HISTORY AND DEVELOPMENT OF VALIDATION WITH THE ESP-r SIMULATION PROGRAM

Strachan P A, Kokogiannakis G and Macdonald I A

Energy Systems Research Unit

University of Strathclyde,

75 Montrose Street,

Glasgow, G1 1XJ, UK

Contact - Paul Strachan: paul@esru.strath.ac.uk

Tel: +44 141 548 2041 Fax: +44 141 552 5105

ABSTRACT

It is well recognised that validation of dynamic building simulation programs is a long-term complex task. There have been many large national and international efforts that have led to a well-established validation methodology comprising analytical, inter-program comparison and empirical validation elements, and a significant number of tests have been developed. As simulation usage increases, driven by such initiatives as the European Energy Performance of Buildings Directive, such tests are starting to be incorporated into national and international standards. Although many program developers have run many of the developed tests, there does not appear to have been a systematic attempt to incorporate such tests into routine operation of the simulation programs. This paper reports work undertaken to address this deficiency. The paper summarizes the tests which have been applied to the simulation program ESP-r. These tests have been developed within International Energy Agency Annexes, within CEN standards, within various large-scale national projects, and by the UK's Chartered

Institution of Building Services Engineers. The structure used to encapsulate the tests allows developers to ensure that recent code modifications have not resulted in unforeseen impacts on program predictions, and allows users to check for themselves against benchmarks.

KEYWORDS

Validation, simulation programs.

1. INTRODUCTION

As documented by several authors (e.g. [1]) there have been many international validation studies. The first part of this paper details a large number of important validation studies involving the simulation program ESP-r [2]. It is clear that the validation process is long-term and continuous. As Bloomfield [1] states, claims such as ‘program X has been validated’ should be avoided unless additional qualifying information is provided to give real meaning to it. This qualifying information should include specific information on the accuracy, purpose and situation corresponding to the tests performed. Given the complex nature of detailed simulation programs, it is impossible to prescribe a set of tests that could conclusively “validate” a program, not least because programs are constantly being developed to model new technologies, to address the complexity inherent in increasing use of passive techniques and to include complex interactions of building occupiers with the building fabric and systems.

The various elements of program validation are well established [1],[3],[4] and comprise the following elements:

- Review of theory
- Code checking
- Analytical verification
- Inter-program comparison

- Empirical validation

The first two of these are necessary for any technical software development. To permit future developments and re-use, high quality comprehensive documentation of the theory and its implementation is an essential element for state-of-the-art programs which are too complex for individuals to develop. Several examples of the other validation techniques are given in this paper. It is important to recognise that all these techniques have their advantages and disadvantages and that they should all be deployed to test specific parts of a program as well as the whole program.

The first part of this paper sets out a summary of significant validation studies in which the simulation program ESP-r has been involved. In each case a brief summary of the type of validation is given, with a reference to detailed reports and papers. An evaluation of the study is also given. It can be seen that the early exercises were mostly focussed on empirical validation as this is the most obvious method to test program validity. However, these early studies pointed out the difficulties with experimental studies – the need for high levels of instrumentation, consideration of all heat and mass flow paths/processes, accurate control and minimisation of uncertainty. Following this, a more balanced view was taken showing the complementary nature of the various validation techniques.

The subsequent sections of the paper describe recent developments to address two problems:

a) The fact that many of the validation tests are not persistent. As an example, analytical conduction tests have been applied many times throughout the course of development of ESP-r; each time, new tests have had to be constructed because the models and results have not been integrated into the program. It is clear that embedding validation tests within the program itself allows the possibility of routine application to ensure that program developments have not affected results, as well as the gradual build-up of a comprehensive set of more complex validation cases.

b) New users to simulation need to be trained. There have been several papers detailing the likelihood of increased errors as the number of inputs increases (e.g. [5]), which may be partially alleviated by good interface design. However, many modelling decisions require experience, so there is also a need for tests that make new users consider different approaches to modelling problems, and a need for tests that will give modellers confidence in their ability to generate 'correct' predictions. As modelling and simulation becomes more routine (e.g. as a result of legislation such as the European Energy Performance in Buildings Directive), more modelling studies and modellers to undertake them will be needed. In due course, accreditation of modellers is likely to be necessary. As a first step in this direction, the UK's Chartered Institution of Building Services Engineers (CIBSE) initiated a pilot project to develop a number of tests for program users. Details of the ideas behind these tests are described.

2. PREVIOUS ESP-r VALIDATION STUDIES

Comparison with Scottish test houses

Two 3-bedroom houses in Livingston, Scotland were monitored (about 50 sensors plus climate data) in summer free-floating conditions. One was unoccupied, the other occupied. Predictions of air and surface temperatures showed good agreement with measurements in the unoccupied house. There was high uncertainty in infiltration rates which were only spot measured [6],[7].

