Summary A methodology for the validation of dynamic thermal models of buildings has been presented. The three techniques, analytical verification, inter-model comparisons and empirical validation have been described and their relative merits assessed by reference to previous validation work on ESP, SERI-RES, DEROB and BLAST. Previous empirical validation work on these models has been reviewed. This research has shown that numerous sources of error have existed in previous studies leading to uncertainty in model predictions. The effects of these errors has meant that none of the previous empirical validation studies would have produced conclusive evidence of internal errors in the models themselves. An approach towards developing tests to empirically validate dynamic thermal models is given.

Empirical validation of dynamic thermal computer models of buildings

N. T. BOWMAN, BSc, PhD, CEng, CPhys, MIEE, MInstP, MCIBSE, MIOA, and
K. J. LOMAS, BSc, PhD.

1 Introduction

Validation is concerned with testing the validity of a model's theoretical basis and its ability to reproduce observed performance. Models, of which there are more than 300 in existence, vary principally in their complexity and the hardware requirements for their use. As a general rule, increasing complexity is directly proportional to the facilities provided and cost of the hardware, and inversely proportional to the ease of use, although there are signs that this situation is changing for the better.

A convenient way of ranking models has been proposed by Burch[1] who suggests that models can be categorised according to the level of detail in the energy problem to be investigated. Three levels are identified: the mechanism level; the building level; and the utilities level. In this study, our central interest is in understanding more about those factors which determine the accuracy to be expected from models which work at the mechanism level. These dynamic thermal models work from first principles, require a very detailed building description, and have a variety of output options that include the variation with time of internal temperatures, heating loads and energy fluxes.

Other types of models are not considered since they are often simplified derivatives of these more complex models and they are likely therefore to be inherently less accurate.

A major research initiative at the Solar Energy Research Institute (SERI), Judkoff[2], has shown conclusively that there is a need for a structured approach to validation. The SERI validation methodology uses three types of investigatory methods, each designed to reveal errors in the modelling process. These methods are: (1) analytical verification; (2) inter-model comparisons; and (3) empirical validation.

This 'SERI methodology' is similar to that adopted by the UK collaborative research project which is investigating validation techniques. The principal research organisations involved in this collaborative research project are:

Leicester Polytechnic, funded by the Science and Engineering Research Council (SERC)—Empirical emphasis;

Nottingham University (SERC funded)—Analytical emphasis;

the *Building Research Establishment*—Analytical emphasis and inter-model comparisons; and

the *Rutherford Appleton Laboratory* (RAL)—Analytical emphasis, sensitivity studies and implementing models.

Further information on this research programme may be found in Bloomfield[3].

The main aim of the UK research is to produce a package of tests which can be used to validate dynamic thermal models of buildings. The package should contain: the data necessary to implement the tests; guidance on any model modifications necessary to implement the tests; a list of the algorithms exercised by the tests; and the answers which the models should produce with a statement about the accuracy to be expected.

An example of a data base of this type, has already been established at the SERI, and contains the model input data and the actual measured building response data from three test cells and five passive solar residences located in four climatic regions of the USA. The data is sufficient to test the ability of a model to predict average monthly performance, Busch[4]. A more detailed set of data would be required for the validation of simulation algorithms.

2 The models

Dynamic thermal models differ principally in their choice of explicit deterministic models for the mechanisms included, and in the numerical method of solving the resulting equations. The choice of models used to develop the validation tools should ideally be representative of the variety of solution techniques used. Other considerations in identifying suitable codes have been:

current use in industry (or research institutions);
source code availability;

documentation standard;
user characteristics;
ease of implementation on the RAL PRIME computer;
previous validation studies; and
energy analysis capabilities.

The codes which are presently being considered, and the solution techniques they use, are:
ESP (Finite Difference, Implicit), Clarke[5];
SERI-RES (Finite Difference, Explicit), Palmiter[6];
BLAST (Response Factor), Control Data Corporation[7];
DEROB (Finite Difference, Implicit), Higgs[8];
HTB2 (Finite Difference, Explicit), Alexander[9].

It is expected that more models will be involved, if it can be demonstrated that they will increase the variety of techniques, although not all codes will be used to the same extent as the research programme evolves.

It is interesting that Littler[10,11], after reviewing some 50 models, concluded that BLAST, DEROB, ESP and SUNCODE (which is essentially the same as SERI-RES), were potentially the most suitable for evaluating passive solar designs. Other studies into passive solar design methods have also adopted ESP (Dupagne[12,13]) and SUNCODE (Dupagne[13]) as large reference simulation models.

The models used in this study have been used primarily by their authors for a wide range of design problems: e.g. ESP (Clarke[14,15]); DEROB (Arumi[16,17]); SERI-RES (De Kieffer[18], Ruyssevelt[19]); and BLAST (Siminovitch[20], Burch[21]). An important research use of the models has been in the development of simplified design tools. The predictions of the large simulations being used as a truth model against which to test the performance of simplified microcomputer of hand calculator methods. Specifically, ESP has been used by the Commission of the European Communities' European Passive Solar Modelling Sub-group, to assess the accuracy of simplified design procedures (Dupagne[22]); a similar use has been made of BLAST (Sullivan[23], Derickson[24]).

Three possible techniques have been identified for assessing the accuracy of dynamic thermal models of buildings namely, analytical verification, inter-model comparisons and empirical validation (Judkoff[2]). The next section briefly examines and describes each technique and outlines some studies in which these techniques have been applied to the models BLAST, DEROB, ESP and SERI-RES, SUNCODE or the earlier version SUNCAT.

3 Analytical verification

In analytical verification, the predictions of the model are compared with carefully designed problems with known analytical solutions. This technique is severely limited because of the small range of problems for which exact analytical solutions can be formulated. The solutions do, however, provide an exact truth model against which to compare the predictions of the models.

