EVALUATION PROCEDURES FOR BUILDING THERMAL SIMULATION PROGRAMS

by

David P. Bloomfield, Head
Systems Performance Prediction
Section, Building Research
Establishment, Watford,
Herts. WD2 7JR, ENGLAND

ABSTRACT

This paper describes the techniques for validating dynamic thermal models devised by collaborating institutions in the United Kingdom. Following a review of past work on Imodel validation, the United States Solar Energy Research Institute (SERI) methodology was used as a starting point. Approximations and errors can arise at all stages of development, revision and use of a program. Emphasis was placed on thorough theoretical reviews of basic physical processes treated by programs and on the actual techniques adopted in some widely used programs. These were used to test the validation methodology. The main processes reviewed were internal longwave radiation, external convection, conduction and shortwave solar radiation.

A new set of analytical tests for conduction was devised. These provide a sensitive tool for detecting errors and for examining the consequences of different numerical solution methods. A review of the treatment of internal longwave radiation uncovered major errors in some existing programs and a set of analytical tests was devised. These were applied to the algorithms in seven current programs and conclusions drawn as to their adequacy. The adequacy of common approximations in convection and solar radiation algorithms was examined and areas where the worst errors could arise identified.

Past validation efforts have concentrated on inter-model comparisons and empirical validation studies. The experience gained from such work was reviewed and several new studies were conducted. In order to avoid past problems a great deal of attention was paid to the definition of the specification of the building and its operation. One study was conducted by a single user with seven commonly used programs. Another study using a very simple structure was conducted as part of an International Energy Agency collaborative project (IEA Solar Task VIII). The latter led to a set of tests for evaluating and selecting programs.

A detailed review of experimental data sets was conducted in order to assess their suitability for validating thermal simulation programs. Attempts were also made to devise improved statistical techniques for comparing measured and predicted quantities. Many of the existing datasets were not adequate for ustablishing the existence of program errors. The best dataset located was used to develop an empirical validation tool and this was tested using three programs.

This paper highlights the main conclusions and describes relevant new activities such as the formation of a research/industry Club in the UK and a new IEA Annex 'The calculation of environmental performance of buildings1 (Buildinga & Community Systems Annex 21).

INTRODUCTION

The word **'validation'** is much misunderstood. It is not feasible to verify the correctness of every path through detailed dynamic thermal simulation programs, to investigate every assumption and approximation, or to take account of every situation in which a program might be used in practice. A working definition of **validation** was adopted: **the testing of the theoretical (physical) correctness of a program and of the mathematical and numerical solution procedures used.**

BRE/SERC VALIDATION

A four year collaborative project funded by the UK Department of the Environment and the Science and Engineering Research Council (SERC) was conducted by the Building Research Establishment (BRE), Leicester Polytechnic (LP), Nottingham University (NU) and the Rutherford Appleton Laboratory (RAL) with BRE acting as co-ordinator [Ref. 1]. The scope of the work was defined to cover detailed simulation programs developed from first principles, which explicitly model the dynamics of buildings and their associated equipment. The overall objective was to improve confidence in thermal modelling procedures by devising and testing a methodology for validation. It was necessary to limit the processes examined in order to make sufficient progress with the available resources. The details of systems, controls, infiltration, air movement and comfort were excluded because of the great increase of complexity that these would have introduced. Although these are of great importance, it was judged better to deal with some of the more basic, fundamental processes first in order that a firm basis could be established for subsequent work. The project was completed in 1988 and a six volume final report will be published shortly.

A review of past validation research showed that most of the results had been inconclusive. No clear separation had been made between mistakes in the code or theory and mistakes and uncertainties arising from the data supplied to the programs. The situations studied were often too complex and this prevented meaningful conclusions from being drawn. Many uncertainties about the building, the climatic data, occupants' behaviour, measurement accuracy etc. often existed. There seemed to have been little recognition of these difficulties and no sound statistical basis had been used to interpret the results. The methodology developed by the SERI [Ref. 2] was the most complete attempt at a rational basis for validation.

