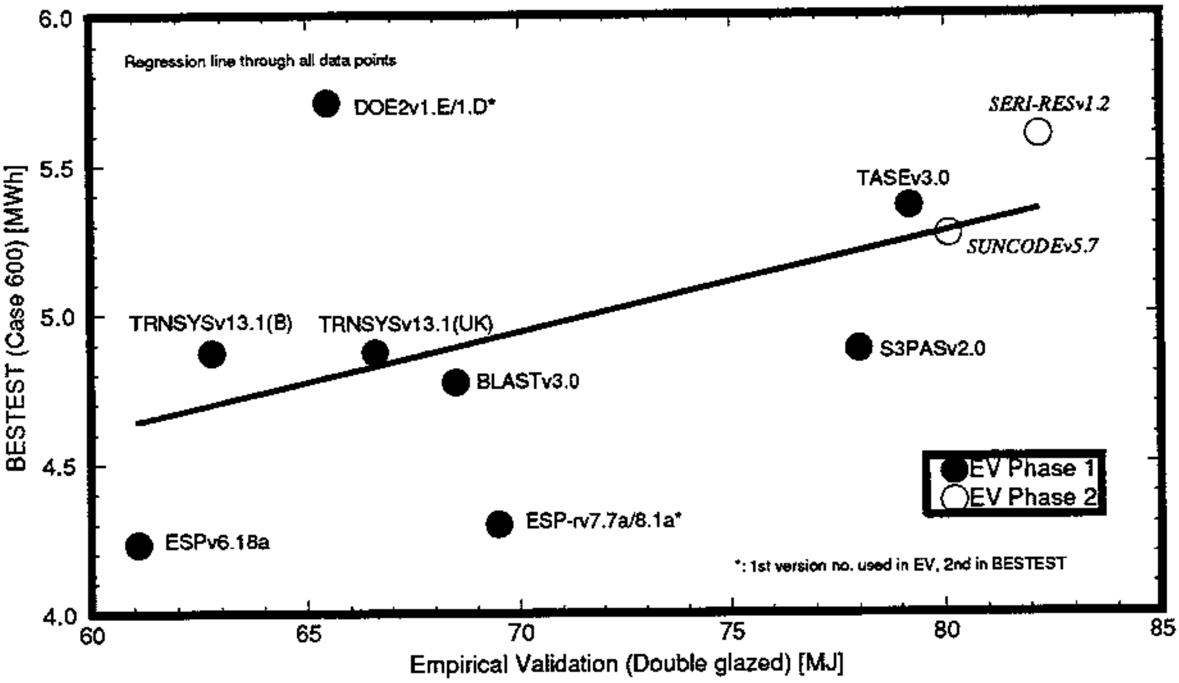


Figure 18: IEA BESTEST and Empirical Validation Heating Energy Correlation

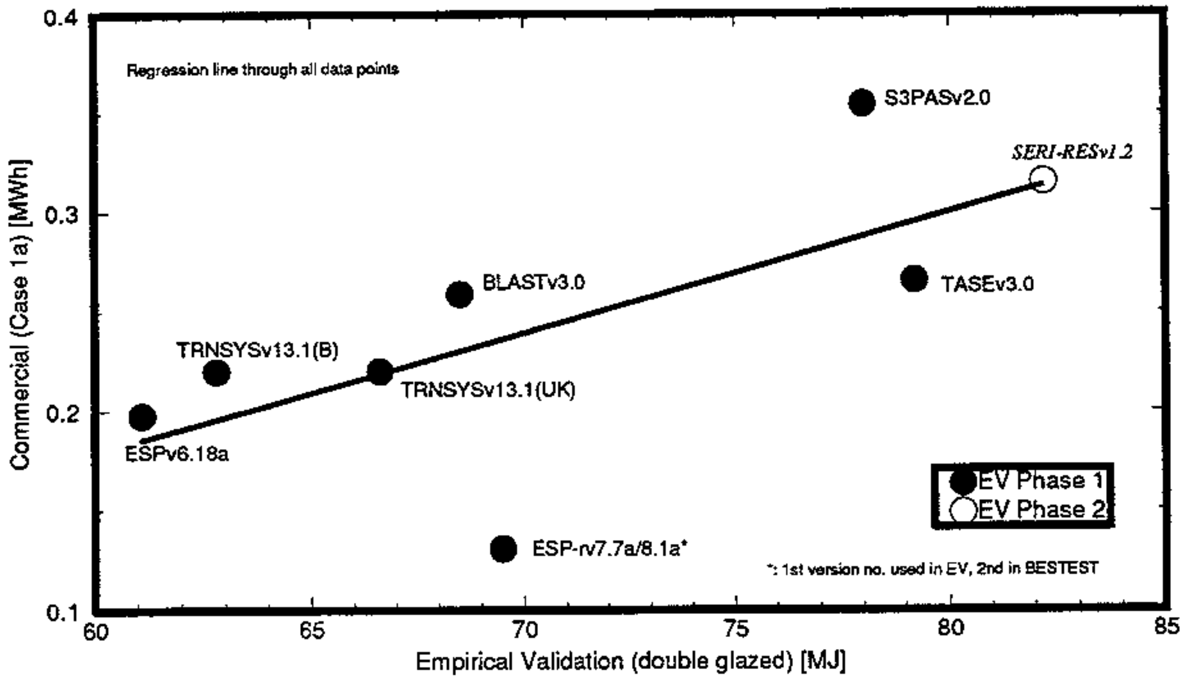


Figure 19: IEA Commercial and Empirical Validation Heating Energy Correlation

Hopefully this is indeed the case, otherwise the thermal analysis of many buildings will become much more complex because detailed air flows must be modelled. In real buildings, fragile convective air-flow patterns can easily be disrupted by furniture, occupants, draughts etc. The relationship between the test rooms and real buildings is discussed further in section 10.

8. PROGRAM PERFORMANCE

Some general observations have already been made about the performance of the programs on various, individual tasks - predicting energy demands, energy savings, maximum and minimum temperatures etc. (section 6.3). In Table 12, an indication is given as to whether the primary parameters predicted by the programs were inside or outside the error bands. It can be seen that none of the programs produced predictions within the estimated error bands for all 12 of the primary parameters. It is equally clear however that some programs perform much better than others.

Even though the predictions of a number of programs fell outside the error bounds for a particular test, the values produced by some of these programs are much closer to the measurements than the predictions of the others. This is illustrated for the primary parameters in Figures 10, 12 and 14, and for the hourly values by the statistics given in Tables C2 and C3 of Appendix C.

8.1. Comparisons with Results from Other IEA Studies

Within IEA BCS Annex 21/SHC Task 12 benchmarks based on inter-model comparisons have been developed based on simple hypothetical buildings. An extensive range of Building Energy Simulation Tests (BESTEST) (13) based on domestic scale buildings and similar tests based on a commercial building (14) were developed. The IEA participants conducted numerous simulations for each of these buildings. The buildings, climatic conditions and operating modes were quite different from the conditions prevailing during the empirical validation experiments. It is, nevertheless, worth comparing the results from the EMC rooms with those obtained in these studies to see if any general trends emerge.

8.1.1. BESTEST Results

Eight of the programs which participated in the empirical validation exercise had also participated in the parallel IEA inter-model comparison study to develop Building Energy Simulation Tests, BESTEST. The BESTEST work used weather for Denver for a whole year, and the buildings were heated by a warm air system and, in most cases, natural infiltration at 0.5ach^{-1} occurred (13).

One of the BESTEST buildings (Case 600) had only a modest amount of thermal mass and South facing double-glazing, similar to the EMC room 1. The comparison of the heating energy demands predicted for the two situations (Fig. 18) reveals some interesting similarities.

(i) The overall variability in the predictions (from the highest to the lowest predictor) is about 30% for both buildings.

(ii) There is a broad measure of agreement in the rankings of 7 of the 9 programs (shown by regression line). The two programs not following the trend are DOE2v1E/D and ESP-rv7.7a/8.1a. The discrepancy for DOE2v1E/D is consistent with the difference between heater type assumed in BESTEST and that installed in the EMC rooms (see Appendix F2).

(iii) In both cases ESP v6.18a predicted results which were lower than those of the other programs, this phenomenon was in fact observed in nearly all the BESTEST cases. (In the final BESTEST report the ESPv6.18a results were not used. Instead results from the later release ESP-rv8.1a were adopted. For BESTEST case 600 ESP-rv8.1a predicted similar results to ESPv6.18a).

(iv) In both tests, TASEv3.0 and SERI-RESv1.2 tended to predict higher values than most of the other programs. This phenomenon also occurred for many of the other BESTEST cases.

(v) The other ESP results do not clearly follow the overall trend, possibly because different versions were used. However, ESP-rv8.1a did not predict much lower values than the other programs in the BESTEST work.

These observations lend further support to the proposition that some of the divergence amongst the results obtained in the EMC test rooms is due to inherent differences in the programs and not due to important 'external errors' either in the experimental design or introduced by the program users.

8.1.2. Commercial Building Results

Six of the programs which participated in the empirical validation exercise were also used to produce results for the hypothetical commercial building. This was a 3 zone building, consisting of 'offices' facing North and South with a corridor between. This module was assumed to be surrounded on all sides by similar modules. The rooms had modest thermal mass, double glazing and were heated by a warm air system in winter and cooled by ventilation in summer. The systems were ideal, i.e. adequately sized with instantaneous response. The building was simulated for a whole year using Denver weather data (14).

(a) The predicted annual heating energy demands for the South facing room show a similar ranking to that obtained in the empirical validation exercise (shown by the regression line).

(b) The heating energy demands predicted by ESPv6.18a for the South facing room were, as in the empirical comparisons, much lower than the results from the other programs. This result adds further support to the proposition that ESPv6.18a predicts heating energy demands which are lower than those predicted by other programs.

(c) As in the BESTEST work, ESP-rv8.1a produced a lower heating energy consumption prediction than all the other programs.

(d) As in the BESTEST work, SERI-RES v1.2 predicted higher results than most of the other programs.

Again, these results are broadly consistent with those obtained in the empirical validation exercise. It seems very unlikely therefore, that the poor performance of ESPv6.18a and ESP-rv8.1a can be attributed to particular features of the empirical validation experimentation.

9. INTERPRETING THE RESULTS

In principle, the interpretation of the empirical validation results is quite simple. Once predictions fall outside the error bands then the probability that the result could be due to uncertainties in the experimentation (an external error in the experiments) is so small (in this case about 1%) that it is reasonable to assume that the program predictions are in fact incorrect. For this interpretation to stand it is essential that all sources of experimental uncertainty have been identified and correctly accounted for. In practice, it is never possible to be absolutely certain that this is the case, however, provided that all the significant error sources have been accounted for, any minor effects will be unimportant. In the EA work we can be confident that all significant experimental error sources have been accounted for.

(i) The experiments, simulations and analyses were undertaken by groups with considerable expertise in this area.

(ii) The experiments were the subject of close scrutiny by by peer review committees when they were originally conducted. They were also the subject of careful examination by the Annex 21/Task 12 participants.

(iii) The experimentation has been the focus of attention by all (25) IEA participants, most of whom are themselves knowledgeable in this area.

(iv) All possible quality assurance checks have been undertaken and documented and were open to scrutiny; none revealed any shortcomings which were not ultimately accounted for in the experimental data analyses.

In short, the IEA experiments stood up to critical and widespread reviewing by experts in thermal modelling and validation.

Poor predictions can either be due to inherent weaknesses in the program (internal errors) or alternatively mistakes by the program user, such as inaccurate modelling or input data blunders (i.e. external errors introduced by users). These error sources are very hard to eliminate, because the data input requirements of DSPs are onerous and diverse and because, for many programs, the data input interface is cumbersome. (Considerable work is under way around the world to rectify this situation). In the EA exercise however, every effort was made to avoid this form of error.

(i) Experts, vendors or authors were asked to produce predictions whenever possible.

(ii) The importance of accurate modelling and quality assurance checking was emphasised to all participants.

(iii) A second Phase enabled users to rectify any mistakes or to improve their modelling procedures.

In short, every effort was made to eliminate this source of external error. This effort, together with the broad agreement with the IEA inter-program comparison work (Section 8), would indicate that all significant influences which the program user may have, have, as far as possible, been eliminated, and that the predictions obtained are a true reflection of the program performance itself. If such user errors remain, then they are also likely to exist when other, less skilled, operators employ the program, under the time constraints of a real design situation. It may be that the manuals, help facilities, user interface structure (which is part of the program) is underdeveloped. Here the division between what is an internal source of error, and what is an external source of error, becomes blurred.

The foregoing arguments indicate that there is only a small possibility that some important external errors remain. One should therefore be suspicious of any predictions which lie outside the error bands. In this regard, it should be noted that no program predicted inside the error bands on all occasions. Those programs for which predictions consistently fall well outside the error bands and, as such, perform less well than other programs, can reasonably be taken as being in error. Programs in this category can be identified from section 8.

There are a number of possible reasons (internal errors) which may cause programs to perform badly.

(i) They omit sub-models or algorithms for thermal processes which are important.

(ii) They contain inappropriate algorithmic assumptions.

(iii) They contain coding errors.

Using information from this IEA exercise alone it is difficult to identify which of these sources of error are causing problems for a particular program. However, it has been noted (Appendix F8 and F2) that TRNSYS and DOE cannot model heating systems with a radiant or part radiant source, this is an internal error of type (i) above. Other programs, ENERGY2v1.0, CHEETAHv1.2, and DEROBv1th also appear to be weak in this regard (Appendix A). This fact may limit the range of applicability of these programs.

Assumptions are made by all the programs and in fact they may be difficult or impossible to avoid. Common assumptions made by all the DSPs are that: conduction is one-dimensional and room air is perfectly mixed (i.e. at a uniform temperature). Many assume no heater/control dynamics and that surface heat transfer coefficients are constant. Other assumptions are program specific, for example, a few need data on night-time cloud-cover for calculating night-time radiation loss to the night sky. In general this is unavailable so assumptions (usually that the daytime cloud-cover persists during the subsequent night) have to be made. Thus, assumptions are made in all programs, and not necessarily because the authors wish to over-simplify, rather that other circumstances (lack of data, need for reasonable solution times, etc) may enforce it. Assumptions which have been explored within the empirical validation exercise cover, for example, diffuse sky radiation distribution (Appx.Fl, F11), external longwave radiation modelling (Appx.F3), the use of simplified convection algorithms (Appx.F6, F10, Fl1) and heater characteristics (Appx.Fl, F2, F4, F6, F8).

One benefit of empirical validation is that it tests the combined effect of all the algorithmic assumptions made in a program. Inter-model comparisons and analytic tests do not usually test the appropriateness of algorithmic assumptions. Thus they do not 'pick up' internal errors of type (ii) above. It is for this reason that programs which have previously done well in such tests may nevertheless perform badly in a rigorous empirical validation exercise. The only way problems of this type can be identified is by focused, algorithm level, comparisons between program predictions and measurements (see section 12).

Coding errors are the most obvious form of internal error. It is generally accepted that such errors must exist in large computer programs, it is an inevitable consequence of the complexities of the programming task. It is when these errors cease to be benign and begin to have a noticeable impact on predictions that they are of concern.

Coding errors are rather difficult to isolate by empirical validation alone because they are compounded by internal errors of type (i) and (ii). However, errors of this type were revealed in DOE2v1D, ESPv6.18a, an earlier version of TASE and TRNSYSv12.1 by the complementary IEA BESTEST work. Some programs performed reasonably well in the empirical validation exercise whilst others, which used many of the same algorithmic assumptions, performed rather poorly. This suggests the existence of critical, but as yet unidentified, coding errors in some programs. In particular, the persistent under-prediction of heating energy demands by ESP is a source of concern.

It is interesting to consider the interpretation which should be made when programs produce good results. Some possibilities are:

(i) the program contains no internal errors;

(ii) any internal errors which do exist are benign, for the particular conditions being tested; and

(iii) internal errors exist but they are counteracted by other, compensating internal errors.

In practical terms, for the situation tested, it does not matter which of the above possibilities is true. Provided the experiment reflects the situation in which the program will actually be used, its credibility will have been, quite rightly, increased for this situation. The problem is defining the 'situations' to which the experimental validation is applicable (see section 10). When used beyond this realm of applicability, any internal errors may no longer be benign or counteracted by other compensatory errors. It is for this reason that validation has to be seen as a continuing process, in which programs are tested under wide ranging, different, but relevant, conditions (see section 12).

The IEA empirical validation exercise has however produced a snap-shot of the capabilities of many programs, for a particular set of circumstances. It has identified some programs which are highly likely to contain significant internal errors but it has provided limited insight into the type or source of the error. Other tests, such as the IEA BESTEST (13) or algorithm level empirical validation tests are needed to do this. The exercise has also identified programs which performed better than the rest and, as a result, the credibility of these programs has, quite rightly, been enhanced.

10. RELEVANCE TO REAL BUILDINGS

The implications of the results of an exercise such as this, for the design of real buildings, can easily be misinterpreted. The key issues concern:

- the programs and the way they are used; and
- the relationship between the test rooms and 'real' buildings; and
- the importance of other issues besides program accuracy.

Throughout this work care has been taken to identify correctly the versions of the programs being used, and via the program questionnaire (Appendix A) to obtain an idea of the way it was being used (some programs provide alternative ways of modelling the same physical process). Although some programs produced roughly the same performance despite different versions being used (e.g. ESP, TRNSYS), this was not the case for BLAST (in Phase 2). Therefore it may be dangerous to extrapolate the results of this exercise to other program versions and modelling options. The empirical validation package (Volume 2) will help users to evaluate different versions of programs. Ultimately it would be beneficial if DSP users had a battery of tests to hand so that new programs/new versions of an existing program, or new ways of using a program, could be rapidly evaluated prior to proceeding with real design work.

The test rooms themselves were chosen to be a good compromise between the needs for realism and experimental accuracy (control of external experimental errors). They deliberately stress fabric heat transfer issues whilst avoiding the vaguenesses of internal gains and infiltration. Because fabric heat gain and loss is an essential feature of all buildings, the programs which performed well in the IEA study are likely to predict this aspect well for real buildings. Thus, confidence in their predictive abilities for real buildings should, quite rightly, be increased.

Conversely, programs which performed badly should be treated rather more cautiously. It may be that internal gains, plant operation and ventilation dominate the performance in a particular building (e.g. in the case of commercial buildings) in which case fabric conduction inaccuracies will be less important. Alternatively fabric issues may dominate in the real building, for example in passive solar house design, in which case programs which performed badly in the IEA exercise should be used with caution.

In practice, issues other than accuracy may also influence the choice of the program to be used. Usually, practitioners use only the program(s) they have available. When this is not the case there is a variety of factors which may influence the choice. Some examples are:

- the modelling capabilities of the program (e.g. can it model HVAC, complex geometries, complex occupancy schedules);
- the ease of data input (e.g. graphical input, menu options);
- the availability of climate or material databases;
- the clarity of data output (graphs, charts, etc.); and
- the speed with which a solution is obtained.

It is also the case that using a DSP can help to generate a better intuitive feel for the thermal behaviour of a building.

It is self-evident that if the predictions are wrong, these other features are of limited worth. Ultimately, it is the accuracy of the result which determines the final building design and its thermal performance. It is this aspect for which thermal modellers are responsible and for which they may be legally liable.

11. REFLECTIONS ON EMPIRICAL VALIDATION

This IEA exercise was the largest program validation exercise ever undertaken. It involved many of the leading modellers from around the world and so it is worth reflecting on the exercise to try and extract pointers to the way future empirical validation should be undertaken.

(i) It is important to realise that empirical validation is extremely hard to do well, much harder than other forms of validation (such as inter- model comparisons or analytic tests). For example, even ignoring the experimental design and data collection, the data analysis and general management of the exercise took around two years. Additionally, each of the participants will have devoted between about 7 and 21 man-days in order to run the programs.

(ii) The results of empirical validation exercises are likely to attract considerable attention. Particularly when, as in the IEA exercise, the data collection and analysis is conducted by a neutral third party with the program authors/vendors/experts producing blind predictions. The results are of great interest because (a) they address the 'bottom line' issue - how well can the programs predict what will actually happen? and (b) well designed and managed validation exercises are rare (because of their cost and technical complexity).

(iii) The above points mean that empirical validation should only be attempted after effort has been expended on the planning stages: the choice of experimental design; the methods for managing the flow of information between the participants; and the approach to analysing the results.

(iv) Good experimental design demands experience in the field. It has to be linked to sound error analysis. Experimental work is hard to do well. When predictions and measurements diverge, it is usually the experimentation which becomes the focus of attention, and experimental errors or mistakes are sought. This is particularly so when divergent results are obtained for programs which are thought to be reliable, even when other programs are demonstrating good agreement with a dataset.

(v) It is very difficult to ensure that all possible sources of experimental error have been identified and correctly accounted for. The widest possible consultation with experts in the field is valuable to assist in critically appraising the experimentation.

(vi) The IEA participants were unanimous in recommending that well designed empirical validation work should be undertaken in two phases. The first phase being conducted blind, i.e. without knowledge of the actual measurements, and the second phase being open, i.e. with all available measurements being given to the modellers.
The first phase gives 'added value'. It can provide a clear, clean snap-shot of program capabilities when used in a similar way to that which is adopted for real design problems. It can identify programs which are in error and the possible areas of weakness. The ambiguities which can be introduced when modellers are able to fit predictions to measurements are avoided. With such a scheme it is important that all modellers are given similar and adequate time in which to make their predictions and do their analyses.

(vii) If a blind validation exercise is attempted it is recommended that a feedback mechanism is set up to assist in resolving problems. The use of a hotline (which must provide a rapid response) and newssheets is one approach. All the IEA participants who chose to comment, felt this was a good approach.

(viii) Only one participant felt that the method of describing the experimental conditions was unsatisfactory. Others felt that it matched the needs of the programs very well. All the IEA participants agreed that it is a good exemplar of how to document a building which is to be modelled by thermal programs. (The IEA BCS Annex 21/SHC Task 12 site handbook is provided in Volume 2 of this final report).

(ix) There was general agreement with the practice of estimating the total uncertainty in the experimentation and the measurements using either Differential Sensitivity Analysis (DSA) or Monte Carlo Analysis (MCA). In practice, the MCA approach has tended to give slightly wider, and hence more conservative - which is desirable, error bars. It is important to ensure the error bars account for all possible sources of uncertainty. This can be complicated when only one program is used to estimate the error bars against which many programs will be assessed. The uncertainty calculated using the particular program should, as in this exercise, be supplemented with uncertainties due to factors which the chosen program does not address.

(x) The total uncertainty in the IEA experiments is indicative of that which will be obtained in carefully designed experiments in passive moms. The total uncertainty for 10 day heating energy was about 16%. This was much less than the range 40% between the program predictions. Well conducted empirical validation experiments can therefore provide a powerful test of thermal programs and can identify programs which are in error.

12. FUTURE EMPIRICAL VALIDATION NEEDS

In broad terms data for empirical validation can be divided into two categories:

(i) Data for whole model evaluation, sometimes called system level data (3). The IEA dataset is a good example of this.

(ii) Data for evaluating individual algorithms, sometimes called mechanism-level data (3). The South facing vertical solar irradiance data in the LEA data set is an example of this.

Whole model data is particularly useful for testing programs as part of a program accreditation process and mechanism-level data has an important role as part of a program development process.

It is possible (as in the IEA work) to collect data of both types in a single experiment. However, it may well be the case that algorithms can be more effectively tested using specialised experiments which do not necessarily reflect realistic building geometries and operating conditions. For example, shortwave radiation algorithms can be tested using just a vertically mounted pyranometer. Algorithms which are currently felt to be in need of improvement by further experimental work include those covering:

(i) internal convection and air circulation;

(ii) external convection; and

(iii) glazing systems.

Work on internal air circulation and glazing algorithms is currently under way in the UK.

Concerning whole model evaluation, the IEA work has, to a large extent, covered single-zone, direct-gain buildings in a heated or unheated mode. Work elsewhere, such as in the EC PASSYS project (7), has also generated data of this type. Further insight may be gained from using data from a room heated by a purely convective source. The air could also be gently stirred so that conditions more closely match the perfectly mixed assumption adopted by most programs. This may, however, introduce other problems, such as atypical surface convection.

Data sets collected from real domestic scale buildings and office modules are likely to be of wider interest. In such buildings it is much harder to eliminate or control sources of experimental errors/uncertainty, and many previous attempts at such work have, for this reason, proved unsatisfactory. Using the lessons learned in test room studies (such as this IEA work), it may be possible to refine the design of such experiments to avoid the types of costly failures which have been observed in the past. Full-scale domestic buildings would stress the algorithms describing: fabric; solar gain; thermal storage; ventilation; and, to a modest extent, internal casual gain algorithms. A non-domestic building would place greater emphasis on internal gain and plant modelling capabilities.

13. CONCLUSIONS

13.1. Empirical Validation Procedures

(a) The empirical validation work in IEA BCS Annex 21/SHC Task 12 was the largest exercise of its type ever undertaken. It included virtually all the major detailed thermal simulation programs (DSPs) currently used within the member countries, and some from non-member countries. All were used by either the program developers, vendors or an expert user.

(b) High quality measured data sets were identified and a tightly documented specification for five empirical validation benchmark tests was produced. This was tested by applying it to 25 different combinations of DSP and user. The benchmarks have been refined and documented for use by others. The document is an exemplar of how to describe empirical validation benchmarks.

(c) It is recommended that a well designed empirical validation exercise should be undertaken in two Phases. The first phase being conducted blind, i.e. without knowledge of the actual (measured) building performance, and the second phase being open, i.e. with all the available measurements being given to the modellers.

(d) The inclusion of a blind-phase gives added value to an empirical validation exercise. It can provide a clear, clean snap-shot of DSP capabilities and under conditions which mimic those which exist when programs are used in a real building design context. A blind phase may also indicate aspects of programs which are weak.

(e) In Phase 1 of the IEA exercise, the program users had available to them any information which they felt was necessary to avoid misinterpretation of the specification and errors in the program input data. This was achieved by the provision of a telephone and fax "hotline". A mechanism for providing such information is recommended if a blind validation exercise is attempted.

(f) In any centrally managed, multi-participant, validation exercise, all participants should be provided with the same information. In the IEA exercise, this was achieved by circulating a Newssheet. As far as possible, all participants should be placed under the same time constraints. This avoids any bias. These time constraints should not, however, preclude accurate use of the DSP.

(g) The methodology of firstly establishing base-case results and then accounting for the total experimental uncertainty was accepted by all participants. In the IEA exercise Monte-Carlo analysis was used in conjunction with SERI-RES to estimate this total uncertainty. This was supplemented by Differential Sensitivity Studies to ensure that all known sources of experimental uncertainty were accounted for. It is time consuming and difficult to identify, and account for, all possible sources of uncertainty which may exist in an empirical validation experiment.

(h) Any empirical validation exercise should ultimately make the measured data available to the program users so that they can explore, and hopefully understand, the reasons for any divergence between the predictions of their program and the measurements. This was Phase 2 of the IEA work.

(i) In Phase 2, many participants undertook sensitivity analyses to determine the consequences of alternative algorithms and modelling techniques. Issues of particular significance in this work, and which are likely to be important in similar exercises, included: external longwave exchange modelling, internal surface convection and air movement, heater dynamics, and edge and corner heat losses.

(j) Complex statistical techniques, such as time-series or cross-correlation analysis, were not employed in this exercise, partly because the duration of the exercise precluded this and partly because sufficient insight was gained without them.

13.2. The IEA Empirical Validation Package

(a) The test rooms used in the IEA study were selected after a thorough review of the available sources of validation data. The rooms were a good thermal compromise between the needs of realism and experimental accuracy. The benchmarks produced can be used to assess the reliability of programs for predicting the heat transfer through the building fabric. This is an essential aspect of the thermal analysis of virtually all buildings.

(b) The collection of data for empirical validation is expensive and time-consuming. Quality assurance checks should be undertaken to ensure that the data is reliable. The rooms used in the IEA exercise were the subject of close scrutiny by those managing the exercise and by the participants. The quality assurance checks included: decommissioning one room, measurement of a room heat loss coefficient, and periodic room matching trials. All the quality assurance checks indicated that the rooms were built exactly as described in the site handbook.

(c) It is difficult to identify all the sources of experimental uncertainty, and to estimate the magnitude of these uncertainties. In Phase 2 of the MA exercise the uncertainty attributed to edge and corner effects, the relative humidity data, and the air temperature variations, were questioned. The initial estimates were, therefore, revised and used to calculate the total (experimental) uncertainty. It is reasonable to assume that all significant sources of experimental error have been accounted for. The uncertainties are included in the IEA Empirical Validation Package.

13.3. Program Performance

(a) The IEA empirical validation exercise has provided a snap-shot of the capabilities of the current state-of-the-art in thermal modelling of buildings. It has provided an indication of the variability in program predictions, indicated those programs which perform reasonably well and those which do not, and succeeded in identifying weaknesses in some DSPs.

(b) In both Phase 1 and Phase 2 the programs predicted seven-day energy consumptions for two simple single-zone buildings which varied by around 50% of the mean value. The predicted energy savings, due to replacing an opaque south facing surface with a double glazed window, varied, in Phase 2, from 13% to 40%. The predicted peak air temperatures in these rooms varied by 11°, which was between 3°C and 14°C above the thermostat set-point.

(c) In the unheated glazed rooms the predicted peak air temperatures varied by 7.9°C to 8.6°C (depending on the glazing type). All the programs correctly predicted that the peak temperatures in the double-glazed room would exceed those in the opaque room. However, the predicted increase in peak temperature varied from about 9.5°C to 17°C. These variations were the same for both Phase 1 and Phase 2.

(d) The variability and rank ordering of the program predictions for the heated rooms was similar to that observed in the complementary lEA inter-model comparison exercises. This, together with other considerations, indicates that the predictions obtained are a true reflection of the performance of the program, i.e. when the influence of the program user has, as far as possible, been eliminated.

(e) Comparisons between the Phase 2 predictions and the actual measurements indicate that: six programs predicted heating energy consumptions to within the estimated total uncertainty band for both rooms; 1 program predicted both maximum and minimum temperatures in the two heated rooms to within the total uncertainty bands; but no programs predicted the maximum and minimum temperatures in all three unheated rooms to within the uncertainty bands.

(0 The heat output from the test room heaters had a large radiant component. The tendency for some programs to underpredict heating energy demands is consistent with the assumption, which is inherent in some programs, that the heater output is purely convective.