To date, the only fully documented truly analytical tests, of which the authors are aware, are those developed by SERI. These tests verify conduction and solar charging of mass mechanisms in the model. They were selected because these mechanisms are important to building energy flow. The tests are reasonably easy to implement but they are limited to the problem of one-dimensional heat transfer[2].

The tests have been performed with SUNCAT-2.4 (Wortman[25]) and SERI-RES itself (Haves[26]). The agreement between the predictions of the models and the exact analytical solution was within the precision to which SERI-RES prints its output.

When DEROB 3 and BLAST 3.0 were tested they showed a slower than expected response in internal temperature to a step-function change in external temperature. A similar slow response to solar radiation was also exhibited by the models. The problem in DEROB 3 was isolated and rectified. This version, and DEROB 4, now accurately reproduce the trends of the analytical solutions. The problem with BLAST 3.0 was being investigated (Wortman[27]).

For the sake of completeness, it is worth noting that the models DOE-2.1 and CALPAS 3 have successfully reproduced the analytical results (Wortman[27], Atkinson[28]).

A quasi-theoretical solution has been derived for the response of a real 1 m test cube to a step change in outdoor temperature. The response curve for inside temperature predicted by ESP closely followed the 'theoretical' curve (Dupagne[12]). The differences between the two curves is, however, as likely to be due to a bad 'theoretical' estimation as due to errors in the model. The 'theoretical' final steady inside temperature and that predicted by ESP was 9.5°C, SUNCODE predicted a value of only 8.5°C.

4 Inter-model comparisons

In inter-model comparisons, the predictions by two or more models of the thermal performance of some hypothetical building are compared. Such studies are sometimes termed software-software comparisons in the USA. The advantages of these types of study are that they are simpler than the other techniques, and any complexity of the building or any climate regime can be chosen. The principal disadvantage is that there is no absolute truth model against which to compare the predictions.

In work for the International Energy Agency (IEA), the ability of ESP, and ten other European models, to predict the hourly heating and cooling loads during a summer and winter period in a large commercial building (the Avonbank Building, Bristol, UK) have been compared (Oscar Faber & Partners[29]). There were marked differences in the predictions by the models, however, because of the complexity of the building the data input to each of the models was not exactly the same. Similar problems complicate the later study of the Collins Building, Glasgow, UK (Department of the Environment[30]).

DEROB has been used to predict the heating loads for a residence, a lightweight commercial building and a heavyweight commercial building each operating in four different modes (Merriam[31]). These predictions were compared with those of TRNSYS, another large simulation model. Although the weekly and monthly loads were in good agreement, hourly predictions varied significantly. The study was, however, clouded by the inability of DEROB to handle the chosen residence explicitly.

A great deal of effort was directed towards ensuring input equivalence among the codes SUNCAT 2.4, BLAST 3.0 and DEROB 4 in the comparative study undertaken by SERI (Judkoff[32,33]). In phase one of their comparative study, there was little difference in the annual cooling loads predicted for fictitious high and low mass buildings subjected to weather data from Madison, Wisconsin. There was some difference, however, in the peak hourly cooling load predictions (even discounting the predictions of the early version of DEROB 3). With slight changes in the buildings' thermal properties the differences in the annual heating and cooling loads of both high and low mass cases became more marked. The greatest differences (23 per cent for annual heating load and 20 per cent for the cooling load) were revealed using weather data from Albuquerque, New Mexico. With the Albuquerque data the solar energy input

dominates rather than conductive heat loss when using the Madison data. Bearing in mind the rigour with which the studies were performed, and that the models successfully passed the analytical tests, the most likely explanation of the discrepancies appears to be due to modelling the dynamic interaction between mechanisms, rather than as a result of mishandling any of the major mechanisms. Since there is no absolute truth model in inter-model comparisons it is not possible to ascertain which, if any, of the models is performing correctly.

5 Empirical validation

5.1 The advantages and disadvantages of empirical validation.

Empirical validation, sometimes called software-hardware comparisons in the USA, is the ultimate stage in any validation process. This technique has the greatest potential for assessing whether the approximations and operations in the model are adequate to predict the measured building response. In principle, the method can provide an absolute truth model but, unlike analytical verification, it is not limited to simple buildings. The potential power of this technique, together with the fact that the process is comparable with those that prevail when the model is used in design studies, means that it is widely used for validating dynamic thermal models; over 130 comparisons of actual building performance with predictions made by thermal models have been disclosed by a recent literature search at Leicester Polytechnic (Lomas[63]).

Simply stated, empirical validation is the comparison of the predictions of the model with physical reality. This simple statement however, belies the problems associated with undertaking such studies. The fundamental difference between empirical validation and the other two techniques is that it involves experimentation, with all its attendant problems. To be considered are such questions as—What shall I measure? Where and how often shall I measure it? How accurately shall I measure it?

The primary thrust of this paper is to propose and justify a procedure for empirical validation. Much can be learned about the attendant problems from a careful examination of the empirical validation undertaken to date on ESP, DEROB, SERI-RES and BLAST. Firstly, however, it is necessary to examine the sources of error in the empirical validation process and give some indication of their relative magnitude.

5.2 Sources of error in empirical validation

The sources of error can be divided into two categories, internal errors, and external errors (Judkoff[2]). Internal errors are due to inaccuracies in the modelling and numerical solution techniques adopted by the model and due to coding errors. External errors occur in gathering the model input data, in transferring this data to the model, in measuring the building response and in comparing measured and predicted values. A listing of possible error sources is given in Table 1.

The greatest uncertainty is probably introduced by the occupants of buildings. Wagner[34] describes studies in which the energy use in nearly identical buildings has varied by as much as 40:1 due to occupant effects. Since the effects which occupants have on the operation of a building cannot at present be modelled explicitly, large discrepancies are likely to arise between the predicted and measured performance in occupied structures. Following a review of over 24 studies of occupied buildings, comprising of about 100 simulations of building energy use, performed by a variety of users with 18 models, each using anywhere

from 1 to 243 buildings, Wagner[34] notes that, 'the availability of accurate and sufficiently complete input data, especially on occupant behaviour, limits the ability of even detailed models to accurately predict energy use, in some cases severely so'.