Theoretical Reviews and Analytical Tests

A questionnaire was developed to enable the theoretical basis of a program to be described in a standard form. This was tested and revised by studying the following programs - DEROB, ESP, HTB2, SERIRES. It was very difficult to obtain clear and accurate information about the theory, assumptions and approximations used. Much of the available information was misleading, incomplete or incorrect in some details. There is little common understanding of the precise meaning of many technical terms. A significant number of mistakes/errors/bugs were uncovered in the programs examined and their documentation. The understanding gained was invaluable in completing the other stages of the validation methodology.

Thorough theoretical reviews of **conduction, external convection, internal longwave radiation, solar radiation** were conducted. Original data sources were consulted, **sensitivity analyses** were conducted to assess the adequacy of approximations and assumptions, and **analytical tests** were devised and executed where possible. **Analytical tests** involve the derivation of 'exact' solutions by analytical means which can be compared with the equivalent program predictions. These could be used for **debugging**, in which case the number of sources of error should be restricted so that a clear interpretation of results can be obtained. Alternatively, they could be designed as **tests of the accuracy obtainable in 'normal' building applications.** Because of the wide range of building types and user applications, general tests of the latter type are difficult to devise. For debugging tests it is advantageous to choose ones which test some relatively extreme conditions. Exploring the conditions under which an algorithm will break down can lead to results which are easier to generalise.

A wide range of Conduction solution techniques were reviewed by BRE. No completely satisfactory theoretical way of choosing the appropriate level of discretisation in order to achieve a given accuracy was found. The studies conducted showed that it is unsafe to leave this choice entirely to the user. The SERI analytical tests were applied to SERI-RES and to ESP, involving the application of step changes in external temperature to a wall. These provide only a weak test because the temperature drop across the thermal mass is less than 5% of the drop across the whole wall. The use of sinusoidal and ramp excitations was explored and the analytical tests were used to explore the effect of different discretisation schemes. More sensitive step excitation tests for the conduction algorithm were devised for a wider range of conditions [Ref. 3]. The weather sampling and reconstruction schemes can have an important effect on the results of analytic tests. The BRE recommended tests aim first to check that the program is free of bugs and then to relax the simplifications, working back to more realistic simulations in order to see how much deliberate physical and numerical approximations contribute to errors. Fig. 1 illustrates the result of using different levels of discretisation in SERI-RES for one of the tests. The number of tests required can be reduced by demanding a high level of agreement for a relatively small number of sensitive tests.