(g) Some programs produced predictions which fell outside the total uncertainty bands on a number of occasions. It is reasonable to assume that these programs contain 'internal errors'. Since the rooms stress the algorithms concerned with modelling fabric heat flows, and since these flows are present in virtually all buildings, caution should be exercised when using such programs to assess the performance of real buildings.

(h) These results are specific to the particular building and climate conditions tested, and so they should not be taken as indicative of the programs' performance in general. However, the test rooms do stress the algorithms describing fabric heat flow, which is a feature of virtually all buildings. The consequences of any errors in the predictions need to be assessed by careful consideration of the purpose for which the performance assessment is being undertaken. Ideally, a more rational approach to risk assessment should be developed.

13.4. Future Work

(a) It is apparent that, whilst some programs perform reasonably well, others are not as good. Developers and vendors should make full use of the validation benchmarks which already exist to ensure that their programs are as reliable as possible, and that the limitations of their program are fully understood and made clear to potential users.

(b) This exercise has shown that it is possible for thermal programs to be evaluated in a reliable and unbiased manner, whilst involving the program authors and vendors. It indicates that an independent program accreditation system could be developed. Such a system would be valuable to enhance the credibility of thermal modelling in general and to give users confidence in the abilities of specific programs.

(c) It may be useful to complement this study with one which is characterised by convective heating and well-mixed zone air. The issues of internal convection and air circulation, particularly over poorly insulated (glazed) surfaces, may also be fruitful areas for future research.

(d) The LEA exercises, and other recent work, indicates that there is a general consensus about how empirical validation should be undertaken and about how to collect reliable data. A few groups exist which are capable of gathering such data. It may, therefore, be plausible to obtain empirical validation benchmarks from whole, multi-zone buildings. This would be a difficult but valuable task and enable the accuracy of whole model predictions for such buildings to be tested for the first time.

REFERENCES

1. Bloomfield, D.P., Lomas, KJ. and Martin, CJ., 1992, Assessing programs which predict the thermal performance of buildings: BRE Information Paper, IP7/92, 3-pp.
2. Lomas, KJ. and Bowman, N.T., 1987, Developing and testing tools for empirical validation, Ch. 14, Vol. IV of SERC/BRE final report on An investigation into analytical and empirical validation techniques for dynamic thermal models of buildings, 79pp, BRE Garston.
3. Judkoff, It et al, 1983, A methodology for validating building energy analysis simulations, Solar Energy Res. Inst. Draft Report, TR-254-1508, 204pp.
4. Lomas, KJ., 1991, Dynamic thermal simulation models of buildings: New method of empirical validation, BSER&T 12 (1), 25-37.
5. Bloomfield, D.P., 1988, An investigation into analytical and empirical validation techniques for dynamic thermal models of buildings, Vol.1, Executive Summary, SERC/BRE final report, BRE Garston, 25pp.
6. EMC, 1991, Detailed model comparisons: An empirical validation exercise using SERI-RES, Contractor Interim Report from Energy Monitoring Company to UK Department of Energy, via ETSU, ETSU 1197-I, 141pp.
7. PASSYS, 1989, Project Phase 1, Subgroup model validation and development final report 1986-1989, CEC Report 033-89-PASSYS-MVD-FP-017.
8. Lomas, K.J., Martin, C.J., Eppel, H., Watson, M. and Bloomfield, D., 1994, IEA BCS Annex 21C/SHC Task 12B, Empirical Validation of Thermal Building Simulation Programs using Test Room Data, Volume 2: Empirical Validation Package, IEA21RR7/94
9. Lomas, KJ (Ed.), 1994, IEA BCS Annex 21C/SHC Task 12B, Empirical Validation of Thermal Building Simulation Programs using Test Room Data, Volume 3: Working Reports, IEA21RN401/94
10. Lomas, KJ., 1990, Data sets for validating dynamic thermal models of buildings, BEPAC Tech. Note 4/90, 82pp, BRE Garston.
11. Mørck, O.C., 1986, Simulation Model Validation using Test Cell Data, IEA Solar heating and cooling programme, Task VIII passive and hybrid solar low energy buildings, Thermal Insulation Laboratory, Tech. Univ. of Denmark, Report No. 176, 158pp.
12. Bland, B.H., 1992, Conduction in dynamic thermal models: analytical tests for validation, BSER&T (13) 4, 197-208.
13. Judkoff, R. and Neymark, J., 1994, IEA SHC 12B/BCS 21C, Building Energy Simulation Test (BESTEST) and Diagnostic Method, US National Renewable Energy Lab, Report NREL/TP-472-6231
14. Haapala, T., Kalema, T. and Kataya, S., 1993, Energy analysis tests for commercial buildings (commercial benchmarks), Tampere University, Finland, draft IEA SHC 12B/BCS 21C report
15. Fauconnier, R., Martinon, M.H. and Dalicieux, P., 1989, La simulation numérique du component thermique des bâtiment; 8 - validation experimentale de la prise en compte des echanges d'humidite dans le mobilier et les parois en tenant compte de la stratification, Fédération Nationale de Bâtiment (Direction de la Recherche) & Electricité de France, Direction des Etudes et Recherches, 48pp.
16. PV-Wave, 1992, Visual Data Analysis Software, Precision Visuals Ltd.
17. Lomas, KJ. and Eppel, H., 1992, Sensitivity analysis techniques for building thermal simulation programs, Energy and Building (19) 1, 21-44.
18. Kreyszig, E., 1988, Advanced Engineering Mathematics, John Wiley & Sons, New York, 1988.
19. Alamdari, F. and Hammond, G.P., 1983, Improved data correlations for buoyancy-driven convection in rooms, BSER&T (4) 3.

Appendix A - Features of Participating Programs

Al Survey of Program Features

During Phase 1 of the exercise participants were asked to complete a brief survey form to describe the features of the programs they were using.

Subsequently, a second and more extensive survey form was submitted (see Newssheet No.8, included in Volume 3: Working Reports), to try and capture much more detail. This survey was also conducted for the programs participating in the development of the IEA 21B/12C Building Energy Simulating Tests (BESTEST).

In addition to providing a good description of the participating programs, the information could also make it possible to identify any relationships between the key features of the programs and their predictive capabilities.

A2 **List of Features**

All the participants completed the brief Phase 1 form, 16 of the 17 distinctively different pro grams were described in the survey forms completed during Phase 2 (Table Al).

It has not been possible to carefully check the absolute accuracy of these entries, although the IEA participants have been encouraged to check the entries for their own programs.

A3 Notes to Table Al

1. In the table, the filled symbol indicates that the option used in the IEA Empirical Validation exercise. Additional open symbols indicate other options which are available within the program but which were not used.
2. Where the term 'fixed' is used to describe an option it implies that a quantity is fixed within the program code. 'U-spec' denotes that the value is fixed during the simulation but specified by the user. 'Calc' means that the program calculates the appropriate value, once or periodically throughout the simulation.

		TSBI3 A22	DOE-2.1E	TAS v7.54	ENERGY2 v1.0	CHEETAH v1.20	3TC v1.0	APACHE v6.6.2	HTB2 v1.2	CLIM2000 v1.1	DEROB-LTH v9301	S3PAS v2.0	BLAST v3.0 lvl193	TASE v3.0	TRNSYS v13.1	SERI-RES v1.2	ESP-r v7.7a
Program status	Public domain		●						●							●	
	Commercial	●		●	●	●		●					●		●		
	Research						●			●	●	●		●			●
Solution method	Explicit fin diff				●				●							●	●
	Implicit fin diff	●						●			●	○					●
	Weighting factors		●														
	Response factor		●	●								●	●	●			
	Other					●	●			●		○			●		
Meteorological data reconstruction scheme	Stepwise	●	●		●				●		●		●	●		●	○
	Lin interpolation			●		●	●	●				●			●		●
	Other									●							
Zone air	Single temp	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	Stratified				○				○								
	Distribution									○							
Heater dynamics	None	●	●	●	●	●	●	●	●	●	●	●	●		●	●	○
	First order							○	○								
	Detailed model							○						●			●
Heater output characteristics	Purely convective		●		●	●					●				●		
	Fixed R/C split															●	
	U-spec R/C split	●		●			●	●	●	●		●	●				○
	Detailed model													●			●
Control temperature	Air temperature	●	●	●			●				●	●		●	●		
	Fixed R/C split					●										●	
	U-spec R/C split	○			●			●	●	●			●				●
	Surface																○
	Inter-construction																○

TABLE Al Features of the participating programs
Part (a): General Information, zones and control temperature

		TSBI3 A22	DOE-2.1E	TAS v7.54	ENERGY2 v1.0	CHEETAH v1.20	3TC v1.0	APACHE v6.6.2	HTB2 v1.2	CLIM2000 v1.1	DEROB-LTH v9301	S3PAS v2.0	BLAST v3.0 lvl193	TASE v3.0	TRNSYS v13.1	SERI-RES v1.2	ESP-r v7.7a
Window heat loss	U-spec resistance	●		●	●	●	●	●		●	●	●			●	●	○
	Dynamic		●						●				●				●
	Other													●			
Airgaps within windows	Fixed resistance											●					
	U-spec resistance	●	○		●	●	●	●	●	●	●		●		●	●	●
	Calc resistance		●											●			
	Separate zone																○
	Other			●													
Window transmission of direct radiation	Calc (explicitly)	●	●	●	●	●	●	●			●	●	●		●	●	●
	Calc (user funct)								●					●			
	Other									●							
Window transmission of diffuse radiation	Fixed altitude					●								●		●	●
	U-spec trans								●	●							
	Mult norm direct	●									●						
	Hemisph integral		●	●								●					
	Other				●		●	●					●		●		
Distribution of solar radiation in zones	Fixed				●	●	●	●		●		○					
	U-spec	●	●						●			○				●	○
	Calc once											●		●	●		
	Calc			●							●		●				●
Diffuse sky model	Isotropic			●		●			●	●	●	●	●				○
	Anisotropic	●	●		●		●	●						●	●	●	●

TABLE Al Features of the participating programs
Part (c): Glazing and solar processes

		TSBI3 A22	DOE-2.1E	TAS v7.54	ENERGY2 v1.0	CHEETAH v1.20	3TC v1.0	APACHE v6.6.2	HTB2 v1.2	CLIM2000 v1.1	DEROB-LTH v9301	S3PAS v2.0	BLAST v3.0 lvl193	TASE v3.0	TRNSYS v13.1	SERI-RES v1.2	ESP-r v7.7a
Heat transfer from external surfaces	R/C combined	●				●	●	●		○	●			●	●	●	
	R/C separated		●	●	●				●	●	○	●	●				●
External convection	Fixed coeffs					●					●					●	
	U-spec coeffs	●			○		●	●		●		●		●			○
	Calc coeffs		●	●	●				●				●				●
	Other														●		
External radiative heat transfer	Assumed to T_{air}	●									●		●		●	●	
	T_{sky} from met										○	●	●				
	Calc T_{sky}		●	●	●	●	●	●	●	●				●			●
	Viewfactors										●	●	●				●

TABLE A1 Features of the participating programs
Part (d): External heat transfer processes

Appendix B - Results and Statistics for Blind Predictions - Phase 1

B1 Primary Parameters

The predictions for the primary parameters predicted by each program are given in Table B1, along with the average for all the Phase 1 predictions. The measured values and the estimated 99percentile uncertainty band are also given.

B2 Hourly Statistics

The hourly measured and predicted values were compared using the following simple statistical measures:

Difference
$$D_t = X_t - R_t \tag{1}$$

Maximum Difference
$$\hat{D} = Max\ D_t \tag{2}$$

Minimum Difference
$$\check{D} = Min\ D_t \tag{3}$$

Mean Difference
$$\bar{D} = \sum_{t=1}^{N} D_t / N \tag{4}$$

Mean Absolute Difference
$$|\bar{D}| = \sum_{t=1}^{N} |D_t / N| \tag{5}$$

Root Mean Square Difference
$$\sqrt{\overline{D^2}} = \sqrt{\sum_{t=1}^{N} D_t^2 / N} \tag{6}$$

where: X_t = Predicted values at hour t
R_t = Reference (measured value) at hour t
N = Total hours in comparison period

The statistics for the air temperatures (May period) were evaluated for the complete 24-hour, 7-day period (N=168). The heating energy statistics (October period) were calculated only for the 7x12-hour periods for which heating was scheduled (N=84). The south facing solar irradiances (both periods) were evaluated only for the periods when the sun was up. The results are in Tables B2, B3 and B4.

	Heated, October Period						Free-floating, May Period						Sfvr	
	Double glazed			Opaque			Double glazed		Single Glazed		Opaque		Oct.	May
Program	E [MJ]	$\hat{T}$ [°C]	$\check{T}$ [°C]	E [MJ]	$\hat{T}$ [°C]	$\check{T}$ [°C]	$\hat{T}$ [°C]	$\check{T}$ [°C]	$\hat{T}$ [°C]	$\check{T}$ [°C]	$\hat{T}$ [°C]	$\check{T}$ [°C]	[MJ]	[MJ]
WG6TCv1992 (Udine, I)	-	-	-	-	-	-	35.6	5.1	35.2	4.9	26.2	4.9	-	84.0
TSBI3v2.0 (DBRI, DK)	85.5	36.1	11.6	95.7	30.0	12.4	28.0	12.5	28.1	11.1	16.4	9.2	72.2	80.8
DOE2.1E (LBL, US)	65.6	40.8	11.7	100.7	30.2	11.5	30.3	11.5	31.7	10.6	16.1	7.6	83.5	83.3
TASv7.54 (DMU/EDSL, UK)	79.8	36.8	9.6	104.7	30.0	11.4	29.0	11.1	28.4	9.4	15.7	7.8	67.5	79.6
ENERGY2v1.0 (Arup, UK)	70.9	37.7	11.3	99.8	30.0	12.6	28.4	12.0	30.8	11.0	16.5	9.1	83.8	81.3
CHEETAHv1.2 (CSIRO, AUS)	85.2	35.1	10.9	103.0	30.0	12.7	29.2	12.1	29.1	10.4	16.7	9.3	71.9	79.4
3TCv1.0 (Facet, UK)	85.1	36.3	12.3	117.1	30.0	14.1	27.0	12.2	27.3	11.3	15.9	9.1	76.0	79.1
APACHEv6.5.3 (Facet, UK)	86.1	36.3	12.6	118.6	30.1	14.8	26.9	12.1	27.5	11.0	15.8	8.9	76.0	79.1
HTB2v1.10 (UWCC, UK)	94.4	33.3	9.8	108.2	30.1	13.0	26.4	10.8	27.1	9.9	16.6	8.3	67.0	79.0
HTB2v1.2 (FHT, GER)	88.5	35.5	10.4	106.3	30.2	13.1	28.6	12.7	28.7	11.6	18.2	9.5	69.8	79.1
CLIM2000v1.1 (EDF, F)	83.3	41.4	10.5	108.9	29.9	12.4	30.8	11.9	29.2	10.0	15.0	7.7	73.9	80.4
DEROBvlth (Lund, S)	57.3	35.4	9.3	82.6	30.0	10.6	35.0	12.6	35.0	10.8	17.8	9.6	68.1	84.7
S3PASv2.0 (Sevilla, E)	78.0	38.8	12.1	105.7	30.0	13.7	29.1	11.9	30.4	11.0	17.4	9.4	79.1	79.0
BLASTv3lvl143 (CSU, US)	83.8	36.4	10.2	123.4	30.0	11.2	28.1	11.1	28.4	9.9	16.8	8.9	76.5	81.3
BLASTv3.0 (Torino, I)	91.5	38.0	9.5	130.8	30.0	10.1	33.4	11.2	33.0	9.9	20.5	9.7	-	-
TASEv3.0 (Tampere, FIN)	79.2	39.4	10.6	101.1	30.1	13.0	32.7	12.6	30.8	10.4	17.1	9.6	75.6	78.1
TRNSYSv13.1 (UWISC, US)	57.1	41.5	12.9	87.3	30.0	13.8	29.1	12.3	30.2	10.8	16.8	9.7	82.5	78.5
TRNSYSv13.1 (Brussel, B)	62.8	36.9	12.3	88.3	30.0	13.2	27.3	12.0	29.2	10.7	16.8	9.6	84.1	81.9
TRNSYSv13 (BRE, UK)	66.6	36.1	11.6	93.4	30.0	12.9	27.7	12.4	29.6	11.0	17.4	9.9	82.4	78.4
TRNSYSv12 (BRE, UK)	71.2	34.7	11.3	93.8	30.0	12.9	27.8	12.5	29.7	11.1	17.4	9.9	72.6	79.4
SUNCODEv5.7 (Ecotope, US)	75.3	36.4	10.6	112.1	30.0	12.4	28.0	11.9	28.2	10.6	16.7	9.0	78.4	81.5
SERI-RESv1.2 (BRE, UK)	83.2	37.9	12.1	104.7	30.0	13.6	28.5	11.7	29.2	10.4	16.6	9.2	77.6	80.0
ESP+v2.1 (DMU/ASL, UK)	55.5	43.8	13.8	92.7	29.9	14.3	31.7	13.6	32.0	11.9	15.5	8.4	76.8	80.7
ESP-Rv7.7a (ESRU, UK)	69.5	40.3	12.4	100.3	30.3	12.8	28.9	11.7	29.5	10.8	15.4	7.9	77.7	78.7
ESPv6.18a (DMU, UK)	61.1	42.7	14.0	94.7	30.0	14.0	32.6	13.5	33.3	12.0	15.7	8.4	69.4	77.5
average	75.7	37.8	11.4	103.1	30.0	12.8	29.6	11.8	30.1	10.5	17.2	8.8	76.0	80.1
measured	89.3	37.8	11.9	117.1	29.8	14.6	31.0	12.2	32.6	12.1	16.8	9.2	81.1	82.8
upper bound	92.7	40.5	13.9	122.3	30.2	16.4	33.4	13.6	35.0	13.6	17.5	10.0	85.5	88.8
lower bound	78.1	36.5	11.5	105.3	29.4	14.0	29.6	11.6	31.2	11.6	15.7	8.6	76.7	76.8

italics indicate programs for which Phase 2 results exist

Table B1: Phase 1 Results

Program	double glazed					opaque				
	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\bar{D}^{-2}$	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\bar{D}^{-2}$
WG6TCv1992 (Udine, I)	-	-	-	-	-	-	-	-	-	-
TSBI3v2.0 (DBRI, DK)	1451	-925	-50	258	409	130	-2549	-260	272	604
DOE2.1E (LBL, US)	338	-1321	-283	322	471	43	-839	-195	197	285
TASv7.54 (DMU/EDSL, UK)	-	-	-	-	-	-	-	-	-	-
ENERGY2v1.0 (Arup, UK)	644	-1141	-219	307	438	82	-833	-206	210	290
CHEETAHv1.2 (CSIRO, AUS)	1085	-1032	-48	273	392	12	-641	-168	168	217
3TCv1.0 (Facet, UK)	807	-706	-50	197	280	231	-101	0	58	74
APACHEv6.5.2 (Facet, UK)	731	-835	-38	173	255	324	-186	18	71	99
HTB2v1.10 (UWCC, UK)	1242	-526	59	217	328	245	-350	-109	125	153
HTB2v1.2 (FHT, GER)	1201	-791	-12	259	380	82	-494	-130	134	166
CLIM2000v1.1 (EDF, F)	1605	-1321	-75	329	515	130	-307	-103	112	136
DEROBvlth (Lund, S)	437	-1625	-381	436	584	52	-1424	-411	412	557
S3PASv2.0 (Sevilla, E)	467	-1141	-134	206	315	86	-537	-136	141	180
BLASTv3lv1143 (CSU, US)	710	-970	-65	175	260	267	-228	75	114	131
BLASTv3.0 (Torino, I)	769	-604	26	139	208	308	-101	163	183	204
TASEv3.0 (Tampere, FIN)	562	-1141	-121	197	307	-29	-830	-191	191	260
TRNSYSv13.1 (UWISC, US)	243	-1184	-383	402	545	-43	-1407	-354	354	476
TRNSYSv13.1 (Brussel, B)	501	-1147	-316	375	496	-29	-1417	-344	344	468
TRNSYSv13 (BRE, UK)	665	-1063	-270	351	460	44	-1346	-282	283	403
TRNSYSv12 (BRE, UK)	822	-1063	-216	355	448	49	-1346	-277	279	400
SUNCODEv5.7 (Ecotope, US)	291	-1285	-167	214	315	174	-234	-60	77	98
SERI-RESv1.2 (BRE, UK)	1499	-1208	-73	330	501	87	-548	-148	152	191
ESP+v2.1 (DMU/ASL, UK)	-	-	-	-	-	-	-	-	-	-
ESP-Rv7.7a (ESRU, UK)	446	-1134	-236	297	410	47	-778	-201	203	287
ESPv6.18a (DMU, UK)	259	-1292	-340	366	512	-14	-1058	-275	275	348
Largest Value	1605	-526	59	436	584	324	-101	163	412	604
Smallest Value	243	-1625	-383	139	208	-43	-2549	-411	58	74
Mean Value	763	-1066	-154	281	401	104	-798	-163	198	274
Range in Values	1362	1099	442	296	377	367	2448	573	354	531

italics indicate programs for which Phase 2 results exist

Table B2: Phase 1 Heating Energy Statistics for October Heating Periods [kJ]

Program	double glazed					opaque					single glazed				
	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\overline{D^2}$	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\overline{D^2}$	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\overline{D^2}$
WG6TCv1992 (Udine, I)	9.6	-10.1	-2.8	5.3	6.1	12.9	-5.8	1.8	4.7	5.6	8.5	-10.6	-3.4	5.3	6.2
TSBI3v2.0 (DBRI, DK)	1.2	-3.2	-0.7	0.9	1.2	0.5	-1.5	-0.6	0.7	0.8	0.6	-4.9	-1.8	1.8	2.2
DOE2.1E (LBL, US)	0.4	-1.1	-0.6	0.7	0.7	-0.5	-2.1	-1.4	1.4	1.5	0.2	-1.8	-1.1	1.1	1.2
TASv7.54 (DMU/EDSL, UK)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ENERGY2v1.0 (Arup, UK)	0.3	-2.8	-0.8	0.9	1.0	0.2	-1.4	-0.6	0.6	0.7	0.7	-2.1	-1.0	1.0	1.1
CHEETAHv1.2 (CSIRO, AUS)	0.6	-2.3	-0.7	0.7	0.9	1.3	-1.1	-0.2	0.5	0.6	-0.2	-3.6	-2.0	2.0	2.1
3TCv1.0 (Facet, UK)	0.2	-4.7	-1.4	1.5	1.9	0.0	-2.4	-1.0	1.0	1.2	-0.1	-6.1	-2.2	2.2	2.7
APACHEv6.5.2 (Facet, UK)	-0.0	-4.4	-1.5	1.5	1.9	-0.1	-2.4	-1.1	1.1	1.2	-0.5	-5.4	-2.4	2.4	2.7
HTB2v1.10 (UWCC, UK)	-0.4	-4.6	-1.9	1.9	2.2	-0.0	-1.8	-1.0	1.0	1.1	-0.7	-5.5	-2.7	2.7	2.9
HTB2v1.2 (FHT, GER)	2.1	-2.9	0.0	1.0	1.2	1.6	-0.9	0.3	0.6	0.8	1.7	-4.2	-0.9	1.4	1.7
CLIM2000v1.1 (EDF, F)	2.9	-1.1	0.0	0.6	0.8	-1.1	-3.1	-2.1	2.1	2.2	1.0	-3.5	-2.0	2.0	2.1
DEROBvlth (Lund, S)	6.7	-0.8	1.1	1.4	2.2	1.3	-0.4	0.3	0.4	0.5	6.1	-3.0	-0.3	1.9	2.3
S3PASv2.0 (Sevilla, E)	0.6	-2.4	-0.9	0.9	1.1	1.0	-0.6	-0.1	0.4	0.4	0.4	-3.1	-1.3	1.3	1.5
BLASTv3lvl143 (CSU, US)	2.1	-3.6	-1.4	1.7	1.9	0.8	-1.3	-0.5	0.6	0.7	1.4	-5.1	-2.4	2.5	2.9
BLASTv3.0 (Torino, I)	3.7	-2.2	-0.2	1.1	1.3	4.9	-0.7	1.4	1.5	2.0	2.1	-3.9	-1.5	1.8	2.1
TASEv3.0 (Tampere, FIN)	4.3	-0.6	0.8	0.9	1.4	1.9	-0.6	0.2	0.4	0.6	1.7	-3.0	-1.3	1.5	1.7
TRNSYSv13.1 (UWISC, US)	0.9	-2.6	-0.7	0.8	1.1	1.3	-0.6	-0.0	0.3	0.4	0.4	-3.8	-1.7	1.7	1.9
TRNSYSv13.1 (Brussel, B)	-0.0	-4.0	-1.3	1.3	1.6	1.3	-0.6	-0.0	0.3	0.4	-0.1	-4.1	-1.9	1.9	2.1
TRNSYSv13 (BRE, UK)	0.2	-3.5	-0.9	0.9	1.3	1.2	-0.5	0.2	0.3	0.5	0.2	-3.4	-1.4	1.5	1.7
TRNSYSv12 (BRE, UK)	0.4	-3.4	-0.8	0.9	1.2	1.3	-0.4	0.3	0.4	0.5	0.4	-3.4	-1.3	1.4	1.6
SUNCODEv5.7 (Ecotope, US)	3.4	-3.9	-0.8	1.3	1.6	0.4	-1.2	-0.6	0.6	0.7	2.6	-5.3	-1.7	1.9	2.3
SERI-RESv1.2 (BRE, UK)	1.1	-3.7	-0.9	1.1	1.5	0.1	-1.9	-0.5	0.5	0.7	0.6	-4.6	-1.6	1.7	2.1
ESP+v2.1 (DMU/ASL, UK)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ESP-Rv7.7a (ESRU, UK)	0.7	-2.3	-1.0	1.1	1.2	-0.8	-2.8	-1.9	1.9	1.9	-0.0	-3.2	-1.8	1.8	2.0
ESPv6.18a (DMU, UK)	3.7	0.0	1.3	1.3	1.4	-0.1	-2.3	-1.4	1.4	1.5	2.8	-1.4	0.0	0.7	0.9
Largest Value	9.6	0.0	1.3	5.3	6.1	12.9	-0.4	1.8	4.7	5.6	8.5	-1.4	0.0	5.3	6.2
Smallest Value	-0.4	-10.1	-2.8	0.6	0.7	-1.1	-5.8	-2.1	0.3	0.4	-0.7	-10.6	-3.4	0.7	0.9
Mean Value	1.8	-3.0	-0.7	1.3	1.6	1.2	-1.6	-0.4	1.0	1.1	1.2	-4.0	-1.6	1.9	2.1
Range in Values	10.0	10.1	4.0	4.7	5.4	14.0	5.4	3.9	4.4	5.2	9.2	9.3	3.4	4.7	5.3

italics indicate programs for which Phase 2 results exist

Table B3: Phase 1 Air Temperature Statistics for Complete Free-floating, May Period [°C]

Program	Free-floating, May period					Heated, October period				
	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\bar{D}^{2}$	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\bar{D}^{2}$
WG6TCv1992 (Udine, I)	128	-119	3	39	50	-	-	-	-	-
TSBI3v2.0 (DBRI, DK)	59	-61	-5	14	20	37	-235	-34	45	73
DOE2.1E (LBL, US)	37	-33	1	9	12	74	-45	9	19	27
TASv7.54 (DMU/EDSL, UK)	-	-	-	-	-	-	-	-	-	-
ENERGY2v1.0 (Arup, UK)	46	-90	-4	20	26	460	-53	10	27	63
CHEETAHv1.2 (CSIRO, AUS)	101	-147	-9	38	50	121	-229	-35	72	96
3TCv1.0 (Facet, UK)	46	-63	-9	19	25	144	-222	-20	61	78
APACHEv6.5.2 (Facet, UK)	46	-63	-9	19	25	144	-222	-20	61	78
HTB2v1.10 (UWCC, UK)	22	-61	-10	15	20	50	-210	-53	59	82
HTB2v1.2 (FHT, GER)	136	-178	-9	41	61	123	-235	-44	69	90
CLIM2000v1.1 (EDF, F)	47	-81	-6	18	25	135	-126	-27	41	56
DEROBvlth (Lund, S)	68	-63	5	19	27	139	-168	-49	59	80
S3PASv2.0 (Sevilla, E)	23	-59	-10	15	20	183	-99	-8	35	51
BLASTv3lvl143 (CSU, US)	121	-168	-4	35	52	154	-133	-17	48	65
BLASTv3.0 (Torino, I)	-	-	-	-	-	-	-	-	-	-
TASEv3.0 (Tampere, FIN)	37	-81	-12	19	27	96	-112	-21	31	44
TRNSYSv13.1 (UWISC, US)	23	-66	-11	16	22	102	-49	5	20	30
TRNSYSv13.1 (Brussel, B)	28	-73	-2	13	18	97	-43	12	18	27
TRNSYSv13 (BRE, UK)	22	-66	-11	16	22	102	-49	5	20	30
TRNSYSv12 (BRE, UK)	22	-58	-9	14	19	32	-128	-32	37	51
SUNCODEv5.7 (Ecotope, US)	170	-227	-3	46	70	205	-160	-10	56	78
SERI-RESv1.2 (BRE, UK)	206	-205	-9	75	89	451	-378	-23	157	200
ESP+v2.1 (DMU/ASL, UK)	-	-	-	-	-	-	-	-	-	-
ESP-Rv7.7a (ESRU, UK)	42	-84	-10	19	27	158	-174	-13	49	65
ESPv6.18a (DMU, UK)	89	-119	-13	30	40	188	-253	-45	73	101
Largest Value	206	-33	5	75	89	460	-43	12	157	200
Smallest Value	22	-227	-13	9	12	32	-378	-53	18	27
Mean Value	66	-97	-7	24	33	154	-163	-21	51	71
Range in Values	184	194	18	66	77	428	335	65	138	173

italics indicate programs for which Phase 2 results exist

Table B4: Phase 1 South Facing Verical Solar Irradiance Statistics [W/m^2]

Appendix C - Results and Statistics for Non-Blind Predictions - Phase 2

Cl Primary Parameters

The predictions for the primary parameters predicted by each program are given in Table Cl, along with the average for all the Phase 2 predictions. The measured values and the estimated 99percentile uncertainty band are also given.

C2 **Hourly Statistics**

The same statistics were used as for the Phase 1 comparisons (see Section B2). The results are in Tables C2, C3 and C4.