The lack of air infiltration measurements has beguiled many empirical validation studies. For example, a data set from a direct gain test cell monitored at the Los Alamos Scientific Laboratory during 1976/1977 lacked such values. Consequently, in validation studies of six models, SUNSPOT (Hunn[35], Wray[36]), DEROB (Arumi[37,38,39]); UWENSOL (Emery[40]), DOE-2 (Schnurr[41]), CALPAS 3 (Atkinson[42]) and a hand calculator model (Goldstein[43]), values had to be assumed. The modellers chose values ranging from 0.25 to 3.0 air changes per hour, and they all reported good agreement between measured and predicted values. Subsequently, sensitivity analyses performed with BLAST, showed that a change of infiltration rate of 50 per cent produced a 4°C change in the predicted internal air temperature (Bauman[44]). Since it is hard even to make a 'best guess' at missing air infiltration rates, claims about the predictive accuracy of models when infiltration rates are absent must be treated with great scepticism.

The effect of missing climate data depends on the parameter in question and the structure being considered. For the Los Alamos cell mentioned above, assumptions about missing wet bulb temperatures, wind direction and pressure measurements had little effect on the predictions of internal air temperature made by BLAST (Bauman[44]). The direct/diffuse split of solar radiation (which also had to be assumed) had, however, a marked effect on the internal temperature of this small, highly glazed, cell.

Using high quality data from a four-zone, single storey ranch house in Colorado, USA, Judkoff[45] demonstrated that input assumptions based on standard engineering references, such as wall conductance, can cause errors in auxiliary load prediction of the order of 60 per cent.

5.3 Previous empirical validation work on ESP, SERI-RES, DEROB and BLAST

5.3.1 ESP

Four empirical validation studies are examined. In the first of these, which was undertaken by the program author as part of its development (Clarke[46,47]), an unoccupied unheated terraced house (number 74) and an adjacent occupied heated terraced house (number 73) in Livingstone, Scotland, were monitored.

In house number 74, the internal air temperature predictions were compared with the measured data for two of the rooms and in general good agreement was reported. Discrepancies between measured and predicted values did occur twice during the 11 day monitoring period. This was thought to be due primarily to over-injection of solar radiation, made possible because the solar data was collected at a site two miles away from the actual houses. In house number 73, surface temperatures in one room were predicted after the model was fed with the measured room air temperatures. This latter study exercises only a small part of the model and only weak conclusions can be drawn.

ESP has been involved in three studies undertaken as part of the IEA Energy Conservation in Buildings and Community Systems work programme. In the Annex 1 study (Oscar Faber & Partners[29]), the predictions of ESP (and 10 other models) were compared with the measured hourly heating and cooling loads of a large, three-floor, occupied, commercial building, the Avonbank Building, located near Bristol, England. The accuracy with which energy usage was measured was such that errors of ± 25 per cent in the daily energy usage and ± 30 per cent in the

Table 1. Sources of external error.

MODEL INPUT DATA	Climate Data	• Some (or all) climate data taken at a remote site. • Frequency of measurement insufficient to define variable. • Finite accuracy of measurements.
	Site Data	• Unmeasured shading objects. • Ground reflected radiation or reflectivity not defined.
	Building Data	• Inadequate description of building geometry and construction. • Uncertain workmanship. • Infiltration and/or advection rates not measured. • Use of handbook rather than measured thermophysical properties. • Temperature of adjacent unmodelled zones not defined. • Finite accuracy of measured values.
	Plant Performance Data	• Uncertain radiant/convective split. • Uncertainty in point of heat input. • Uncertain thermostat tolerance.
	Occupancy	• Interference with the building system. • Uncertain wild gains from appliances. • Ill-defined occupancy profile • Uncertainty in modelling furnishings.
	User Interface	• Blunders when entering data. • Interpretation of poorly documented input data. • Assuming values to replace missing data. • Modification to building description so that it can be modelled. • Amended program coding so that the building can be modelled.
BUILDING RESPONSE DATA	Logging	• Noisy, missing or spurious data. • Frequency of measurement insufficient to define variable. • Finite accuracy of recording system.
	Interference	• Internal features of structure altered by monitoring equipment.
COMPARISON PROCEDURE	Data Comparison	• Transcription of measured data from charts etc. • Differences between measured and predicted parameters. • Point of measurement and prediction differ.

peak requirement were estimated. The accuracy of the model input data was also suspect primarily because:

(a) the climate data was collected 15 km from the site of the building;

(b) no air infiltration measurements were made; and

(c) the internal total fresh air flow rate and the distribution of this air to separate zones was assumed from the design values.

In the Annex IV study (Department of the Environment[30]) the predictions of ESP and eight other models were compared with data measured in another three-floor commercial building, the Collins Publishers building in Glasgow, Scotland. Half of the second floor was partitioned off, occupied and monitored. The selection of input data for ESP proved to be a big problem because:

(a) some of the solar and sky radiation measurements were unusable;

(b) the measured inter-zone air flow rates contradicted the information given in the building specification;

(c) the infiltration rates were again largely unknown; and

(d) there was some doubt as to whether the air temperature probes were sufficiently protected against radiation effects.

Numerous other difficulties also existed and the weak building specifications resulted in large differences in values used by the various participants for such fundamental properties as, external opaque wall areas, internal volume and wall U-values. In addition to these problems, the building was too large to be modelled precisely by many of the models, including ESP, without introducing artificial zoning. One of the participants in the study, Lebrun[48], concluded that because of the uncertainties in the model input data the results of the Collins Building project should not be used as reference data for validating any simulation method.