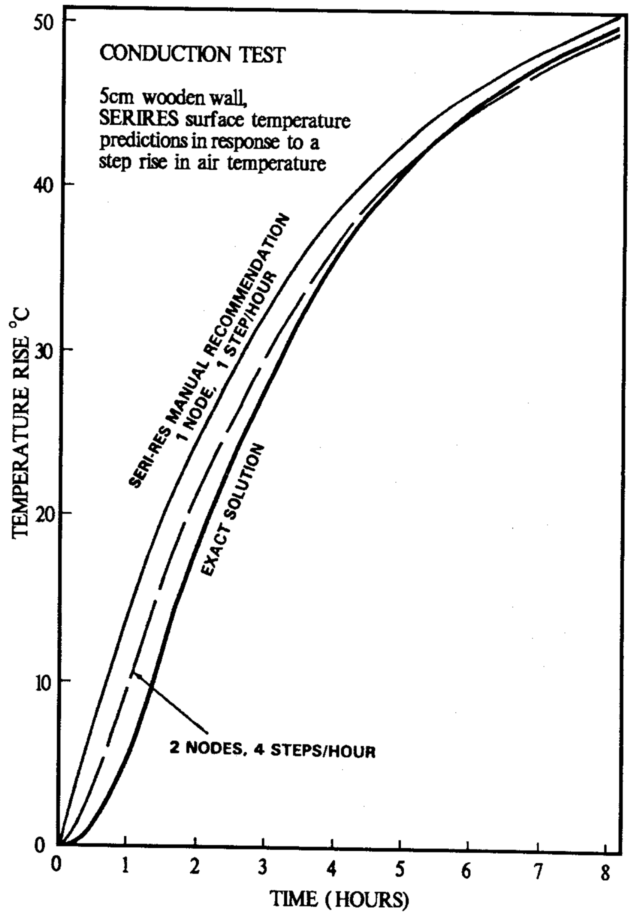

FIG. 1 Results of BRE Analytical Test for SERIRES Conduction Algorithm

Internal longwave radiation algorithms were reviewed by BRE. There was a wide range in complexity, and hence in computational requirements, between the methods chosen. An exact method of calculation for steady state exchanges was used to devise analytical tests, the three major variables considered being surface temperatures, emissivities and room geometry. Fig. 2 illustrates the range of geometries chosen. Eight representative algorithms were examined using these tests and Fig. 3 illustrates some results for geometry F. The four cases correspond to different wall and radiator situations. The more detailed methods usually performed better, but one due to Walton [Ref. 4], which does not explicitly consider geometry, performed almost as well as the most accurate method. Some of the methods examined showed very high errors and this led to the identification of misunderstandings in the development of the algorithms and to their resolution.

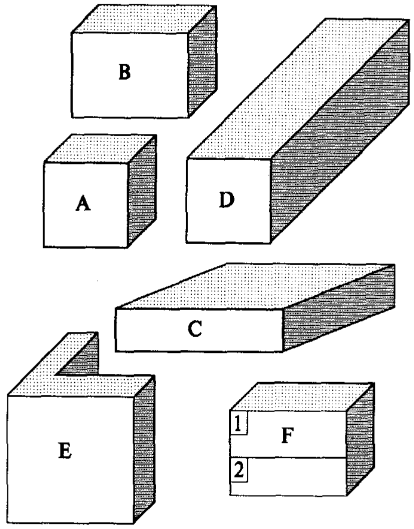

FIG. 2 Geometrical Configurations for BRE Analytical Tests for Internal Longwave Radiation

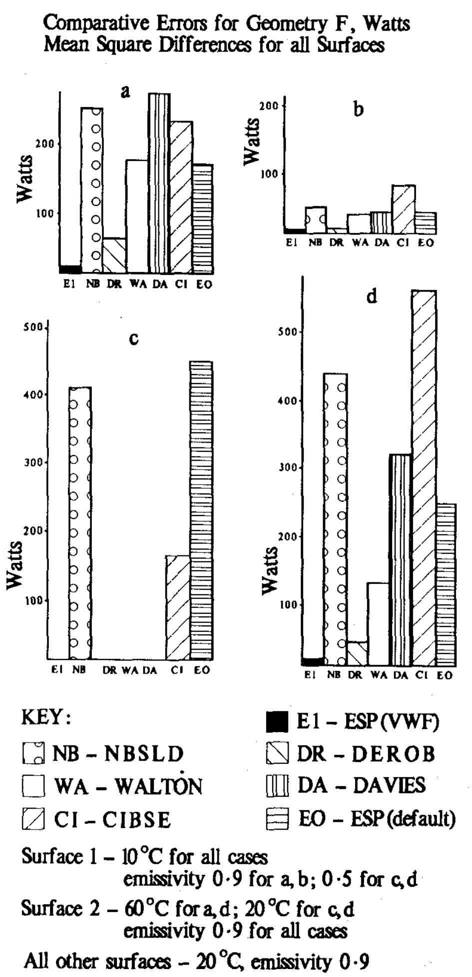

FIG. 3 Results of BRE Analytical Tests for Internal Longwave Radiation Algorithms, Geometry F

A review of **solar processes** was conducted by NU [Ref. 5]. This included glazing properties, shading by the building facade, one dimensional approximations and internal distribution of solar radiation. Commonly available algorithms were studied and guidance developed as to their applicability. The worst errors in solar gain were concluded to arise from:

-incorrect specification of building orientation
-modelling a window as a pure conductance
-neglecting shading due to setback of glazing from the facade
-neglecting shading of incident diffuse radiation.

Each of these effects could give rise to a maximum error of about 5-10% Under the worst case conditions the solar gain could be overestimated by 20% A simplified representation of the overall simulation process was used to assess the conditions under which the more detailed algorithms are necessary. These are strongly dependent on the type of building, climate and application purpose for which the program is being used. It was found to be difficult to devise analytical tests which could be applied without detailed knowledge of the program code. The use of statistical tests was explored by NU and these may hold promise for the future but would need further development.