	Heated, October Period						Free-floating, May Period						Sfvr	
	Double glazed			Opaque			Double glazed		Single Glazed		Opaque		Oct.	May
Program	E [MJ]	$\hat{T}$ [°C]	$\check{T}$ [°C]	E [MJ]	$\hat{T}$ [°C]	$\check{T}$ [°C]	$\hat{T}$ [°C]	$\check{T}$ [°C]	$\hat{T}$ [°C]	$\check{T}$ [°C]	$\hat{T}$ [°C]	$\check{T}$ [°C]	[MJ]	[MJ]
WG6TCv1992 (Udine, I)	-	-	-	-	-	-	*33.8*	*13.4*	*33.7*	*11.5*	*18.0*	*10.2*	-	*84.0*
TSBI3v2.0 (DBRI, DK)	*80.8*	*37.3*	*12.0*	*111.1*	*30.0*	*14.1*	*29.1*	*12.7*	*29.0*	*11.2*	*16.6*	*9.4*	*80.9*	*81.1*
DOE2.1E (LBL, US)	65.6	40.8	11.7	100.7	30.2	11.5	30.3	11.5	31.7	10.6	16.1	7.6	83.5	83.3
TASv7.54 (DMU/EDSL, UK)	*83.3*	*36.5*	*10.8*	*109.4*	*30.0*	*12.9*	*28.5*	*11.6*	*30.4*	*11.0*	*16.1*	*8.3*	*72.5*	*79.6*
ENERGY2v1.0 (Arup, UK)	70.9	37.7	11.3	99.8	30.0	12.6	28.4	12.0	30.8	11.0	16.5	9.1	83.8	81.3
CHEETAHv1.2 (CSIRO, AUS)	85.2	35.1	10.9	103.0	30.0	12.7	29.2	12.1	29.1	10.4	16.7	9.3	71.9	79.4
3TCv1.0 (Facet, UK)	85.1	36.3	12.3	117.1	30.0	14.1	27.0	12.2	27.3	11.3	15.9	9.1	76.0	79.1
APACHEv6.5.2 (Facet, UK)	86.1	36.3	12.6	118.6	30.1	14.8	26.9	12.1	27.5	11.0	15.8	8.9	76.0	79.1
HTB2v1.10 (UWCC, UK)	94.4	33.3	9.8	108.2	30.1	13.0	26.4	10.8	27.1	9.9	16.6	8.3	67.0	79.0
HTB2v1.2 (FHT, GER)	*82.1*	*36.0*	*10.3*	*103.9*	*29.7*	*12.8*	*28.4*	*12.1*	*28.7*	*11.1*	*18.2*	*9.3*	*75.1*	*79.5*
CLIM2000v1.1 (EDF, F)	83.3	41.4	10.5	108.9	29.9	12.4	30.8	11.9	29.2	10.0	15.0	7.7	73.9	80.4
DEROBvlth (Lund, S)	57.3	35.4	9.3	82.6	30.0	10.6	35.0	12.6	35.0	10.8	17.8	9.6	68.1	84.7
S3PASv2.0 (Sevilla, E)	78.0	38.8	12.1	105.7	30.0	13.7	29.1	11.9	30.4	11.0	17.4	9.4	79.1	79.0
BLASTv3lvl143 (CSU, US)	83.8	36.4	10.2	123.4	30.0	11.2	28.1	11.1	28.4	9.9	16.8	8.9	76.5	81.3
BLASTv3lvl203 (Torino, I)	*68.5*	*41.1*	*13.3*	*97.3*	*30.1*	*14.5*	*32.5*	*13.1*	*32.5*	*11.6*	*16.9*	*10.0*	-	-
TASEv3.0 (Tampere, FIN)	79.2	39.4	10.6	101.1	30.1	13.0	32.7	12.6	30.8	10.4	17.1	9.6	75.6	78.1
TRNSYSv13.1 (UWISC, US)	57.1	41.5	12.9	87.3	30.0	13.8	29.1	12.3	30.2	10.8	16.8	9.7	82.5	78.5
TRNSYSv13.1 (Brussel, B)	62.8	36.9	12.3	88.3	30.0	13.2	27.3	12.0	29.2	10.7	16.8	9.6	84.1	81.9
TRNSYSv13 (BRE, UK)	66.6	36.1	11.6	93.4	30.0	12.9	27.7	12.4	29.6	11.0	17.4	9.9	82.4	78.4
TRNSYSv12 (BRE, UK)	71.2	34.7	11.3	93.8	30.0	12.9	27.8	12.5	29.7	11.1	17.4	9.9	72.6	79.4
SUNCODEv5.7 (Ecotope, US)	*80.1*	*36.4*	*10.7*	*111.9*	*30.0*	*12.5*	*28.8*	*11.9*	*29.2*	*10.6*	*16.7*	*9.0*	*72.8*	*79.3*
SERI-RESv1.2 (BRE, UK)	*82.2*	*36.8*	*11.1*	*103.8*	*30.0*	*13.2*	*28.9*	*11.8*	*29.5*	*10.6*	*16.7*	*9.2*	*73.5*	*77.6*
ESP+v2.1 (DMU/ASL, UK)	55.5	43.8	13.8	92.7	29.9	14.3	31.7	13.6	32.0	11.9	15.5	8.4	76.8	80.7
ESP-Rv7.7a (ESRU, UK)	69.5	40.3	12.4	100.3	30.3	12.8	28.9	11.7	29.5	10.8	15.4	7.9	77.7	78.7
ESPv6.18a (DMU, UK)	61.1	42.7	14.0	94.7	30.0	14.0	32.6	13.5	33.3	12.0	15.7	8.4	69.4	77.5
average	74.6	38.0	11.6	102.4	30.0	13.1	29.6	12.2	30.1	10.8	16.6	9.1	76.2	79.9
measured	89.3	37.8	11.9	117.1	29.8	14.6	31.0	12.2	32.6	12.1	16.8	9.2	81.1	82.8
upper bound	92.7	40.5	13.9	122.3	30.2	16.4	33.4	13.6	35.0	13.6	17.5	10.0	85.5	88.8
lower bound	78.1	36.5	11.5	105.3	29.4	14.0	29.6	11.6	31.2	11.6	15.7	8.6	76.7	76.8

italics indicate Phase 2 results

Table C1: Phase 1 and Phase 2 Results

Program	double glazed					opaque				
	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\overline{D^2}$	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\overline{D^2}$
WG6TCv1992 (Udine, I)	-	-	-	-	-	-	-	-	-	-
TSBI3v2.0 (DBRI, DK)	*1055*	*-1141*	*-105*	*243*	*374*	*130*	*-281*	*-77*	*97*	*115*
DOE2.1E (LBL, US)	338	-1321	-283	322	471	43	-839	-195	197	285
TASv7.54 (DMU/EDSL, UK)	*620*	*-679*	*-72*	*176*	*246*	*150*	*-484*	*-92*	*106*	*144*
ENERGY2v1.0 (Arup, UK)	644	-1141	-219	307	438	82	-833	-206	210	290
CHEETAHv15.2 (CSIRO, AUS)	1085	-1032	-48	273	392	12	-641	-168	168	217
3TCv1.0 (Facet, UK)	807	-706	-50	197	280	231	-101	0	58	74
APACHEv6.5.2 (Facet, UK)	731	-835	-38	173	255	324	-186	18	71	99
HTB2v1.10 (UWCC, UK)	1242	-526	59	217	328	245	-350	-109	125	153
HTB2v1.2 (FHT, GER)	*936*	*-845*	*-88*	*244*	*339*	*85*	*-604*	*-159*	*163*	*203*
CLIM2000v1.1 (EDF, F)	1605	-1321	-75	329	515	130	-307	-103	112	136
DEROBvlth (Lund, S)	437	-1625	-381	436	584	52	-1424	-411	412	557
S3PASv2.0 (Sevilla, E)	467	-1141	-134	206	315	86	-537	-136	141	180
BLASTv3lvl143 (CSU, US)	710	-970	-65	175	260	267	-228	75	114	131
BLASTv3lvl203 (Torino, I)	*238*	*-1141*	*-248*	*279*	*407*	*-29*	*-894*	*-236*	*236*	*317*
TASEv3.0 (Tampere, FIN)	562	-1141	-121	197	307	-29	-830	-191	191	260
TRNSYSv13.1 (UWISC, US)	243	-1184	-383	402	545	-43	-1407	-354	354	476
TRNSYSv13.1 (Brussel, B)	501	-1147	-316	375	496	-29	-1417	-344	344	468
TRNSYSv13 (BRE, UK)	665	-1063	-270	351	460	44	-1346	-282	283	403
TRNSYSv12 (BRE, UK)	822	-1063	-216	355	448	49	-1346	-277	279	400
SUNCODEv5.7 (Ecotope, US)	*528*	*-718*	*-110*	*188*	*268*	*167*	*-234*	*-62*	*78*	*98*
SERI-RESv1.2 (BRE, UK)	*948*	*-1062*	*-85*	*248*	*371*	*70*	*-551*	*-159*	*161*	*199*
ESP+v2.1 (DMU/ASL, UK)	-	-	-	-	-	-	-	-	-	-
ESP-Rv7.7a (ESRU, UK)	446	-1134	-236	297	410	47	-778	-201	203	287
ESPv6.18a (DMU, UK)	259	-1292	-340	366	512	-14	-1058	-275	275	348
Largest Value	1605	-526	59	436	584	324	-101	75	412	557
Smallest Value	238	-1625	-383	173	246	-43	-1424	-411	58	74
Mean Value	691	-1053	-166	276	392	90	-725	-171	190	254
Range in Values	1367	1099	442	263	338	367	1323	486	354	483

italics indicate Phase 2 results

Table C2: Phase 2 Heating Energy Statistics for October Heating Periods [kJ]

Program	double glazed					opaque				
	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\bar{D}^{-2}$	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\bar{D}^{-2}$
WG6TCv1992 (Udine, I)	-	-	-	-	-	-	-	-	-	-
TSBI3v2.0 (DBRI, DK)	*1055*	*-1141*	*-105*	*243*	*374*	*130*	*-281*	*-77*	*97*	*115*
DOE2.1E (LBL, US)	338	-1321	-283	322	471	43	-839	-195	197	285
TASv7.54 (DMU/EDSL, UK)	*620*	*-679*	*-72*	*176*	*246*	*150*	*-484*	*-92*	*106*	*144*
ENERGY2v1.0 (Arup, UK)	644	-1141	-219	307	438	82	-833	-206	210	290
CHEETAHv15.2 (CSIRO, AUS)	1085	-1032	-48	273	392	12	-641	-168	168	217
3TCv1.0 (Facet, UK)	807	-706	-50	197	280	231	-101	0	58	74
APACHEv6.5.2 (Facet, UK)	731	-835	-38	173	255	324	-186	18	71	99
HTB2v1.10 (UWCC, UK)	1242	-526	59	217	328	245	-350	-109	125	153
HTB2v1.2 (FHT, GER)	*936*	*-845*	*-88*	*244*	*339*	*85*	*-604*	*-159*	*163*	*203*
CLIM2000v1.1 (EDF, F)	1605	-1321	-75	329	515	130	-307	-103	112	136
DEROBvlth (Lund, S)	437	-1625	-381	436	584	52	-1424	-411	412	557
S3PASv2.0 (Sevilla, E)	467	-1141	-134	206	315	86	-537	-136	141	180
BLASTv3lvl143 (CSU, US)	710	-970	-65	175	260	267	-228	75	114	131
BLASTv3lvl203 (Torino, I)	*238*	*-1141*	*-248*	*279*	*407*	*-29*	*-894*	*-236*	*236*	*317*
TASEv3.0 (Tampere, FIN)	562	-1141	-121	197	307	-29	-830	-191	191	260
TRNSYSv13.1 (UWISC, US)	243	-1184	-383	402	545	-43	-1407	-354	354	476
TRNSYSv13.1 (Brussel, B)	501	-1147	-316	375	496	-29	-1417	-344	344	468
TRNSYSv13 (BRE, UK)	665	-1063	-270	351	460	44	-1346	-282	283	403
TRNSYSv12 (BRE, UK)	822	-1063	-216	355	448	49	-1346	-277	279	400
SUNCODEv5.7 (Ecotope, US)	*528*	*-718*	*-110*	*188*	*268*	*167*	*-234*	*-62*	*78*	*98*
SERI-RESv1.2 (BRE, UK)	*948*	*-1062*	*-85*	*248*	*371*	*70*	*-551*	*-159*	*161*	*199*
ESP+v2.1 (DMU/ASL, UK)	-	-	-	-	-	-	-	-	-	-
ESP-Rv7.7a (ESRU, UK)	446	-1134	-236	297	410	47	-778	-201	203	287
ESPv6.18a (DMU, UK)	259	-1292	-340	366	512	-14	-1058	-275	275	348
Largest Value	1605	-526	59	436	584	324	-101	75	412	557
Smallest Value	238	-1625	-383	173	246	-43	-1424	-411	58	74
Mean Value	691	-1053	-166	276	392	90	-725	-171	190	254
Range in Values	1367	1099	442	263	338	367	1323	486	354	483

italics indicate Phase 2 results

Table C2: Phase 2 Heating Energy Statistics for October Heating Periods [kJ]

Program	double glazed					opaque					single glazed				
	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\overline{D^2}$	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\overline{D^2}$	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\overline{D^2}$
WG6TCv1992 (Udine, I)	*4.5*	*0.3*	*1.7*	*1.7*	*1.9*	*2.1*	*0.2*	*0.9*	*0.9*	*1.0*	*3.6*	*-1.3*	*0.3*	*0.8*	*1.1*
TSBI3v2.0 (DBRI, DK)	*1.3*	*-2.2*	*-0.5*	*0.7*	*0.9*	*0.8*	*-1.3*	*-0.4*	*0.6*	*0.7*	*0.2*	*-4.1*	*-1.7*	*1.7*	*2.0*
DOE2.1E (LBL, US)	0.4	-1.1	-0.6	0.7	0.7	-0.5	-2.1	-1.4	1.4	1.5	0.2	-1.8	-1.1	1.1	1.2
TASv7.54 (DMU/EDSL, UK)	*0.5*	*-2.9*	*-1.1*	*1.1*	*1.4*	*-0.3*	*-2.0*	*-1.2*	*1.2*	*1.3*	*0.4*	*-3.1*	*-1.3*	*1.3*	*1.5*
ENERGY2v1.0 (Arup, UK)	0.3	-2.8	-0.8	0.9	1.0	0.2	-1.4	-0.6	0.6	0.7	0.7	-2.1	-1.0	1.0	1.1
CHEETAHv15.2 (CSIRO, AUS)	0.6	-2.3	-0.7	0.7	0.9	1.3	-1.1	-0.2	0.5	0.6	-0.2	-3.6	-2.0	2.0	2.1
3TCv1.0 (Facet, UK)	0.2	-4.7	-1.4	1.5	1.9	0.0	-2.4	-1.0	1.0	1.2	-0.1	-6.1	-2.2	2.2	2.7
APACHEv6.5.2 (Facet, UK)	-0.0	-4.4	-1.5	1.5	1.9	-0.1	-2.4	-1.1	1.1	1.2	-0.5	-5.4	-2.4	2.4	2.7
HTB2v1.10 (UWCC, UK)	-0.4	-4.6	-1.9	1.9	2.2	-0.0	-1.8	-1.0	1.0	1.1	-0.7	-5.5	-2.7	2.7	2.9
HTB2v1.2 (FHT, GER)	*2.2*	*-2.9*	*-0.2*	*1.0*	*1.2*	*1.6*	*-1.1*	*0.1*	*0.6*	*0.8*	*1.7*	*-4.1*	*-1.1*	*1.6*	*1.8*
CLIM2000v1.1 (EDF, F)	2.9	-1.1	0.0	0.6	0.8	-1.1	-3.1	-2.1	2.1	2.2	1.0	-3.5	-2.0	2.0	2.1
DEROBvlth (Lund, S)	6.7	-0.8	1.1	1.4	2.2	1.3	-0.4	0.3	0.4	0.5	6.1	-3.0	-0.3	1.9	2.3
S3PASv2.0 (Sevilla, E)	0.6	-2.4	-0.9	0.9	1.1	1.0	-0.6	-0.1	0.4	0.4	0.4	-3.1	-1.3	1.3	1.5
BLASTv3lvl143 (CSU, US)	2.1	-3.6	-1.4	1.7	1.9	0.8	-1.3	-0.5	0.6	0.7	1.4	-5.1	-2.4	2.5	2.9
BLASTv3lvl203 (Torino, I)	*2.7*	*-0.9*	*0.7*	*0.8*	*0.9*	*1.5*	*-0.8*	*0.2*	*0.4*	*0.5*	*1.4*	*-2.7*	*-0.7*	*0.9*	*1.1*
TASEv3.0 (Tampere, FIN)	4.3	-0.6	0.8	0.9	1.4	1.9	-0.6	0.2	0.4	0.6	1.7	-3.0	-1.3	1.5	1.7
TRNSYSv13.1 (UWISC, US)	0.9	-2.6	-0.7	0.8	1.1	1.3	-0.6	-0.0	0.3	0.4	0.4	-3.8	-1.7	1.7	1.9
TRNSYSv13.1 (Brussel, B)	-0.0	-4.0	-1.3	1.3	1.6	1.3	-0.6	-0.0	0.3	0.4	-0.1	-4.1	-1.9	1.9	2.1
TRNSYSv13 (BRE, UK)	0.2	-3.5	-0.9	0.9	1.3	1.2	-0.5	0.2	0.3	0.5	0.2	-3.4	-1.4	1.5	1.7
TRNSYSv12 (BRE, UK)	0.4	-3.4	-0.8	0.9	1.2	1.3	-0.4	0.3	0.4	0.5	0.4	-3.4	-1.3	1.4	1.6
SUNCODEv5.7 (Ecotope, US)	*0.6*	*-2.6*	*-0.9*	*1.0*	*1.2*	*0.3*	*-1.2*	*-0.6*	*0.6*	*0.7*	*-0.2*	*-3.8*	*-1.8*	*1.8*	*2.1*
SERI-RESv1.2 (BRE, UK)	*0.2*	*-2.6*	*-1.1*	*1.1*	*1.2*	*0.4*	*-1.4*	*-0.5*	*0.5*	*0.6*	*-0.3*	*-3.5*	*-1.8*	*1.8*	*1.9*
ESP+v2.1 (DMU/ASL, UK)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ESP-Rv7.7a (ESRU, UK)	0.7	-2.3	-1.0	1.1	1.2	-0.8	-2.8	-1.9	1.9	1.9	-0.0	-3.2	-1.8	1.8	2.0
ESPv6.18a (DMU, UK)	3.7	0.0	1.3	1.3	1.4	-0.1	-2.3	-1.4	1.4	1.5	2.8	-1.4	0.0	0.7	0.9
Largest Value	6.7	0.3	1.7	1.9	2.2	2.1	0.2	0.9	2.1	2.2	6.1	-1.3	0.3	2.7	2.9
Smallest Value	-0.4	-4.7	-1.9	0.6	0.7	-1.1	-3.1	-2.1	0.3	0.4	-0.7	-6.1	-2.7	0.7	0.9
Mean Value	1.5	-2.4	-0.5	1.1	1.4	0.6	-1.4	-0.5	0.8	0.9	0.8	-3.6	-1.5	1.7	1.9
Range in Values	7.0	4.9	3.6	1.3	1.5	3.2	3.2	3.0	1.8	1.8	6.8	4.8	2.9	2.0	2.1

italics indicate Phase 2 results

Table C3: Phase 2 Air Temperature Statistics for Complete Free-floating, May Period [°C]

Program	Free-floating, May period					Heated, October period				
	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\bar{D}^{2}$	$\hat{D}$	$\check{D}$	$\bar{D}$	$\lvert\bar{D}\rvert$	$\bar{D}^{2}$
WG6TCv1992 (Udine, I)	*128*	*-119*	*3*	*39*	*50*	-	-	-	-	-
TSBI3v2.0 (DBRI, DK)	*27*	*-40*	*-4*	*10*	*14*	*77*	*-53*	*-1*	*20*	*25*
DOE2.1E (LBL, US)	37	-33	1	9	12	74	-45	9	19	27
TASv7.54 (DMU/EDSL, UK)	*23*	*-60*	*-8*	*14*	*19*	*21*	*-181*	*-33*	*37*	*55*
ENERGY2v1.0 (Arup, UK)	46	-90	-4	20	26	460	-53	10	27	63
CHEETAHv15.2 (CSIRO, AUS)	101	-147	-9	38	50	121	-229	-35	72	96
3TCv1.0 (Facet, UK)	46	-63	-9	19	25	144	-222	-20	61	78
APACHEv6.5.2 (Facet, UK)	46	-63	-9	19	25	144	-222	-20	61	78
HTB2v1.10 (UWCC, UK)	22	-61	-10	15	20	50	-210	-53	59	82
HTB2v1.2 (FHT, GER)	*41*	*-87*	*-8*	*17*	*24*	*158*	*-174*	*-23*	*53*	*69*
CLIM2000v1.1 (EDF, F)	47	-81	-6	18	25	135	-126	-27	41	56
DEROBvlth (Lund, S)	68	-63	5	19	27	139	-168	-49	59	80
S3PASv2.0 (Sevilla, E)	23	-59	-10	15	20	183	-99	-8	35	51
BLASTv3lvl143 (CSU, US)	121	-168	-4	35	52	154	-133	-17	48	65
BLASTv3lvl203 (Torino, I)	-	-	-	-	-	-	-	-	-	-
TASEv3.0 (Tampere, FIN)	37	-81	-12	19	27	96	-112	-21	31	44
TRNSYSv13.1 (UWISC, US)	23	-66	-11	16	22	102	-49	5	20	30
TRNSYSv13.1 (Brussel, B)	28	-73	-2	13	18	97	-43	12	18	27
TRNSYSv13 (BRE, UK)	22	-66	-11	16	22	102	-49	5	20	30
TRNSYSv12 (BRE, UK)	22	-58	-9	14	19	32	-128	-32	37	51
SUNCODEv5.7 (Ecotope, US)	*22*	*-58*	*-9*	*14*	*19*	*33*	*-127*	*-32*	*36*	*51*
SERI-RESv1.2 (BRE, UK)	*132*	*-111*	*-13*	*34*	*46*	*187*	*-212*	*-29*	*81*	*103*
ESP+v2.1 (DMU/ASL, UK)	-	-	-	-	-	-	-	-	-	-
ESP-Rv7.7a (ESRU, UK)	42	-84	-10	19	27	158	-174	-13	49	65
ESPv6.18a (DMU, UK)	89	-119	-13	30	40	188	-253	-45	73	101
Largest Value	132	-33	5	39	52	460	-43	12	81	103
Smallest Value	22	-168	-13	9	12	21	-253	-53	18	25
Mean Value	46	-75	-7	18	25	126	-137	-19	43	59
Range in Values	110	135	18	30	40	439	210	65	63	78

italics indicate Phase 2 results

Table C4: Phase 2 South Facing Verical Solar Irradiance Statistics [W/m^2]

Appendix D - Issues Raised in Model User Reports

Dl Introduction

Many of the Model User Reports (MURs) questioned the reliability and appropriateness of the experiment on which this IEA exercise was based.

(i) Reliability of the data; and

(ii) Modelling issues.

These issues are discussed in this Appendix.

D2.1 Reliability of the Data

Uncertainties in the information which was supplied to modellers can be divided into two categories: uncertainties in the description of the buildings and uncertainties in the measured (e.g. climate) data which was supplied to operate the models. A vital distinction separates the two types of data. We can, if necessary go back to the rooms and check any details which might be in doubt. This is unfortunately not the case with the climate data, where the given uncertainties have to be accepted.

Fortunately, none of these uncertainties need compromise the conclusions of a well conducted empirical validation test. Techniques exist for establishing the impact of uncertainties from all sources, thus allowing firm conclusions to be drawn from an experiment.

D2.2 Overall simplifications within the building specification

The data in the handbook describing the buildings was presented in two forms: drawings of the rooms, and tables of surface areas, constructions and material properties. In principle modellers could have been supplied only with drawings, and left to input these into each model. Thermal properties could have been obtained by each modeller from the reference sources which he or she normally uses. This approach might be considered preferable to the one adopted, in that it would test the model in a more realistic way, under conditions which are even closer to those occurring in a real design application. In the IEA exercise, however, the aim was to reduce to a minimum the likelihood that (external) user errors would significantly influence the results. Therefore, a decision was made not to make model users take details from drawings, or thermal properties from diverse sources.

Inevitably, simplifications were introduced in the process of presenting the information from the drawings. Decisions were made to omit certain features of the room construction. These decisions were based on a mixture of experience and the results of previous sensitivity studies.

Sensitivity analyses by DMU (using ESPv6.18a) showed that the simplifications to the building description identified by some participants (Appx.F3, Fl1) which included the large uncertainty in the roofspace infiltration rate (1 to 3 ach) would have no noticeable effect on any of the primary parameters for the double glazed, heated room. They are very unlikely to have an effect on the other rooms either.

The provision of drawings in the description document of the test cells meant that modellers were aware that these simplifications had been made, and meant that they could model the additional features if they desired. Any information required for this but not given in the documentation could, of course, have been obtained from the hotline.

The measured UA-value of the rooms and their internal convective regime could not have been supplied (even if they had been available at the start of the exercise), because they are not program inputs. They are thermal characteristics which programs are supposed to be able to model (see Appx.F3, F10). Such details would not be available when programs are used for real design problems.

By reducing the building description to a series of tables it was considered that the specification could be standardised among the participating modellers, reducing the spread of results between models by removing an important source of user variation. Given the spread in the results which was observed even using this approach it seems that the decision was appropriate.

The way of describing a building for thermal program validation adopted in this IEA study is a good exemplar which should be adopted in similar work in the future. In fact, it is essential to simplify the description of a building to those aspects which are thermally significant if results, which accurately reflect the performance of the program, with a minimum of user effects, are to be produced.

With such documentation, it is important to provide contact addresses (and/or telephone/fax. numbers) so that additional, crucial, data which may have been overlooked can be provided.

All the participants who chose to comment felt that this management approach had been very useful, very good or excellent.

D2.3 Fabric heat loss

D2.3.1 Material properties

During the course of the validation exercise the use of manufacturers' data on the thermal properties of the materials in the test rooms has been criticised (e.g. Appx.F8).

The uncertainties quoted for the material properties stem from four sources:

- there are uncertainties in the manufacturers' measurement of the properties of their materials,
- there may be variations in the properties of materials with age,
- there may be variations with operating conditions (most notably the variation of insulation material k-values with operating temperature), and
- there may be small variations between different batches of the same material.

Carrying out tests on the actual materials used in the test room construction would serve to eliminate only the last of these sources of uncertainty. In doing so it would allow a very small reduction in the total uncertainty in each property. This would in turn have the effect of slightly reducing the overall uncertainty associated with a simulation. This would increase the power of the datasets as a test of simulation models. Given the small number of programs which lay within the uncertainty bands in the existing test it seems that this increase in power is not, for the moment at least, required.

D2.3.2 Edge effects

It has been suggested that the edge effects in the test rooms may have been significantly underestimated (Appx.F3). This would explain the general underprediction of room energy consumption.

It was acknowledged in the original building description that the edge effect correction had been derived from only two 2-dimensional calculations. Other edges had either been assumed to be similar in construction, or had been ignored altogether. It was therefore clear that the neglected edges would contribute further to the building heat loss, and that the corrections given were likely to be underestimates. The concerns expressed fall into two categories:

- some of the edges which were neglected may have made a significant contribution to the 2-dimensional heat loss effects,
- some of the edges which were assumed to have losses comparable to those calculated may in fact have had rather higher losses. In particular the vertical south edge of the party wall may have losses greater than the 'Type A' calculated edge loss which was attributed to them.

Further 2-dimensional analysis would clarify this situation, but was not practical within the time available. For the purposes of this exercise the error bar on the edge effect calculation was increased to -0 +50% to account for the effects described. This has been incorporated in the overall error bands plotted in this report.

D2.3.3 Overall UA measurements

One way of determining whether the construction of the test room, the material characteristics and the edge effects are broadly in line with the specification given, is to measure the overall heat loss of the real room, and compare with the corresponding calculated value.

Whilst Phase 2 was underway, the UA-value of one of the test rooms was estimated from data which was collected approximately two weeks after the IEA heated data set was produced. The analysis of this data is described in full in Appendix E. It results in an estimate of the UA-value of the test room fitted with opaque infill panel of 10.94 W/K ±5%

A calculation of the building UA-value using the values distributed in the test mom description, with their corresponding estimated uncertainties, yields a value of 10.13 W/K -7/+8%

It is clear that the uncertainties associated with these two values overlap, and that no difference between the measured and calculated values can be conclusively demonstrated. It is, however, still instructive to look at possible causes of the small discrepancy which has been observed.

Part of the discrepancy observed can be explained by the fact that the experiment to determine the UA-value was carried out with the room continuously heated to a relatively high setpoint. This has the effect of modifying the k-values of the insulation materials used in the room. The adjustment of the material properties to incorporate the effect of the slightly higher than normal mean operating temperature is described in Appendix E. The calculated UA-value becomes 10.55 W/K -7/+8%

It is clear that the corrections for operating temperature serve to account for a large part of the discrepancy observed. The remaining discrepancy is only 3.6%.

This test does not conclusively demonstrate that the buildings are exactly as described, since an error in one area could be cancelled out by a second error elsewhere. However, it does provide one further piece of evidence that the rooms are constructed as described in the site handbook, and that any errors in the calculated edge effects are very small.

Further details about quality assurance on the EMC test site can be found in Volume 2 (Site Handbook). All the checks described there confirm that the rooms were all the same and that they were built as described in the Site Handbook.

D2.4 Relative Humidity measurement

It is clearly regrettable that relative humidity data is missing from the heated dataset. This weakens the power of the dataset for testing models, but it does not 'negate the exercise as an appropriate validation' as suggested in one MUR (Appx.F11). In Newssheet 9 it was suggested that users who were concerned about this could determine the sensitivity of their model to variations in Rh of 55 to 100%, this being the range encountered in the UK in October. Clearly for some models this would increase the width of the error bars associated with the simulation, thus weakening the test.

Sensitivity analyses by DMU (using ESPv6.18a) showed that a decrease in Rh from 99% to 55% increased the predicted energy consumption in the double glazed room by 3.1%. (This is close to the 3.3% increase predicted by the TASv7.54 user (Appx. F3)). The maximum and minimum air temperatures decreased by 0.4°C and 0.2°C respectively. These effects have been incorporated in the overall uncertainty bands plotted in this report.