In the fourth study, which was not undertaken by the program authors, Gough[49] used data from a two-zone test building at the Division of Building Research, National Research Council of Canada (Barakat[50]); the work was part of IEA Task VIII. The building zones were heated by electrical baseboard heaters to a thermostatically controlled temperature of 20°C. The building had small windows and was carefully constructed and operated. The main sources of uncertainty were:

(a) in deriving the direct and diffuse solar fluxes;

(b) the ground reflectivity;

(c) the inter-zone advection rate;

(d) the convective/radiant split of heat supplied; and

(e) the thermal properties from standard handbook values.

For the 14 day winter period over which the comparisons were made, the internal air temperatures rarely rose above the thermostat set point values. The most meaningful comparisons were therefore between predicted and measured auxiliary heating power. The error at peak and trough loads was found to be about 16 per cent.

5.3.2 SERI-RES

Three validation exercises involving SERI-RES, its equivalent SUNCODE or its predecessor SUNCAT are examined. In the first exercise (Wheeling[51]) the results were compared with measurements made on a simple direct gain test cell at the National Centre for Appropriate Technology in Butte, Montana. The major sources of uncertainty in this validation were:

(a) the approximate air infiltration measurements;

(b) the unknown extinction coefficient of the glazing; and

(c) the use of handbook thermal parameters, although these were checked and supplemented by heat flux measurements.

In the second study (Dalrymple[52]), the test building was effectively a pair of adjacent Los Alamos Cells with their fronts removed and a glass conservatory attached. A wall of water-filled drums was also installed. The study was part of the IEA Task VIII work. Problems were encountered in obtaining an adequate building description and the measured temperature data also contained obvious errors. Assumptions had to be made about inter-zone air flow rates, thermostat set points, and ground reflectivity. The model had to be re-run a number of times to achieve a reasonable fit between the zone and sun space measured and predicted temperature trends. A major problem in the study was choosing a value for the conductivity of the water in the drums. When using the normal value for 'static' water, 0.62 W/mK, the predicted and measured water temperatures differed significantly. The fit was improved by increasing the conductivity to 10 W/mK.

Study three (Judkoff[45]) was undertaken by the SERI using data from their validation bungalow. This is probably the most complete and rigorously gathered data set currently documented. The care with which the structure was monitored and the thoroughness of the modelling procedure were all of the highest calibre. The monitored building was a four-zoned, single storey ranch house over a crawl space. Any conclusions are thus limited to residential scale, skin-load-dominated buildings. Using the most accurate input data for a six-day period, the total auxiliary heating load predicted by SERI-RES was 10 per cent below the measured value.

5.3.3 DEROB

All five comparisons between DEROB predictions and measured data were undertaken by the model developers. With the exception of the Gaithersburg study (Arumi[53]) all the structures were located in the arid conditions of New Mexico.

The Balcomb residence (Arumi[54]) is a complex two-storey hybrid structure with adobe walls. The triangular courtyard between the two wings is enclosed by a two-storey glass house. Hot air is taken by fans from the top of the glass house through two underfloor rock stores and recirculated. Auxiliary heating is by electrical baseboard heaters. The main sources of uncertainty in the modelling were:

(a) the structure was 'simplified' to five zones;

(b) the missing climate data was made up from records of data gathered at a remote site;

(c) differences in positions of sensors measuring temperatures and the positions at which DEROB predicts temperatures; and

(d) the structure was occupied.

The Bruce Hunn residence (Northrup[55]) is a hybrid two-storey rectangular concrete building with a single storey wing attached on the western side. The Trombe wall and rock store heat collection and storage system are supplemented by an auxiliary ducted warm air installation. The main problems with this study were:

(a) constructional defects prevented the Trombe wall system from functioning correctly;

(b) the residents who occupied the structure for part of the monitoring period had adjusted the thermostats;

(c) a reduced number of zones were assumed to simplify modelling; and

(d) no infiltration measurements were made.

The Williamson House (Wysocki[56]) is a single storey adobe, concrete and brick residence with large south facing windows for direct solar gain. Uncertainty in the results stems primarily from:

(a) a reduction in the number of zones to simplify modelling;

(b) occupancy effects;

(c) the effects of internal heat sources, i.e. the open fire which was not modelled; and

(d) the globe temperatures which were measured, were compared with predicted air temperatures.

Seven of the twelve passive solar test cells built at the Los Alamos Scientific Laboratory were used in the fourth study (Arumi[37]). The 1.5 × 2.4 × 3.0 m high cells are arranged in pairs, side-by-side, with a common well insulated partition to inhibit inter-cell heat flow. The cells are de-coupled from the ground using wooden supports and each has a single south facing window. The cells differ mainly in their internal passive solar heating system. The types of system included were water walls, direct gain, Trombe walls (vented and unvented) and a collector loop detector with water storage tubes. The main sources of uncertainty in the results were due to:

(a) major simplifications to the structure of the complex cells to facilitate modelling; and

(b) unskilled modellers conducting the simulations.

The fifth study (Arumi[53]) involved the six thermal mass buildings at the US National Bureau of Standards site at Gaithersburg, Maryland. These buildings are not strongly solar driven and have small windows. They were built primarily to assess the influence of mass on energy consumption and not to validate models. This study exercised greater control over external errors than previous DEROB studies and the remaining uncertainty was due to:

(a) ground coupling;

(b) incomplete solar radiation measurements at the site; and

(c) inaccuracies in the external air temperature measurements.

5.3.4 BLAST

Of all the models, it appears that the predictions of BLAST have been compared with measured data from the widest range of building types. They range from a 20,000 m² six-storey office block (Yuill[57]) to a 4 m² test cell (Bauman[58]).

The first study (Yuill[57]) was designed to compare the measured monthly energy usage of HVAC systems, in each of two occupied commercial buildings, with the predictions of BLAST. Major problems were encountered in defining the plant schedules and occupancy profiles; it was also necessary to make major simplifications to model the HVAC systems. This study does, however, provide a valuable insight into the problems to be expected if large occupied buildings with mechanical services systems are used for validation purposes.