IEA Annex 1 (1977-80)

This was an inter-program comparison of 19 different computer programs, including ESP-r, used to simulate the thermal load and energy requirements of commercial buildings. It was the first major international exercise of the type. Although predictions showed wide variations (typically to within ±25% in daily values and ±30% in peak load), there were useful lessons regarding modelling methodology and the level of detail required in the building specification. One of the buildings, the Avonbank building in Bristol was monitored (although not by project participants). Accuracy of model inputs was suspect, so no firm conclusions could be drawn [8],[9].

IEA Annex 4 (1979-82)

This Annex involved the comparison of predicted with measured data from a commercial office building (Collins Publishers Headquarters) in Glasgow. The building was open-plan and air-conditioned (VAV system), monitored with over 500 sensors including automatic tracer gas for infiltration measurements. This was the first major empirical validation exercise, extending over 4.5 years. Nine simulation programs including ESP-r were involved in the study. Agreement was better between programs than between predicted and measured data. Problems in specification and in measurement data were identified. The importance of duct heat transfers, inter-zone airflow and the performance of systems and control in practice were also identified. There were many useful lessons concerned with difficulty of empirical validation on real buildings, but the Annex concluded that because of uncertainties in input data, results could not be used for validation of simulation models [10].

IEA Task 8

ESP-r was compared with 10 other programs against test cell data gathered at the Passive Solar Test Facility of the National Research Council of Canada. ESP-r predictions were within 8% of measured heating energy consumption over a 2-week period. Local overheating maximum temperatures were within 1°C in most cases. In addition, 5 detailed simulation programs were compared for a series of benchmark tests based on residential buildings (precursor to BESTEST - see below). A "reasonably narrow set of ranges in loads and peak temperatures was obtained" [11],[12].

IEA Annex 10 (1984-86)

This Annex was concerned with inter-program comparison of HVAC system simulation programs. The Annex had 2 aims: firstly, to develop a database of component models for air-conditioning and hydronic heating systems; secondly, to undertake simulation exercises on realistic configurations to demonstrate simulation program capabilities. No comparisons were made with measured data, although simulation exercises were based on actual systems. Many programs and studies were involved in the Annex, but typically only 3 or 4 programs for any particular study. Results reported varied in detail. One example is of boiler modelling. Results from 6 models (including ESP-r) gave annual energy consumption within 2.8% of each other, and similar trends were observed for changes to boiler configuration [13],[14],[15].

Comparison with Australian test houses

This was an early comparison of measured and predicted data, for two houses in Australia, one in Townsville (elevated and free-running), the other in Melbourne (heated). The Melbourne house simulation was reasonable compared with the measured data, but the Townsville house gave poor agreement, thought to be due to uncertainties in the modelling of the ventilation [16].

EC Study: various analytical tests

This study was undertaken by EEC experts as part of a selection process for the European reference model in the area of passive design. Dupagne 1983 reported a quasi-theoretical solution for the response of a 1m test cube to a step change in outdoor temperature; ESP-r predictions of internal temperature closely followed the calculated response [17],[18].

SERC validation project (completed 1988)

This was a large project involving 3 programs: ESP-r, HTB2 and SERI-RES, undertaken by the Universities of De Montford (then Leicester Polytechnic) and Nottingham, the Rutherford Appleton Laboratory and the Building Research

Establishment. Validation work included review of theory (focussing on each algorithm and its implementation), analytical verification (solar processing, conduction, convective exchange, view factor calculation and internal longwave exchange), sensitivity analyses and a review of available test data sets [19].

Applicability Study I

This 7 person-year research project was funded by the Energy Technology Support Unit of the UK Department of Energy as part of the Passive Solar Programme. It was undertaken by De Montfort University, with BRE as the major sub-contractor. The project focussed on inter-program comparisons between ESP-r, HTB2 and SERI-RES for passive solar houses. Results indicated that the 3 programs predict similar trends for energy use as geometry, construction type, heating system, thermostat set-point, window type and window orientation are varied. For double glazing or better, the programs predicted annual energy savings to be made by varying window area, orientation and type to be within a resolution of about 7%. Algorithms describing internal heat transfer coefficients and the windows were identified to be primarily responsible for inter-program variability [20].

IEA Annex 21 (1988-93)

This was a comprehensive study concerned with analytical verification, inter-program comparisons (BESTEST) and empirical validation based on data from test rooms. Simulations using ESP-r were undertaken by ESRU (empirical validation) and De Montfort University (other validation studies).

a) Empirical data from small well-controlled and monitored outdoor test rooms were compared with predictions from 17 different programs [21]. Predictions and measurements were made of total energy consumption, maximum and minimum temperatures, vertical solar radiation and hourly temperature profiles. ESP-r predicted within the error bands of measurement for vertical irradiance and maximum and minimum temperatures, but underpredicted for heating energy consumption. Although some programs predicted energy consumption within the error band assigned to measurements, most programs underpredicted. Some causes suggested for this included heater dynamics/interaction with internal convective heat transfer, underestimation of edge losses in the test cell and non-uniform room air temperature. Sensitivity studies in this study (and others) indicated, for such test rooms, the importance of internal convection coefficients. Work on this topic has been addressed in ESP-r [22].