The rather unusual form of the measured humidity data has been queried, and doubts have also been expressed as to the rather high mean value (Appx. F10, Fl 1). As described in the site handbook these measurements were made using wet and dry bulb thermometers. This is a 'first principle' approach to the measurement of humidity, and is considered highly trustworthy. It is certainly far more reliable than electronic sensors which are highly sensitive to contamination by atmospheric pollutants.

As part of his Phase 2 investigations the TASv7.54 user (Appx.F3) obtained humidity data from a meteorological station in a nearby town (Bedford). This showed the same characteristics as the data supplied for the unheated period: a relatively high mean value (87.8% during the October period), and long periods at 100% humidity. The same user has also noted that the name of the site, Cranfield, derives from Crane Field. Cranes are long legged wading birds which normally nest in bogs and swamps!

D2.5 Solar radiation data

The solar radiation data supplied for the blind validation exercise has also been questioned (Appx.F1, F3, F4, F10, Fl1). Some participants felt that some of the direct normal values calculated from the site measurements of direct and diffuse horizontal were unreliable. The TASv7.54 user (Appx.F3) also queried the way in which shadow band data has been corrected and the TSBI3 user (Appx.Fl) suggested that the given south vertical radiation was not compatible with the previously supplied global and diffuse horizontal measurements. These problems were particularly severe when the sun is near the horizon, just after sunrise and just before sunset.

Diffuse radiation is measured on the test site using a pyranometer fitted with a shadow ring. As well as obscuring the direct component the ring obscures part of the diffuse sky, and the reading of the pyranometer has to be corrected for this. To make this correction (which varies with shadow band position and hence with time of year) it is necessary to assume a distribution for the diffuse sky radiation. The distribution normally assumed when the data was processed for the IEA task was uniform, in keeping with normal meteorological practice. This raises a problem of consistency when feeding the diffuse data to a model which assumes an alternative, anisotropic, distribution . Also, the real distribution is often far from uniform, so the correction process may introduce significant errors into the estimate of the total diffuse radiation component at certain times.

Previous data supplied to the TASv7.54 user as part of a separate project had used an alternative correction which takes account of the part of the circumsolar component obscured by the shadow band. This accounts for the discrepancy which he observed between his earlier data and the IEA climate files. The method of correcting may cause the diffuse irradiances to differ by up to 6% (TASv7.54 user, Appx.F3). Sensitivity analyses by DMU (using ESPv6.18a), however, indicated that this would have virtually no impact on the primary parameters predicted in the glazed rooms (less than 0.5MJ in heating energy and 0.1°C in temperatures). Nevertheless, the effect was included in all the uncertainty bands calculated in this report.

The process of deriving direct normal from global and diffuse measurements is, under some circumstances, susceptible to high levels of error. One of the worst times is when the sun is close to the horizon. At this time the direct component strikes the horizontal pyranometers at an incidence angle of almost 90°. This leads to some serious measurement problems:

- small errors in the alignment of the instruments become significant,
- the cosine response of the instruments is likely to be unreliable at these incidence angles, and may vary from instrument to instrument, and
- moisture on the outside of the pyranometer domes may cause refraction effects at the surfaces of the glass.

When the difference between the direct and diffuse horizontal radiation measured at low solar altitudes is taken a spurious result may thus appear. Furthermore, the sine of the solar altitude is very small at this time, and dividing by it serves to magnify the erroneous reading. To overcome this problem we set the direct normal component to zero when the solar altitude falls below 5°. This avoids the generation of spuriously large radiation values on East and West surfaces, and causes minimal errors on the South wall. This may be one explanation for the discrepancies observed in the data.

For the EMC test rooms the problem is not too serious - the only glazing is oriented towards the south, and for a large part of the year it receives low altitude direct normal radiation at a very oblique incidence angle, or not at all. The other surfaces are painted white, and this reduces the sensitivity to the (potentially uncertain) radiation which lands on them.

It is accepted that measurements of global and diffuse solar radiation horizontal are not sufficient to allow the direct normal component to be reliably estimated at all times. (Data from a normal incidence (tracking) instrument, and diffuse measurements using a tracking disk rather than a shadow ring are probably also necessary if models of solar radiation on other surfaces are to be comprehensively tested). As discussed above, this does not present a problem for the test rooms, as the only important surface is the south vertical and direct normal irradiance can generally be obtained when this surface is illuminated, and the radiation on the surface was measured directly. However, modellers who routinely use global and diffuse horizontal data to predict the radiation on East or West facing surfaces should perhaps consider carefully the accuracy they can expect to obtain.

The TSBI3 user (Appx.Fl) suggests that the measured diffuse and global radiation may be wrong, in that they appear to be inconsistent with levels on the south vertical early in the morning. In part the problem may be due to the instrumentation limitations described above, and it also appears that he also uses a different solar position algorithm to the Energy Monitoring Company. However, it is not possible to state that the given global and diffuse irradiance values are inconsistent with the measured radiation on the south vertical, as the distribution of diffuse radiation is completely unknown. Early in the morning there is likely to be considerable 'horizon brightening '. At its most extreme, this could yield a positive radiation level on the south face whilst the global and diffuse measurements on the horizontal plane were still zero! The measurements would not be wrong: they would simply be inconsistent with the conventional assumption of a uniform diffuse sky.

In response to these problems the calculated direct normal radiation values have been removed from the meteorological file distributed with the validation package. Model users who require it can recreate it using the solar position algorithm used within their model. In this way predictions of diffuse radiation (if global and direct normal are required by the model) or of global horizontal radiation (if diffuse and direct Formal have been supplied) will be consistent with the data provided.

D2.6 Areas in which too much information was supplied

In the original test room description the amount of solar radiation believed to be reflected back out of the test rooms was supplied, as was the distribution of the remaining radiation. This information is not an aspect of the room construction, and should not have been supplied to modellers. Instead it should have been calculated by those models which have this facility, and estimated by other participants in their Formal way. Accordingly this information has been relegated to an appendix in the revised Site Handbook (Volume 2).

D2.7 Air temperature uncertainty

At least one participant (Appx.F8) suggested that the uncertainty attributed to the measured room air temperatures and the heater setpoint was too low. This source of uncertainty arises because the programs assume that the air temperature in rooms is uniform throughout, whereas in fact (see section D3.3) it varies.

To obtain error estimates, the three measured air temperatures (located in the vertical rake) were plotted and compared with the mean temperature of the three. This showed that, during heated periods, the mean temperature was almost always within the ±0.2°C of the middle temperature on which the heater was controlled. The uncertainty in the thermostat setpoint temperature was therefore unchanged at ±0.2°C.

The measured maximum and minimum air temperatures against which the predictions are compared were the average of the three vertical temperature measurements. The measurement uncertainty in this value was taken as the difference between this average and the single temperature which was most different from it. The biggest observed difference occurred at the time of the peak temperature in the double glazed room and was ±1.2°C (October period). For all other parameters this difference was less than ±0.4°C. These uncertainties have been included in all the uncertainty bands plotted in this report.

D3 Modelling issues

A range of modelling issues have been raised in the three page model user reports. Some of these describe features of the test rooms which could not be represented in certain codes, most notably heater output characteristics and dynamics. Other researchers have identified effects which they feel could be the cause of discrepancies between their simulation results and the measured building performance. These effects concern the behaviour of the air within the test room, in particular the convective heat transfer coefficients at the room surfaces and the air movement within the room.

D3.1 Heater modelling

The most common heat source in the UK domestic building stock is the hot water filled radiator. Normally these are supplied from a central boiler. The oil-filled electric heaters used in the test rooms were chosen because they closely matched the output characteristics of their water filled counterparts, but could be controlled and metered much more accurately.

Modellers were given information about the characteristics of these heaters, in the form of a calculated radiative/convective split, and a measured first-order time constant. The expressions used to calculate the radiative convective/split were also used to determine the rise in heater surface temperature for a given power output. The calculated value was found to agree closely with that measured, giving a degree of confidence in the quoted R/C split (Dl).

Inspection of Appendix A (Table Al), which describes the features available within the models, reveals that 5 of the models have no facility to represent a radiative component in the heater output: they represent the heater as 100% convective. This would be expected to produce systematic overestimation of the room air temperature, and corresponding under-estimation of the room energy consumption. Since many real heat sources have a significant radiative output this appears to be a significant shortcoming of these models. The DOE2v1E user recognized this limitation (Appx.F2).

The participant using HTB2v1.2 (Appx.F4) has discovered that by increasing the convective output of the heater it was possible to obtain a better fit to the room air temperature, but not to overall energy consumption. It therefore seems likely that this adjustment, which would be expected to raise the room air temperature, is simply compensating for another cause of prediction error and not improving the overall quality of the simulation.

Further inspection of Table Al shows that all but 4 of the models lack the ability to represent the dynamic behaviour of the room heat source. Once again, this limitation would be expected to produce under-estimates of the room energy consumption, as energy used to charge the mass of the heater emerges after the heating period, when it is not required. Since dynamic behaviour is a feature of many realistic heat sources the lack of this facility in a model is also considered to be a significant shortcoming.

The ESP-rv7.7a user (Appx.F10) has queried the use of a 22 minute heater time constant, given that climate data was supplied only on a hourly basis. The thermal processes which combine to determine the overall behaviour of a building operate over timescales which range from 10^{-8} seconds (for radiative processes) to 10^{5} seconds. In the case of buildings which are coupled to the ground this upper limit rises to 10^{10} seconds. In any practical modelling exercise time constants below a certain level are ignored (these processes are assumed to occur instantaneously) and time constants above a certain level are also ignored (assumed to be constant over the duration of a simulation). Deciding whether processes fall into one of these categories is a matter of modelling judgement. However it seems clear that when hourly predictions are of interest, a process with a time constant of 22 minutes cannot be ignored.

D3.2 Internal convection: heat transfer coefficients at internal surfaces

The users of ESP-rv7.7a and ESP+v2.1 (Appx.F10 and Fl1) have observed that the use of greatly increased convective heat transfer coefficients, up to 7 W/m2K compared to more conventional values of around 3 W/m2K (D2), produces an improvement in the predictions of ESP.

The default convective heat transfer algorithms in ESP are those of Alamdari and Hammond. The use of the increased coefficients has been justified by the suggestion that the test rooms are not 'typical' of real buildings, and that as a result heat transfer coefficients are much higher. Initially this argument is hard to understand. The rooms have realistic surface heat flows, realistic surface finishes, and are heated by a realistic heat source. The principal senses in which they are unrealistic is their size, and the lack of any ventilation.

The convective heat transfer from a vertical surface (for example a wall) depends only on the height of the surface (D3), and the height of the test rooms is the same as that of a real room. The heat transfer coefficient for a flat plate (for example the test room floor) heated to a temperature T above that of the room air can be approximated by [HOLMAN]:

$$h = 1.32\left[\frac{\Delta T}{L}\right]^{\frac{1}{4}}$$

where L is the length of the surface. Thus for smaller installations the heat transfer coefficient would indeed be expected to increase. However, this dependence is very weak: doubling the dimensions of the test room would suffice to yield a realistic room size, but would decrease the heat transfer coefficient on the floor by less than 20%.

In practice the heat transfer within an enclosure is much more complex than suggested by simple equations of the above type, as complex flow patterns result from the interactions between the surfaces of the room. However this very simple analysis does suggest that the dependence of convective coefficients on room size is unlikely to be as large as suggested.

Some algorithms take account of this dependence on surface dimensions. In particular, the Alamdari- Hammond correlations includes the height or length of the surface, and this effect should therefore have been automatically accounted for in the original simulations carried out with the ESP default algorithm.

The fact that the test rooms are tightly sealed, unventilated and unoccupied is likely to lead to lower convective heat transfer rates. In a 'real' building infiltration, ventilation and the movement of occupants will produce a certain amount of additional forced convection. Thus ventilation and occupancy considerations would tend to lead to the conclusion that lower coefficients should be used, not higher.

A third effect may provide the explanation of why the use of higher than normal heat transfer coefficients is plausible. This is the interaction of the heat source with the room air and fabric. The heat source is of a realistic size, and is located in a small room. It may be that this serves to exaggerate the effect of the plume of air rising from the heater on local heat transfer coefficients.

Recent experiments carried out in another test room at the EMC site (D4) suggest that neither the presence of ventilation or exaggerated heater interaction can justify the increase of heat transfer coefficients which has been suggested. In that work convective coefficients were estimated from measurements of temperature profile within the boundary layers at the room surfaces. That test room was ventilated at 2 ac/h, and was heated by a fan assisted convector, both of which would be expected to increase the heat transfer coefficients above the values observed in the rooms used for the IEA work. An increase in internal convective coefficients was observed when the heater was in operation, but this served only to bring the initially rather low coefficients into line with Alamdari- Hammond values.

D33 Internal convection: air movement within the rooms

Without exception, the models used in this study all employed a single air temperature, that is they effectively assumed that the air within the rooms was well mixed. Only three of the models which took part in the exercise had the option to use a more sophisticated air model: ENERGY2v 1.10 and HTB2v1.2 offer a stratified model, and CLIM2000v1.1 offers a simplified distribution model.

It has been suggested within the IEA group that there may be a body of very slow moving air in the middle of the room, with a boundary layer circulating at a much higher speed. This hypothesis could explain the time delays which have been observed between the coupling of the heater to the air temperature in the test room. It may also go some way towards explaining the general under-prediction of energy consumption, since the outer loop of air may become overheated whilst waiting for the thermostat temperature to respond.

The issue was presented by DMU at the IEA meeting in Fontainbleau, and as a result other modellers (Appx.F6 and F8) have re-stated it in their Model User Reports.

Subsequent CFD calculations carried out at De Montfort University (using Flow-3D) have shown that for heated periods the stagnant core is less obvious. Figure D1 shows the resulting flow pattern and air temperature distribution when the room heater is operating. Internal long-wave radiation has not been modelled. The TASv7.54 user suggested (Appx.F3), following C11) analysis, that the still core, moving boundary pattern occurs during the unheated periods only.

The question which remains unanswered is the extent to which the effect of air temperature distribution is reduced in real buildings, where infiltration, ventilation and possibly even occupants stir up the room air. At first it seems that this wide variety of effects might serve to eliminate the effect entirely. However, there are many air temperature measurements carried out in 'real' buildings which indicate that levels of, for example vertical stratification, are sometimes of the order 1 to 3°C: at least as large as those observed in the test rooms.

References

(Dl) Martin, C., Detailed Model Comparisons: an empirical validation study using SERI-RES, ETSU Report 1197-I, 1991, 141pp.

(D2) CIBSE Guide, Section A3, The Charted Institute of Building Services Engineers, London, 1986.

(D3) J P Holman, Heat Transfer, McGraw Hill Book Company, 1986.

(D4) S R Delaforce, E R Hitchin and D M T Watson, Convective Heat Transfer at Internal Surfaces, Building and Environment, Vol 28, No. 2, p 211-220, 1993.

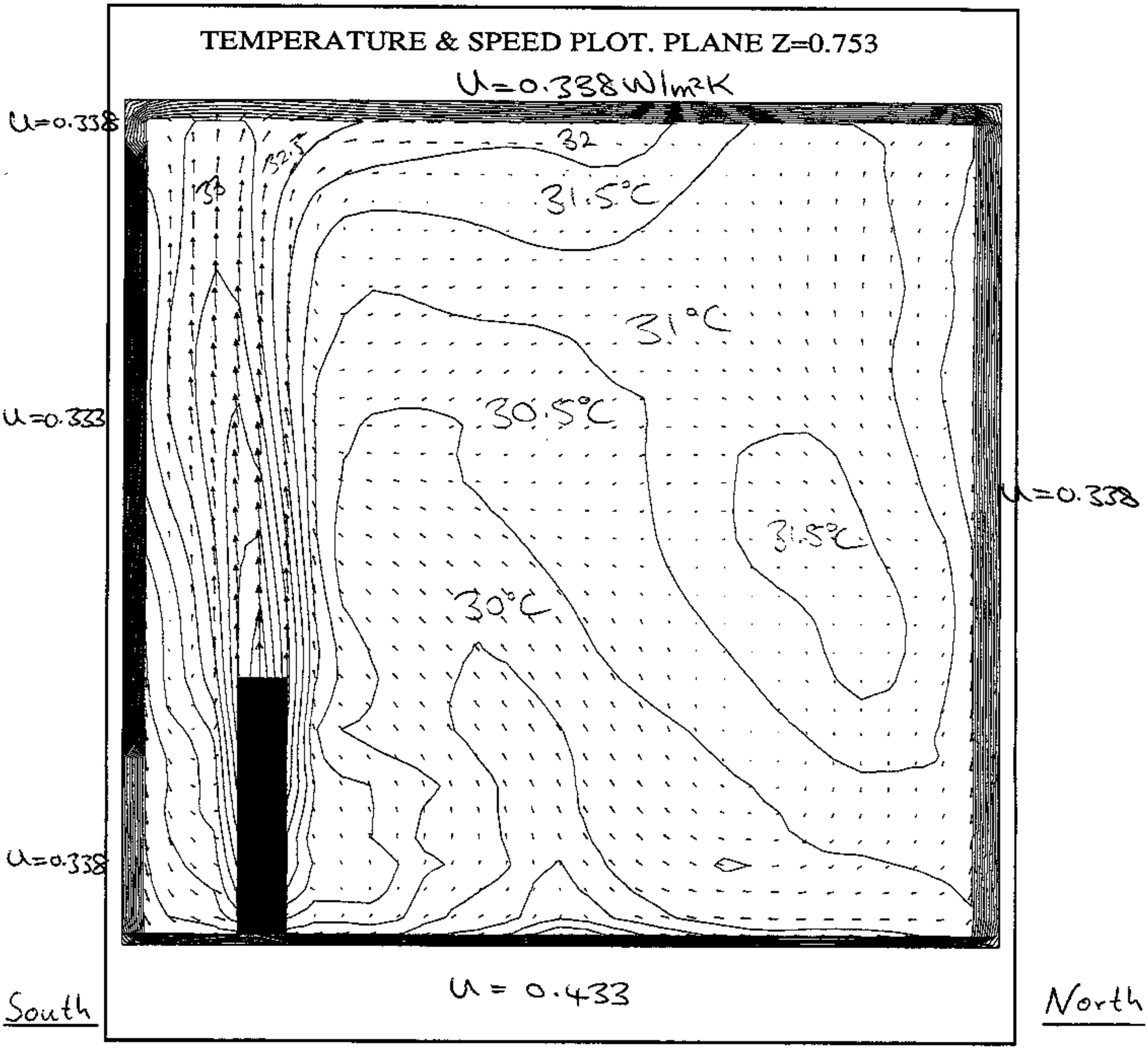


Figure D1a: Vertical mid-room section showing predicted airflows and temperature distribution

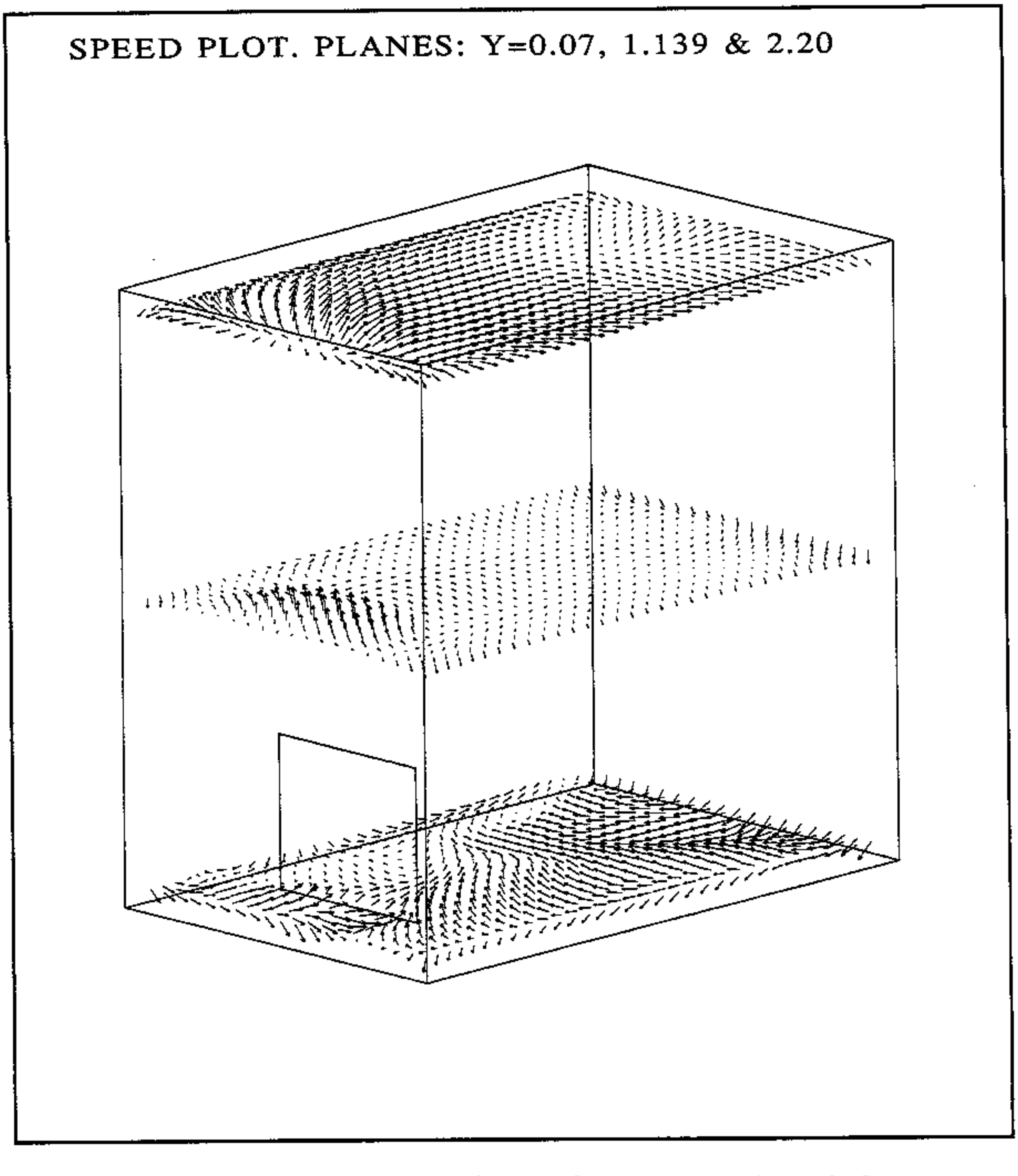


Figure D1b: Predicted air speed at various horizontal planes through the room

THE UA-VALUE OF AN EMC TEST ROOM:

Comparison between measured and calculated values

El INTRODUCTION

The Energy Monitoring Company (EMC) test rooms have been used in a variety of applications: investigating the physical processes which control the thermal performance of buildings (El), testing new building components (E2), and generating data for empirical model validation (E3). In this latter application it is necessary to supply the simulation model user with a detailed description of the construction and material properties of the test rooms. For the EMC rooms this takes the form of a site handbook (E4). If the conclusions of such validation work are to have credibility it is clearly of the utmost importance that the description in that handbook is as close as possible to the physical reality of the rooms and their instrumentation.

A wide range of checks and tests has been implemented to ensure that this objective is consistently achieved (E5). This short note describes a further check. The specific heat loss (or UA-value) of a test room envelope is determined from an experiment in which a room without a window was held at a constant temperature for a ten day period. The resulting estimate of the UA-value is then compared with the value calculated from the information contained in the site handbook.

E2 THE DATA SET

During November 1988 a ten day data set was collected from the EMC test rooms, during which the rooms were continuously heated to an air temperature of 30°C. This dataset was collected immediately after the dataset featuring intermittent heater operation which was used in the IEA validation exercise. One of the rooms was equipped with an opaque infill panel in place of the glazing. Figure El shows the temperatures inside and outside the test room, and the power consumed by the heater.

Figure E2 shows the global and diffuse horizontal radiation measured during the trial, and reveals that this was a heavily overcast period. This makes the dataset ideal for determining the UA-value of the test room fabric, as the amount of solar radiation falling on the external surfaces is small. Furthermore, the overcast conditions imply sky temperatures which are relatively close to the measured site air temperature.

E3 DATA ANALYSIS AND RESULTS

E3.1 Temperature drop across the room fabric

The mean external and internal air temperatures throughout the test are easily found from the recorded data. The internal air temperature is calculated as the spatial average of the three temperatures recorded inside the test room. The results of these calculations are a mean external air temperature of 7.80°C over the ten day trial, and a mean internal air temperature of 29.67°C, giving a mean temperature difference across the building fabric of 21.87°C.

E3.2 Electrical energy delivered to the room

The total energy delivered to the room by the electric heater is likewise simple to calculate, and for the total ten day experiment it is 57.68 kWh, or an average power input of 240.3 W.

E3.3 Solar radiation absorbed on the external faces of the room

Although the room analysed is not equipped with a window it is still expected to have a small response to solar radiation, due to absorption at the external surfaces. Only the front and rear surfaces will be considered here, since the external side wall of the room is extensively shaded by the adjacent room, and the room ceiling is shielded from solar radiation by the vented roofspace.

The solar radiation on the south wall of the room was measured as part of the experiment, and the total over the ten day trial was 13.2 kWh/m^2, an average flux of 54.9 W/m^2. The radiation on the north face of the rooms was not measured, but at this time of the year no direct radiation falls on this surface. The total incident radiation can thus be estimated from the measured diffuse. The vertical north wall has a view factor of the sky of 0.5. The corresponding view factor to the ground, which has an assumed reflectance of 0.2, is also 0.5 and thus a simple estimate of the diffuse radiation on the north wall is 0.6 times the diffuse radiation on the horizontal. Over the ten day trial this amounts to 3.5 kWh/m^2 or 14.8 W/m^2.

A proportion of the shortwave radiation incident on the external faces of the test rooms is reflected back. The remainder is transferred partly to the external air, and partly to the interior of the room. The proportion transferred into the room is given by the ratio of the external film resistance to the combination of the wall resistance and the internal film coefficient. In practice the different constructional elements that make up the wall have different resistances, but a reasonable estimate can be obtained using an appropriately area weighted mean value. The total energy transferred to the room by solar radiation incident on the external surfaces is therefore given by:

$$\frac{S \times A \times (1-q) \times r_{external}}{(r_{wall} + r_{internal})}$$

where:

S is the incident radiation on the surface
A is the area of the surface (equal to 3.43 m^2 for the north and south walls)
q is the surface reflectivity (equal to 0.84 for the EMC test rooms)
$r_{external}$ is the external film resistance (here taken as 0.06 m^{2}K/W)
r_{wall} is the wall resistance (1.8 m^{2}K/W for the north wall and 1.2 m^{2}K/W for the south), and
$r_{internal}$ is the internal film resistance (here taken as 0.12 m^{2}K/W).

This expression evaluates to 0.3 W for the north wall, and 1.4 W for the south wall, a total transfer to the test room of 1.7 W. This represents a very small fraction of the room energy balance, and the effect of the significant approximations in the above calculation on the calculated U-value is therefore negligible.

E3.4 Longwave radiation exchange between the external faces of the room and the sky

In practice the external temperature which the test room experiences is not exactly equal to the external air temperature, as some energy is lost directly to the sky by radiation from the external surfaces of the building. Once again, we will consider only the north and south faces of the room. These have a view factor of 0.5 to the surrounding ground, which we assume to be at the same temperature as the ambient air, a reasonable assumption for grass. They also have a view factor of 0.5 to the sky, the temperature of which was not measured. The heat transfer from these surfaces is therefore to a temperature which is a combination of ambient air and sky temperatures: all the convective and half the radiant transfer being to ambient, and the remainder of the radiative transfer being to the sky temperature.

Thus the effective temperature with which the room is exchanging energy can be expressed as:

$$T_{external} = \frac{(h_{conv} + \frac{1}{2}h_{rad})T_{ambient} + \frac{1}{2}h_{rad}T_{sky}}{(h_{rad} + h_{conv})}$$

where:

T_{amb} is the ambient air temperature
T_{sky} is the sky temperature
h_{conv} is the convective film coefficient at the external surfaces (here taken as 12 W/m²K)
h_{rad} is the radiative coefficient (equal to 4.5 W/m²K at these temperatures and emissivities)

using these values the effective temperature to which these two surfaces are exposed evaluates to:

$$0.87\ T_{amb} + 0.13\ T_{sky}$$

However, these two walls represent only 36% of the total area through which the test room is losing heat. The remaining surfaces exchange heat only with the ambient air temperature. Thus the effective temperature seen by the whole room is:

$$0.95\ T_{amb} + 0.05\ T_{sky}$$

This can alternatively be expressed in terms of the sky temperature depression, that is the difference between sky temperature and ambient air temperature:

$$T_{amb} - 0.05\ D_{sky}$$

Mean sky depressions are typically of the order 10°C, and this has been confirmed by measurements made over other periods at the EMC site. However, over the course of this experiment the weather was predominantly overcast, and therefore the sky temperature is likely to be closer to ambient air temperature. It is proposed that a depression of 5°C is assumed, and that the uncertainty in this value is taken to be ±5°C. This yields an effective external temperature of 7.55 ± 0.25°C.

E3.5 Calculation of the room UA-value

The room UA-value is calculated by dividing the mean energy input to the room by the mean internal external temperature difference:

$$UA = \frac{Q_{heater} + Q_{solar}}{T_{internal} - T_{external}}$$

Inserting the figures derived above in the previous sections into this expression yields a UA-value of 10.94 W/K.

E3.6 Analysis of systematic errors

The uncertainty in the measurement of the power delivered to the test room is ±2% (E4).

The uncertainty in the measurement of the internal and external temperatures used in the above calculation is ±0.2°C (E4). The uncertainty in the temperature drop measured across the room is therefore ±0.4°C or a further ±2%. An additional uncertainty is added to this by the estimation of sky temperature in order to estimate the effective external temperature. As discussed in Section 3.5 this is estimated to add a further ±0.25°C or ±1%.

We conclude that the overall systematic error in the determination of test room UA-value is ±5%.

E3.7 Analysis of random errors

As well as the systematic errors described above there is likely to be a random error component in the derived UA-value: that is if the experiment was repeated many times the results would show a degree of random variation. The most obvious way to determine the magnitude of this variation would be to repeat the experiment many times and observe the variations directly. However, as only one dataset is available this approach is not feasible here.