A single storey, timber frame residence (Kusuda[59]) has been monitored to obtain comprehensive hour-by-hour measurements to assess the effectiveness of summer attic ventilation. Occupancy was simulated by electric lamps and the hourly air temperatures and cooling loads, necessary to maintain the set internal temperatures, were measured and compared with the BLAST predictions. The main sources of error were:

(a) the complex plans and internal furnishings were simplified to simple rectangular zones;

(b) the hourly latent cooling loads were estimated from daily measurements;

(c) the use of artificial weather data; and

(d) measured temperatures and cooling loads were 'averaged' prior to the comparison with predicted values.

The ability of BLAST to accurately predict building and HVAC electricity usage was assessed using data gathered from two US Army buildings (Herron[60]). Each building was single storey, one had 20 rooms while the other had 40 rooms. Both data sets were gathered during the summer when the electrical cooling plant was subject to its greatest demand. Large differences between measured and predicted parameters were observed, the main areas of uncertainty being:

(a) occupancy effects;

(b) building data gathered from drawings; and

(c) no ventilation/infiltration measurements.

In the Los Alamos direct gain test cell study (Bauman[58]) which compared the predictions from a research version of the program, BLAST/MRT, with hourly measurements of air temperature in the cell, the main areas of uncertainty were:

(a) solar radiation data (hourly direct normal, diffuse horizontal and ground reflected radiation) were estimated from measurements on a vertical south facing surface; and

(b) no air infiltration measurements were made.

A further study has been undertaken on a single zone, massive, well insulated concrete structure (Bauman[44,58]) located inside the Environmental Chamber at the US National Bureau of Standards. This study excluded solar and wind effects. Measurements of air temperature, surface temperature, heat fluxes and cooling loads were compared with the predictions of BLAST 3.0. The structure was subjected to a variety of daily cyclic temperature regimes. Differences of up to 25 per cent between measured and predicted daily heat transfer were observed. The main areas of uncertainty were in the choice of:

(a) the surface convection coefficient for one wall;

(b) the conduction coefficient of the glazing; and

(c) in measuring the latent cooling component.

The most recent validation of BLAST has been undertaken by SERI (Judkoff[45]) using data from their validation bungalow. This validation attempt is probably the most rigorous so far reported. Preliminary results gave close agreement between the air temperatures in one zone predicted by BLAST and those predicted by SERI-RES (Judkoff[61]). Using the most accurate input data for a six-day period, the total auxiliary heating load predicted by BLAST was about 17 per cent below the measured value and the RMS air temperature difference was about 0.4°C (Judkoff[45]). The SERI-RES predictions differed by 10 per cent (Section 5.3.2) and further studies are being undertaken to pinpoint the source of the inaccuracies (Judkoff[2]).

5.4 Summary and conclusions of previous validation work

Previous attempts to validate the models ESP, SERI-RES, DEROB and BLAST have been reviewed and many sources of external error have been revealed; these are summarised in Table 2. The review has led to the following general conclusions.

(a) Numerous sources of error may exist in the data input to dynamic thermal models. These errors are propogated through the model leading to uncertainty in the predictions.

(b) The presence of external errors (and the consequential uncertainty in model predictions) has meant that none of the empirical validation studies undertaken on ESP, SERI-RES, DEROB and BLAST would have produced conclusive evidence of internal errors in the models themselves.

(c) Simple sensitivity analyses have succeeded in indicating the uncertainty in the predictions due to uncertainty in the model input data.

Table 2. External errors in previous empirical validation work on ESP, SERI-RES, DEROB and BLAST.

Program name, and version	Name and description of building. Building location, year of monitoring and reference source.	Class of monitoring (Wagner[34])	Weather data	Building specification	Uncertain workmanship	Infiltration, advection	Thermophysical properties	Adjacent zone temperatures	Complex plant	Occupancy effects	Furnishings	Amending program code	Altered building description	Ground coupling	Data recorded on charts
ESP Single-zone	Avonbank multi-storey office building, Bristol, England 1973 29	B-	●		○	●	●		●	●	●			●	●	
ESP Single-zone	Unheated, two-storey, terraced house, Livingstone, Scotland 1978 46,47	B+	●		○	●	●	●						●	●	●
ESP Multi-zone	Passive solar experimental building, Ottawa, Canada 1980> 49	A					●	●								
ESP Multi-zone	Collins' Publishers multi-storey building, Glasgow, Scotland 1982 30	A					●	●	●		●	○	●	●	●	
SUNCAT	Unheated, direct gain, passive solar cell, Butte, MT, USA 1978/9 51	B+					●	●								
SUNCODE	Passive solar experimental building, Los Alamos, NM, USA 1981 52	B+			●		●	●						●		
SERI-RES	Validation research test house, Golden, CO, USA 1982 45	A												○	●	
DEROB	The Balcomb, hybrid solar, residence, Santa Fe, NM, USA 1976→8 54	B+	●		○	○	●				●			●	●	●
DEROB/ PASOLE	Seven, unheated, passive solar test cells, Los Alamos, NM, USA 1978 37	B+	○		●	●	●							●		○
DEROB	The Bruce Hunn, hybrid solar, residence, Los Alamos, NM, USA 1978 55	B+	●		○	●	●		●	●	●	●		●	●	
DEROB	The Williamson, passive solar, house, Santa Fe, NM, USA 1978/9 56	B+			○	●	●		●	●	●			●	●	
DEROB	Six, thermal mass, experimental buildings, Gaithersburg, MD, USA 1982 53	A	●												●	
BLAST 2.0	A six-storey and a two-storey office building, Unspecified city, Canada 1976 57	C	○		●	○	●		●	●	●			●	●	
BLAST 2.0	Conventional, single-storey, test house, Houston, TX, USA 1977 59	B+	●		○	○	●				●			●	●	
BLAST 2.0	U.S. Army dental clinic and battalion H.Q. Fort Hood, TX & Fort Carson, CO, USA 1978 60	A		○	○		●			●	●			●	●	
BLAST/MRT	Unheated, direct gain passive solar cell, Los Alamos, NM, USA 1978 58	B+	●				●	●								○
BLAST 3.0	Indoor, high mass, experimental building, Gaithersburg, MD, USA <1979 44,58	A						●						●		
BLAST 3.0	Validation research test house, Golden, CO, USA 1982 45	A												○	●	

KEY: ● SOURCE OF EXTERNAL ERROR   ○ PROBABLE SOURCE OF EXTERNAL ERROR

(d) The absence of a clearly stated methodology prior to performing empirical validation studies has led to many problems. For example,

(i) inadequate and inaccurate data has been input to the model, and

(ii) the building selected has not necessarily been suitable.