A review of **convection from external surfaces** was conducted by NU [Ref.8]. Conclusions were drawn on the best algorithms available and on possible improved modelling approaches. There are three possible air flow regimes and one major source of uncertainty arises from the determination of the point of transition from one to another. Together with uncertainties in wind velocity and surface roughness, the convection coefficient is likely to be subject to uncertainties of the order of 50% This will have the largest effect on the calculation of heat loss through glazed openings. Typical uncertainties for dwellings in the UK were estimated to be 4% in heating loads. The use of combined convection radiation coefficients was judged to be suitable for steady state calculations but not for the assessment of short-term dynamic effects. Insufficient data exists for the effects of wind direction to be properly assessed. Modelling studies to examine the effects of orientation, windbreaks etc should therefore be treated with some caution.

Inter-model Comparisons

The comparison of building thermal performance predicted by one program with that predicted by others is one of the most frequently used techniques in past validation work. There are, however, several important concepts that must be understood if the results of such exercises are to be correctly interpreted.

The specification of buildings and simulation problems are necessarily in terms of simplifications of the real world, so that some realism is already lost. These simplifications will not match those made in all simulation programs. A program user must therefore translate a simplified description into the real world problem and then simplify it again in a different way to match the constructs in the program that is to be used. Because there is no unique way of making this transformation, errors are unavoidable. This effect has been demonstrated to be of great importance where users of the same program produced a large range in results [Ref. 7]. BRE conducted an exercise designed to overcome some of these problems. A single user designed a detailed building specification in such a way that compatible data could be provided for seven programs. Fairly large differences in predicted heating loads for this half duplex house were obtained [Ref. 8]. The range in results was even larger for predictions of the effects of variations in the original design. This exercise demonstrated a number of important points:

(1) the difficulties of specifying a problem to a prediction program
(2) confirmation that there are real and important differences in the quality of advice that would be obtained by a designer using these programs
(3) the difficulty of obtaining firm conclusions as to the adequacy of any single program from such inter-model comparisons.

An alternative approach to inter-model comparison was developed within this project and within IEA Solar Task VIII [Ref. 9] in order to address (3) above. Reference building specifications were devised and documented in detail. These were variants of a basic building, more complex features being added, one at a time. The functioning of different algorithms could be judged by examining differences in predictions between pairs of reference cases. Simulations were conducted with five detailed simulation models which had already been subjected to evaluation and the ranges in predictions were found to be reasonably small. Considerable effort was expended to check the accuracy of the data input for these simulations, and the differences between results were judged to be explicable in terms of the different assumptions and approximations made in the programs. The ranges in results were used as 'targets' against which the performance of other programs could be judged. Although they can not be taken as definitive 'correct' results, they could be argued to represent the best information available to the modelling community and can be used as a helpful aid to program developers and to users selecting a program. A methodology for using the set of 25 IEA VIII cases in the most efficient way was attempted and this will be developed further within two new IEA projects - IEA Annex 21 and IEA Solar Task 12.

Empirical Validation

This technique should be the most convincing one. It provides quantitative information on how well the program can predict building performance in practice. There are, however, some difficulties:

(1) It is expensive, difficult and time-consuming to monitor a building in sufficient detail and with sufficient accuracy to provide information suitable for validating a dynamic program.
(2) It is difficult to generalise results from one particular combination of building type, climate and operating conditions to others.
(3) Measurements of both building performance parameters and program input parameters have a finite accuracy. The uncertainty in program input parameters (air change rates, material properties, occupant behaviour etc.) leads to uncertainty in the program predictions quite apart from the adequacy of the algorithms employed within the program.
(4) Empirical validation of the dynamic performance of programs therefore involves the comparison of pairs of time series, each of which are subject to uncertainties. Soundly based statistical procedures for performing such comparisons and for conducting sensitivity analyses have yet to be fully developed and proven [Ref. 11]. Non-linearities and correlation between inputs need to be allowed for.