Instead the ten day dataset described above has been analysed on a daily basis. Figure E3 shows the result of plotting daily mean heater power against mean internal-external temperature drop. This very simple analysis takes no account of the dynamics of the test room, and can thus be considered to provide an upper limit on the random error. Standard regression gives the slope of the line shown on Figure E3 as 10.94 W/K, the same as the result obtained above, with a 95% confidence interval of ±1.2%.

This random error is statistically independent of the systematic error derived in the previous section, and it is combined with that error by addition in quadrature, that is:

$$e_{total} = \sqrt{e_{systematic}^{2} + e_{random}^{2}} = \sqrt{5^2 + 1.2^2} = 5.1\%$$

E4 DISCUSSION AND CONCLUSIONS

The UA-value of an EMC test room equipped with an opaque infill panel has been determined experimentally, and found to be 10.94 W/K ±5%, that is it lies between 10.39 and 11.49 WIK. The value calculated from the constructional details given in the description supplied to modellers and standard values for internal and external surface coefficients and airgap resistances (E6) is 10.13 W/K, and when the estimated uncertainties in the properties supplied are taken into account the uncertainty in this calculated UA-value is found to be -7/+8%, that is the value lies between 9.42 and 10.94 W/K. It is clear that these error bars overlap by a large margin, and that it is impossible to confirm from these results that there is any discrepancy between the measured and calculated values. However, there is a difference of 7% in the two values, and it is interesting to speculate what might be causing this.

There are two physical effects which are known to be present in the test room which may account for the observed discrepancy:

- The first relates to the variation of material properties with temperature. The quoted book values are normally measured between temperatures of zero and 20°C, a mean material temperature of 10°C. Because the room was continuously heated to a relatively high setpoint during the test analysed above the mean material temperatures are rather higher than this. During the test the internal temperature was 29.7°C and the mean external temperature was approximately 8°C, giving a mean material temperature of approximately 19°C. The k-value of the insulating materials used in the test room can be expected to increase by up to 8% for every 10°C increase in material operating temperature (E7). Clearly this will not affect all of the materials used in the test room. Re-calculation of the room UA-value with the insulation k-values increased in line with the results in (E6) yields a UA-value of 10.55 W/K. This effect alone brings the calculated value within 3.6% of the value measured at an internal temperature of 30°C.

- The second possible effect concerns the treatment of the three dimensional heat conduction at the edges and corners of the rooms. This additional heat loss was estimated using information from a limited number of two dimensional heat transfer calculations, and the material properties specified for the corner constructions in the site handbook were modified accordingly. In the estimation process, which predicted a 7% increase in heat loss due to edge effects, a number of edges at which the three dimensional effects were believed to be small were ignored. The corrections given are therefore slight underestimates of the overall extra heat loss due to edge effects. This is acknowledged in the documentation describing those corrections. It is considered quite possible that this underestimation of edge losses could account for part, or even all, of the remaining 4% discrepancy between the calculated and measured values. However the uncertainty bands in the calculated and measured UA-values overlap, and it is therefore not possible to draw firm conclusions about the magnitude of the underestimation of the edge effects.

A further effect which has been proposed to account for observed test room heat losses slightly higher than the calculated values is air movement around and through the glass fibre quilt used in the rooms. Other studies have failed to detect this effect in a hot box (E8), and from the above discussion it seems that it is also unlikely to be occurring in the EMC test rooms.

REFERENCES

El TEST CELL STUDIES 1: Solar Distribution. C J Martin and D M J Watson. November 1986. EMC Report to ETSU.

E2 A Field Test of the Performance of an Opaque wall clad with Transparent Insulation Material. C J Martin and D M J Watson. December 1990. EMC Report to ETSU S1197-P8.

E3 Detailed Model Comparisons: An Empirical Validation Exercise using SERI-RES. C J Martin. June 1991. EMC Report to ETSU S1197-P9.

E4 Site Handbook: Empirical Validation Data Sets 099 and 110 from the EMC Test Rooms. K J Lomas and C J Martin. March 1992. IEA Report number IEA21RN197/92.

ES High Quality Data for Empirical Model Validation: Quality Assurance at the EMC Test Room Site. C J Martin and D M J Watson. March 1993. IEA Report number: IEA21RN330/93.

E6 Thermal properties of building structures. Chartered Institute of Building Services Engineers. A3.

E7 PASSYS II - Final Report of the UK Consortium. P Baker et al. December 1992. ETSU Report S/DS/00107/REP

E8 Effects of Air Movement on Performances of Loft Insulants. A G Guy and J A Nixon. January 1990. Building Technical File. Number 28. p 39-45.

Determination of the UA-value of opaque test room

Conditions in opaque test room

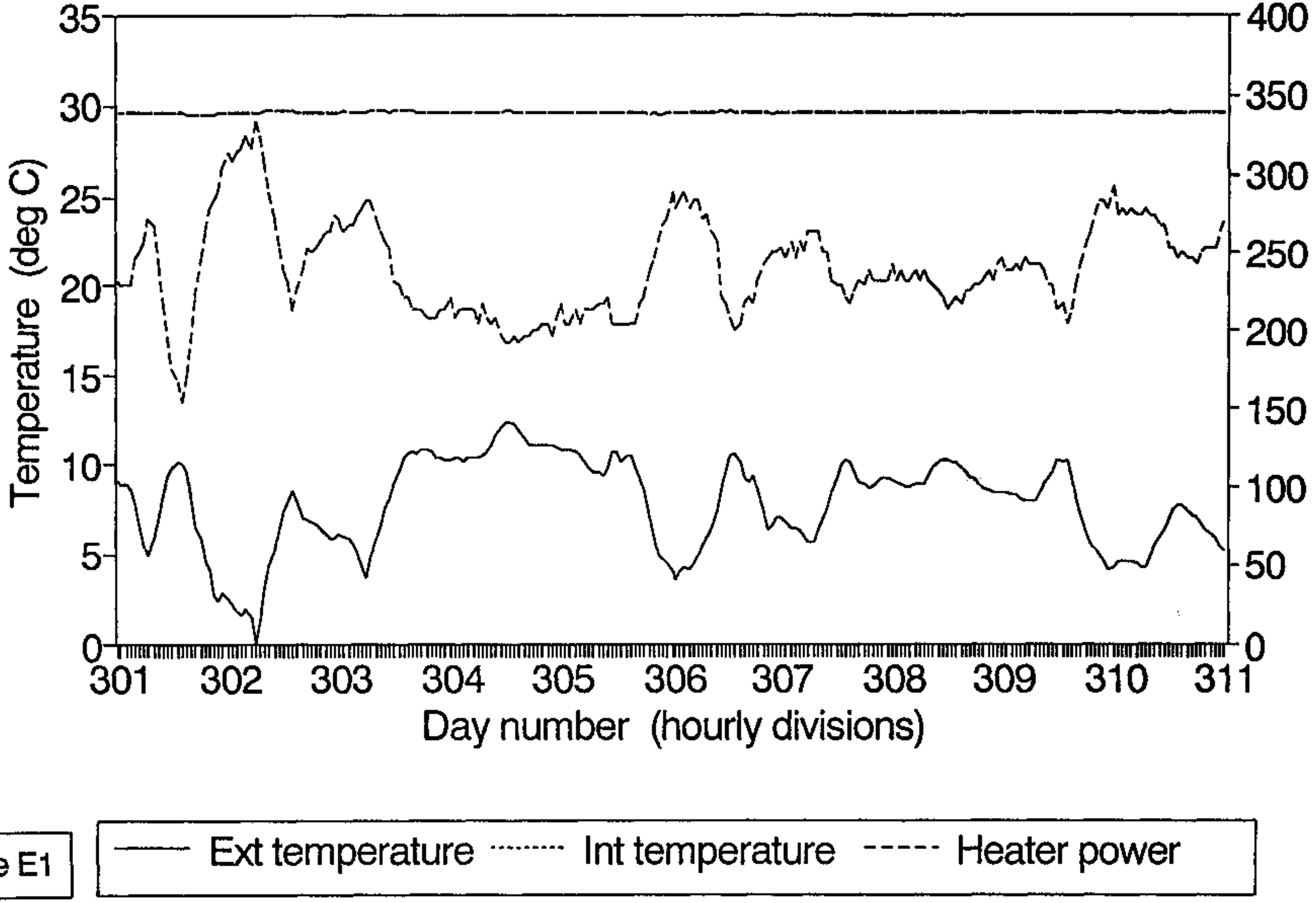


Figure E1

Determination of the UA-value of opaque test room

Solar radiation during trial

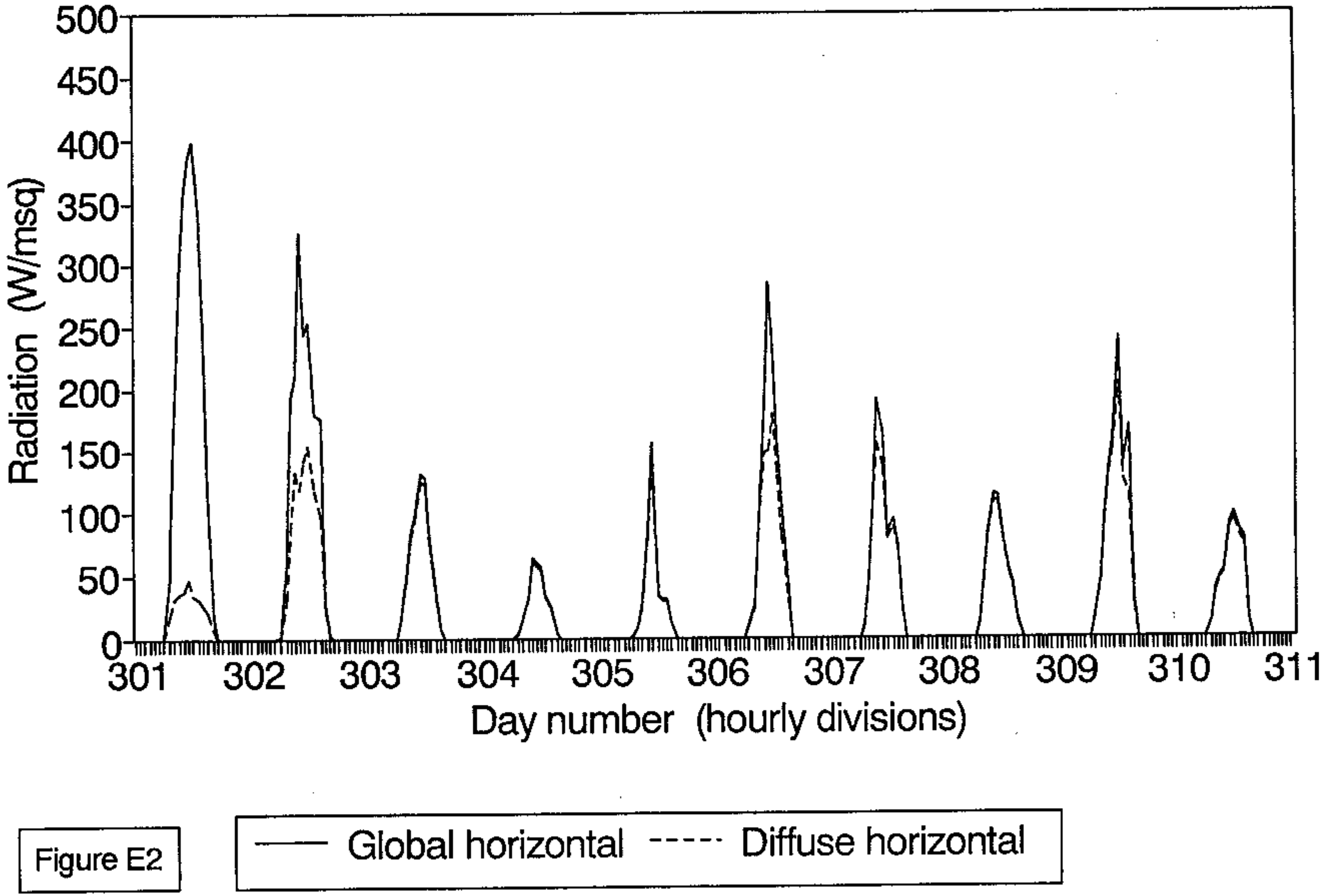


Figure E2

Determination of the UA-value of opaque test room

Analysis of daily average data

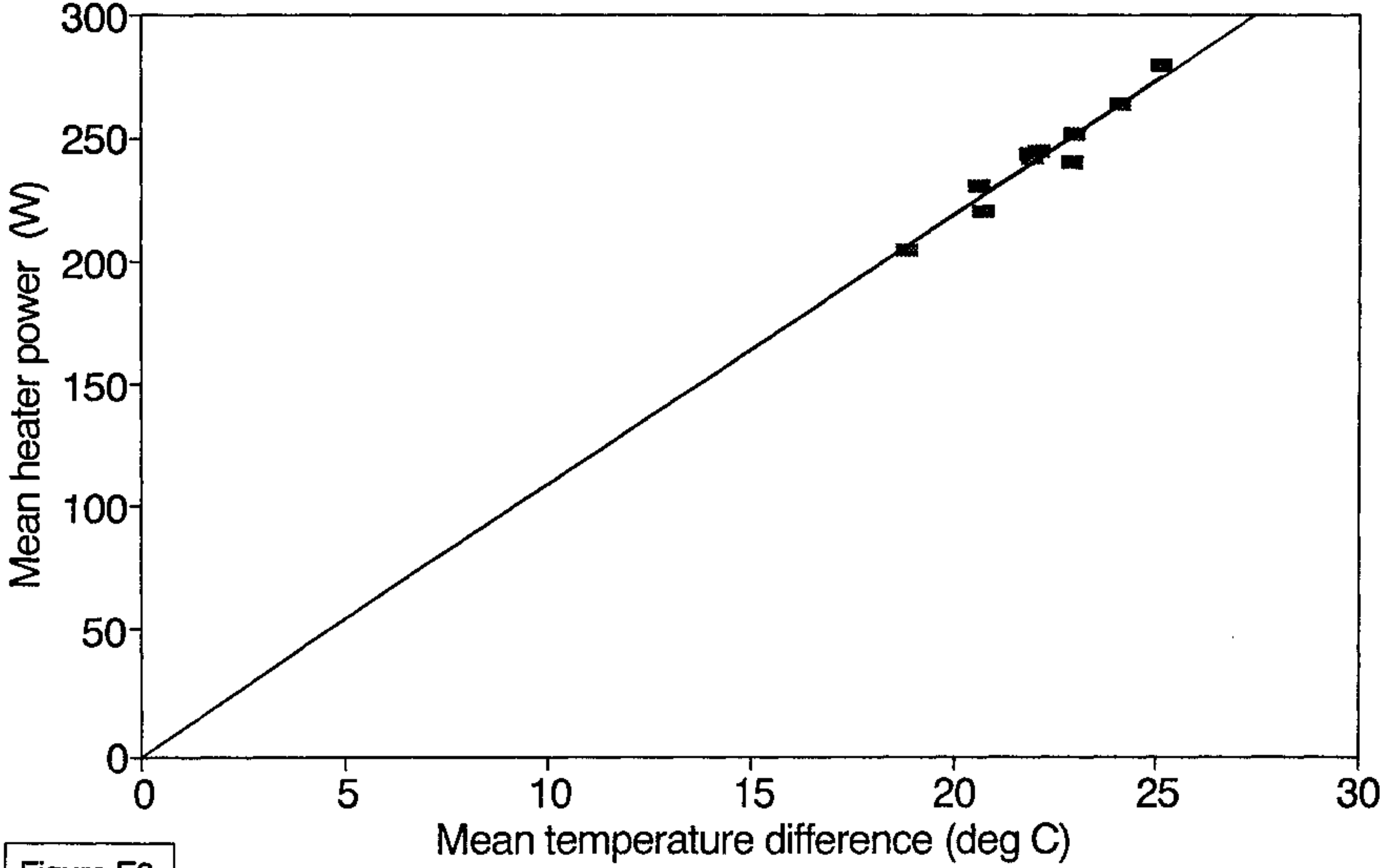


Figure E3

Appendix F - Model User Reports

Model User Reports were submitted by 11 participants. They are reproduced here as submitted, without any modifications. They are ordered in the sequence used for plotting throughout this report.

Fl	TSBI3v2.0
F2	DOE2v1E
F3	TASv7.54
F4	HTB2v1.2
F5	DEROBv1th
F6	BLASTv3.01v1203
F7	TASEv3.0
F8	TRNSYSv13.1
F9	SERI-RESv1.2
F10	ESP+v2.1
F11	ESP-rv7.7a

All participants were asked to submit no more than 3 pages. A small number slightly over-ran this limit.

Appendix Fl: TSBI3v2.0 - K.Johnsen - SBI, DK

Background for participation

The Danish Building Research Institute (SBI) participates in the IEA Task 12: Building Energy Analysis and Design Tools for Solar Applications, however, not in the model evaluation sub-task (B). For some unclear reasons we were only asked to participate in the Empirical Validation at a very late stage of the whole exercise, and with very short notice. Therefore we only had very limited time and within one week in December 1992 the six data models were set up and all runs made, i.e. in about 40 man-hours.

Representation of test room models in *tsbi3*

The very detailed descriptions of the test rooms were, however, extremely helpful and we only had to ask a few questions to the hot-line service for clarification of minor ambiguities, of which the most important was about inconsistencies in the solar parameters of the weather data files.

There was no problems in setting up the data models, in fact the way the room geometry and the constructions were described matched perfectly to the data structure of *tsbi3.* After having seen the measured results we would only make one change in the building defmition: We would model the floorspace as a separate zone, in stead of assuming that the floor faces the outdoor air (through some air-gap resistance).

Calculations and a few mistakes in the *tsbi3* runs.

As mentioned the *tsbi3* runs were made in short time, and during the short period of input data checking we only found one minor error in the input data for one case (the radiative/convective fractions of heat from the electric heater).

During the interpretation of the calculation results and the comparison with the measured data we have, however, discovered two more severe input mistakes (1. and 2.).

1. In the simulation of room 3 (case HO, Volume 110) the schedule for the heater was wrong, so that the heater was only ON during the first 6 days and OFF on the last day.
2. For all the 6 cases the default value of skyline (cut-off angle for direct radiation at low solar heights) of *7 degrees* was used, in stead of the specified *free* horizon. This was in fact a minor bug in the program, which made it somewhat tricky for the user to 'get rid of' the standard value.

From the comparisons of the solar radiation values in the weather data files, the measured solar radiation on the south wall, and the calculation results, it has furthermore become clear that there are some inconsistencies in the measured solar radiation values:

3. For the May data the best agreement between the weather data and the measured radiation on vertical south wall is achieved when using the data for *global and diffuse* radiation, while for the October data the best agreement is achieved when using the *direct normal and diffuse* radiation. Since the values for global and diffuse radiation originally were used to create the weather data file for *tsbi3* we investigated the importance of using the normal radiation data instead for the heated cases (documented below).

All of the three observations (1., 2. and 3.) have a significant impact on the results for the heated cases, while the default horizon skyline is almost insignificant for the free floating cases in May (because of the sun's position in East at sunrise). However, new runs were made for all six cases and the fmal results were send to the Empirical Validation Hotline.

Conclusions

Investigations of the deviations in temperatures, energy consumption, and solar irradiance, uncovered two severe input mistakes and some inconsistensies is the delivered weather data. After correction of these mistakes new simulations were performed for the three heating cases.

The final results of all the calculations show exellent agreement with the measurements, and in all the comparisons, set up by De Monfort University, *tsbi3* seems to be among the few programs that gives results within the expected error band for solar irradiance as well as energy consumption and temperatures in the heated cases.

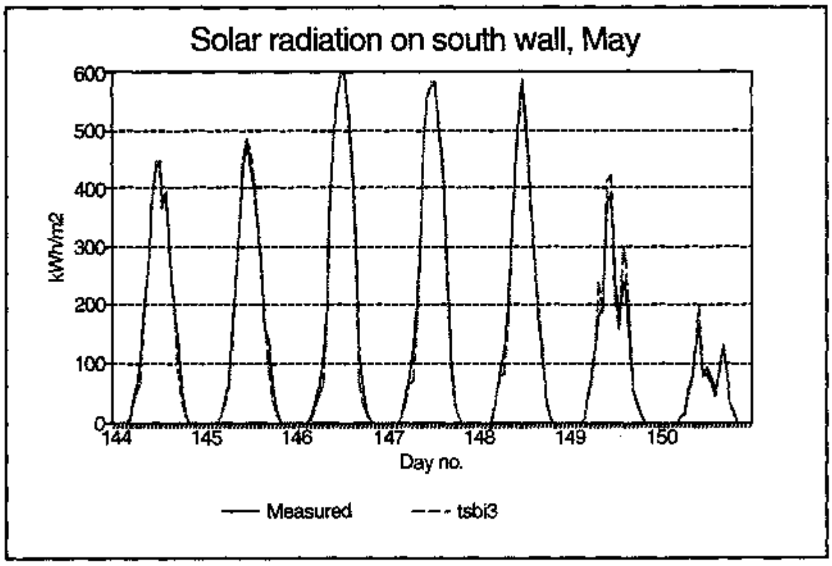
Discrepancies in the climate solar data and/or measured solar radiation?

tsbi3 uses normal and diffuse solar radiation as input. By our standard program for weather data conversion we converted the data for global and diffuse radiation to normal radiation. This seems to indicate some problems with the measured/calculated data for normal radiation, especially at sunrise hours. For several days of the May period the weather data for radiation are like the following:

Day no.	Hour no.	Global on horizontal	Diffuse	Normal	Our Normal Rad
144, 145	5	21	12	0 !	111, 125

If the global radiation on horizontal is greater than the diffuse radiation there must be some direct radiation and, according to our model, in these cases more than 100 W/m^2. Therefore the conclusion is that the data for direct, normal radiation must be wrong (too low).

We were recommended to use our own values for the normal radiation, i.e. based on global and diffuse radiation. So we did.



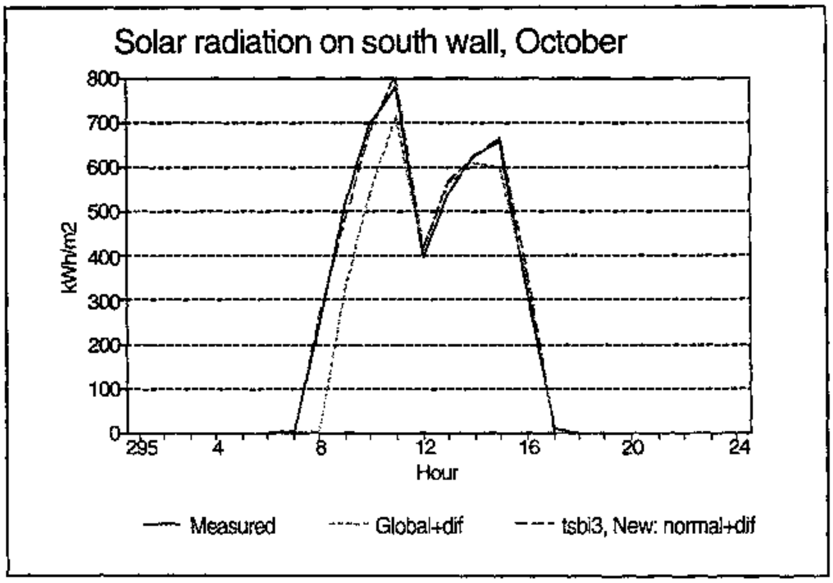
The figure shows the *tsbi3* calculations compared to the measured values for the solar radiation on the south wall. The differencies are very small, although there seem to be some minor deviations at days or hours with partly clouded sky (e.g. day 149).

For the solar data in the October period the problem seemed to be the opposite, i.e. to high values in the first hours after sunrise:

Day no.	Hour no.	Global, horiz	Diffuse	Normal	Our Normal Rad	Measured on South
294, 295	8	112, 99	35, 11	614, 732	281, 267	229, 240

In hour 8 the Solar - SouthWall azimuth is aroud 72 deg. and the solar radiation on the wall will on clear days be around 1/3 of the direct normal radiation (cos 72° = 0.31). It is therefore obvious that the data for global radiation (as we used in the original runs) are wrong, or at least inconsistent with the measured irradiation values. The weather data for the normal radiation are, however, in good agrrement with the measured irradiances on the vertical south wall, and therefore we have to assume that these data are 'correct'.

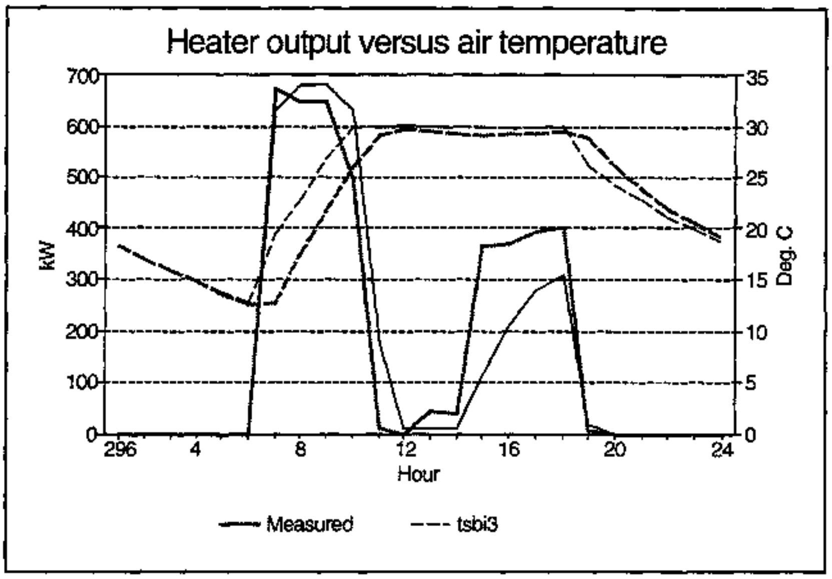
The calculation results (or the solar radiation models) of *tsbi3* seem to confirm this assumption.



The figure shows the original and the new calculations of solar radiation on the south wall for day 295 i October. In the new calculation the new (correct or at least consistent) weather data are used, and a free horizon is assumed.

Function of sensor and thermostat of the electric heater

In the hourly values the most important reason for differences between measured and predicted temperatures and heating demands is the lack of models for the dynamics of the oil-filled electric heater and it's control system. *tsbi3*, as well as all other programs, predict a much faster temperature rise than measured, and a longer heating-up period because of the assumption of an 'ideal' control which will keep the power on, until the set-point is actually reached. In reality the 'accelerator' of the thermostat will 'fool' the heater and make it switch *off* long time before the set-point is reached (for the test-cases typically at 25-26 °C, see figure). For the heating-up period the specifications of the controller seem to match the thermal mass of the heater (time constant 22 mins.) quite well, and explain the differences, but it is difficult to explain the afternoon heating period where the heater seems to be *on* for too long time and with to much power. The steep rise of the curve is explainable but it is strange that the set-point is never really reached, although the power is on for 4-5 hours. Probably the interplay between the dynamics of the heater and of the room, and the unspecified local heating of the wall behind the heater could be some reasons for the actual performance, but we would welcome any physical explanation for the experienced performance.



Sensitivity of solar transmission and distribution

The maximum air temperatures in the free floating cases of the glazed rooms is somewhat on the low side compared to the measured values. Since there is extremely good agreement between calculated and measured solar irradiance on the south wall, the following explanations for the low maximum temperatures could be given:

a) bad model for solar transmittance, i.e. too small transmittance of solar energy through the glazing,
b) poor model for the thermal dynamics of the rooms and the constructions, and
c) wrong assumption on the split of the solar energy: lost / to air / to surfaces

The uncertainty on c) is probably much bigger than on a) but both of these problems deals with the actual transmission of solar energy to the room. We have therefore performed a study of the sensitivity of the maximum air temperature to the solar transmission.

For both cases (single and double glazing) we have found that:

> 1 % increase of the solar energy increases the maximum temperature by 0.17 K. For example if the 'lost' fraction is only 10 % in stead of 20 %, the maximum temperature will be 1.7 K higher.

We also investigated variations of the solar energy fraction to the room air, and this shows little sensitivity:

> 5 % increase in the fraction to the air increases the maximum temperature by 0.1 K, i.e. a change of the fraction to the room air from 5 % to 20 % (default value in *tsbi3)* will only give an increase of about 0.3 K.

Other reasons for the deviation in maximum temperatures could be that the solar data of the weather files contains unreliable (high) values and that the solar absorption on the exterior envelope is unrealistic low (0.16).

Other sources of uncertainties

Mainly the description and the actual performance of the floorspace and the roofspace.

Appendix F2: DOE-2.1E - F.Winkelmann - LBL, USA

1. **Problems encountered in representing the test rooms within the model.**
   See 6 below.

2. **Problems encountered with the documentation provided.**
   None.

3. **How useful was the hotline?**
   Very.

4. **How useful were the newssheets?**
   Very.

5. **How was Quality Assurance organised?**
   The information in the DOE-2 input verification reports was cross-checked with the building data from the Site Handbook. The DrawBLD program, which plots a 3-D view of the building, was used to check the input geometry.

6. **Results and conclusions from sensitivity studies.**
   The major problem with DOE-2 relative to this exercise is that the heat from the radiator is assumed to be 100% convective. The result is (1) no time delay between heat output from the radiator and corresponding increase in air temperature, and (2) the room surfaces do not absorb IR radiation from the radiator so the resulting conduction loss to ambient from these surfaces does not take place - which leads to an under-estimation of radiator output needed to maintain 30°C inside air temperature. This effect is primarily responsible for DOE- 2.1E's 17-25% under-prediction of heating energy.
   A sensitivity analysis was performed in which room weighting factors were applied in DOE-2 to the convective-plus-radiative heat output from the radiator. (These weighting factors are similar to those used in DOE-2 to calculate loads from instantaneous heat gains from solar, lights, people, plug loads, etc.). With these weighting factors, the DOE-2 heating energy results are within 5% of the measurements for Room 3 - Opaque and Room 1 - Double Glazed. The resulting air temperatures agreed quite well with the measured air temperatures.
   It has turned out to be more effort than we anticipated to incorporate an improved radiator model in DOE-2 that accounts for the radiative/convective split, time constant, PID control, etc. for the radiator that was used. Radiators are rare in the US, which is why DOE-2 is so deficient in this area. We probably will put a better model in at some future date or perhaps some European user can be convinced to do this.

7. **Were any bugs found in the model as a result of this exercise?**
   No.