(e) More thought needs to be directed towards measured and predicted values, selecting the parameters which are to be compared, and how this comparison will be made.

(f) Only the very highest quality building construction and data gathering techiques can hope to produce conclusive evidence of internal errors in dynamic thermal models.

(g) It is difficult, expensive and time-consuming to obtain the high quality data needed for validating dynamic thermal models.

6 Developing tests to validate dynamic thermal models

6.1 Validation philosophy

It has been suggested by Bauman[44,58] that, for full validation, the model and algorithms should be compared with experimental data for the full range of buildings, boundary conditions, excitations and climates that could be met. Clearly, this is unpractical and Cohen[62] has therefore suggested that the model should predict the performance of a statistically significant set of buildings and climates. Since the collection of data is expensive and time consuming, it is extremely unlikely that it will be possible to apply the models to a sufficient number of buildings and climates to achieve complete statistical confidence. By a careful examination of extant data sets throughout the world, it should be possible to test the accuracy of the principal algorithms, sub-models and solution techniques used in the whole models over a range of climates and building configurations.

The emerging philosophy is to devise 'tests' which proceed sequentially from simple to more complex situations. Simple situations involving one or two of the more important mechanisms can be investigated by means of analytical tests. The collaborative research group are configuring tests to check the accuracy of algorithms for simulating dynamic conduction heat transfer through opaque surfaces which are exposed to different driving functions. Other algorithms being examined in isolation, are those which model the distribution of solar radiation over internal surfaces, the internal and external surface convection coefficients and the longwave radiation exchange between internal surfaces.

More rigorous tests must be based on measured data from real structures. In designing experiments for use in empirical validation, the main challenge is to identify and control the many thermal mechanisms which operate in the structure so that the causes of divergence between predicted and measured values can be identified. Clearly, this can most easily be accomplished in simple test cells. However, even in the simplest cell, extremely detailed monitoring at the mechanism level will be necessary to identify the reason for divergence between the measured values and those predicted by the models.

By using data from progressively more complex structures the number of algorithms exercised will increase, as will the complexity of the mechanisms involved and the interactions between them. In the larger structures, such as commercial buildings, it is, however, virtually impossible to identify and fully quantify all the interacting mechanisms; validation is therefore likely to be restricted to comparisons between the model predictions and measurements which reflect performance at the building or utilities level.

Even tests based on the simplest test cell experiments are inherently much more complex than the most sophisticated analytical tests, but tests of intermediate complexity can probably be devised on the basis of inter-model comparisons.

6.2 Data sets for empirical validation

A fundamental requirement of any suitable high quality data set is that the measurements must have been taken from a building that has been constructed and operated in a manner which can be explicity modelled. This requirement may be most easily met in experimental buildings. Structures, such as residences and commercial buildings are likely to contain undiagnosed heat flow paths, such as thermal bridges and air leakage paths, or features which cannot be explicitly modelled, such as furnishings and occupants. At present, the validation research group is interested in the accuracy with which models predict thermal loads, rather than their ability to simulate HVAC systems. Any structure with a controlled internal environment should only contain simple items of plant for which the nature and point of heat input/extraction to each individual zone can be defined and, where appropriate, controllers should have a simple operating strategy and respond to a clearly definable stimulus.

It is equally important that the precise data input requirements of the models are satisfied by the data sets. In general terms this means that they should contain: an adequate physical description of the building; details of the building operation and control schedule; comprehensive climate records taken at the site; and measurements of the building response parameters of interest. The time varying measurements should have been taken over short time intervals, typically, one hour or less.

The necessary attributes of data sets, which are to be used for empirical validation, are discussed more fully elsewhere (Lomas[63]). These requirements have lead to the development of assessment criteria which have enabled the authors to isolate, from approximately 200 extant data sets, those which appear to be the most useful for validation purposes.

6.3 Comparing measured and predicted values

Three major issues have to be dealt with when comparing the measured and predicted values. These are:

(a) which variables, measured at which point in the structure, and over what time period are to be compared?

(b) how is the fit between measured and predicted values to be determined?

(c) how are the uncertainties in the model input and building response data to be accounted for?

An interesting approach has been suggested by Hunn[64] who has outlined a procedure employing Monte Carlo methods.

6.4 Presentation of empirical tests

It is the intention of the authors to make the chosen data sets available to the modelling community in the form of an adaptive package. Each package should contain:

(a) a Site Handbook which will describe all input parameters necessary to model the building structure in question;

(b) a Data Tape/Disk containing all of the measured climatic variables and building response parameters; and

(c) Guidelines on how to conduct the tests and interpret the results.

References

1 Burch, J., Data requirements for validation of building energy mechanism level simulations, Draft Report, Solar Energy Research Institute (SERI), Golden, CO, USA (1984).

2 Judkoff, R., et al, A methodology for validating building energy analysis simulations, Draft Report, SERI/TR-254-1508, Golden, CO, USA (1983).

3 Bloomfield, D., Appraisal techniques for methods of calculating the thermal performance of buildings, BSER&T, 6 (1), 13-20 (1985).