LP conducted an extensive literature search of experimental datasets which might be suitable for validation [Ref.12]. Out of the 589 monitored structures identified throughout the world, 130 had been used for empirical validation. An examination of these showed that many were subject to unacceptably high uncertainties in data required as input to programs. Only 6 appeared to be of sufficiently high quality to enable an accurate evaluation of the predictive capabilities of the three models of main interest. Two datasets were obtained and one, the Polytechnic of Central London (PCL) test set, was used to investigate empirical validation as a technique. Predictions were performed 'blind', in ignorance of the actual measured performance, for the programs ESP, HTB2 and SERI-RES. Statistical techniques for assessing the agreement between program and measurements were investigated by RAL and NU. Both differential and stochastic sensitivity analyses were conducted to assess the effects of uncertainties in the data input. It was concluded that in the absence of a more complete theory, sufficiently accurate estimates of the error bands could be obtained by performing separate simulations in which input data values were varied about their best estimates one at a time, and then by adding the effects of these errors in quadrature. Even for this very well monitored building, the total uncertainties in data inputs, and hence in model predictions, was considerable. Very little mechanism level data was available. Despite these limitations, the dataset did prove capable of detecting errors in the programs tested. The dataset has been thoroughly documented and is available from LP in the form of a data disk, site handbook and users' guide.

Influence of the Program User

One of the main factors that has become apparent in the course of this work is the importance of the program user. The opportunities for misunderstandings, mistakes and misuse of models are legion, quite apart from the inherent difficulties arising from the need to perform pre-modelling to fit the real world to the model. It is therefore very difficult to completely separate validation of a program from other issues such as the documentation of the building, of the program itself, the human computer interface and the interpretation of results. An investigation was conducted to assess the effect of the many assumptions that the user has to make [Ref.13]. This demonstrated the difficulties facing a user and highlighted the need for more research and the production of guidance in how best to approach a particular application problem. The need for standardisation has already been mentioned. Important advances could be made relatively quickly by:

(1) producing and disseminating guidance on how models should be used
(2) making appropriate data available for use in particular applications, and adopting standard formats
(3) improving the human computer interface, perhaps standardising it
(4) subjecting all programs used for e.g. regulatory purposes to a minimum set of evaluation tests.

NEW DEVELOPMENTS

In the UK an Industry/ Research Club has been formed with BRE providing the secretariat. The **Building Environmental Performance Analysis Club (BEPAC)** attracted some 100 members in its first year (1988). The objective of the club is **'To improve the quality of building performance by encouraging the use and development of environmental prediction methods for buildings'**. Four task groups have been formed so far and these are currently defining a list of priorities within their respective fields. These are **Air Movement**, **Lighting**, **Controls** and **Standards**. In addition to two full members' meetings each year, workshops on important technical topics or problems are planned.

The Buildings and Community Systems group of the **IEA** sponsored a workshop in May 1988 which led to the formation of a new project - **Annex 21 - The Calculation of the Environmental Performance of Buildings**. Over the past year 10 countries have been conducting a joint feasibility study led by the author to define the work programme for a three year collaborative project in this field.

CONCLUSIONS

A thorough review of past validation work showed that results had been inconclusive, mainly due to the use of situations that were too complex, to inadequate treatment of uncertainties and to insufficient attention being paid to statistical analysis of results. The program user can play a significant part in influencing the quality and adequacy of the results produced. It is essential that user effects are considered carefully in any validation work.

The accuracy or adequacy of a model is dependent on the purpose for which it is being used - the **application**. Any statement that a program has been validated is meaningless unless accompanied by more detailed information describing the application and the circumstances for which the 'validation' has been performed.

If program results are to have any meaning it is essential that the user has a good understanding of the main program assumptions and approximations. For validation work a detailed knowledge of the theoretical basis is necessary. Such information did not exist or was not documented in a suitable form for the programs examined. The format for documenting such information needs to be developed and standard terminology should be established within the modelling community. Detailed questionnaires were prepared for four models within the BRE/SERC work. Automatic intelligent systems to aid in documentation of theory and assumptions show promise and will be explored within IEA Annex 21.

Thorough reviews of theory and available algorithms for conduction, internal longwave radiation and solar radiation have been conducted and guidance given for future model developers. Analytical tests have been devised and executed for conduction and internal longwave radiation. These are suitable for verifying that the algorithms are working as intended by the developer and have proved successful in identifying errors and examining the adequacy of numerical solution schemes. Reviewing theory and examining source code is tedious but was shown to be effective in detecting implementation errors. More automatic, statistical techniques were explored but need further development. Symbolic algebra approaches to software 'proving' are not currently at a stage such that they can be applied to dynamic thermal simulation models.