8. **Further comments.**
   We wish to thank the organisers for doing such an excellent job on this project. It's the best of its kind that we know of. We were involved back in 1978 or so on the first IEA inter-program comparison (IEA Annex 1) and can appreciate the care and effort that you have put in. We saw factors of two and three differences in the various programs at that time, so it is good to see that the programs are getting better - the programs overall are showing much better agreement in your exercise. Plus we have the advantage of your careful measurements; we wish we had had them 15 years ago. It is ironic, however, that after all this time no one seems to be able to model a "simple" radiator!

Appendix F3: TASv7.54 - M.Gough - EDSL, UK

1. Background

EDSL's involvement with the EMC datasets which form the basis of this study dates from 1990, when a blind validation trial on Tas was carried out by Chris Martin. This concluded: "Tas has been found capable of reproducing the performance of the test rooms well. In absolute terms it has performed better than either of the models previously tested against this data set. In terms of predicting the trends across different glazing options Tas performs about as well as those models." (Martin, C. A blind validation trial of the model Tas using data from the EMC passive solar test rooms. EMC, February 1991.) Since then, EDSL has used the EMC datasets to gain valuable insights into numerous aspects of modelling. The main results from this work are summarised below.

The IEA exercise set out to use the same datasets in a blind trial involving many more programs. EDSL accepted the invitation to participate, and owing to its earlier involvement with the dataset, agreed that the Tas simulations should be done by DMU. This meant that EDSL was not able to provide detailed input to the study at this stage.

Phase 1 of the IEA work seems to have concentrated on attempts to use the trial results to prove the existence of 'internal errors' in the participating programs and thus to cast doubt on their usefulness. This approach is of dubious value. In the first place, it is quite hard to prove the existence of internal errors from studies of this kind, and is questionable whether the IEA study has achieved this. But even if it had, it would tell us nothing new, since it is common knowledge that programs do indeed contain 'internal errors', in the sense that they work with assumptions and approximations. Finally, recent experience has shown that public statements which emphasize program errors can be misinterpreted by the press and do serious damage to the image of thermal simulation.

Setting aside these reservations about basic approach, there are weaknesses in the methodology. The argument central to Phase 1 rests on two assumptions: 1) that all input errors have been identified and eliminated, and 2) that the sensitivity analysis has correctly established reasonable error bands for the predicted quantities. These assumptions are open to doubt. Leaving aside the data input errors (later corrected) which were introduced by DMU when entering the building into Tas, the sources of error and uncertainty can be summarised thus:

1. Edge losses were underestimated in the data supplied to modellers, possibly by as much as a factor of two.
2. Various details of the test room were overlooked. For example no mention is made of the wooden battens to which the plasterboard walls are fixed. These battens cause significant thermal bridging around the door, and where they intersect with the structural woodwork. A 4mm hardboard sheet was omitted from the party wall description. The stated area of the concrete tiles was about 7% too high.
3. The roof and floor spaces are only sketchily described, apparently because they were incorrectly regarded as thermally unimportant. In the roof space significant elements of thermal mass such as joists were omitted, an insulating panel on the south face was described as plywood, and no mention was made of the fact that the partition wall is not of full height. Infiltration rates in the roof and floor spaces had to be guessed as there was no accurate measured data on flow rates or apertures.
4. Relative humidity data is missing from the weather data for the heated runs, and this gives rise to uncertainty in the prediction of sky radiation. This, together with other uncertainties in the weather data (eg. cloudiness), makes a significant contribution to simulation error which has nothing to do with 'internal errors'.

The lesson to be learned from this experience is that in studies of this kind it is unlikely that all the relevant issues will emerge in advance of the simulation runs, and the temptation to rush into print with hard statements about model validity (or invalidity) purely on the basis of blind trials should be resisted. The most productive part of any validation exercise is that in which one attempts to understand precisely what combination of input, measurement and modelling errors lies behind any observed discrepancies, and this can only be achieved with full visibility.

2 Results of Tas Studies using EMC datasets

2.0 Modelling Assumptions

The Tas model used by EDSL was based on the description given in the Site Handbook, but with suitable corrections. Edge losses were modelled by entering the correct width of the walls etc. in 3D-Tas, joists being included as a percentage of wall area, and wood conductivity corrections ignored. Other errors in the test room, roof space and floor space construction descriptions, described above, were also corrected.

As compared with the DMU Tas model, the following refinements were also introduced. The Tas external shading facility was used to model the effect of the neighbouring test cell, and the internal shading facility was used to track sun patches around the room interior. To avoid guesswork about the behaviour of the roof and floor spaces, these spaces were conditioned to reproduce the measured air temperatures there. The version of Tas used was the same as that used in Phase 1 by De Montfort University (Tas7.54, released April 1991) with one modification: Tas's clock was adjusted by half an hour to ensure consistency with the measured data.

For these studies EDSL used the complete EMC datasets. These cover longer periods than the IEA 21. sets, and give data on more variables. One discrepancy has been identified between data from the two sources: diffuse radiation data on the EMC dataset is up to 6% greater than that supplied by DMU, owing to the use of a different shadow band adjustment algorithm.

The EDSL runs used the version of the weather data based on the 'US' timing convention. The DMU runs, by contrast, used the 'UK' timing convention data: this introduced errors, particularly for the heated runs for which the lack of sub-hourly measurements made the conversion process inaccurate. In order to reduce the uncertainty arising from the lack of relative humidity data for the heated runs, RH data measured concurrently at a site near Bedford was obtained from the Met. Office and included in the Tas weather file. The mean measured RH over the 7 day heated period was 87.8%.

Simulation results based on this model have been submitted to DMU for the Phase 2 analysis and are described below for the four cases so far investigated. Accounts are also given of studies aimed at accounting for observed discrepancies by refining the thermal model in various ways. Space does not permit discussion of investigations into external convection, solar position, diffuse solar algorithms and the influence of the adjoining cell, or a series of tests confirming the correctness of the basic Tas heat transfer algorithms.

2.1 Unheated Opaque Case

This case shows a good overall match between the simulation and the measurements. The predicted inside air temperature is usually within 1K of the measured value, and the surface temperatures show even better agreement. This is over a 50 day period when the outside temperature varies between 3.7 and 23.5 C and which contained a good mix of sunny and cloudy days. The largest errors occur during the 7 day subset of this period used for the IEA study.

Within this encouraging picture, certain consistent patterns of discrepancy were, however, apparent, and these needed to be accounted for. EDSL's efforts to account for and reduce the 1K typical error are described below.

A noticeable feature was that the predicted temperatures tended to lie a little below the measured values. This is almost certainly due to somewhat low estimates of downward long-wave flux generated by Tas7.54's sky radiation model (which uses Brunt's formula modified by Boltz's cloud correction, as advocated by Kondrat'yev). Since the release of Tas7.54 EDSL has reviewed the literature on sky radiation and improved the Tas algorithm accordingly. The current release of Tas7.64 uses Swinbank's formula (generally regarded as an improvement on Brunt's), and further significant refinements based on more recent work will be incorporated in the next release.

A second effect revealed by the comparison of predicted and measured temperatures was a phase-lead effect in the simulated air temperature. Since the surface temperatures were on the whole in excellent agreement with the measurements, the investigation centred on convection processes.

The effect to be accounted for can be described as an unexpected lagging of measured air temperature behind mean surface temperature. On the basis of normal convection coefficients and the known thermal capacity of the air in the room, there should, in the absence of gains and ventilation, be a lag of no more than a few minutes between these two temperatures, yet the observed lag was often more than an hour.

The reason for this effect appears to be the development of a core of rather still air which becomes thermally isolated from convection currents circulating around it. This picture has emerged from simulations with the dynamic CFD program AMBIENS 2, using surface temperatures generated by Tas. A stable pattern develops in which air circulates round the cell in a thin boundary layer, but in the core air velocities are very low (typically of the order 1 mm/s, corresponding to perhaps one circuit of the cell per hour). Because the flow in the core is slow and circulatory, mixing is mainly by diffusion, and this sluggish process effectively introduces a large thermal resistance between the core air and the boundary layer.

To test this hypothesis, runs were done with reduced convection coefficients, and with suitably chosen values the simulation showed excellent agreement with measured air temperature. Care is needed in drawing wider inferences from this result, however, for three reasons. Firstly, the use of low convection coefficients does not properly represent the fairly vigorous convection which occurs in the boundary layer, and which transfers heat between walls. Secondly this may be an effect which is seen only in small rooms. Thirdly, it can be expected that any small disturbance from infiltration, occupants, gains, down-draughts etc., such as would occur in any occupied building, would disrupt the stable flow field and justify the standard 'stirred tank' approximation.

Despite these reservations, the low convection coefficient assumption seemed to represent well the behaviour of the unheated opaque cell, and improved the fit sufficiently for other, smaller effects to become apparent in the data. One of these was a tendency, on some mornings, for the inside temperature to remain near its minimum value for longer than Tas was predicting. Since this effect tended to follow clear nights, condensation was suspected as the cause. A modification of the code to model condensation and evaporation of moisture at the outer surfaces of the structure produced the desired effect: on cold, clear nights moisture condensed on the walls and roof of the test cell, and later (usually on the following morning) the moisture evaporated, causing a cooling effect which held the inside temperature down in the way observed in the measured data.

Air and surface temperatures were at this stage being matched to an accuracy typically in the region of 0.3K, with rare instances of errors in the range 0.5-1.0K. The remaining small discrepancies can at least in part be attributed to inadequate information about sky conditions. Tas estimates overnight cloudiness on the basis of conditions the previous day, and sometimes this assumption is wrong: a cloudy day may be followed by a clear night or vice versa. Instances where this type of error occurred were identified with the help of Met. Office records.

2.2 Unheated Double Glazed Case

The predicted performance for this case showed a level of agreement with measurement similar to that seen in the Unheated Opaque case, with an air temperature error typically less than 1K. With normal Tas convection relationships the phase lead effect is still apparent overnight, but is much less marked than in the opaque cell.

Investigations were again carried out to account for the various small discrepancies observed. The measured surface temperatures are somewhat puzzling at first sight. Compared to the predicted values they are often low (sometimes by as much as 3K) during the day. In fact the day-time measured surface temperatures (which are available for all surfaces except the south wall), are on the whole lower than the measured air temperature. This is surprising in view of the fact that the air must be being heated principally by convection from these surfaces. The apparent paradox is resolved if one assumes that the surface temperature sensors are mainly at sites shielded from direct (and perhaps to some extent from diffuse) sunlight, and on sunlit surfaces they therefore give a low measurement of average surface temperature. This hypothesis is supported by the observations that 1) the surface temperature discrepancy for the walls is greatest on clear days, and 2) on such days the floor sensor shows a sudden rise at about noon, when presumably the sensor begins to be directly illuminated by the sun.

A tendency to over-predict ceiling temperature by about 1K, notably on bright but cloudy days, can be put down to Tas's diffuse solar distribution algorithm, which in this room apportions a little too much radiation to this surface. Finally, one cause of a somewhat high prediction of surface temperature may be the omission from the simulation of a polystyrene screen designed to shield a temperature sensor, which diverts a proportion of the solar energy directly to the air.

2.3 Heated Opaque Case

This case produced a good match of energy consumption, both in terms of the total over the 7 day period (109.4 MJ, or 94% of measured) and the daily profiles. Surface temperatures, too, follow the measured curves well in general, though they tend to be about 1K low in the early part of the night and rise a little too abruptly in the morning. Air temperature shows the by now familiar phase-lead phenomenon. Because of this the minimum temperature during the 7 day period is lower than measured (12.9C as against 14.6C).

The phase-lead effect appears to have two main causes. Overnight the cell appears to settle into the 'isolated core' convection mode seen in the unheated case. Supporting this hypothesis is the observation that the air temperature match is much improved if convection coefficients are reduced when there is no heater input or solar gain. An anticipatory temperature rise seen at system switch-on is probably in part due to a temperature gradient between the front and back of the cell. This effect can thus be considered to some extent an error in measuring average room temperature. It is, however, also partly caused by an increase in wall convection driven by heater input, and encouraging results are being obtained from initial efforts at modelling this process.

A third source of error in the temperature prediction was the failure to model the dynamics of the heater. The introduction of a 22 minute time constant into the heater model caused it to continue to deliver heat for an hour or so after switch-off, which lifted both the air temperature and the surface temperatures towards the measured curves in the early hours of the night. The best fits of air and surface temperature, however, are obtained with a time constant of about twice this value.

2.4 Heated Double Glazed Case

Predicted energy consumption for this case over the 7 day period was 83.3 MJ, or 94% of the measured value. The minimum predicted temperature was 10.8C (as against 11.9C measured). The under-prediction of minimum temperature is attributable mainly to the 'isolated core' convection phenomenon, with the Tas7.54 sky radiation model also playing a role.

This case shows the expected combination of features. Surface temperatures, though a good match overnight, show the effects observed in the unheated case, and for the same reasons. Simulated air temperature anticipates the morning rise and the night-time fall in the expected way. The daily heat input profiles show all the main features of the measured data, although the rapidly varying solar gain points up some minor deficiencies in the simulation of the heater control function.

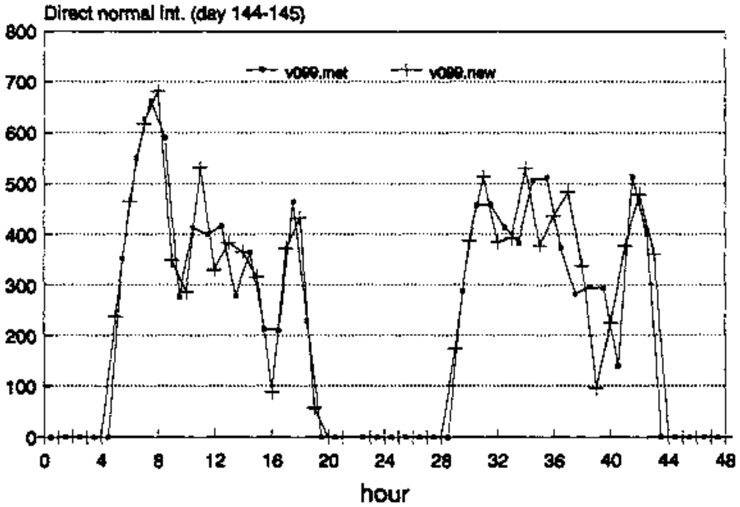
With a night-time convection correction, allowance for the dynamics of the heater and refinements to the sky radiation model, measured air and surface temperatures are reproduced well. Due probably to a stirring effect caused by a down-draught from the window, the double glazed cell requires a much smaller degree of correction for night-time convection conditions than the opaque cell - an encouraging result in view of the fact that windows are a feature of most occupied buildings.

Appendix F4: HTB2v1.2 - P.Pfrommer - FHT Stuttgart, D

There were no great problems in representing the test room within HTB2. The adiabatic effect of the west wall was achieved by placing a highly insulating layer behind the construction. However, there is no possibility in HTB2 to consider a "Solar to air" factor. The 5% radiation portion was assigned to the floor.

HTB2 had to be altered to work with the EMC weather data. Secondly, to get an appropriate data output on an hourly base (averaging of the output values on the time step base of 30 seconds) a new output possibility was created. This was also necessary to present the irradiance onto the vertical south glazing. These aspects were realised by changes in the HTB2 source code. At first the new program output didn't work very well and there were some dubious output values. It was through the personal feedback from Leicester on the first result set that these inconsistencies were uncovered. The mistakes could be corrected for the final version of the result set.

The hotline was also useful concerning the discussion about the weather data formats. The HTB2 version 1.2 had been altered to read weather data in the UK format (hour centred) starting with the period 00:30-1:30 (this was originally done to permit SERIRES compatibility), but the original climate data correspond to the US convention (half-hour centred). Therefore, alternative climate files with hour centred values (*v099.new* and *vll0.new*) were used for the simulations. The UK-format *vll0.new* file was created by an averaging method converting the original file. The UK-format *v099.new* file was constructed directly from five minutely measurement data. The presented graphs of the Sfv Radiation (Case fd Day 144) in Hotline Newssheet No. 8 showed a clear weakness of the HTB2_fht results. The daily course was represented very poorly. The minimum at 2 pm was not represented. An investigation showed that this inaccuracy is not a result of any bad HTB2 algorithms but an effect of the climate file *v099.new*. The following figure shows the hour-centred and half-hour centred climate files v099 for the days 144 and 145 comparing the course of the direct normal radiation intensity. The values are presented at the hour or half-hour centred time points.



The Figure demonstrates that the agreement of the two files is very poor. For example at day number 144 the radiation minimum at 2 pm is only shown by the US-format file *v099.met***.** As a conclusion, the comparison of simulation results is only possible between models using the same climate file.

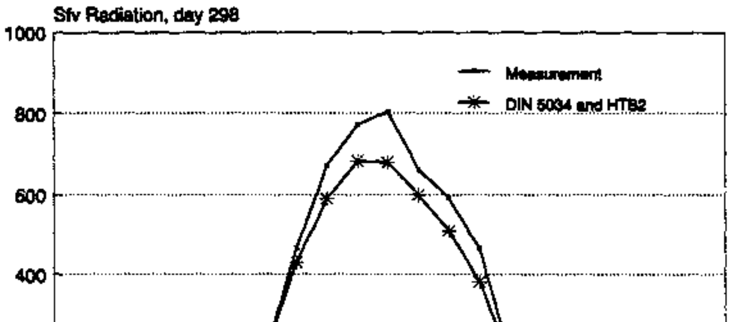
The HTB2_fht simulation results contain two inconsistances which were identified during the results analysis.

a) The solar radiation algorithms of HTB2 version used for this study calculated the sun location with a one hour time-shift. That means, it calculated the highest sun position and south azimuth at 1 pm (one hour too late). Till now, this problem was hidden by using climate files starting with the period 23.30 - 0.30 which assigned the irradiation data to a correct sun position, but assigned it to the subsequent hour (01.00).

b) HTB2 automatically (without user instruction) switches at the 25th of May to summertime which led to an additional one hour time shift for the days 145 - 150. For these days the sun position is calculated two hours to late.

The mistakes are now corrected and the whole solar radiation process has been verificated very carefully. A comparison with the solar radiation calculation principles of the German DIN 5034 showed a very good agreement (although the basic algorithms to predict the sun-height, sun-azimuth and incidence angle are totally different). The thermal simulations of the Empirical Validation were repeated with the correct time conventions. As a result the Sfv Radiation at day 147 (Final Report, Figure 6) now lies within the range of the other programs.

The investigations described above led to the additional conclusion that the measurement data of the Sfv Radiation for the heated period are dubious. The following Figure shows the measured and calculated radiation intensity for the day number 298.



The HTB2 results and the calculated values according to DIN 5034 agree and are therefore represented as one line. The calculations assume an isotropic diffuse sky and a ground reflectivity of 0.2. The deviations to the measurement values are quite high and can only be explained by assuming an non-isotropic sky and/or a higher ground reflectivity. An investigation evaluated the corresponding ground reflectivities which would lead to a better agreement. The ground reflectivities varied between 0.4 and 0.8 dependent on the time of day. This is a very unsatisfying result, because such an high variation of the ground reflectivity is very unlikely. The influence of a non-isotropic sky luminance distribution was examined at the Fachhochschule für Technik in Stuttgart in connection with a different project. The impact was found to be quite small especially for vertical glazing. Hence, there is no satisfying explanation of this phenomenon.

Further analysis of the measurement values concentrated on the simulation of the heating system. To adapt the HTB2 simulation results to the measurement values the following attempts were made.

i. variation of the time constant;
ii. variation of the thermostat type (on/off, perfect proportional, band proportional);
iii. variation of the convection/radiation split.

i. A heater time constant of 22 minutes led to a slight improvement of the simulated air temperatures but simultaniously worsen the simulated heating power. HTB2 only gives information about the heating power output but does not show the heating power input which was measured. Hence, the time-constant describes the response of the room to the heater more correctly but the resultant heating power values are not comparable with the measured system input power values.
ii. The variation of the thermostat type had nearly no influence on the results.
iii. When the radiative output of the heater was increased from 60% to 90% considering an addiational radiative connection to the window, the inside air temperature during the heating period became very close to the measued values. Particularly, the time lag between the modelled and measured air temperatures during the heating phase in the early morning could be reflected. This suggests a special convective condition occures in the small (non-ventilated) room which leads to a high energy transport from the heater to the surfaces but to a very low convective connection with the zone air.
iv. However, although the air temperatures were predicted quite accurately the energy performance could not be improved, in fact, it was slightly worse. Hence, in the final result set, the original heating conditions were used.

Appendix F5: DEROBv1th - M.Wall/P.Wallenten - Lund Univ., S

Maria Wall Petter Wallentén
Dept. of Building Science, Lund University,
P.O. Box 118, S-221 00 Lund, Sweden.

1 Introduction

The program DEROB-LTH has been used in the Empirical Validation Test constructed by IEA group 12B/21 C. Until now, the calculations have been made without any knowledge about the measurements, but in this paper we had the opportunity to compare calculations with measurements.

2 DEROB-LTH

DEROB, which is an abbreviation for Dynamic Energy Response of Buildings, is a family of 6 modules calculating energy consumption for heating, cooling and ventilation. The program was originally developed by Francisco Arumi-Noé at the Numerical Simulation Laboratory, School of Architecture, University of Texas, Austin. Since 1985 the DEROB modules have been further developed to suit the local needs at the Department of Building Science at Lund University. This program version is called DEROB-LTH.
DEROB uses a RC network which is solved by the Crank-Nicholson method with a fixed time step of one hour. The matrixes are solved by the Newton-Raphson method. DEROB can simulate buildings with arbitrary geometries and interprets the presence of shading devices. The program can not simulate a radiator system. The heating demand is calculated as if the heat goes directly to the air node without any time delay and can not be divided into one radiative and one convective part.

3 Describing the buildings

The buildings were basically modelled as described in the Site Handbook. However, as mentioned above we could not simulate the radiator.
The documentation provided together with the hotline and newssheets were all together good.

4 Results

Figure 1 and 2 show examples of the original blind calculations. In Figure 1 the measured and calculated temperature in the unheated opaque room are shown. We can see a phase difference were the calculated temperature rises about one hour earlier than the measured. This difference in phase exists in almost all comparisons of free floating temperatures in the empirical validation test. We have made a lot of comparisons between measurements and calculations before but never had a phase difference. Trying to find an explanation to this was of course of interest to us. The calculated temperature during the last but one day is suddenly too high due to a higher wind speed this day which will affect the outside film coefficient and the air infiltration. In DEROB the outside film coefficient is a constant and the air infiltration was zero as told in the specifications.

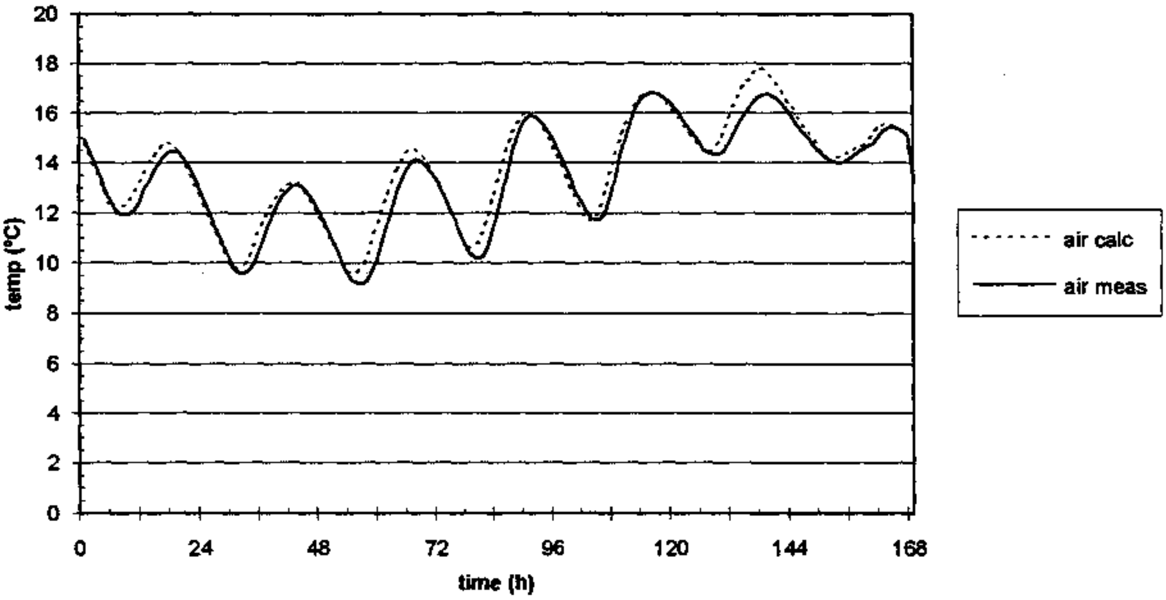


Figure 1 Air temperature in the free floating opaque room.

In Figure 2 the air temperature in the heated double glazed room is shown. In DEROB the air will be directly heated up to 30 °C and it is obvious that the radiator with 60/40% radiative/convective split will not work like this. The estimated power and energy will in this case be wrong. To include a radiator system into DEROB is not a quick and easy work. Our sensitivity analysis has therefore been concentrated on the free floating cases.

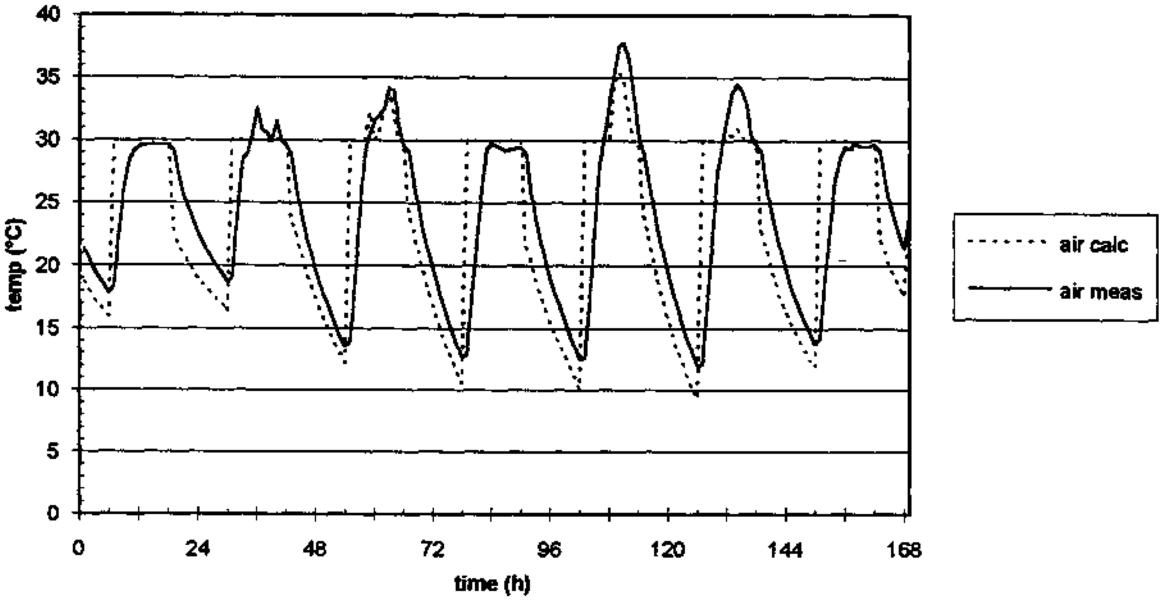


Figure 2 Air temperature in the heated double glazed room.

Listed below are most of the parameters studied, and the effect of changing them.

- The calculated incident solar radiation on the south facade is very similar to the measured radiation in May. In October the calculated radiation is lower than the measured. DEROB calculates the solar declination only for the middle day of each month. The hourly angles of the sun is calculated in the end of each our, true solar time. If more correctly, the angles are calculated for the actual day and in the middle of the hour, the calculations are even better in agreement with measurements.

- If the solar declination is calculated on the 27th of May instead of the 16th, the maximum air temperature in the free floating opaque room will be about 0.5 °C lower. This will also move the phase 10 minutes later in time, mostly before noon. If the solar angles are calculated in the middle of the hour instead of in the end, this will move the phase about 15 minutes later in time, mostly before noon. Both these changes will move the phase less than half an hour. This can not alone explain the noticed difference in phase.

- DEROB has a fixed time step of one hour. We tricked the program to get a time step of half an hour. The heat capacities were doubled and so was the climate file, and the calculations had to be without solar radiation. This halved time step had almost no effect.

- The solar absorptivity on the outside was altered from 0.16 to 0.10. This lowered the free floating temperature about 0.3 °C. It had no effect on the phase difference. 0.16 seems to be a more correct value for the absoptivity.

- The outside film coefficient was fixed in DEROB to 15 $W/m^2°C$. By changing it to 21 $W/m^2°C$ the free floating temperature in the room was lowered about 0.5 °C. It had no effect on the phase.

- The film coefficient on the inside is calculated in DEROB and varies in time. We tested different fixed values on the convective part but this gave a very small effect on the temperature.

- The long wave radiation to the sky was not calculated in this version of DEROB. This will have some effect. If you compare the measured outside temperature with the measured air temperature in the roofspace, the temperature in the roofspace is colder than outside during the nights.

We did not find any bugs in the program. We will of course continue working with the program and probably change the time when the solar angle is calculated, to the middle of the hour instead of in the end. We will also calculate the solar position more than once a month. Furthermore the possibility of taking into account a radiator system will be studied.

1.

Appendix F6: BLASTv3.01v1203 - V.Bocchio/A.Mazza - Poli.Torino, I

Problems

It was impossible to model the heater time constant of 22 minutes.

Documentation

The documentation provided was very detailed and complete. Some explanations were found in the Hotline Newssheets. General organisation and data management were excellent.

Sensitivity Studies

A first analysis of measured data discovered an error due to the batch program, used to automatically perform multiple runs and to provide results in the requested format, that exchanged two materials. This error and other minor errors (solar absorptivity of roof and shading due to east side buildings) were corrected and the new results are considered as base results for sensitivity studies. Effects are evaluated due to:

- use of a new command for the detailed calculation of convective heat exchange on inside and outside surfaces (available in the last BLAST release). Both BLAST releases (lvl 193 and lvl 203) have given identical results in the simulation of base cases, so it was possible to use the newest release;
- thermal conductivity increase;
- change in the heater radiative-convective split.