4 Busch, R. D., The new SERI data base for validating passive system computer models, Proc. Systems Simulation and Economic Analysis Conference, San Diego, CA, USA, 349-353 (1980).

5 Clarke, J. A., ESP Manual, ABACUS, University of Strathclyde, 17 Sections (1982).

6 Palmiter, L. and Wheeling, T., SERI residential energy simulator version 1.0, SERI, Golden, CO, USA (1983).

7 Control Data Corporation, Cyber BLAST 3.0 User Information Manual, Control Data Corporation, Alexandria, VA, USA (1980).

8 Higgs, F. S., et al, DEROB-NBRI 2.0: Vol. 1, Users' manual, draft version; Vol. 2, Notes about the programs; Vol. 3, Appendices on the theoretical basis of DEROB, National Building Research Institute, Council for Science and Industrial Research, Pretoria, S. Africa (1983).

9 Alexander, D. K. and Lewis, P. T., HTB2—a model for the thermal environment of buildings in operation, Technical Reference Manual rev. 0.0 (draft), Welsh School of Architecture, UWIST, Cardiff (1984).

10 Littler, J. G. F., Overview of some available models for passive solar design, Computer Aided Design, 14 (1), 15-18 (1982).

11 Littler, J. G. F., Review of some promising simulation models for passive solar design, Sun at Work in Britain, 16, 11-18 (1983).

12 Dupagne, A. and Hauglustaine, J. M., European passive solar modelling sub-group, Proc. European Community Contractors Meeting, Meersburg, FRG, in Solar Energy R & D in the European Community, Series A, Vol. 2, Solar Energy Applications to Dwellings, 34-40 (Reidel, 1983).

13 Dupagne, A., Comparison of passive solar design methods within the EEC programme, Proc. Conf. on Design Methods for Passive Solar Buildings, UK Section of the International Solar Energy Society (ISES), 1:1-1:7 (1983).

14 Clarke, J. A., Computer simulation of energy exchanges in buildings, Proc. Second International Symposium on Energy Conservation in the Built Environment, Copenhagen, Denmark, 145-158 (1979).

15 Clarke, J. A., An in-depth investigation of the operation and performance of off-peak electrical storage equipment in factory workshop units, Main Report, ABACUS, University of Strathclyde, Scotland (1982).

16 Arumi, F. and Hourmanesh, M., Energy performance of solar walls: a computer analysis, Energy and Buildings, 167-174 (1977).

17 Arumi, F. and Kim, J. J., Parametric studies on the thermal response of a direct gain room to the distribution of massive elements on the walls, Proceedings, 5th National Passive Solar Conference, Amherst, MA, USA, 96-100 (1980).

18 DeKieffer, R., et al, Passive solar manufactured buildings, development and analysis, Proceedings, Passive and Hybrid Solar Energy Up-date, Washington, DC, USA, 186-193 (1982).

19 Ruyssevelt, P., et al, Optional strategies for upgrading windows in existing housing, Design Methods for Passive Solar Buildings, UK Section of the ISES, 7:59-7:68 (1983).

20 Siminovitch, M. J., The design and comparative computer simulation study of a variable thermal barrier in a passive building application, Proceedings, 5th National Passive Solar Conference, Amherst, MA, USA, 995-999 (1980).

21 Burch, D. M., et al, The effect of thermal mass on night temperature setback savings, ASHRAE Transactions, 90, Pt. 2 (1984).

22 Dupagne, A. and Hauglustaine, J. M., European passive solar modelling sub-group, Proceedings, European Community Contractors Meeting, Brussels, Belgium, in Solar Energy R & D in the European Community, Series A, 4, Solar Energy Applications to Dwellings, 23-29 (Reidel, 1983).

23 Sullivan, P. W., et al, Incremental cooling load determination for passive direct gain heating systems, Proc. Ann. Meet. of American Section of the ISES, 4.2, 977-980 (1981).

24 Derickson, R. G. and Holtz, J. M., The residential energy manual: an example of use and comparison of results to detailed computer simulation, Progress in Passive Solar Energy Systems, American Solar Energy Society, 99-104 (1983).

25 Wortman, D., et al, The implementation of an analytical verification technique on three building energy analysis codes, Proc. ASME, Solar Energy Division 3rd Ann. Conference, Reno, NV, USA, 268-276 (1981).

26 Haves, P., SERI-RES: A thermal simulation model for passive solar and low energy buildings, Design Methods for Passive Solar Buildings, UK Section of the ISES (C34), 4:28-4:36 (1983).

27 Wortman, D., et al, An overview of validation procedures for building energy analysis simulation codes, Proc. Ann. Meet. of American Section of the ISES, 4.2, 1021-1023 (1981).

28 Atkinson, B. A., et al, Validation of CALPAS3 computer simulation program, Proc., 6th National Passive Solar Conference, Portland, OR, USA, 358-361 (1981).

29 Oscar, Faber- & Partners, Results and analyses of Avonbank building simulation, Report for Int. En. Agency Annex I on Energy Conservation in Buildings and Community Systems (1980).

30 Department of the Environment, Building Research Establishment (UK), Glasgow commercial building monitoring project, Final Report for Annex IV on Energy Conservation in Buildings and Community Systems (1984).

31 Merriam, R. L. and Purcell, G. C., A comparative analysis and evaluation by test of solar heating and cooling system computer programs, Proc. Systems Simulation and Economic Analysis Conference, San Diego, CA, USA, 453-456 (1980).

32 Judkoff, R., et al, A comparative study of four passive building energy simulations: DOE-2.1, BLAST, SUNCAT-2.4, DEROB-III, Proc., 5th National Passive Solar Conference, Amherst, MA, USA, 126-130 (1980).

33 Judkoff, R., et al, A comparative study of four building energy simulations: phase II DOE-2.1, BLAST3.0, SUNCAT-2.4, and DEROB-4, Proc., 6th National Passive Solar Conference, Portland, OR, USA, 367-371 (1981).