The description of a real world problem to a program has to be made in terms of simplifications, partially forced by the program, its user interface and approximations, and partially by pre-modelling simplifications made by the user. There is no unique way to describe the real world problem in terms which are suitable for all models. This is an important conclusion which influences how inter-model comparisons should be conducted and their results analysed. A set of well-specified realistic building specifications was produced and tested for several programs. Significant differences between results were obtained. A set of simplified building specifications was devised and tested as part of IEA Task VIII. The programs used had all been subjected to previous validation of some sort and represented the best available techniques in the opinions of the participating experts. Reasonable agreement was obtained between the program results, and a set of 'benchmark' tests was devised. These were structured to successively test different program features and to aid in the interpretation of results. This methodology has been adopted by some of the participating countries and is under consideration by an ASHRAE Task group. Two new IEA groups have also decided to adopt this approach as part of an evaluation procedure.

Past experience of empirical validation was reviewed and a thorough survey of existing experimental datasets suitable for validation was conducted. Data from only six sites were judged to be adequate. A test cell dataset was compared with predictions from ESP, HTB2, and SERIRES without prior knowledge of the experimental results. An analysis of uncertainties in the data input values was undertaken. This showed the existence of some errors in the programs tested, highlighted the need for improved algorithms for the calculation of surface coefficients, for more mechanism level experimental data and for more high quality datasets for a wide range of building types, climates etc.

A great improvement in the quality and credibility of thermal modelling could be made by adopting the following measures:

-improving the documentation of program theory and assumptions
-producing and disseminating guidance on how programs should be used
-making appropriate data available
-improving the human computer interface
-subjecting programs to a basic set of evaluation tests.

Research on these topics will be undertaken within IEA Annex 21 and the results will be disseminated and promoted with the help of Industry/Research clubs like BEPAC.

ACKNOWLEDGEMENTS

The BRE/SERC validation work was conducted by several collaborating institutions (BRE, LP, NU & RAL). The BRE work forms part of the research programme of the UK Department of Environment and is published by permission of the Director.

REFERENCES

1. Allen EJ et al, Analytical and Empirical Validation of Dynamic Thermal Models, Int. Building Energy Simulation Conf., Seattle, August 1985
2. Judkoff R. et al, A Methodology for Validating Building Energy Analysis Simulations, Solar Energy Research Institute Report SERI/TR-254-1508
3. Bland BH, Experience with Analytic Tests for the Validation of Dynamic Thermal Models, Proc. USER1, Ostend, Sept. 1988, SCS Europe, Coupure Links 653, B-9000 Ghent, Belgium
4. Walton GN, A new Algorithm for radiant interchanges in room loads calculation, ASHRAE Trans 86(2), pp 190-208
5. Allen JP & Whittle JK, How simple can solar algorithms be?, Bath, July 1986 (see above)
6. Allen JP & Whittle JK, The Appropriateness of Heat Transfer Coefficient Algorithms, Bath July 1986 (see above)
7. Jones L, The analyst as a factor in the Prediction of Energy Consumption, Proc. 2nd Int. CIB Symposium on Energy Conservation in the Built Environment, Danish Building Research Institute, Copenhagen 1979
8. Allen EJ & Bloomfield DP, Improving Confidence in Thermal Calculation Procedures, Proc. CLIMA 2000, Copenhagen Dk, 1985
9. Judkoff R et al, International Energy Agency Design Tool Evaluation Procedure, Proc. Solar 88, 13th National Passive Solar Conference, Cambridge, Ma, USA, July 1988
10. Bloomfield DP et al, Improving the Design of Low Energy Passive Solar Houses, Proc. Solar 88, 13th National Passive Solar Conf., Cambridge, Ma, USA, July 1988
11. Irving AD & Allen JP, Stochastic Sensitivity Analysis of Dynamic Thermal Models, CIB.86 10th Triennial Congress 'Advanced Building Technology', Vol 3
12. Bowman NT & Lomas KJ, The Evaluation and Use of Existing Data Sets for Validating Dynamic Thermal Models of Buildings, Bath, July 1986 (see above)
13. Bloomfield DP, The Influence of the User on the Results obtained from Thermal Simulation Programs, Bath, July 1986 (see above)