The cases considered are as follows

1. Base case. Primitive case after the correction of wrong materials and other minor changes
2. Base case + detailed calculation of convective heat exchange on internal surfaces
3. Base case + detailed calculation of convective heat exchange on external surfaces
4. Base case + detailed calculation of convective heat exchange on internal and external surfaces
5. Base case + thermal conductivity increase (conductivity of thermal bridge materials Wood A, B and C +50%)
6. Base case + thermal conductivity increase + detailed calculation of convective heat exchange on external surfaces
7. Base case + thermal conductivity increase + detailed calculation of convective heat exchange on external surfaces + 70%-30% radiative-convective split
8. Base case + thermal conductivity increase + detailed calculation of convective heat exchange on external surfaces + 80%-20% radiative-convective split
9. Base case + thermal conductivity increase + detailed calculation of convective heat exchange on external surfaces + 90%-10% radiative-convective split

Air Temperature Statistics for Free-floating case

	double glazed				opaque			
case	T^+	T^-	$\lvert D \rvert$	$\sqrt{D^2}$	T^+	T^-	$\lvert D \rvert$	$\sqrt{D^2}$
1	32.50	13.15	0.77	0.94	16.92	9.97	0.37	0.48
2	32.88	12.64	0.73	1.00	17.10	9.57	0.40	0.52
3	32.09	12.58	0.49	0.69	16.79	9.37	0.43	0.50
4	32.51	12.17	0.65	0.85	17.00	8.91	0.56	0.64
5	32.24	12.78	0.58	0.76	17.02	9.85	0.36	0.47
6	31.83	12.19	0.47	0.62	16.86	9.22	0.46	0.52

Heating Energy Statistics for Heating case

	double glazed					opaque				
case	T^+	T^-	Q	$\lvert D \rvert$	$\sqrt{D^2}$	T^+	T^-	Q	$\lvert D \rvert$	$\sqrt{D^2}$
1	41.11	13.30	68.5	279	407	30.01	14.53	97.3	237	317
6	39.24	10.75	80.6	180	276	30.08	12.42	109.8	97	157
7	39.43	10.81	82.6	163	257	30.08	12.52	111.8	72	107
8	39.59	10.85	84.9	157	248	30.08	12.66	113.5	73	117
9	39.39	10.79	73.7	365	608	30.24	12.81	74.5	645	791

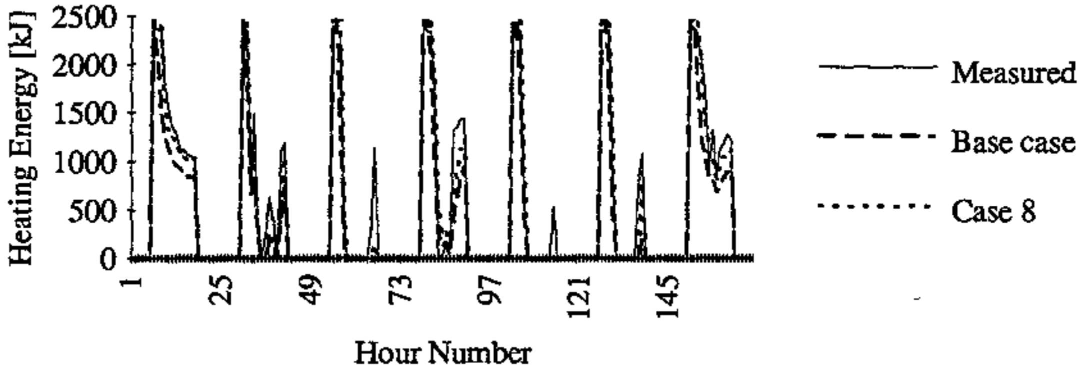
Note: Air temperature statistic values are calculated for a 168 hours period. Energy values are calculated for a 84 hours period

T^+, T^- = Maximum, minimum air temperatures (ºC), Q = Heating Energy (MJ), $\lvert D \rvert$ = Mean absolute difference, $\sqrt{D^2}$ = Root Mean Square Difference between predicted and measured values

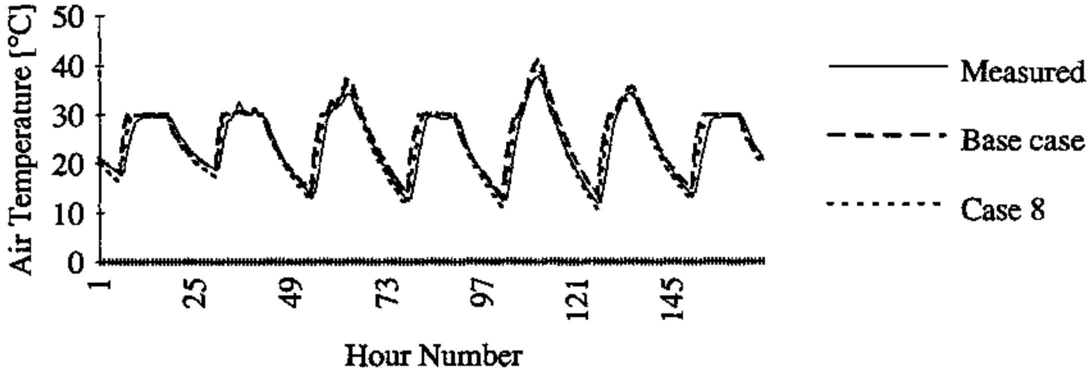
4. Discussion of results

The floating temperature simulations were used to check the results for different convective heat transfer correlations and increased U-value of the envelope. The most detailed heat transfer correlations available in the latest version of BLAST show better agreement with the experimental temperature values, when applied to the external surfaces, probably due to better treatment of wind velocity effects. Improved convective heat transfer correlations applied to internal surfaces give very little changes in the results with moderately worse agreement with measurements. Therefore, in the remaining cases examined, detailed treatment of convective heat exchange was applied only to external surfaces. A moderate increase of the envelope U-value (+1%) obtained by a 50% increase of the thermal conductivity of wood in thermal bridges gives even better agreement with experimental values. Case 6 can be considered the best result achievable without unreasonable changes in input values, in the floating temperature simulations. For the heating simulations, starting from case 6, with the standard 60-40 radiative-convective split of radiator power, the R/C split was increased to 70-30%, 80-20% and 90-10%. The reason for doing this is the attempt to fictitiously simulate the isolation of the air temperature sensor in the middle of the test cell from the convective power emitted by the radiator; CH) calculations in fact showed that the temperature sensor is in a stagnant position, surrounded by the convective air loop induced by the radiator, and the air in the cell is far from being well mixed. Therefore it is not surprising that case 8 with 80-20% R/C split gives better results under heating conditions.

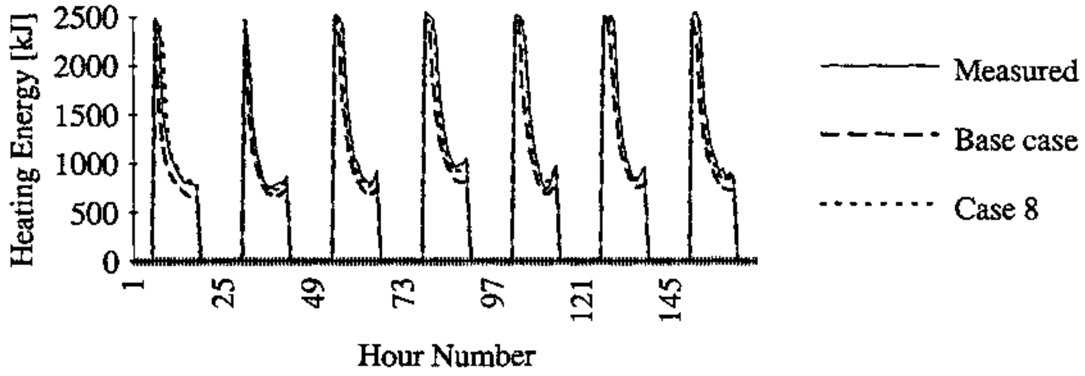
BLAST appears to be very accurate in the floating cases with results always in the error band. The heating cases are undermined by the fact that BLAST (and probably all the other programs) assume uniform air temperature in the room, while in reality this is not the case.



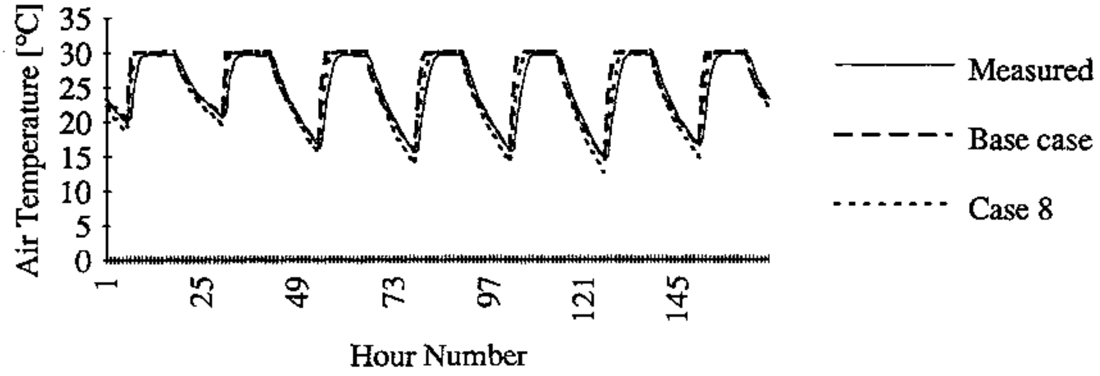
Heating Energy, double glazed



Air Temperature, double glazed



Heating Energy, opaque



Air Temperature, opaque

Appendix F7: TASEv3.0 - T.Haapala/T.Kalema - Tampere Univ., FIN

The sensitivity analysis for the empirical validation exercise was made according to the report IEA21RN359/93 (Hotline Newssheet No. 9). There was given a list of uncertainties for the input parameters of the thermal analysis and the limits in which the parameters can vary. Our sensitivity analysis of the TASE program included all the parameters given except the latitude and the longitude coordinates, the external surface emissivity, the R/C split, which is modelled internally, and the time constant of the heater, which is excluded from the TASE program. In addition to this the calculation of the extinction coefficient of glass and the cleanliness of glazing were joined together.

The sensitivity analysis were made using the free-floating and the heated cases with double glazed windows. The total effect of all the uncertainties studied was estimated by the quadrature sum of the calculated results of the individual uncertainties.The upper extreme were made by choosing the positive alternatives of effects of all the calculated individual uncertainties, calculating the total uncertainty and adding this to the value of the base case. The lower extremes were made in similar way using the negative alternatives of the effects.

The results are presented in Figure 1, Table 1 and Table 2. In the Figure l the calculated results and their uncertainty ranges for heating energy and maximum and minimum room air temperatures are compared with the measured data. The measured heating energy is a little greater than calculated upper extreme. For the heated case the measured maximum room air temperature is located in the lower uncertainty range of the calculation, but the measured minimum room air temperature is a little greater than upper extreme of the calculation. For the free-floating case both the measured maximum and minimum room air temperature are located in the uncertainty ranges of the calculation.

Table 1 shows that the largest individual uncertainty of the free-floating case for the mean air temperature and the maximum room air temperature was due to the external surface absorptivity, but to the minimum room air temperature it was rockwool thickness.To the heated case, Table 2, the largest individual uncertainty for the heating energy and the mean room air temperature was rockwool thickness and to the maximum and the minimum room air temperature the concrete heat capacity.

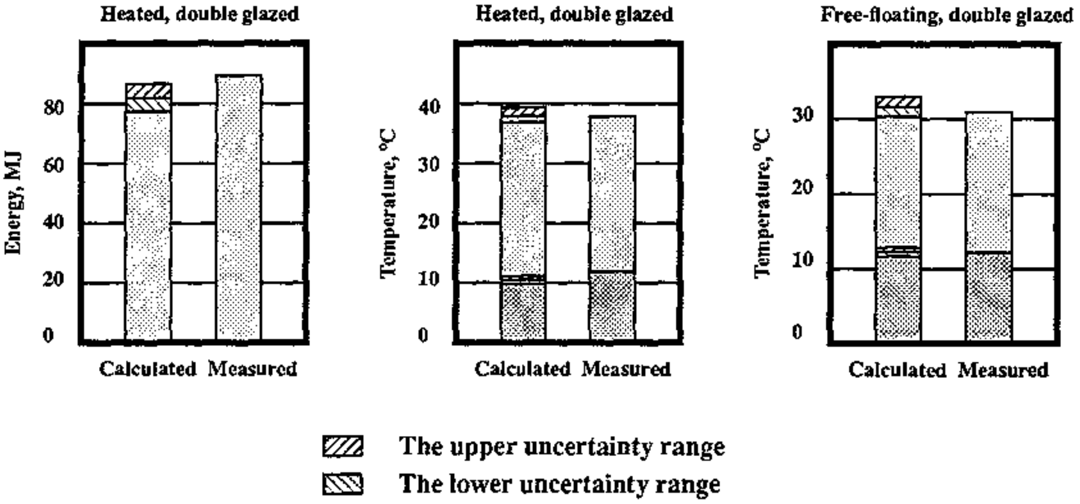


Figure 1. The calculated results, their uncertainty ranges and the measured data for the heating energy and the maximum and the minimum room air temperatures.

Table 1: Individual effects of uncertainties to the room air temperature of the double glazed free-floating case on May 24th ... 30th.

		mean		maximum			minimum		
		T °C	ΔT °C	T °C	ΔT °C	date d-h	T °C	ΔT °C	date d-h
	The base case	20.77		31.61		27-17	12.26		25-06
1.	**Site Details**								
1.1	Ground reflectivity, 0.2								
	a) 0.15	20.40	-0.37	30.88	-0.73	27-17	12.05	-0.21	25-06
	b) 0.25	21.15	+0.38	32.34	+0.73	27-17	12.46	+0.20	25-06
1.2	Glazing orientation, 9 to West from South								
	a) 8.5	20.77	0	31.53	-0.08	27-17	12.26	0	25-06
	b) 9.5	20.78	+0.01	31.70	+0.09	27-17	12.26	0	25-06
2.	**Test Room Surface Finishes**								
2.1	External surface absorptivity, 0.16								
	a) 0.1, 0.75 in ceiling	20.47	-0.30	30.98	-0.63	27-17	12.12	-0.14	25-06
	b) 0.3, 0.95 in ceiling	21.28	+0.51	32.62	+1.01	27-17	12.51	+0.25	25-06
2.2	Internal floor absorptivity, 0.5								
	a) 0.4	20.66	-0.11	31.42	-0.19	27-17	12.14	-0.12	25-06
	b) 0.6	20.88	+0.11	31.80	+0.19	27-17	12.37	+0.11	25-06
2.3	Internal other surface absorptivity, 0.16								
	a) 0.14	20.73	-0.04	31.52	-0.09	27-17	12.24	-0.02	25-06
	b) 0.18	20.81	+0.04	31.70	+0.09	27-17	12.27	+0.01	25-06
2.4	Internal emissivity, 0.9								
	a) 0.85	20.78	+0.01	31.60	-0.01	27-17	12.28	+0.02	25-06
	b) 0.95	20.76	-0.01	31.62	+0.01	27-17	12.24	-0.02	25-06
3.	**Material Properties**								
3.1	Styrofoam conductivity, 0.027 W/mK								
	a) 0.025 W/mK	20.85	+0.08	31.70	+0.09	27-17	12.34	+0.08	25-06
	b) 0.033 W/mK	20.56	-0.21	31.38	-0.23	27-17	12.03	-0.23	25-06
3.2	Concrete heat capacity, 920 J/kgK								
	a) 828 J/kgK	20.76	-0.01	31.95	+0.34	27-17	12.04	-0.22	25-06
	b) 1012 J/kgK	20.79	+0.02	31.28	-0.33	27-17	12.48	+0.23	25-06
3.3	Rockwool conductivity, 0.043 W/mK								
	a) 0.040 W/mK	20.90	+0.13	31.70	+0.09	27-17	12.43	+0.17	25-06
	b) 0.046 W/mK	20.65	-0.12	31.53	-0.08	27-17	12.10	-0.16	25-06
3.4	Rockwool thickness								
	a) -10 mm	20.56	-0.21	31.47	-0.14	27-17	11.98	-0.28	25-06
	b) +10 mm	20.96	+0.19	31.74	+0.13	27-17	12.51	+0.25	25-06
3.5	Plasterboard heat capacity, 1090 J/kgK								
	a) 981 J/kgK	20.76	-0.01	31.90	+0.29	27-17	12.05	-0.21	25-06
	b) 1199 J/kgK	20.78	+0.01	31.33	-0.28	27-17	12.46	+0.20	25-06
3.6	Wood conductivity, 0.125 W/mK								
	a) 0.1 W/mK	20.94	+0.17	31.79	+0.18	27-17	12.44	+0.18	25-06
	b) 0.15 W/mK	20.64	-0.13	31.48	-0.13	27-17	12.11	-0.15	25-06
3.7	Edge effects								
	a) 0 %	20.77	0	31.61	0	27-17	12.26	0	25-06
	b) +50 %	20.64	-0.13	31.47	-0.14	27-17	12.11	-0.15	25-06
4.	**Glazing Properties**								
4.1	Glazed area, 1.5 m^2								
	a) 1.48 m^2	20.72	-0.05	31.47	-0.14	27-17	12.26	0	25-06
	b) 1.52 m^2	20.83	+0.06	31.76	+0.15	27-17	12.26	0	25-06
4.2	Glass extinction and cleanliness, transmission, 1.0								
	a) 0.95	20.62	-0.15	31.28	-0.33	27-16	12.17	-0.09	25-06
	b) 1.03	20.62	-0.15	31.19	-0.42	27-16	12.22	-0.04	25-06
5.	**Test Room Characteristics**								
5.1	Ventilation, 0.0 1/h								
	a) 0.0 1/h	20.77	0	31.61	0	27-17	12.26	0	25-06
	b) 0.05 1/h	20.68	-0.09	31.49	-0.12	27-17	12.17	-0.09	25-06
	Total Effect								
	Upper Extreme	21.48	+0.71	32.99	+1.38		12.84	+0.58	
	Lower Extreme	20.11	-0.66	30.34	-1.27		11.63	-0.63	

Table 2: Individual effects of uncertainties to the heating energy and the room air temperature of the double glazed heated case on October 20th ... 26th.

		mean				maximum			minimum		
		Q MJ	ΔQ MJ	T °C	ΔT °C	T °C	ΔT °C	date d-h	T °C	ΔT °C	date d-h
	The base case	81.87		24.33		37.96		24-15	10.58		24-06
1.	**Site Details**										
1.1	Ground reflectivity, 0.2										
	a) 0.15	82.54	+0.67	24.28	-0.05	37.73	-0.23	24-15	10.56	-0.02	24-06
	b) 0.25	81.20	-0.67	24.37	+0.04	38.19	+0.23	24-15	10.59	+0.01	24-06
1.2	Glazing orientation, 9 to West from South										
	a) 8.5	81.74	-0.13	24.33	0	37.94	-0.02	24-15	10.58	0	24-06
	b) 9.5	82.00	+0.13	24.33	0	37.97	+0.01	24-15	10.58	0	24-06
2.	**Test Room Surface Finishes**										
2.1	External surface absorptivity, 0.16										
	a) 0.1, 0.75 in ceiling	83.11	+1.24	24.23	-0.10	37.44	-0.52	24-15	10.53	-0.05	25-06
	b) 0.3, 0.95 in ceiling	79.53	-2.34	24.53	+0.20	39.09	+1.13	24-15	10.61	+0.03	24-06
2.2	Internal floor absorptivity, 0.5										
	a) 0.4	82.07	+0.20	24.23	-0.10	37.91	-0.05	24-15	10.44	-0.14	25-06
	b) 0.6	81.73	-0.14	24.42	+0.09	38.02	+0.06	24-15	10.64	+0.06	24-06
2.3	Internal other surface absorptivity, 0.16										
	a) 0.14	82.19	+0.32	24.31	-0.02	37.80	-0.16	24-15	10.58	0	24-06
	b) 0.18	81.58	-0.29	24.34	+0.01	38.09	+0.13	24-15	10.57	-0.01	24-06
2.4	Internal emissivity, 0.9										
	a) 0.85	81.51	-0.36	24.33	0	37.88	-0.08	24-15	10.58	0	24-06
	b) 0.95	82.21	+0.34	24.32	-0.01	38.03	+0.07	24-15	10.57	-0.01	24-06
3.	**Material Properties**										
3.1	Styrofoam conductivity, 0.027 W/mK										
	a) 0.025 W/mK	81.30	-0.57	24.39	+0.06	38.06	+0.10	24-15	10.66	+0.08	24-06
	b) 0.033 W/mK	83.45	+1.58	24.15	-0.18	37.68	-0.28	24-15	10.30	-0.28	25-06
3.2	Concrete heat capacity, 920 J/kgK										
	a) 828 J/kgK	81.18	-0.69	24.25	-0.08	38.52	+0.56	24-15	10.23	-0.35	24-06
	b) 1012 J/kgK	82.54	+0.67	24.40	+0.07	37.40	-0.56	24-15	10.93	+0.35	24-06
3.3	Rockwool conductivity, 0.043 W/mK										
	a) 0.040 W/mK	80.47	-1.40	24.46	+0.13	38.16	+0.20	24-15	10.76	+0.18	24-06
	b) 0.046 W/mK	83.24	+1.37	24.20	-0.13	37.76	-0.20	24-15	10.40	-0.18	24-06
3.4	Rockwool thickness										
	a) -10 mm	84.25	+2.38	24.09	-0.24	37.63	-0.33	24-15	10.20	-0.38	25-06
	b) +10 mm	79.89	-1.98	24.53	+0.20	38.24	+0.28	24-15	10.85	+0.27	24-06
3.5	Plasterboard heat capacity, 1090 J/kgK										
	a) 981 J/kgK	80.96	-0.91	24.22	-0.11	38.28	+0.32	24-15	10.24	-0.34	24-06
	b) 1199 J/kgK	82.75	+0.88	24.42	+0.09	37.64	-0.32	24-15	10.90	+0.32	24-06
3.6	Wood conductivity, 0.125 W/mK										
	a) 0.1 W/mK	79.82	-2.05	24.48	+0.15	38.21	+0.25	24-15	10.76	+0.18	24-06
	b) 0.15 W/mK	83.60	+1.73	24.20	-0.13	37.75	-0.21	24-15	10.43	-0.15	24-06
3.7	Edge effects										
	a) 0 %	81.87	0	24.33	0	37.96	0	24-15	10.58	0	24-06
	b) +50 %	83.53	+1.66	24.20	-0.13	37.75	-0.21	24-15	10.43	-0.15	24-06
4.	**Glazing Properties**										
4.1	Glazed area, 1.5 m^2										
	a) 1.48 m^2	81.95	+0.08	24.32	-0.01	37.80	-0.16	24-15	10.61	+0.03	24-06
	b) 1.52 m^2	81.79	-0.08	24.33	0	38.13	+0.16	24-15	10.54	-0.04	24-06
4.2	Glass extinction and cleanliness, transmission, 1.0										
	a) 0.95	81.66	-0.21	24.30	-0.03	38.04	+0.08	24-15	10.55	-0.03	24-06
	b) 1.03	82.01	+0.14	24.34	+0.01	37.91	-0.05	24-15	10.59	+0.01	24-06
5.	**Test Room Characteristics**										
5.1	Ventilation, 0.0 1/h										
	a) 0.0 1/h	81.87	0	24.33	0	37.96	0	24-15	10.58	0	24-06
	b) 0.05 1/h	82.78	+0.91	24.26	-0.07	37.82	-0.07	24-15	10.50	-0.08	24-06
5.2	Heater power, 680 W										
	a) 640 W	80.90	-0.97	24.23	-0.10	37.82	-0.14	24-15	10.57	-0.01	24-06
	b) 720 W	82.73	+0.86	24.41	+0.08	38.06	+0.10	24-15	10.59	+0.01	24-06
5.3	Setpoint, 30.0 °C										
	a) 29.8 °C	80.87	-1.00	24.22	-0.11	37.88	-0.08	24-15	10.52	-0.06	24-06
	b) 30.2 °C	82.87	+1.00	24.43	+0.10	38.04	+0.08	24-15	10.63	+0.05	24-06
	Total Effect										
	Upper Extreme	86.55	+4.68	24.73	+0.40	39.38	+1.42		11.19	+0.61	
	Lower Extreme	77.41	-4.46	23.87	-0.46	36.88	-1.08		9.82	-0.76	

Appendix F8: TRNSYSv13.1 - R.v.d.Perre/P.Verstraete - VUB, B

0. Model used

TYPE 56 (the building model within TRNSYS version 13.1) has been used
the user selected (lacking a detailed knowledge of the experiment) the following submodels :

- external convection (including external longwave) : bookvalues
- internal convection / airgap resistances : bookvalues
- solar distribution : conform to specifications
- solar processor : anisotropic sky-model
- window transmission : angle dependent - calculated

1. Problems encountered in representing the test rooms within the model

1.1 it is impossible to represent in TYPE 56 a radiative/convective split of a heater
- a convective heater was assumed (this may cause substantial 'errors')

1.2 it is impossible to represent in TYPE 56 any heater dynamics
- an ideal heater was assumed (this may cause substantial 'errors')

1.3 difficulties in choosing values for internal and external surface film coefficients
- in IBA 21 RN 345/93 (March '93) we proved (with ESP simulations) that the selection of an appropriate model for surface film coefficients was crucial for this particular testcell experiment (leading within one program to a range of 47% of predicted heating need) .

1.4 edge-effects are known to be very important in well-insulated testcells.
- the provided 'corrected' conductivity values for use in 1D calculations were taken into account

1.5 temperature dependent conductivity can be important in high temperature heated testcells
- this cannot not be modelled in TYPE 56, but is supposed to have only a minor effect compared to 1.1-1.2-1.3 and 1.4

2. Problems with the documentation provided

2.1 insufficient basic information for quality assessment
- basic information allowing a quality assessment of the constructed models, and of the physical experiments (reference model) was unavailable, or insufficiently documented, or misleading.
- In April '93 VUB asked for global measured testcell characteristics (such as UA-values), and requested global quality information about measured data (experimental uncertainties additional to sensor accuracy), amongst several other items. Global measured testcell characteristics could not be provided, while experimental uncertainties on non-constructional model inputs/outputs were stated to be equal to sensor accuracies (which is extremely unlikely for this kind of experiment - see also 2.3).
- (The lack of) such basic information (global measured characteristics, experimental uncertainties) should be reported by the time of (not) selecting a dataset

2.2 reliability of bookvalues
- if an empirical validation exercise relies mainly on book values for material (and other) characteristics, such EV exercise may easily turn out to become a data-validation exercise rather than a model validation exercise. Differential sensitivity analysis or Monte-Carlo techniques are very useful statistical techniques, but can never replace factual knowledge about the system under study.

2.3 quality and relevance of measurements
- assuming an experimental uncertainty equal to sensor accuracy, implies that the provided data describes unambiguously and completely the physical experiment, (applied to 'room air control' this implies that the elementary air volume around the controller is representative for the overall room air)

- previous experiences in EV (PASSYS) indicate that such experimental uncertainty assumption, for this type of experiment and level of provided data / documentation, may not only be invalid, but more-over might make the measurement irrelevant for empirical whole (non-CFD) model validation
- quality and relevance of measurements are essential requirements for High Quality Datasets, and must be assessed by the modelvalidators prior to any simulation, (even prior to the 'blind' validation)

3. How useful was the hotline

- an eventual strong need for, or benefit from, a hotline support in a (model validation) exercise may indicate that the quality of the documentation sent around might be subject to improvement. Quality of documentation is a key-issue for high-quality datasets in empirical validation.

4. How useful were the newssheets

- the new ssheets were very useful

5. How was quality assurance organized

- see §2

6. Results and conclusions from sensitivity studies

- Almost every program in the EV exercise underestimates the heating need. This might suggest potential problems with the testcell envelope characteristics, or with the temperature control. A look to the measurements indicates that the internal air responds extremely slow to changes in heating regime. When the heater turns on or off, it takes MORE THAN ONE HOUR before a change in air temperature is noticed. This might suggest potential problems with the air-mixing in the testroom.
- Various simulations (for the opaque tescell) were undertaken to explore both suggestions

00. blind simulation
01. as 00 but steady state in order to derive UA-value
02. as 01 but with CIBSE surface film coefficients
    03. as 02 but with 15% increase of UA-value compared to 01 (defining a resistive edge-surface)
    04. as 03 but transient simulation with real climate and imposed temperature profiles (conv. heater)
05. as 03 but with real climate and imposed temperature setpoint file (convective heater)
    06. s 03 but with real climate and imposed (rad/conv) internal gain equal to measured heating
07. as 06 but with increased air-capacitance (adding 150 kg to internal air-mass)

	Qaux	Tair	Tfloor	Tback	Tceiling
measurement	194 W	24.9	23.9	23.6	23.8
simul 00	146 W	24.8	22.6	22.9	22.9
simul 04	171 W	24.0	21.7	20.0	21.2
simul 05	181 W	24.9	22.5	20.7	21.9
simul 06	194 W	24.5	23.5	21.5	22.8

- The blind simulation (sim 00) underestimates the heating need by 48 W (25%), predicts a correct average internal airtemperature, while underpredicting internal surface temperatures. Substituting ASHREA bookvalues for surface film coefficients by CIBSE bookvalues (01-02) does not affect very much calculated global UA-values. Simulating the testroom with an increased UA-value but unchanged capacity (sim 04), moves the energy need in the right direction, but the air and surface temperatures in the wrong direction. From the hourly simulations (well mixed air - convective heating) one can see that for both cases (00-04), the air-temperatures rise or fall immediately after switching the heater on or off. In the heating period (especially in the beginning) air temperatures are largely over-estimated, in the non-heating period (at night) air-temperatures are under-estimated. At the other hand simulated internal surface temperatures (almost) always underestimate measured values (in especially for sim 04), but without much variation. This might suggest the presence of an inert air body in the centre of the testroom, with boundary flows along surfaces in the heating period, (in contradiction to the experiment specification which limit the air temperature uncertainty to +/- 0.2 °C)

- Imposing the measured air-temperature as a setpoint in a testcell with a 15% increased UA-value, and equiped with a convective heater (sim 05), indicates that in the non-heating period predicted surface temperatures almost become identical to the measured values, while in the heating period simulations drastically underpredict internal surface temperatures. This confirms the suggestion of an inert air body in the centre of the testroom, with a strong boundary flow along surfaces in the heating period. Despite the convective nature of the heater in the simulation, predicted energy consumption is pretty close to the measured value (13 Watts on 194 Watts or 07% underprediction).
- Defining the heat input as a radiative/convective gain in a 'modified' testcell (sim06), leads again to overpredictions for the simulated air-temperature in the heating period, and underpredictions in the non-heating period, which is reflected in the predictions for the internal surface temperatures.
- Assuming that one does not know exactly the control temperature within a range of +/- 1°C, and inputting this uncertainty in a setpointfile (sim05), does not bring the measured surface temperatures within the simulated uncertainty range during the heating period. At the other hand, defining an uncertainty of +/- 10% on the inputted flux in sim06, (this is doubling the previous uncertainty range), brings the measured air and surface temperatures pretty close or even within the simulated uncertainty band.
- Assigning a fictive capacity to the (in reality maybe inert) airmass improves drastically the prediction of the air temperature, without affecting too much the predictions for the internal surface temperatures

7. were any bugs found in the model as a result of this exercise

7.1 no bugs were (could be) revealed in TRNSYS as a result of this exercise (but BESTEST did !)