b) BESTEST: This was an inter-program comparison exercise of passive solar spaces. The work included a diagnostic method, based on incremental changes to a base case model, as well as comparisons between predictions from a number of detailed public domain programs from the US and Europe (qualification tests). One of the BESTEST diagnostics identified a problem in ESP-r with internal solar absorptance. Although this had already been identified and corrected, it showed the ability of BESTEST to identify potential sources of program error. In the qualification tests, ESP-r predicted relatively low annual heating loads for some tests. Sensitivity studies showed that the differences with other programs are largely a result of different algorithms for calculating internal surface convection coefficients. Since no definitive algorithms exist, ESP-r results were used in setting reference ranges for the qualification tests [23]. The tests have been incorporated into ASHRAE Standard 140 [24].

c) Benchmarks for Commercial Buildings: This study was an inter-program comparison with 6 programs modelling a simple module of a commercial building in various configurations. Output parameters were annual heating and cooling, hourly integrated peak heating and cooling, peak room air temperatures, and heat losses for windows, exterior walls and ventilation. For annual and peak heating, ESP-r gave approximately 20% smaller values than the mean of all programs. For the majority of other parameters, ESP-r was close to the mean of predictions from all programs [25].

d) Analytical testing. Analytical tests were applied to ESP-r and SERI-RES (using simple zone models) to test for steady state and dynamic conduction, the incidence of direct solar radiation on external surfaces of arbitrary orientation, and the transmission of direct radiation through simple glazing systems. ESP-r calculated energy consumption of unventilated buildings in the steady state correctly, and the worst error in external heat flux due to a step change in temperature was 1.0% in the dynamic response tests. Small errors appeared in ventilated buildings due to ESP-r not taking account of variation of air density with temperature for calculating ventilation heat loss. Good accuracy in calculating solar position and incident direct radiation was reported. Errors up to 0.02 in transmission coefficient (only at high incidence angles) were found, resulting from ESP-r's interpolation algorithm [26].

EC PASSYS project (1986-93)

PASSYS, sponsored by the European Commission (EC), was a large-scale project involving teams from several European countries. The focus was on developing outdoor test cell facilities, with model validation forming a major component of the work. The Model Validation and Development subgroup built on previous work to develop a validation methodology comprising literature review, code checking, sensitivity studies, inter-program comparison, analytical verification and empirical validation. This methodology was applied to ESP-r.

Phase I. Validation studies were undertaken by teams throughout Europe studying individual processes and their implementation within ESP-r: they involved review of algorithms, code checking, inter-program comparison, analytical verification and sensitivity studies and limited process-level experiments. For example, in the case of internal long-wave exchange, the work built upon the BRE/SERC study referred to above. It included a review of different theoretical methods for calculating internal long-wave exchange, analytical tests, sensitivity studies, and an empirical side-by-side experiment. For external longwave processes, the literature review resulted in the implementation of the Berdahl and Martin algorithm in ESP-r.

Phase II. Empirical whole-model validation was undertaken, based on the PASSYS test cells located at 14 test sites in 11 countries throughout Europe. These unoccupied room-sized test cells provided a realistically-sized test environment. However, because of the large thickness of insulation, 2-D and 3-D conduction was found to be important and data from a calibration wall was used to calibrate the ESP-r model. Passive solar components tested included a reference component (double glazed window in insulated wall), the reference component with added mass, a conservatory, transparent insulation, different glazing types and a Trombe wall. Work focussed on developing the methodology for such tests in terms of design of experiments, high levels of instrumentation, quality control on data, and production of high quality data sets. It included uncertainty analysis on measured and predicted data, and residuals analysis (to attempt to explain the causes for differences between measured and predicted data). As an example, in the case of a conservatory experiment of a 15 day test with the conservatory in buffer mode, the mean value of the residuals between measurements and ESP-r predictions for conservatory air temperature was 0.56°C [4],[27],[28].

Comparison of duct system computer models

This project focussed on the selection of public domain computer modelling software for simulating the complex behaviour of ducted air distribution systems used for space conditioning in residential and small commercial buildings. Five programs were selected and subjected to a series of analytical evaluations (3 duct-system-only and one integrated system). Of these, 3 programs, including ESP-r operated by the Florida Solar Energy Center (FSEC), passed the criteria set. For the various tests, ESP-r showed agreement varying from acceptable to excellent; the worst discrepancy observed was 6%.

The three programs were then used in whole building simulations in inter-comparison mode, with simulations undertaken by ESRU and analysis by FSEC. ESP-r air flows and pressures were very well predicted for each simulation; however, some problems were reported regarding predicted energy penalties and delivery and distribution efficiencies. The authors also remarked on the difficulty of ensuring input equivalencing for such complex inter-program comparisons [29].