34 Wagner, B. S., Comparison of predicted and measured energy use in occupied buildings, ASHRAE Trans., 90, Pt. 2 (1984).

35 Hunn, B. D., et al, Applications of DOE-1 to passive solar heating of commercial buildings: preliminary results, Proc., 3rd National Passive Solar Conference, San José, CA, USA, 159-163 (1979).

36 Wray, W. O., et al, Sensitivity of direct gain performance to detailed characteristics of the living space, Proc., 5th National Passive Solar Conference, Amherst, MA, USA, 92-95 (1980).

37 Arumi, N. F., Field validation of the DEROB/PASOLE system, Proc., 3rd National Passive Solar Conference, San José, CA, USA, 152-158 (1979).

38 Arumi, N. F. and Wysocki, M., DEROB: A system for simulating the dynamic energy performance of passive solar structures, ASME/AIChE National Heat Transfer Conference, Orlando, FL, USA (1980).

39 Arumi, N. F., The continued development and physical validation of computer program DEROB, DoEnv. Passive & Hybrid Solar Energy Project Up-date Conference, Washington, DC, USA, 2.69-2.72 (1980).

40 Emery, A. F., et al, The simulation of building heat transfer for passive solar systems, Energy & Buildings, 3, 287-294 (1981).

41 Schnurr, N. M. et al, Applications of DOE-2 to direct gain passive solar systems: implementation of a weighting factor calculation technique, Proc., 4th National Passive Solar Conference, Kansas City, MO, USA, 182-186 (1979).

42 Atkinson, B. A., et al, Validation of CALPAS3 computer simulation program, Proc., 6th National Passive Solar Conference, Portland, OR, USA, 358-361 (1981).

43 Goldstein, D. B., Modelling passive solar buildings with hand calculators, Proc., 3rd National Passive Solar Conference, San José, CA, USA, 164-169 (1979).

44 Bauman, F., et al, Verification of BLAST by comparison with measurements of a solar dominated test cell and a thermally massive building, Proc. Am. Soc. Mech. E. Solar Energy Division 3rd Annual Conference on Systems Simulation, Economic Analysis, Solar Heating & Cooling Operation Results, Reno, NV, USA, 299-308 (1981).

45 Judkoff, T., et al, Empirical validation using data from the SERI Class A validation house, Proceedings, Annual Meeting of the American Section of the ISES, Minneapolis, MN, USA, 705-710 (1983).

46 Clarke, J. A., Environmental systems performance, PhD Thesis, University of Strathclyde (1977).

47 Clarke, J. A., Energy implications in building design: a thermal simulation design model, Proc., 3rd International Symposium on the Use of Computers for Environmental Engineering related to Buildings, Banff, Alberta, Canada, 3-20 (1978).

48 Lebrun, J., et al, Summary of the Belgian contribution in the Collins building analysis, Appendix C3 (in reference No. 30), 113-118 (1984).

49 Gough, M., ESP simulation of Canadian direct gain test cell, Report for the International Energy Agency Task VIII, Sub-task B Validation Exercise (unpublished) (1984).

50 Barakat, S. A., Passive solar heating studies at the division of building research, National Research Council of Canada, Building Research Note No. 188 (1982).

51 Wheeling, T., et al, Performance of passive test units during the 1978-79 heating season, Proc., ISES Silver Jubilee Congress, Atlanta, GA, USA, 1589-1593 (1979).

52 Dalrymple, G. J. and James, R., A report on the IEA Task VIII test cell simulation LASL cells 3 and 4 (14-26 February 1981), Energy Technology Support Unit, Department of the Environment, UK (prepared by CAPS Scientific Ref. 4406/IEA/R1, unpublished) (1983).

53 Arumi, N. F. and Burch, D. M., DEROB simulation of the NBS thermal mass test building, pre-print for ASHRAE Transactions, 90, Pt. 2 (1984).

54 Arumi, N. F. and Northrup, D. O., A field validation of the thermal performance of a passively heated building as simulated by the DEROB system, Energy and Buildings, 2, 1, 65-75 (1979).

55 Northrup, D. O. and Arumi, N. F., The Bruce Hunn residence as simulated by the DEROB system, ASHRAE Transactions, 86, Pt. 2, 873-887 (1980).

56 Wysocki, M. D., et al, The Williamson house as simulated by the DEROB system: a field validation, Proc., Annual Meeting of the American Section of the ISES, Phoenix, USA, 3.2, 794-798 (1980).

57 Yuill, G. K. and Phillips, E. G., Comparisons of BLAST program predictions with energy consumptions of two buildings, ASHRAE Trans., 87, 1, 1200-1206 (1981).

58 Bauman, F., et al, Verification of BLAST by comparison with measurements of a solar dominated test cell and a thermally massive building, Trans. ASME, Journal of Solar Energy Engineering, 105, 207-216 (1983).

59 Kusuda, T., et al, Comparison of calculated hourly cooling load and attic temperatures with measured data for a Houston test house, ASHRAE Trans., 87, 1, 1185-1198 (1981).

60 Herron, D., et al, Comparison of building loads analysis and system thermodynamics (BLAST) computer program simulations and measured energy use of army buildings, US Army Construction Engineering Research Laboratory, Champaign, IL, USA, Interim Report. E-161 (1980).

61 Judkoff, R., Empirical validation of building energy analysis simulation programs: a status report, Proceedings, Passive and Hybrid Solar Energy Up-date, Washington, DC, USA, 33-40 (1982).

62 Cohen, T. S., Statistical problems in design technique validation, SERI Report RR-721-377, SERI, Golden, CO, USA (1980).

63 Lomas, K. J. and Bowman, N. T., Data sets for the empirical validation of dynamic thermal models of buildings (in press) (1985).

64 Hunn, B. D., et al, The DOE passive solar Class A performance evaluation program: preliminary results, Proceedings, Passive and Hybrid Solar Energy Up-date, Washington, DC, USA, 42-49 (1982).