7.2 the test might suggest the presence of 'internal errors' in the pseudo-truth model (dataset) or in its documentation : model input and outputs may not to correspond with the physical experiment (e.g. UA-values, or the meaning of room air control and its related experimental uncertainty) .

8. Conclusions

The essence of model validation is the acceptance that 'simulation' is just another 'experiment', and that validating a particular simulation model requires (basically) :

- the **definition** and execution of two experiments :

one experiment with the model under study, for one or several 'typical' (realistic ??) cases, and a second experiment with a pseudo-truth model, (which implements identical or superior abstractions of the real world), - by using other model results, analytical results, or empirical results -, for (ideally) exactly the same test-cases (e.g. for the the same model inputs / outputs)

- the **comparison** of both experiment results, giving rise to preliminary questions such as :

what accuracy are we looking for ? what to compare ? how to compare ?

- the **diagnostics** of eventual differences between both experiment results :

where are the differences ? what causes the differences ? how relevant are the differences ?

- if both experiments implement two different test-cases, result comparisons become very difficult, and model validation may remain inconclusive
- results comparison and/or diagnostic requirements influence greatly the experiment definition : a bad experiment definition may lead to wrong or inconclusive results
- a 'validation' exercise without diagnostics is no validation, but rather qualification (cfr. BESTEST)

validation is always based on a comparison of experiment results from TWO models : in EV the physical experiment, -the 2nd model-, is always unique its kind, therefore diagnostics are essential in revealing eventual errors/shortcomings in the physical experiment

More in general, Model Validation can be defined as :

substantiating that the experimental model, within its domain of applicability, behaves with satisfactory accuracy consistent with the study objectives

More specifically, we can conclude that in this EV exercise :

- TRNSYS was tested outside its domain of applicability, but nevertheless succeeded in following rather well the transient respons of the testcell fabric.
- Potential problems were suggested with regard to the experiment description (global testcell characteristics), as well as with the quality assessment of the dataset itself (experimental uncertainty)
- The quality of the dataset might be insufficient for a conclusive whole-model validation exercise

Appendix F9: SERI-RESv1.2 - S.Hammond - BRE, UK

1. Problems encountered in representing the test rooms within the model

1.1 Accuracy of data.

In SERI-RES 1.2, wall and floor areas, together with some length values, can be input only to one decimal place accuracy. This means lengths are accurate only to 5 cm and areas to .05m ^{2}.

1.2 Sloping roof.

For the sloping roofed upper space, a height was needed for calculation of the total volume. This was taken as the mean of the two wall heights. Also for calculation of solar incidence, the roof slope was taken as 10.1°.

1.3 Representation of the floor.

Since the floor construction was so close to the ground, there was some query over representing it as an external wall with the outer surface conditions as "AMBIENT". However, for the short period of the study, it was decided that the thickness and insulation of the floor construction was such as to make the inaccuracy small, so the model described above was used.

1.4 Consideration of the next-door cell.

The West wall connected to the neighbouring cell was modelled according to Table 5.6 of the construction. This was the alternative preferred method, and reduced the error in area approximation (see 1.1). It was taken as adjacent to a volume of constant temperature 20°C for both May and October.

1.5 Heating units.

SERI-RES cannot model anything more complex than the amount of energy required to heat, cool or ventilate any cell, so the radiator power of 680W was given without any radiative convective split, and with a simple control of cut-off at 30°C between 6:00 and 18:00.

2. Problems encountered with the documentation provided.

2.1 Solar distribution on internal walls.

The split between surfaces was supplied on p 12 of the specification, but took no account of whether there was single or double glazing. According to Performance Assessment Method Documentation for SERI-RES, the glazing affects the heat lost through the window, and hence the fractions on the surfaces. The values supplied were, however, used as given.

2.2 Surface coefficients.

These were calculated using the CIBSE guide A3, giving the following values:

Internal:	Vertical	8.33
	Floor	7.14
	Ceiling	10.00
External	Vertical	16.60
	Roof	25.00

2.3 Air gaps.

SERI-RES requires a resistance for air gaps. Following the recommendations of the CIBSE Guide for air gaps above 20mm, this resistance was taken as 0.18 m^{2}K/W.

2.4 Glazing.

SERI-RES requires U-values to be given for each glazing type. These were found using the CIBSE Guide and the data given, and were taken as 3.4 W/m ^{2}K for double glazing and 5.6 W/m^{2}K for single glazing.

2.5 Climate data set.

The data given were for the hour previous to the time of recording, whereas SERI-RES 12 requires the data centred on the recorded hour, starting at midnight for any day. A program was written to convert the given data, but subsequently a revised data set for SERI-RES was supplied by De Montfort.

2.6 Climate data checks.

There was no graph of the direct normal solar incident radiation, so this could not be checked for use in SERI-RES 1.2. SERI-RES 1.2 also needs wet bulb temperature, but this is used only for latent heat calculations, and it was not considered worth deriving it from relative humidity and other given data.

Usefulness of hotline

The hotline was found to be useful in two main directions.

3.1 Missing data.

Where data were not given, it was useful to be able to call up and check if one had missed something in the specification, and get advice on alternative sources of information. This had a two-way effect if more than one participant needed extra data.

3.2 Errors in the specification.

The very small number of errors in construction details which were found could be confirmed quickly without waiting for the newssheets.

Usefulness of newssheets.

The newssheets were most useful for the following.

4.1 Following up hotline information.

Confirmation of errors, extra data information and changes was essential, and the newssheets were good vehicles for this.

4.2 Deadlines.

The newssheets were useful for reminders to participants, including any minor changes or additions.

Organisation of quaity assurance.

[illegible]

Since the tests were performed blind, only the input data could be checked. The climate file was checked initially using the procedure proposed in the Validation Guidebook IEA21RN196/92, but after the new SERI-RES 1.2 format file was supplied it had to be assumed that this file was correct. In fact there was a query on the validity of the first pair of revised climate files sent. They were sent again and runs repeated. Input data for SERI-RES 1.2 were checked in-house, and the input files were also sent to De Montfort University with the results.

Results and conclusions from sensitivity studies.

None were performed for SERI-RES.

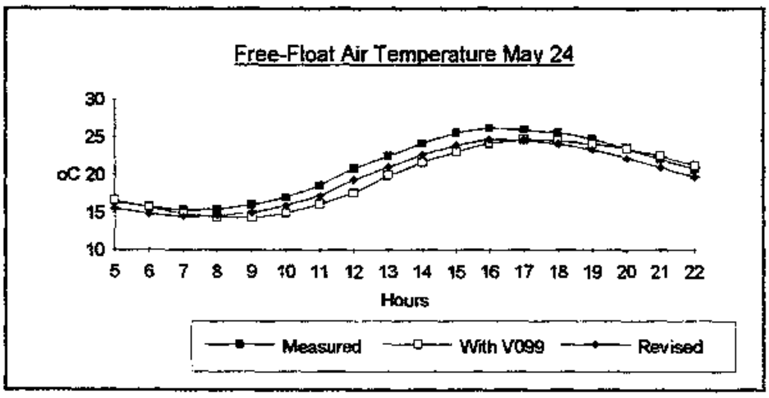
Finding of bugs in the model as a result of this exercise.

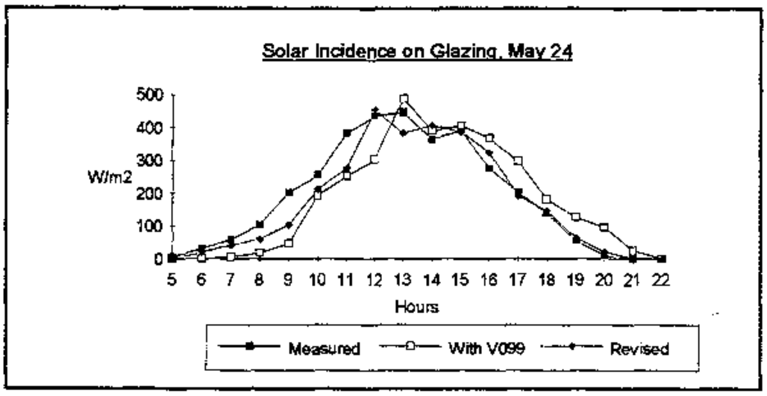
Although bugs have been found in SERI-RES 1.2 during other exercises, for example with sidefin shading during the IEA21/C BESTEST studies, none were found as a result of this exercise, with the possible exception of solar incident radiation as described in the following section.

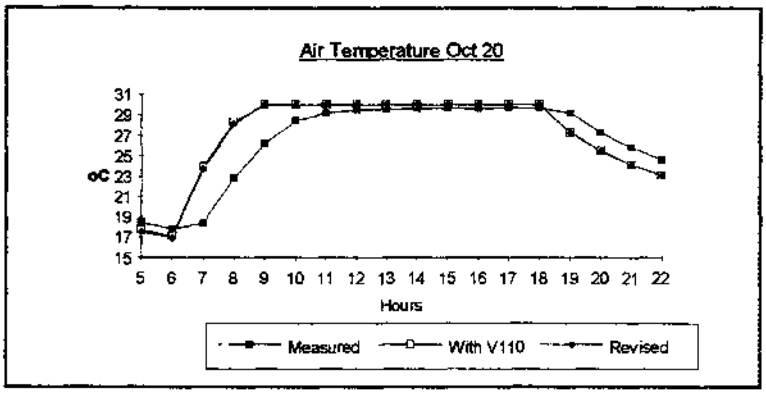
Comparison of results with measurements.

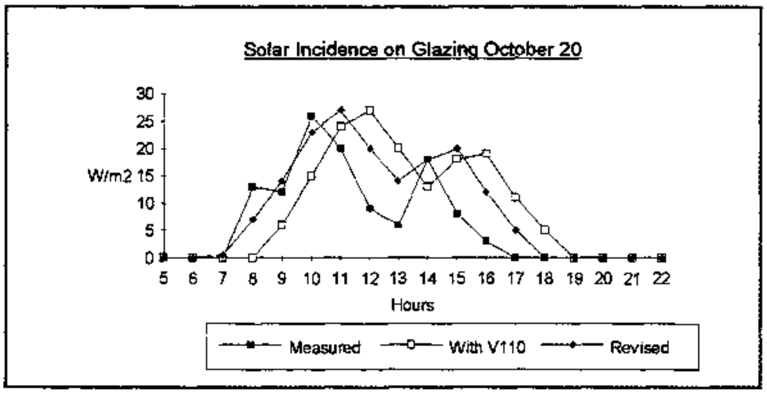
The major problems with SERI-RES 1.2 occur in connection with the modelling of solar radiation. The weather data as supplied, which start at midnight as the program requires, give solar radiation values which appear to be an hour behind those measured for the test room glazing. The SERI-RES runs were all done again modifying the weather data by deleting the first record. Though this made the solar incidence match more closely with regard to timing, there was wide variation in values of both this and air temperature.

For example, consider Cell 1 for both free-float and heated conditions.



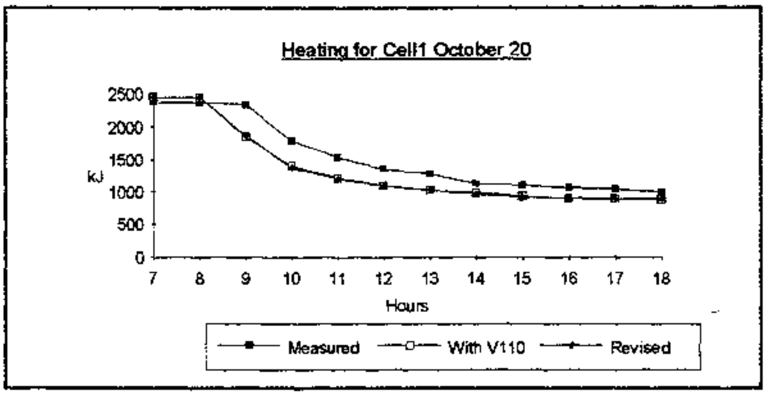






There appear to be grounds for investigating the algorithms used in SERI-RES 1.2 for calculating solar radiation, and for checking the midnight start time for data. There was nothing in the specification to indicate that daylight saving times were used for May, and this would not account for similar errors in October.

As described in Section 1.5, the heating unit calculations in SERI-RES 1.2 are purely for the energy requirements for the conditions required, and do not go into detail of HVAC type, etc. The results are consistently lower than those measured, except at the beginning of the day, when the immediate application of full heat in SERI-RES will not correspond very well to the reality of a radiator which takes some time to warm up, as well as responding to higher solar radiation values. This may be noted in the air temperature, which is 28°C for SERI-RES while the measured value is only 23°C at 8.00. Another feature which SERI-RES cannot model is the radiative/convective split in the heat input, and this will have some influence. There is little difference between the results from the original V110 data and those revised with regard to solar radiation.



In free-float cases, back wall and ceiling temperatures are almost equal in SERI-RES 1.2 results, and are very close to the corresponding air temperatures. In the measured results, back wall and ceiling temperatures are not so close, and are generally lower than the air temperatures. Floor temperatures found by SERI-RES 1.2 are generally higher at night than those measured, and lower during daylight. This may be due partly to the solar radiation differences, and partly due to the modelling of the raised floor. In the heated cases, SERI-RES 1.2 surface temperature results reflect the differences in application of the heating system which has an immediate response in SERI-RES without warm-up. Again, SERI-RES 1.2 back wall and ceiling temperatures follow the air temperature closely, and floor temperature values are generally greater than those measured, including during the day.

Appendix F10: ESP+v2.1 - G.Stuart - ASL Sterling, UK

Introduction

On review of the exercise conducted by De Montfort University, a number of issues are raised relating to the accuracy of the ESP+ model of the test cell, and also to the use of the test cell as being representative of a building.

It should be stated at the outset, we were surprised and dissappointed at the initial reported results from ESP+. However we were also surprised at the apparent generally poor performance of the majority of the models in the exercise; this leads us to give serious consideration to the possibility of other errors in the exercise outwith the scope of the "internal" errors concluded by De Montfort University.

We prepared a draft technical rebuttal in relation to our concerns, and have presented this to De Montfort University for comment and review. At this point in time, De Montfort have indicated they feel there may be some errors in our draft, but no details are available to us. We expect details to be made clear in future communication with De Montfort staff, so that our Rebuttal can be published in full.

The enclosed three page rebuttal is based on the work carried out for our draft technical rebuttal, and relates to our concerns on the modelling of the test-cells. We feel the test-cells are not as "simple" in modelling terms as has been implied, and here we attempt to relate how we would tackle such an exercise, based on our own modelling experience in real applications, together with the consequences on the recorded model results.

References

References are made in this report to other sources. These are quoted below.

[1] Khalifa A. and Marshall R. (1992) Validation of Heat Transfer Coefficients on Interior Building Surfaces Using a Real Sized Indoor Test Cell (Solar Energy Unit, University of Wales, Cardiff).

[2] Pilkington (1988) Glass and Transmission Properties of Windows, 7th Edition. (Pilkington Environmental Advisory Service).

Perceived Modelling Deficiencies

The energy processes in the test-cell should be modelled with greater care than is normal. This is emphasised in the Validation Guidebook. Below are illustrated some of the features which we feel were not modelled well enough to reflect these considerations.

- The model was not created using the interface provided with ESP+, rather it is an edited version of the files used in other ESP versions. This led to problems early in the study when the input parameters for site elevation (100m) and building orientation (9 degrees) were confused, resulting in extremely erroneous results.
- Errors were found in the climate file used compared to that which was measured. This is illustrated in the accompanying graph of climate data for the last day of the simulation period. Also, all relative humidity values were set to 99.0 in the measured data; this must be in error and will lead to significant error in the calculation of long-wave radiation in ESP which uses RH to develop the sky conditions for its external long-wave radiation routines. Correction of these errors will undoubtedly increase energy use.
- The model created had reflected and diffuse solar radiation being received on the underside of the floor. This error could have been remedied by setting the external solar absorption coefficient of the floor construction to 0.0, thus accounting for the close proximity of the external surface of the test cell floor to the ground. This was not done despite a specific reference to this problem in the Site Handbook. Had this been done an increase in energy use would be expected.
- The application of 1AC/h to the void zone seems arbitrary at best, and casts doubt on the true nature of airflow in this zone. Variations in this value will undoubtedly affect energy use.
- Although no information on heat transfer coefficients was provided in the site handbook, research carried out on test-cells in *Wales*[1] suggests internal convective heat transfer coefficients higher than the Alamdari-Hammond correlations used as default in ESP. There is the facility to allow

the user to change these coefficients, but this was not done. The modification to these coefficients will increase energy use.

- The distribution of radiant heat from the emitter will mean much of its output will fall on the south facade, this will in turn affect the heat transfer coefficients on the south wall and window, thereby increasing localised convective, radiant, and conductive heat transfer processes associated with these surfaces. Again, no information on the actual distribution of heat is supplied but if accounted for, an increase in energy would be expected, particularly in the double glazed case.
- Unless a detailed model of the emitter exists it is difficult to capture the performance of the emitter in reality; as stated earlier any assumptions relating to the emitter take on great significance because of the particular geometries involved in the test-cell. Currently, ESP+ does not have such a detailed model, although methods are available to emulate its operation. Had a sufficiently detailed model been available, an increase in energy demand might be expected.
- Surrounding test-cells will shade parts of the building at certain times. ESP+ provides detailed shading analysis software (Insolare) which could have modelled these effects. No account of this was included. This will increase energy use in the test cells.
- The air-gap resistances in constructions were all assumed as default values of 0.18W/m2K. Calculation based on emissivities and air-gap thicknesses based on calculation procedures from *Pilkington* yields different values, particularly for the air-gap between the double glazed panel (0.12W/m2K). Also, slightly different values for transmittance and absorptivity for the glazing were calculated. The inclusion of these corrected values will increase energy use.
- Although no time-series insolation information was presented from the monitored tests, constant values (which will not occur in reality) show the walls would be insolated also. In ESP+ if no time series insolation values are present, default insolation planes can be assigned; it is a simple modification to place solar on the rear wall as well as the floor. This was not done, but if incorporated, an increased energy use might be expected.

It should be noted ASL Sterling is partly responsible for some of the deficiencies in the above model. Having originally indicated we could not participate in the exercise unfunded due to a lack of resources, the model was subsequently created by De Montfort University and ASL Sterling staff asked to comment on it. Whilst some deficiencies were easily corrected by ASL Sterling in the time allowed by the resources available, it was not possible to conduct a detailed critique of the De Montfort Model.

Revised Simulations

To investigate the potential effects of the above parameters on the model a number of simulations were conducted as follows:

1) As a starting point we have the De Montfort model.
2) To eliminate any error due to De Montfort not using the ESP+ interface, the model was reconstructed from the data provided using the interface. Note that whilst we consider this reconstructed model to be subject to a lack of sophistication, simulations conducted on it form the "base" case against which changes in parameters we consider inaccurate could be judged.
3) External solar radiation was removed from the underside of the floor of the test-cell.
4) External convective heat transfer coefficients were increased by 10% to reflect the greater exposed area of the test cell relative to its internal dimensions.
5) Corrected climate data was applied to the model by using the measured data originally provided by De Montfort on a disk.
6) A solar shading analysis was conducted on the model and the results applied to the ESP+ calculation.
7) Constructional information (air-gap resistance etc.) was altered to better reflect the performance of the test cell at the prevailing temperatures, and improve the quality of the model.
8) Internal solar radiation was re-distributed to insolate the walls of the test-cell as well as the floor.

9) Internal convective heat transfer coefficients were applied from the Welsh study.
10) Internal convective heat transfer coefficients were applied to better reflect the distribution of heat to the south facade of the model from the emitter.
11) The combined effects of all the above modifications were simulated.
12) Finally simulation 11 was repeated, this time with no assumptions about maximum load from the emitter.

The table below lists each of the results simulated with the appropriate modification applied to the base case.

Simulation	Opaque			Double Glazed		
	Energy MJ	Max. Temp. C	Min. Temp. C	Energy MJ	Max. Temp. C	Min. Temp. C
1. De Montfort	93.0	29.95	14.5	55.0	43.9	13.7
2. Base	95.3	29.95	13.5	59.0	43.3	13.7
3. Floor Solar	95.6	29.95	13.5	59.4	41.7	13.6
4. External Convection	95.4	29.95	13.5	59.0	43.1	13.7
5. Corrected Climate	97.1	29.95	13.5	60.5	43.2	14.0
6. Solar Shading	96.2	29.95	13.5	59.5	43.1	13.7
7. Constructions	95.3	29.95	13.5	62.7	42.3	13.2
8. Solar Distribution	95.3	29.95	13.5	59.5	42.0	13.6
9. Welsh Internal CHTC	101.4	29.95	13.8	63.2	43.6	13.6
10. Emitter CHTC	95.5	29.95	13.5	59.7	41.8	13.6
11. Combined	105.2	29.95	13.8	70.7	40.9	12.7
12. Combined Ideal	106.7	30.04	13.6	79.1	42.6	13.0

Note that simulations 3 through 10 relate to individual modifications applied to the Base simulation (2).

Concerns Not Simulated

It should be stressed all the above modifications have a strong justification in providing a more realistic model than that reported by De Montfort University; there are undoubted other factors (local airflow regimes, external long-wave exchanges, variable conductivities etc.) which we have not yet investigated because we feel at this point there is insufficient data available to make a valid judgement on their likely values at the time of data collection.

It should be noted however that we are convinced these other factors will only serve to strengthen our contention that ESP+ is more than capable of modelling the test cells accurately when sufficient attention is paid to these details. Note that the results shown in simulation 12 remain (just) outside the De Montfort error bands, and we believe this reflects the uncertainty remaining in these unaccounted factors.

It should also be stressed we feel the inaccurate measurement of the RH (given its effects on the long-wave prediction routines in ESP) is of its-self sufficient cause to negate the exercise as an appropriate validation of ESP. The effect of this will be particularly relevant to the double glazed model, as the glazing will be particularly susceptible to long-wave inaccuracies due to its low thermal mass compared to the rest of the test-cell structure.

Conclusions

At this stage, no final conclusions are drawn. When further discussions with De Montfort have been completed to ascertain and correct any particular concerns they have, we shall take their comments on board and present a full rebuttal.

Appendix F11: ESP-rv7.7a - P.Strachan - ESRU, UK

Program	ESP-r
Version	7.7a
Operator	Dr. P. Strachan Energy Simulation Research Unit University of Strathclyde Glasgow

Introduction

This report summarizes the sensitivity studies that were carried out by ESRU within the validation exercise. It should be noted that most of these were carried out before the measured data were released, and included with the original submission. In particular, the sensitivity of the test cells to the internal convection coefficients was recognised at an early stage in this exercise.

Unfortunately, with no resources for carrying out the work, it has not been possible to carry out any detailed analysis of the residuals (difference between measured and predicted data) or a complete differential sensitivity analysis. It is hoped that there will be useful feedback from the organisers of this exercise as to possible reasons for disagreements between measured and predicted data, as a *quid pro quo* for undertaking the simulations.

Sensitivity Analyses

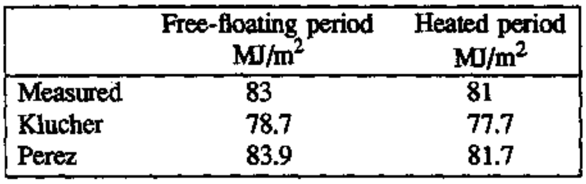
Internal Convection Coefficients. One major area of uncertainty is in the choice of convection coefficients. This is known to be important in the simulation of small test cell configurations (e.g. from work undertaken within PASSYS and the BESTEST programme). In the absence of any supplied information, simulations were carried out with the default ESP-r treatment which uses the Alamdari and Hammond correlations for buoyancy driven flow. However, it may be that the convection coefficients will be substantially higher when a radiator is placed close to a wall as is the case here. For this reason, alternative result sets were submitted for the case of the heated cells (double glazed and opaque cases), using algorithms developed in experimental work carried out at Cardiff University (Khalifa and Marshall, 1990). A statement was made in the associated report that the "assumed coefficients have an important impact on the predictions". No information was supplied in the exercise on the convective regime within the cells (although it was supplied for solar distribution, for example).

Further simulations have now been undertaken using upper limits for internal convection coefficients $(7W\ Im^{2}K)$ identified in a study by Halcrow (1987). The following table summarizes the results.

	Opaque Cell MJ	Double-Glazed Cell MJ
Measured	117. +/- 9	89.5 MJ +/- 7
Average Predicted	103.	76.
ESP-r (Alamdari)	100.8	69.6
ESP-r (Cardiff)	105.9	74.9
ESP-r (Halcrow)	111.4	80.3

Clearly the results, as expected from previous research, show that the use of higher convection coefficients increases predicted energy consumption significantly. It is thought that the different assumptions made by programs regarding convection coefficients may well lead to a significant part of the observed variation between programs. However, it is worth noting that with the upper limit of assumed convection coefficient, the predicted energy consumption is still just below the lower uncertainty band for measured data from the double-glazed cell (although it is within the uncertainty band for the opaque cell).

Solar Modelling. The anisotropic solar distribution algorithm assumed was the current default in ESP-r, the Klucher modeL A small sensitivity study was carried out using an alternative algorithm, the Perez model, available as an option in ESP-r, and which is being considered as a replacement default. The overall effect on heating requirements is small - for the double-glazed cell, the Perez model resulted in a prediction of 67.7 MJ (against 69.6 for the Klucher model). However, it is interesting to note that the Perez model gives better agreement with the measured south facing vertical irradiance (integrated over the 7 days), as shown in the following table.



	Free-floating period MJ/m^2	Heated period MJ/m^2
Measured	83	81
Klucher	78.7	77.7
Perez	83.9	81.7

As can be seen from close study of the attached graph showing measured against the Perez and Klucher predictions, the hour-by-hour predictions with the Perez model are slightly better than with the Klucher model. However, a firm conclusion as to the most suitable algorithm cannot be made from this data alone.

A sensitivity study has also shown that the assumed external absorptance can have a significant impact on the predicted internal air temperatures. Changing the external absorptance over the range 0.10 to 0.30 *(as* suggested in the uncertainties given in Newsletter No. 9) resulted in a change in energy consumption from 70.2 to 67.7 MJ for the double glazed cell (against the base case estimate of 69.6 MJ). A study of internal absorptance resulted in an uncertainty in energy consumption of ±0.7 MJ for the uncertainties given in Newsletter No. 9.

Weather Data. This note is taken from the original submission. The revised weather data sets with tabulated values centred on the hour were used. It should be noted that the way in which the revised data set v110 (October 1987) has been constructed will have significantly smoothed out the peaks and troughs in the data, particularly the solar data. This will have some impact on the time series comparison with measured data, but in the absence of correctly averaged hourly data, this is not possible to quantify. With regard to humidity, in data set v110 the values were missing. Instead, the mean and range of humidity values from the Kew example year for the month of October were used to construct sinusoidally-varying humidity values. For the v099 climate set, the humidity values are obviously incorrect (100% with high temperatures and solar), but these were used as requested. A sensitivity study was carried out with the humidity values reduced by 20%. This decreased the predicted temperatures in the test room and the roof space by 02°C to 0.3°C. This factor, which affects the calculated external longwave exchange, could therefore be of some significance, and may be worthy of further sensitivity studies.

With regard to the supplied solar data, two ESP-r climate files were made up for the revised v110 data set - one using the supplied global horizontal radiation and the other with the supplied direct normal radiation. The resulting predictions were slightly different. It is thought that this difference is caused by an incorrect calculation in the supplied data (probably of the solar altitude). An assumption was made that the measured solar radiation was global horizontal and that direct normal was derived. The simulations were therefore run with the climate data containing global horizontal data.

Heater The original specification requested that the heater be modelled as accurately as possible. This was done by modelling the radiator as a plant component, although it was noted that the time constants of the heater (22 minutes) conflicted with the relatively smoothed solar and other climatic data (hourly averaged data further smoothed with a 2 hour moving average). Two sensitivity runs were undertaken. It was observed that changing the radiator heat output from 680W to 700W caused an increase of $0.2 MJ$ to the total energy consumption; this is not significant. It was also found by conducting similar analysis that the effect of changing the radiator time constant has very little effect on the overall energy consumption.

As mentioned in the original submission, it is known that the position of the radiator can be important in the room response. It is possible to take this into account in ESP-r by subdividing the room and setting up an air flow network. However, this would have involved a substantial increase in the time required for carrying out the study, so unfortunately it was not undertaken - again, further investigation of this aspect would be desirable.

Conclusions

The (limited) sensitivity studies that have been carried out suggest that the the most important uncertainty is associated with the unknown internal convective regime inside the test cells. This factor, well-known to be important from previous studies, may account for a large part of the differences in predictions between the simulation programs. It points to a need for further experimental work to provide guidance on suitable values. In the case of BESTEST we have concluded that ESP-r is operating with a significantly different external convective/radiative regime compared with many other programs.

Additional sensitivity studies have shown the sensitivity of the results to external long-wave exchange, and internal and external shortwave surface absorption. The importance of time-dependent conductivity has not been investigated as yet, although with the high temperatures recorded in the test cells, they may have some influence (as has been found in the case of the PASSYS test cells).

One suggestion for continued work is to take two or three of the simulation programs that give significantly different predictions, and extract the main energy fluxes in order to determine why differences in predictions are so large.

References

Khalifs and Marshall (1990). 'Validation of Heat Transfer Coefficients on Interior Building Surfaces Using a Real-Sized Indoor Test Cell'. *Solar Energy Unit Report,* University of Wales, Cardiff.

Halcrow (1987). 'Heat Transfer at Internal Building Surfaces'. *Report to ETSU,* Sir William Halcrow and Partners Ltd. England.