BRE/EdF validation project: EMC test cells

This empirical validation study was undertaken by BRE and De Montfort University, based on data from the Energy Monitoring Company test cells, with 4 simulation programs used: Apache, Clim2000, ESP-r and SERI-RES. Although ESP-r predictions were slightly closer to measured data than other programs, analysis showed problems with modelling of heater dynamics and with the internal convective heat transfer coefficients. Stratification in the rooms was not modelled [30].

BRE/EdF empirical validation study: BRE office

The study involved a comparison of the monitored performance of an office on the BRE site in Garston against predictions made by several French (CA-SIS and CLIM2000) and UK (Apache, 3TC and ESP-r) programs. ESP-r simulations were undertaken by BRE. The study was conducted in several phases, in the first of which the modellers had no knowledge of the measured building performance. Two separate studies were conducted, both of a pair of unheated offices. There were no window blinds in the operation in the first stage; in the second stage a blind was added to one of the pair of office rooms.

Uncertainties in input values were used to produce prediction error bars for the room temperatures in the no-blinds case, with the width of the band varying from approximately ±2 to ±4°C. Errors observed lay within these error bands. Good agreement among program predictions and measured data was obtained. For the second study, ESP-r predicted a maximum temperature difference between rooms with and without blinds of 2.1°C (measured 3.1°C) and a mean difference of 0.4°C (measured 0.3°C). Uncertainties in internal convective heat transfer coefficients were shown to make the largest contribution to the overall error band [31].

BRE/EdF empirical validation study: BRE house

The study involved a comparison of the monitored performance of a house on the BRE site in Garston against predictions made with the same programs as in the previous study. Again there was a blind validation first stage plus several sensitivity studies.

Overheating produced by the combination of casual gains and solar radiation was reproduced well, with close agreement between measured and predicted peak temperatures. Cooling performance was also well represented, suggesting that thermal and heat loss effects are represented in the correct ratios. However, there were clear differences between program predictions for whole house energy consumption. Of interest is that in some cases, energy consumptions in the upstairs and downstairs zones were acceptable only due to fortuitous cancellations of errors occurring in the two zones, and that cancellation also occurred between errors on successive days. For example, ESP-r overpredicted whole house consumption by 9%; however, downstairs consumption was underpredicted by 4% and (the smaller) upstairs consumption was overpredicted by 44%. The cause is likely to be due to incorrect modelling of infiltration and air movement in the house, but as only whole house infiltration was measured, this could not be confirmed [32].

BRE/EdF empirical validation study: Lisses house

The study involved a comparison of the monitored performance of the Valériane house at Lisses in France against predictions made by French (CLIM2000) and UK (Apache, 3TC and ESP-r) programs. ESP-r simulations were undertaken by BRE; again there was a blind validation first stage, plus several sensitivity studies.

Comparison of whole-house energy consumption over the complete experimental period (more than two winter months) revealed errors ranging from -4% to +26%. Agreement was considered quite reasonable, but again disaggregated figures for upper and lower floors gave less satisfactory agreement; so processes connecting upstairs and downstairs zones are not so well modelled. A detailed sensitivity analysis indicated an uncertainty band of approximately ±12%. All programs predicted values outside these ranges when results from individual phases and zones were considered [33].

Daylighting study

A study was carried out by the Fraunhofer Institute in Freiburg, Germany, involving the prediction of daylight distribution. Six methods were compared with a reference case. ESP-r performed very well for the office space, but a recommendation was made that ground daylight coefficients would improve ESP-r's predictions for certain more complex building geometries [34].

IEA Task 22 - RADTEST

A set of inter-program comparison tests were developed as an extension to the BESTEST suite of tests, to assess the ability of 5 programs (including ESP-r) to model radiant heating and cooling systems. For some tests there were considerable variation between program predictions, although changes from one case to another were generally uniform [35].

HERS BESTEST

ESP-r was applied to The Home Energy Rating System inter-program comparison tests [36, 37]. The agreement in predictions with those of the other programs (SERI-RES/SUNCODE 5.7, DOE-2.1E, BLAST 3.0) was good, with major differences attributed to different simulation inputs.

IEA Task 22 – HVAC BESTEST

A series of inter-program test cases was developed to assess simulation modelling of steady-state [38] and transient [39] performance of unitary vapour-compression air-conditioning systems. Very good agreement was reported for both test sets ([40] and [41] respectively) between Hot3000/ESP-r predictions and those from the other five simulation programs included in the testing. Initial simulation results for two test cases with very high outside air intake resulted in an improvement in the program. Other differences in simulation programs were reduced by using smaller simulation timesteps.

Tests were also carried out on furnace modelling with the Hot3000/ESP-r program and compared with Energyplus and DOE-2.1E, and for some cases, analytical solutions [42]. The results obtained indicate a high correlation between all programs and the analytical results. For the cases without an analytical solution, there was slightly more variation between program predictions.

Network airflow modelling

Three multi-zone airflow models, COMIS, CONTAM and ESP-r, were compared with each other, against experimental data from a controlled test environment and against field measurements from a residential building. For the controlled environment, there was good agreement between the three programs, but sometimes significant differences between predicted and measured airflow rates from one zone to another and sometimes for total airflow rates in each zone. However, for the field tests, good agreement was found between the three programs and the measured data [43].

CEN standards

In the forthcoming CEN standard for cooling loads [44], a comprehensive set of tests are set out in an Appendix. The intention is that programs will have to demonstrate compliance with the tests (within the given tolerance bands) for them to be acceptable for use in predicting cooling loads. ESP-r was run on the proposed tests during their development. Other CEN standards, for example for calculation of energy use for space heating and cooling, will follow a similar strategy in setting out tests for program compliance checking.

Validation as part of PhD theses

Validation has formed an essential element in the development of much of the functionality of the ESP-r program, as shown in the following list.

- Analytical verification of an individual plant component (oil-filled radiator) and a plant network and inter-program comparison of a cooling coil model [45].
- Analytic test and (qualitative) comparison against empirical data of an adaptive convection algorithm. Inter-program comparisons, sensitivity studies and comparison with empirical data of the implementation of the zero-equation turbulence model and alternative wall functions in the CFD domain [22].
- Analytical testing on a network airflow model, inter-program comparison of a boiler model and an empirical validation of a radiator [46].
- Analytical and inter-program comparisons for ESP-r's electrical power flow model [47].
- Analytical verification of building and plant-side controllers, inter-program comparisons of building-side, plant-side and global controllers and an empirical validation of the control of an AHU chiller [48].
- Analytical, inter-program and empirical validation of adaptive gridding; inter-program and empirical validation of variable thermo-physical properties; analytical test of combined heat and moisture transfer [49].
- Analytical verification for 2-D flow in a duct and inter-program comparison for natural (2-D and 3-D) and forced (2-D) ventilation [50].
- Analytical, comparative and empirical validation of contaminant modelling [51].

3. ENCAPSULATED TESTS FOR PROGRAM VALIDATION

As mentioned in the introduction, despite the large effort that has gone into validation, the studies and tests have not persisted. In most cases, program developers or users who wish to investigate specific tests would need to re-establish models – often a laborious task in validation studies where each input parameter must be carefully considered. In addition, most simulation programs are under constant development, so although the program may have "passed" at one stage, there is no guarantee that subsequent changes have not affected predictions.

A structure for encapsulating validation tests was developed within ESP-r, to act as a quality assurance tool for simulationists and as a validation check for program developers after code modifications [52]. To demonstrate the usefulness of the facility, the authors added a general analytical test for dynamic heat transfer through opaque multi-layered building constructions. Solutions were presented for a step change in internal or external temperatures. Either surface or air temperatures or adiabatic conditions can be specified for the inside or outside conditions. Users can define construction multi-layer thermophysical properties, initial conditions, boundary types, simulation duration, timesteps per hour and the location monitored. From this information, a thermal zone is automatically created, a simulation performed and results extracted for comparison with the analytical solution.

The inter-program BESTEST comparison test suite discussed above has now been added to the structure developed by Ben-Nakhi and Aasem. The tests involve the creation of a progressive series of simple models which are formulated to test specific algorithms of simulation programs. For example, two separate models are created, one with low internal emissivity and one with high internal emissivity. The difference in predictions from the two models can be used to evaluate internal longwave calculations.

The main results obtained for each test case are the annual cooling and annual heating load, and the peak heating and peak cooling load. The tests are grouped into high mass and low mass cases, and classed as either diagnostic tests or qualification tests. Some tests require more specific data (either annual or hourly for a specific date) and there are four free float tests where maximum, minimum and average annual temperatures are compared instead of loads.

Using ESP-r, the user can access the tests where they have the choice to run a specific group of tests, run individual tests or run all the tests. After selecting the models to be run, simulation and results analysis are automatically invoked. Results can be displayed or sent to an external file. They are shown with the published maximum and minimum limits as a check for users (see Figure 1). It is also possible to display the results from a previous validation run so that program developers can use the tests as a benchmark for checking whether code modifications have had any impact on predictions.

4. USER TRAINING TESTS

Although it is clear that program developers should always carry out code checking and validation, this is not always possible for users. For example, some analytical tests may be difficult to implement as they can often require specialist knowledge to reproduce the exact boundary conditions. Similarly, empirical validation is best performed by specialists because of the difficulty of representing the exact experimental conditions and because of the time-consuming nature of such studies. Inter-program comparisons can be suitable for users, but because of the number of tests required to comprehensively examine all aspects of a program, only a subset of available tests may be suitable. Also, inter-program comparison tests are usually based on simple buildings so they do not test a user’s ability to conceptualise the best way of modelling a given building.

Therefore, it is believed that user tests should have a different focus – they should be relatively simple to apply, they should be tests for which suitable tolerance bands can be established (probably by inter-program comparison studies), and they should include some examples where users need to consider different modelling approaches (e.g. multi-zone problems).

CIBSE carried out market research on publications and design software and found that a sizeable proportion of members and non-members believe that the software they use accords with CIBSE methods that are set out in the CIBSE Guides. Some also believe the software they use is accredited by CIBSE. These findings prompted CIBSE to develop standard tests to assess design software packages. These are intended to provide a means by which members could test for themselves that the software they use is producing results consistent with those produced by CIBSE methods and with good practice.

CIBSE therefore commissioned the development of a suite of tests, with standardised input data, example results and expected tolerances [53]. This set of simple tests is intended to develop a culture of software testing and validation in the industry. The main focus is on thermal performance of buildings. Since the target audience is program users, the set of tests were developed with the intention of finding a balance between comprehensiveness and ease of application. It is likely that in the future, CIBSE will expand and update the tests. Such tests could also form the basis of user accreditation schemes.

The tests include solar position, basic thermal calculations, solar shading, glazing properties, solar cooling loads, psychrometric properties, interstitial condensation, steady state heat loss, dynamic cooling loads, infiltration and ventilation, and summertime temperatures. To ease the burden on the user, climate, constructions etc used in earlier tests are re-used in later tests.

5. CONCLUSIONS

Validation of building energy simulation programs is a continuing process. Much work has been carried out, as evidenced by the series of studies summarized in this paper. However, there is an urgent need to address two problems to help improve confidence: firstly, to enable validation tests to become persistent by embedding them within the simulation programs, and secondly, to develop a series of user tests to help train, and perhaps in the future to accredit, new users.

A number of benchmark tests with acceptable tolerance bands are being generated within European (CEN) standards (e.g. for peak cooling load predictions) and adopted within US standards (e.g. BESTEST). It is important that program developers make it easy for users to confirm that programs conform to such standards, and this paper has described the first steps towards embedding such a facility within the program ESP-r. Future extensions are planned, in particular using the multi-zonal inter-program comparison and empirical validation tests being developed in the current IEA Task 34/43.

There are also clear indications that modelling and simulation will have a more central role in the design of energy efficient buildings, notably with the adoption of the European Energy Performance of Buildings Directive. This will create the need for more modellers and therefore their training. To this end, some progress has also been made, as reported in this paper, to create a series of user tests.

REFERENCES

[1] Bloomfield D P, An Overview of Validation Methods for Energy and Environmental Software, ASHRAE Trans. Vol 5, Part 2, SE-99-6-1, 1999.

[2] ESRU, ESP-r, http://www.esru.strath.ac.uk, 2005.

[3] Judkoff R, Wortman D, O'Doherty R and Burch J, A Methodology for Validating Building Energy Analysis Simulations, SERI/TR-254-1508, Golden, Colorado, USA: Solar Energy Research Institute, now National Renewable Energy Laboratory, 1983.

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

[5] Chapman J, Data Accuracy and Model Reliability, Proc. Building Energy Performance, Canterbury, April 1991.

[6] Clarke J A and Forrest I, 'Validation of ESP against Test Houses', ABACUS Occasional Paper No 61, 1978.

[7] Clarke J A, Energy Implications in Building Design: A Thermal Simulation Design Model, Proc 3rd Int. Symp. Use of Computers for Environmental Engineering related to Buildings, Banff, Canada, pp3-20, 1978.

[8] Oscar Faber and Partners, IEA Annex 1 Computer Modelling of Building Performance: Results and Analyses of Avonbank Simulation, Oscar Faber and Partners, St Albans, UK, 1980.

[9] US Dept of Energy, IEA Annex 1: Comparison of Load Determination Methodologies for Building Analysis Programs, DOE/CE/20184-1, US Dept Energy, Washington DC, 1981.

[10] BRE, Glasgow Commercial Building Monitoring Project: Final Report, IEA Annex 4, Building Research Establishment, Watford, ISBN 085125 070X, May 1984.

[11] Morck O C (ed), Simulation Model Validation using Test Cell Data, Thermal Insulation Laboratory, Technical University of Denmark, June 1986.

[12] Bloomfield D P, Test Cases for Evaluating Building Thermal Prediction Programs, BEPAC Research Report, 1990.

[13] McLean D J and Clarke J A, 'Results of the Collins Building VAV Air Conditioning Simulation', IEA Annex 10 Interim Report, Apr 1985.

[14] Lebrun J and Liebecq G, Annex 10 System Simulation Synthesis Report, University of Liege, Report No. AN10 881020-RF, 1988.

[15] Morant M-A, Annex 10 System Simulation: The La Chaumiere exercises. Report AN10 860327 - 01, 1986.

[16] Williamson T, 'The Goodness of Fit between ESP Predictions and Monitored Data from Two Australian Test Houses', ABACUS Occasional Paper, University of Strathclyde, 1984.

[17] Archard P and Gicquel R (eds), 'Basic Principles and Concepts for Passive Solar Architecture', European Passive Solar Handbook, CEC, 1987.

[18] Dupagne A, Comparison of Passive Solar Design Methods within EEC Program, Proc Conf on Design Methods for Passive Solar Buildings, UK ISES, Section 1, pp1-7, 1983.

[19] BRE (Building Research Establishment) and Science and Engineering Research Council, An Investigation into Analytical and Empirical Validation Techniques for Dynamic Thermal Models of Buildings, Vols 1-6, 1988.

[20] Lomas K J, Applicability Study I: Executive Summary, ETSU Report S1213, 1992.

[21] Lomas K J, Eppel H, Martin C and Bloomfield D, Empirical Validation of Thermal Building Simulation Programs using Test Room Data', IEA Annex 21/Task 12 Project, Final Report, Vols 1,2 and 3, Sept 1994.

[22] Beausoleil-Morrison I, The Adaptive Coupling of Heat and Air Flow Modelling within Dynamic Whole-Building Simulation, PhD Thesis, University of Strathclyde, 2000.

[23] Judkoff R and Neymark J, IEA Building Energy Simulation Test (BESTEST) and Diagnostic Method, IEA Annex 21/Task 12 Co-operative Project, Final Report, Feb 1995.

[24] ANSI/ASHRAE, Standard 140-2001, Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs, ASHRAE, Atlanta, Georgia, 2001.

[25] Haapala T, Kalerna T and Kataja S (eds), Energy Analysis Test for Commercial Buildings (Commercial Benchmarks), IEA Annex 21/Task 12 Co-operative Project, Report 101, Tampere, Finland, 1995.

[26] Martin CJ, Seale CF and Eppel H, Analytical Testing of Detailed Thermal Simulation Programs, Building Environmental Performance '94, University of York, ppS-27-32, April 1994.

[27] Strachan P, Model Validation Using the PASSYS Test Cells, Building and Environment, 28(2), pp153-165, 1993.

[28] Jensen S O, Empirical Whole Model Validation Case Study: the PASSYS Reference Wall, Building Simulation '93, Adelaide, Australia, pp335-341, 1993.

[29] Gu L, Cummings J E, Swami M V, Fairey P W and Awwad S, Comparison of Duct System Computer Models That Could Provide Input to the Thermal Distribution Standard Method of Test, FSEC-CR-929-96, ASHRAE Project 852-RP, 1996.

[30] Bloomfield D, Candau Y, Dalicieux P, Dellile S, Hammond S, Lomas K J, Martin C, Parand F, Patronis J and Ramdani N, New Techniques for Validating Building Energy Simulation Programs, Building Simulation '95, Madison, USA, pp596-603, 1995.

[31] Bloomfield D, BRE Office Empirical Validation Study, Report v2-breoff-5a, BRE, Watford, UK, 1999.

[32] Bloomfield D, BRE House Empirical Validation Study, Report v2-bre18a, BRE, Watford, UK, 1999.

[33] Bloomfield D, Lisses House Empirical Validation Study, Report v2-brelisses-3a, BRE, Watford, UK, 1999

[34] Reinhart C F and Herkel S, An Evaluation of Radiance Based Simulations of Annual Indoor Illuminance Distributions due to Daylight, Building Simulation '99, Kyoto, Japan, Sept. 1999.

[35] Zweifel G and Achermann M, RADTEST – The Extension of Program Validation Towards Radiant Heating and Cooling, Building Simulation '03, Eindhoven, Netherlands, pp 1505-1511.

[36] Haddad K H and Beausoleil-Morrison I, Results of the HERS BESTEST on an Energy Simulation Computer Program, ASHRAE Trans. Paper CI-01-10-1, 2001.

[37] Judkoff R and Neymark J, Home Energy Rating System: Building Energy Simulation Test (HERS BESTEST), User's Manual, Volume 1, Tier 1 and Tier 2 Tests, National Renewable Energy Laboratory, Golden, Colorado, USA, 1995.

[38] Neymark J and Judkoff R, International Energy Agency Building Energy Simulation Test and Diagnostic Method for Heating, Ventilation and Air-Conditioning Equipment Models (HVAC BESTEST), Volume 1, Cases E100-E200, NREL/TP-550-30152, National Renewable Energy Laboratory, Golden, Colorado, USA, 2002.

[39] Neymark J and Judkoff R, International Energy Agency Building Energy Simulation Test and Diagnostic Method for Heating, Ventilation and Air-Conditioning Equipment Models (HVAC BESTEST), Volume 2, Cases E300-E545, NREL/TP-550-36754, National Renewable Energy Laboratory, Golden, Colorado, USA, 2004.

[40] Haddad K H, An Air-conditioning Model Validation and Implementation into a Building Energy Analysis Software, ASHRAE Trans., 110(2), 4697, June 2004.

[41] Haddad K H, Application of the IEA HVAC BESTEST Suite of Test Cases for the Validation of an Air-conditioning Model, Proc. of the Canadian Building Energy Simulation Conference (eSIM), Vancouver, British Columbia, June 2004.

[42] Purdy J and Beausoleil-Morrison I, HVAC BESTEST: Fuel-fired Furnace Test Cases, IEA Task 22 Subtask C Report, NRCan, Ottawa, 2003.

[43] Haghighat F, Development of a Procedure to Evaluate the Air Leakage Distribution from Fan Pressurization Test – Validation of Three Airflow Models, Digital Library of Construction Informatics, paper w78-2003-137, 2003.

[44] CEN EN xxxx: 2006, Thermal Performance of Buildings – Sensible Room Cooling Load Calculation – General Criteria and Validation.

[45] Aasem E O, Practical Simulation of Buildings and Air-Conditioning Systems in the Transient Domain, PhD Thesis, University of Strathclyde, 1993.

[46] Hensen J L M, On the Thermal Interaction of Building Structure and Heating and Ventilating System, Doctoral dissertation, Eindhoven University of Technology, 1991.

[47] Kelly N J, Towards a Design Environment for Building-Integrated Energy Systems: The Integration of Electrical Power Flow Modelling with Building Simulation, PhD Thesis, University of Strathclyde, 1998.

[48] MacQueen J, The Modelling and Simulation of Energy Management Control Systems, PhD Thesis, University of Strathclyde, 1997.

[49] Nakhi A E, Adaptive Construction Modelling within Whole Building Dynamic Simulation, PhD Thesis, University of Strathclyde, 1995.

[50] Negrao C O R, Conflation of Computational Fluid Dynamics and Building Thermal Simulation, PhD Thesis, University of Strathclyde, 1995.

[51] Samuel A A, On the Behaviour of Contaminant Behaviour Prediction within Whole Building Performance Simulation, PhD Thesis in prep, University of Strathclyde, 2005.

[52] Ben-Nakhi A and Aasem E O, Development and Integration of a User-Friendly Validation Module within Whole Building Dynamic Simulation, Energy Conversion and Management, 44 (1), pp53 – 64, 2002.

[53] Macdonald I, Strachan P and Hand J, CIBSE Standard Tests for the Assessment of Building Services Design Software, CIBSE TM33, ISBN 1 903287 48 0, 2004.

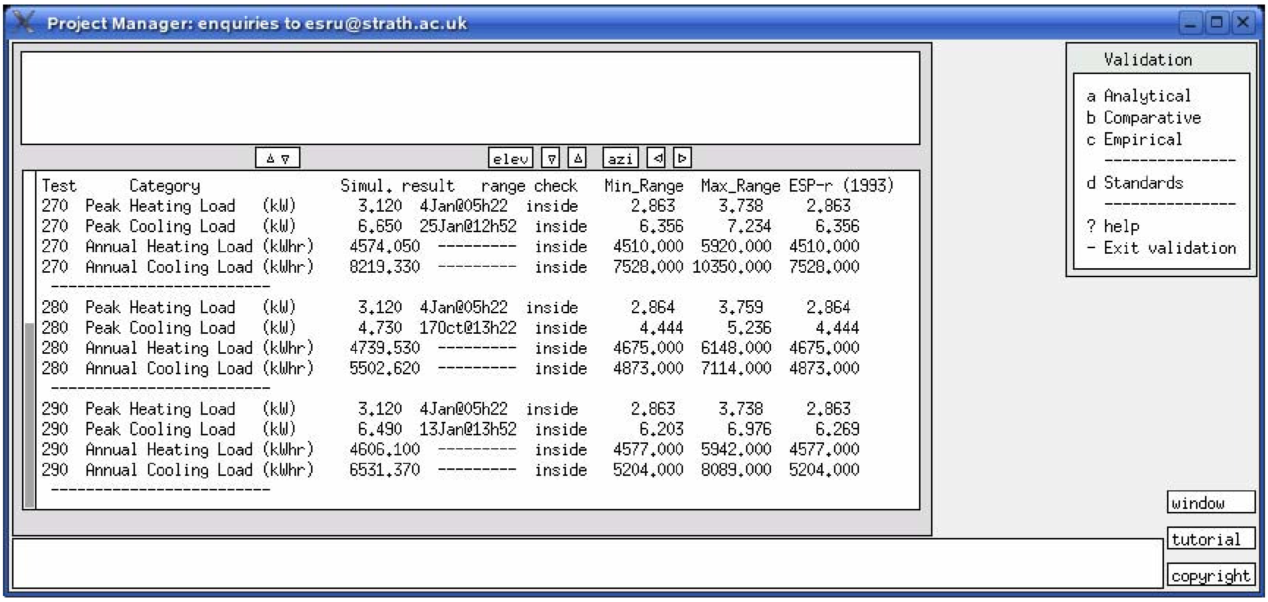

Figure 1 Encapsulation of BESTEST Tests within ESP-r's Project Manager